Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 22 (2020), Issue 3 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Proposal

Viewpoints

Original Papers

Corrigenda and Addendas

Review

Delivery of Compassionate Mental Health Care in a Digital Technology–Driven Age: Scoping Review

Jessica Kemp[1,2*]; Timothy Zhang[1,2*]; Fiona Inglis[3], MA, MI; David Wiljer[3,4,5], PhD; Sanjeev Sockalingam[3,6], MD, MHPE; Allison Crawford[3,6], MD, PhD; Brian Lo[2,3,5], BHSc, MHI; Rebecca Charow[4,5], MSc; Mikayla Munnery[5,7], BCom; Shuranjeet Singh Takhar[5], BSc, MSc (oxon); Gillian Strudwick[2,5,7], RN, PhD

[1]Faculty of Science, University of Waterloo, Waterloo, ON, Canada
[2]Information Management Group, Centre for Addiction and Mental Health, Toronto, ON, Canada
[3]Office of Education, Centre for Addiction and Mental Health, Toronto, ON, Canada
[4]Education Technology and Innovation, University Health Network, Toronto, ON, Canada
[5]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[6]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[7]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada
[*]these authors contributed equally

Corresponding Author:
Gillian Strudwick, RN, PhD
Campbell Family Mental Health Research Institute
Centre for Addiction and Mental Health
1001 Queen St W
Toronto, ON, M6J 1H4
Canada
Phone: 1 4165358501 ext 39333
Email: gillian.strudwick@camh.ca

Abstract

Background: Compassion is a vital component to the achievement of positive health outcomes, particularly in mental health care. The rise of digital technologies may influence the delivery of compassionate care, and thus this relationship between compassion and digital health care needs to be better understood.

Objective: This scoping review aimed to identify existing digital technologies being used by patients and health professionals in the delivery of mental health care, understand how digital technologies are being used in the delivery of compassionate mental health care, and determine the facilitators of and barriers to digital technology use among patients and health professionals in the delivery of compassionate mental health care.

Methods: We conducted this scoping review through a search of Cumulative Index to Nursing and Allied Health Literature, Medical Literature Analysis and Retrieval System Online (MEDLINE), MEDLINE In-Process and EPub Ahead of Print, PsycINFO, and Web of Science for articles published from 1990 to 2019.

Results: Of the 4472 articles screened, 37 articles were included for data extraction. Telemedicine was the most widely used technology by mental health professionals. Digital technologies were described as facilitating compassionate care and were classified using a conceptual model to identify each digital intersection with compassionate care. Facilitators of and barriers to providing compassionate care through digital technology were identified, including increased safety for providers, health care professional perceptions and abilities, and the use of *picture-in-picture* feedback to evaluate social cues.

Conclusions: Implementing digital technology into mental health care can improve the current delivery of compassionate care and create novel ways to provide compassion. However, as this is a new area of study, mental health professionals and organizations alike should be mindful that compassionate human-centered care is maintained in the delivery of digital health care. Future research could develop tools to facilitate and evaluate the enactment of compassion within digital health care.

(J Med Internet Res 2020;22(3):e16263) doi:10.2196/16263

KEYWORDS

compassion; mental health; medical informatics; psychiatry; health information technology; nursing informatics

Introduction

Background

The use of digital technology in mental health care delivery has increased significantly in recent years [1-5]. Improved patient access to mental health services is a common metric used to endorse the use of technology in mental health care through technologies such as telepsychiatry [6]. Owing to the emerging uptake and expansion of digital technologies into a variety of traditional and nontraditional care settings, such as in-home use of digital technologies or the provision of care in virtual environments, there exists a greater need to understand best practices surrounding digital technology use to ensure quality patient-centered care is delivered through these modalities [7-11]. Fostering the delivery of compassionate care has been identified as an important need because compassion has been shown to positively influence the experience of both patients and health professionals alike [7]. Without an adequate understanding of the best practices and uses of digital technology for the delivery of compassionate mental health care, these technologies may detract from compassionate care and hinder health professional-patient relationships, which are of great importance in the context of mental health care. However, when employed appropriately, these same technologies may facilitate and strengthen compassionate mental health care and create new means for relationships between mental health professionals and patients [12].

Compassion in Health Care

Compassion encompasses a wide array of meanings [13,14]. The working definition of compassion used for the purpose of this review defines five dimensions of compassion: (1) awareness of another’s experience of suffering or need, (2) feeling *moved*, (3) recognizing this feeling as a response to the other’s need, (4) making a judgement that the other is suffering, and (5) engaging in a behavior in an attempt to alleviate the suffering [15]. The importance of compassion in mental health care is central; for many patients, receiving compassionate care throughout the process of diagnosis, treatment, and recovery can improve their perceived quality of care [16-18]. When compassion is present in mental health care settings, there can be a greater therapeutic alliance (the quality of the relationship between provider and patient), increased openness of the patient which improves health professionals’ understanding of a patient’s experiences, and greater experiences of empathy as part of the health professional-patient relationship, ultimately supporting patient-centered care [17-22].

Delivery of Compassionate Mental Health Care

The delivery of compassionate mental health care can take many forms, with the foundation being to remain patient focused, establishing interactions based on trust, and ensuring physical and emotional safety [23]. Compassionate care may be subjectively experienced; however, the literature suggests that it is commonly delivered by providing safe and comfortable spaces for health professional-patient interactions and is rooted in an understanding by health professionals of the lived experiences of patients [19,24,25]. It is important to recognize that not all patients will experience the feeling of compassion or build a compassionate relationship in the same way; however, digital technology has the potential to meet a wide range of patient needs and provide more personalized care due to the adaptability of technology [26].

Digital Technology Use and Compassionate Mental Health Care

Many digital technologies are currently being used in mental health care contexts, including (but not limited to) mobile apps, patient portals, electronic health records (EHRs), instant messaging, telemedicine, and virtual reality [27]. This review arose out of the motivation to understand what is known about the suitability of these technologies for facilitating or enhancing compassionate care and whether any evidence can guide best practices for use.

Purpose

The purpose of this review was to identify the ways in which compassionate care can be delivered in mental health care through and with the use of digital technologies, as well as across the continuum of mental health care settings and processes. To develop an understanding of the intersection between digital technology and compassionate mental health care, this review examines 3 research questions (RQs):

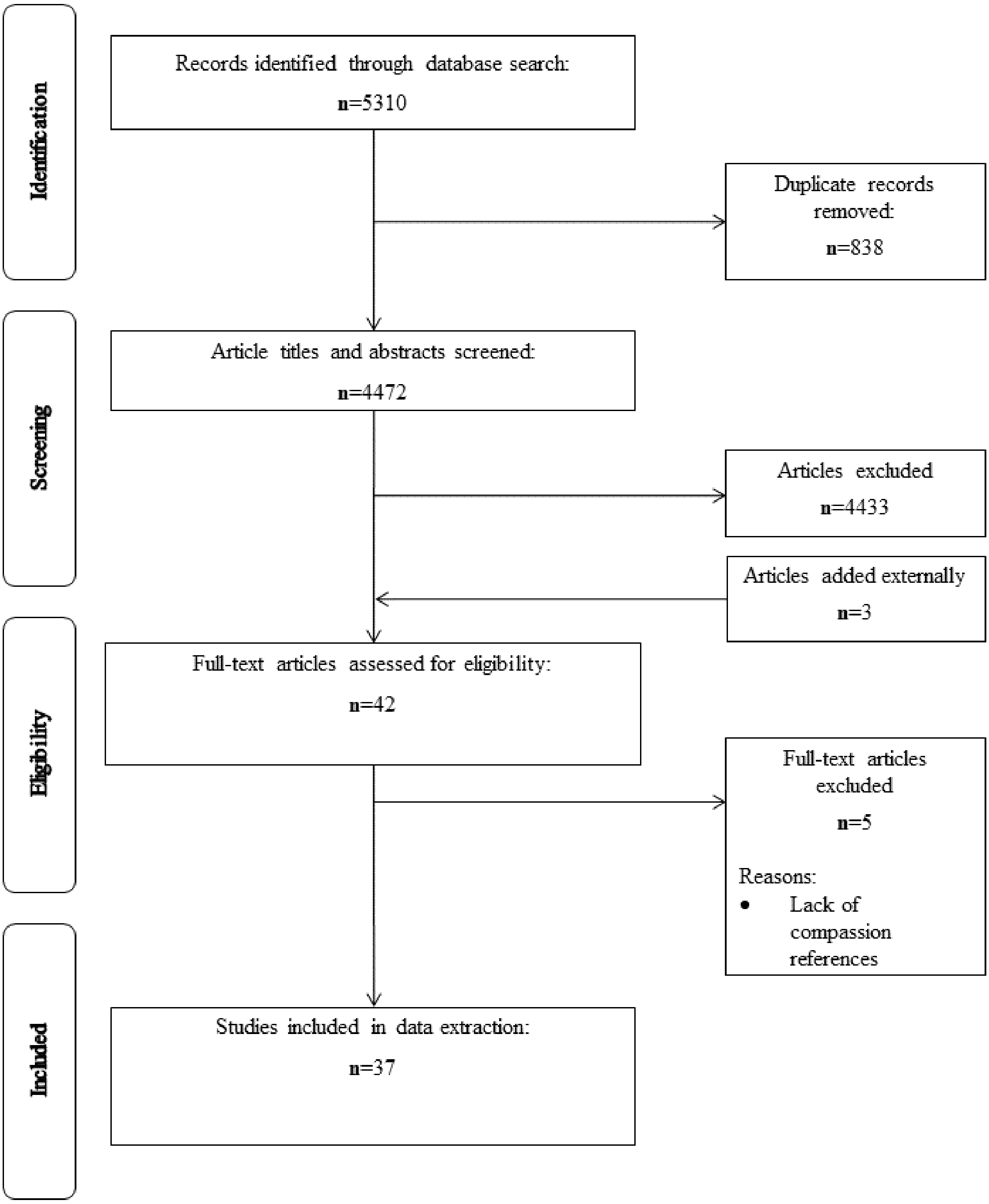
1. What existing digital technologies are most commonly used among patients and/or health professionals in the delivery of mental health care?
2. How are existing digital technologies being used among patients/health professionals in the delivery of compassionate mental health care?
3. What are the perceived facilitators of and barriers to using digital technology among patients and/or health professionals to deliver compassionate mental health care?

Methods

Overview

This review was conducted following the methodological framework for scoping review studies proposed by Arksey and O’Malley [28] and refined by Levac et al [29]. To illustrate the scoping review process, the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) diagram [30], shown in Figure 1, was used as well as the PRISMA-scoping review checklist which outlines the key milestones of a scoping review [31] (Multimedia Appendix 1). A detailed protocol for this scoping review titled *Delivery of Compassionate Mental Health Care in a Digital Technology-Driven Age: Protocol for a scoping review* was published in *BMJ Open* [32]. The following sections provide a brief overview of the methodology utilized in this scoping review.

Figure 1. Preferred Reporting Items for Systematic Review and Meta-Analysis flow diagram of scoping review results.



Stage 1: Identifying the Research Questions

For the purpose of this scoping review, the aforementioned RQs were identified to better understand the delivery of compassionate mental health care through and with the use of digital technology.

Stage 2: Identifying Relevant Studies

All types of research studies including quantitative, qualitative, observational, and literature reviews from any country of origin published in English from 1990 to 2019 were included. Given the nature of the topic being investigated, grey literature was not included. All areas of mental health care, diagnoses, digital technologies used, and age groups were included. Studies were selected if they addressed at least 1 of the 3 RQs and involved the use of digital technology in mental health care in relation to compassionate care. As compassion is a difficult concept to define, the working definition of compassion described earlier was used to guide the identification of eligible articles involving compassionate care. Although all types of digital technology were eligible, some imaging and measurement technologies such as those intended to measure emotion, blood pressure, or conduct body scans were excluded [32].

A total of 5 databases were searched: Cumulative Index to Nursing and Allied Health Literature, Medical Literature Analysis and Retrieval System Online (MEDLINE), MEDLINE In-Process and EPub Ahead of Print, PsycINFO, and Web of Science. A research librarian (FI) completed the search strategy and database searches. As a part of working with a librarian, extensive use of synonyms, Boolean operators, combinations of search terms, and MeSH headings were employed. The complete search strategy for MEDLINE is available in the published protocol for this scoping review [32].

Stage 3: Study Selection

All identified articles were screened independently by 2 reviewers (TZ and AM), concluding with an interrater reliability of 99.22% agreement and a Cohen kappa of 0.59. Disagreements which could not be resolved between TZ and AM were discussed with the greater research team, as outlined in the study protocol [32]. The screening process was facilitated by Covidence (Veritas Health Innovation), a literature review streamlining software recommended by Cochrane [33].

While identifying the relevant studies for the scoping review through the screening process, the authors selected articles that either directly facilitated the delivery of compassionate care or prepared for the delivery of compassionate care while addressing 1 or more of the 5 dimensions of compassion. It is important to note that compassion was not always explicitly brought up in some articles and the professional judgement of the authors had to be used to identify appropriate studies. Upon further research and completion of data extraction, it was evident that there was a greater divide among the relevant studies. The authors chose to use the digital intersections with compassion to further clarify the role/dimension each technology played in the delivery of compassionate care.

Stage 4: Data Items and Data Collection Process

During the process of data extraction, the following article summary information was charted: title, authors, year of publication, country of origin, research design, RQs addressed, and answers to the applicable RQs. Data were charted using Microsoft Excel 2010. The data extraction table is available upon request from the corresponding author.

Stage 5: Synthesizing and Reporting the Results

Both quantitative and qualitative methods were used to analyze the results of the RQs. A descriptive quantitative analysis (descriptive statistics) was used for RQ1, and a qualitative content analysis was used for RQ2 and RQ3. To understand what existing digital technologies are most commonly used among patients and health professionals in the delivery of mental health care, the results of RQ1 were organized using the World Health Organization's (WHO) classification of digital health interventions v1.0 [34]. This classification system organizes digital technologies used in health care based on the user of each intervention.

Stage 6: Consultation

The consultation phase for this review was completed through discussions with mental health and digital health researchers, mental health professionals, and various health care professionals in Ontario, Canada, selected through the Associated Medical Services (AMS) health care community. More specifically, these stakeholders were consulted at Waypoint Centre for Mental Health Care, the Centre for Addiction and Mental Health, the University of Toronto, and Western University. The consultation process was important for the organization of results and to ensure the strategies used for knowledge translation were appropriate. These discussions also supported the identification of important topics to include in the Discussion section of this paper.

Results

Search Results

A total of 37 articles were included in the final review. Details regarding the screening process are described in Figure 1.

Study Characteristics

Table 1 describes the characteristics of the studies included in this review. Studies were identified from 7 countries with 57% (21/37) of these publications originating from the United States. Given the novelty of digital technology use in mental health care, 51% (19/37) of articles were published between 2016 and January 2019. A research focus on a specific mental health diagnosis was uncommon in the selected articles; only 27% (10/37) of the articles were related to a specific diagnosis. Articles that did not specify a mental health diagnosis, and rather addressed mental health care as a single entity or did not specify the diagnoses of patients, were categorized as *unspecified*. Table 1 also includes the methods that were used in the identified articles.

Table 1. Study characteristics.

Article characteristics	Value (N=37), n (%)	References
Country of publication		
United States	21 (57)	[24,35-54]
United Kingdom	8 (22)	[19,20,55-60]
Australia	4 (11)	[61-64]
Canada	1 (3)	[65]
China	1 (3)	[66]
Israel	1 (3)	[26]
The Netherlands	1 (3)	[67]
Research method		
Literature review	16 (43)	[20,26,37,40,42,44-48,50,54,62-64,67]
Questionnaire/survey	8 (22)	[24,49,52,53,56,61,65,66]
Mixed method	5 (14)	[35,55,57-59]
Semistructured interview	4 (11)	[38,41,43,60]
Other[a]	4 (11)	[19,36,39,51]
Year of publication		
2016-2019	19 (51)	[24,38,40,41,47-55,57,59,61,64-66]
2010-2015	13 (35)	[19,20,26,35,37,43-45,56,58,60,62,67]
2000-2009	5 (14)	[36,39,43,46,63]
Mental health diagnosis[b]		
Unspecified	27 (73)	[20,24,26,35,37-48,50,54,55,57,59,62-67]
Schizophrenia and psychosis	3 (8)	[51,53,61]
Anxiety and depression	3 (8)	[19,52,58]
Trauma and stress disorder	2 (5)	[36,49]
Alzheimer and dementia	2 (5)	[56,60]
Addictions/substance use	0 (0)	—[c]
Developmental disabilities	0 (0)	—
Problem gambling	0 (0)	—
Mood and personality disorders	0 (0)	—

[a] Other research methods include group therapy sessions and personal essays written by health professionals.

[b] Categories of mental health diagnosis based on the Centre for Addiction and Mental Health's Mental Illness and Addiction Index [68].

[c] No articles were identified.

Research Question 1: Digital Technology Use in Mental Health Care

Of the 37 articles, 15 [19,24,39,41,42, 45,46,48,51,52,57,58, 64-66] and 22 [20,26,35-38,40,43,44,47,49,50,53-56,59-63,67] articles were specific to digital technology usage by patients and health professionals, respectively. Patient and health professional digital health interventions were then divided into detailed categories, as shown in Table 2. On-demand information services were the most common digital technology used by patients (eg, educational resources), including websites [52,65], mobile phones [51], and apps [57,64,66]. Targeted patient communication technologies [19,24,39,45] had the second highest frequency among patients. This category comprises personalized information that is delivered to individuals or groups of patients from health professionals and can be unidirectional (a message can only be sent by the health professional; patients do not have the option to reply) or bidirectional (patients can reply to messages from the health professional) [69]. It is important to note that the definition of this category proposed by the WHO is only limited to unidirectional communication from the health professional [34]. Some examples of targeted patient communication observed in this review included humanoid animated agents (computational artifacts used to develop human-like relationships with patients through the development of trust, rapport, and therapeutic alliance [45]) used for the purpose of computerized cognitive behavioral therapy (CBT) [45], email communication [39], and websites [24]. In all, 2 articles included examples of patient-to-patient communication via digital technology, including online peer-support groups and chatrooms [42,46]. Personal health tracking interventions included patient portals and EHRs, with the primary function of self-monitoring [41,48]. Untargeted patient communication (generalized communications distributed to a large patient population in which all recipients receive identical messages [34]) was the least common digital health intervention, consisting of a computerized CBT program with a singular set of responses generated for all users [58].

There were 22 cases of digital health interventions being used by health professionals including telemedicine, health professional training, and patient health records. Telemedicine (providing health care from a distance through the use of technology [34]) made up 78% (17/22) of all types of digital health interventions used by health professionals [20,26,35,36,38,40,44,47,50,54-56,59,62,6367]. In this review, telemedicine was observed through the use of videoconferencing [20,26,36,38,40,44,50,54,62,6367], apps [47], telephone communication [55], gaming [35,56], and virtual reality [20,59] to provide patient care. Digital technology was also commonly used among health professionals for training purposes through the use of virtual reality [53,60,61] and apps [49]. The last digital health intervention used among health professionals included in this review was the use of EHRs during patient appointments [43].

Table 2. Digital health interventions.

The World Health Organization classification of digital health interventions[a]	Frequency	References
1.0 Patients	**15**	**—[b]**
1.1 Targeted patient communication	4	[19,24,39,45]
1.2 Untargeted patient communication	1	[58]
1.3 Patient to patient communication	2	[42,46]
1.4 Personal health tracking	2	[41,48]
1.5 Citizen-based reporting	0	—
1.6 On-demand information services to patients	6	[51,52,57,64-66]
1.7 Patient financial transactions	0	—
2.0 Health Professionals	**22**	**—**
2.1 Patient identification and registration	0	—
2.2 Patient health records	1	[43]
2.3 Health professional decision support	0	—
2.4 Telemedicine	17	[20,26,35-38,40,44,47,50,54-56,59,62,63,67]
2.5 Health professional communication	0	—
2.6 Referral coordination	0	—
2.7 Health worker activity planning and scheduling	0	—
2.8 Health professional training	4	[49,53,60,61]
2.9 Prescription and medication management	0	—
2.10 Laboratory and diagnostics imaging management	0	—

[a] The WHO classification system terminology employs clients and health care providers; in the context of this review, patients and clients will be interchangeable as well as health care providers and health care professionals.

[b] No articles were identified.

Research Question 2: Delivery of Compassionate Mental Health Care Through and With Digital Technology

Owing to the subjectivity of the definition of compassionate care, a conceptual model titled *Digital Intersections with Compassionate Care*, shown in Figure 2 and definitions in Table 3, from the textbook chapter *Caring in a Digital Age: Exploring the Interface of Humans and Machines in the Provision of Compassionate Healthcare* [70] was used to understand and organize the unique roles of digital technology in the delivery of compassionate mental health care. This model illustrates the intersections between the 6 main components of compassionate care and digital technology [70]. An additional category was created for the purpose of this review to account for articles that proposed digital technology use in mental health care may detract from compassionate care (Table 4).

The digital intersections (definitions are included in Table 3) addressed in this review include numerous examples of online interventions, training and coaching, compassion-oriented technologies, and artificial emotional intelligence, as shown in Table 4.

Figure 2. Digital intersections with compassionate care. AEI: artificial emotional intelligence.

Table 3. Definitions of the digital intersections with compassionate care.

Digital intersection with compassionate care	Definition
Awareness of suffering	Developing an awareness of one's suffering through the use of digital technology (ie, experiences shared via digital technology increase awareness of one's suffering)
Mediated response	Utilizing digital technology to mediate or influence one's response to suffering
Online intervention	Responding to suffering through an online intervention
Training and coaching	Digital tools used to increase health professional expertise or patient knowledge to ensure the delivery of compassionate care (ie, through digital storytelling, online forums, and messaging systems used to share knowledge and experiences)
Compassion-oriented technologies	Digital technologies created specifically to assist in or facilitate the delivery of compassionate care
Artificial emotional intelligence	Artificial intelligence used to facilitate compassionate interactions with patients

Table 4. Digital intersections with compassionate care.

Digital intersection with compassionate care	Frequency	References
Awareness of suffering	0	—[a]
Mediated response	0	—
Online intervention	11	[19,20,36,38,40,42,50,57-59,63]
Training and coaching	8	[24,49,51-53,60,65,66]
Compassion-oriented technologies	14	[26,35,37,41,44,46-48,51,54,56,62,64,67]
Artificial emotional intelligence: algorithms of compassion	1	[45]
Detractions from compassionate care	3	[39,43,55]

[a] No articles were identified.

Online Intervention

The development of online interventions to respond to suffering (ie, responding to suffering was a direct goal or result of the intervention) was observed in 11 articles [19,20,36,38,40,42,50,57-59,63]; this included the use of online therapy programs [42,57,5863], virtual reality programs to portray lived experiences [20,59], email and instant messaging to respond to patient suffering [19,20], and most commonly, the use of videoconferencing for telemedicine [20,36,38,40,50].

Training and Coaching

Using digital technology to provide training and coaching to increase compassion demonstrates how health professionals can leverage digital technology to better understand the suffering experienced by their patients and thus respond appropriately. Virtual reality coaching was often used for health care professionals to experience simulated positive symptoms of schizophrenia and psychosis [53,61], as well as complex difficulties experienced by patients suffering from dementia [60]. Additionally, digital training and coaching was also used by patients to learn the skills and importance of mindfulness through mood tracking, tips for overall well-being, and scheduled reminders to encourage session completion, all of which were used to respond to suffering and improve care [49,52,65,66]. One article discussed digital technology used to provide training to veterans to increase their understanding of their mental health notes made accessible to them through a patient portal, reducing misinterpretations and improving provider-patient relationships [24].

Compassion-Oriented Technologies

Digital technologies that were classified as *compassion-oriented technologies* based on the conceptual model shown in Figure 2 were the most commonly cited. This digital intersection with compassionate care involves technologies that support compassionate care and are used by health professionals and/or patients, including uses such as shared gaming time between health professionals and patients to facilitate bonding time [35,56], mental health apps [26,47,64], and patient portals [41,48].

Artificial Emotional Intelligence

Artificial emotional intelligence use was infrequently documented in the delivery of compassionate mental health care but was observed in one instance through the use of humanoid animated agents as part of a computerized CBT program [45]. A humanoid animated agent simulates a face-to-face conversation and utilizes verbal and nonverbal social cues to form human-like relationships [45].

Detractions From Compassionate Care

A final category was created to distinguish articles that proposed digital technologies that may detract from compassionate mental health care. A total of 3 articles [39,43,55] were included in this category and included concerns regarding the effect of nonresponses (to email and instant messages) on patients [39], as well as claims from physicians who felt that relationships equivalent to those formed in-person simply could not be achieved through the use of digital technology [55].

Research Question 3: Facilitators of and Barriers to Compassionate Mental Health Care Delivery Through Digital Technology

All articles discussed multiple facilitators of and barriers to compassionate mental health care delivery through the use of digital technology, as shown in Table 5. Facilitators included feedback on social cues, training/education for health professionals, increased safety, multilevel participation, peer-support, improved accessibility, and optional anonymity. Barriers included limitations because of health professionals' perceptions and abilities, impersonal automated responses, lack of social cues, effect of non-responses, group size, computer use during patient encounters, poor quality of technology, and inappropriate uses of technology at various stages of illness.

Tables 6 and 7 compare functions of digital technologies, as identified in RQ3, and the digital health interventions that facilitate each function. The criterion for each category was based on the evidence provided in the articles included as part of the review; digital health interventions were only confirmed to facilitate a function if specifically mentioned in the literature. Any facilitated functions that were not applicable to a particular digital health intervention are indicated as N/A.

Table 5. Facilitators of and barriers to delivery of compassionate care through digital technology.

Facilitators and barriers	Support
Facilitators	
Picture-in-picture functionality	The ability to view oneself on screen while interacting with a patient via videoconferencing; can allow for the evaluation of one's own facial expressions and response to social cues [62].
Physical distance	Patients may feel more at ease when communicating with a health professional through technology from a distance, and it also provides the opportunity to titrate the experience of distance [44].
Training/education	Digital technology can be used for training of health professionals as well as to convey lived experiences [53,60].
Safe for providers	Digital technology can allow health professionals to provide care without safety concerns in settings such as prisons [38,59].
Multilevel participation	Some digital health interventions allow users to simply observe functionalities with no mandatory participation, allowing for an easier transition into care [46].
Social/peer-support connections	Technology connects users to people with shared experiences, creating feelings of understanding and connectedness [42,46].
Convenience/accessibility	On-demand use and the ability to reach rural and remote areas through the use of technology [36,42,50,54,62].
Increased privacy/anonymity	Technology can allow for increased privacy and, in some cases, complete anonymity; this may decrease feelings of judgement and reduce stigma for patients [46,64].
Barriers	
Health professionals' perceptions and abilities	Some health professionals are reluctant to integrate technology into patient care because of personal perceptions and abilities [24,44].
Impersonal	Users may receive similar resources from apps despite varying mental health concerns [42].
Lacking social cues	The use of email and instant messaging does not allow the user to convey or evaluate tone of voice or facial expressions [19,39,42].
Nonresponses	Patients may feel neglected because of nonresponses to emails and/or instant messages [19].
Group size	Online self-help groups comprising large numbers of users may decrease attentiveness to patient needs and detract from individual compassionate relationships [46].
Use of computers during patient encounters	Obstructive positioning of computers used by health professionals during a patient encounter may lead to disengagement and distraction [43].
Quality of technology	Issues with lagging, audio problems, and poor video quality can affect the patient experience during videoconferencing used to provide telemedicine [67].
Stage of illness	The use of some types of digital technology may only be appropriate for specific stages of illness or mental illnesses (ie, not appropriate for a crisis situation) [51].

Table 6. Comparison of facilitated function and associated digital health intervention.

Digital health intervention	Facilitated function				
	Evaluation of social cues	Physical distance	Personalized care	Training for health professionals	Increased safety
Telemedicine	✓	✓	✓	✕	✓
Personal health tracking	✕	✓	✓	✕	✓
Targeted client communication	✕	✓	✕	✕	✓
Untargeted client communication	✕	✓	✕	✕	✓
Client health records	✕	✓	✓	✕	N/A[a]
On-demand information services to clients	✕	✓	✓	✕	✓
Health care provider training	✕	✓	✓	✓	N/A
Client-to-client communication	✕	✓	✓	✕	N/A

[a] Not applicable.

Table 7. Comparison of facilitated function and associated digital health intervention.

Digital health intervention	Facilitated function				
	Multilevel participation	Risk of nonresponses	Peer support	Accessibility	Reduced stigma
Telemedicine	N/A[a]	N/A	✓	✓	✓
Personal health tracking	✓	N/A	N/A	✓	✓
Targeted client communication	✓	✓	N/A	✓	✓
Untargeted client communication	✓	N/A	N/A	✓	✓
Client health records	N/A	N/A	N/A	✓	✓
On-demand information services to clients	✓	N/A	N/A	✓	✓
Health care provider training	N/A	N/A	N/A	✓	✓
Client-to-client communication	✓	N/A	✓	✓	✓

[a] Not applicable.

Discussion

Digital Technology Use in Mental Health Care

This review sought to examine the relationship between the emerging use of digital technology and its effect on the delivery of compassionate care in a mental health context through 3 RQs. Implications are discussed as follows in light of the findings.

In addition to the primary findings that technologies are widely incorporated into mental health care, with an emphasis on health care delivery methods such as telemedicine (Table 3), the majority of digital technologies examined in the identified articles were not targeted toward a particular mental health diagnosis (Table 2). This finding may be because of the fact that some interventions commonly facilitated through digital technologies are applicable to multiple mental health diagnoses. For instance, while the use of CBT has typically been associated with the treatment of anxiety and depression, existing research has established that it can also be effectively tailored to treat other anxiety disorders (eg, phobias and panic disorder), schizophrenia, trauma-related disorders, and bipolar disorders [71]. A computerized CBT intervention would thus be classified as *unspecified* because users with a wide variety of needs may be able to access support and benefits through the same platform. However, emphasis should be put on the fact that such increased reach would not be possible without the delivery medium of digital technology. Similarly, the high prevalence of telemedicine use, observed in this review (Table 3), is also used as a medium to deliver varying types of mental health care rather than standing as a tailored intervention for a specific diagnosis in itself. The relatively high representation of mental health care delivery methods as opposed to specifically tailored mental health interventions for a diagnosis in relation to compassionate care may be an indication of the current infancy of the state of this area. Future research will be required to understand if the delivery of compassionate care through digital technologies varies depending on the mental health diagnosis of patients. Further, future research methodologies should include economic analysis to understand the return on investment of delivering compassionate care between mental health treatment needs.

Digital Technologies Enabling Compassionate Care

The evolution of digital technologies is fueling the emergence of new types of health interventions. Although the decrease of the *in-person* experience may have been associated with a reduction in compassionate care [39,43,55], there are instances where a long-distance delivery of mental health care provides an improved experience for both the health care professional and patient. This review was able to substantiate that compassion is often a core aspect of digital health delivery. In fact, these new modes of intervention enable novel enactments of compassion and means to teach or train health care professionals to provide compassionate health care which would not be previously possible without digital technology. For example, for individuals requiring mental health care in correctional facility settings, escorted transportation to a satellite care site or conducting care in a monitored, secure meeting area with physical barriers may hinder the ability to build a compassionate relationship [38]. Leveraging telemedicine in situations such as this can not only cut down on resource use but also provide a more comfortable environment for both the patient and the health care professional, without which it would be difficult to deliver compassionate care [38].

In addition to areas of opportunity for improved patient experiences, emerging tools are also enhancing health professional education in fostering the delivery of compassionate care in practice. For instance, Ozelie et al present an immersive virtual reality system which offers learning through shared experiences by providing insight into the experience of a person with schizophrenia through simulated hallucinations [53]. This initiative is greatly in line with existing research that demonstrates such access to lived experiences is a highly valued resource, as shared experiences are inherently different from simply speaking or hearing about the experiences of persons with mental illness [72]. Lived experience is foundational to building relationships with others in recovery, particularly in peer-delivered services [72]. Vividly experiencing even a small portion of their patients' experiences can allow health care professionals to better understand the patient perspective, contributing to their awareness of another's experience of suffering or need.

Concurrently, the use of virtual reality also potentially presents itself as a natural advancement in telemedicine. Moving beyond the limitations of a 2D computer screen, virtual reality can allow for a more *in-person* experience while still capitalizing on the benefits of long-distance care [12,20,37,61]. However, unique considerations in the delivery or enablement of compassionate care specifically through virtual reality remains an area for future exploration.

Digital Technologies Detracting From Compassionate Care

In all, 3 studies identified in this review depicted digital technologies as detracting from compassionate care [39,43,55]. The articles that did discuss this aspect focused on the use of provider-based technologies (based on the WHO Classification of Digital Technologies), and primarily gathered information from the provider perspective. A greater understanding of provider and patient differences in their experiences and perceptions surrounding the role of digital technology in health care is necessary to fully understand the role of digital technologies in contributing to compassionate care in practice.

Considerations for Digital Technology Implementation

Tables 6 and 7 present a summary of facilitators and barriers associated with each type of digital technology identified in this review. To our knowledge, this is the first review of its kind to appraise digital technologies in relation to compassionate care. Given the limited resources available at health system and organizational levels, investing in implementing a new digital technology can be a significantly resource-intensive undertaking. This summary can aid in the evaluation of digital technologies to ensure decision makers are investing in technologies that are aligned with organizational values and principles that relate to the provision of person-centered and compassionate care, and help to audit existing technologies in relation to delivering compassionate care.

Limitations

Owing to the nature of scoping reviews, the quality of each identified article was not assessed. Although every effort was made to ensure all articles which may involve compassion in mental health care were included, the subjective nature of compassion may mean some articles were not captured in this review. As both compassion and the intersection between digital technologies and compassion are relatively understudied fields, the models leveraged in this review to classify the types of technologies identified may not be the appropriate taxonomies of organization. Grey literature was also omitted in this review.

Future Steps

Ultimately, the successful use of digital technologies to facilitate compassionate mental health care requires health care organizations to invest the time and resources to leverage implementation science. In addition, health care professionals need to adapt to environmental and contextual factors to appropriately choose technologies to meet needs at the levels of patient, organizational, and population health needs. Future research should focus on expanded implementation of digital technologies in mental health care and identifying both technologies and specific settings where compassionate care would not be possible without digital technology. This information can then be used by digital technology developers and institutions to inform the creation and development of technologies that result in the best outcomes for both health care professionals and patients. Furthermore, identifying how to teach health care professionals to meaningfully use technologies in ways that convey compassionate care should be explored.

In addition, future knowledge translation plans include traditional techniques such as presenting at conferences and giving lectures to those practicing in the mental health field. Other plans include engaging practicing mental health professionals and students in a discussion about the topic to increase awareness of digital compassion.

Conclusions

This review's inquiry into the intersectionality between contemporary digital technology and compassion is a highly relevant and emerging topic, particularly in mental health care. The current state of digital technology in mental health care lends itself well to facilitate compassionate care delivery, particularly when used to serve patients who may not have had the chance to receive health care previously, or who may be uncomfortable or restricted in direct face-to-face interactions. Although there is still much to understand and uncover, health care organizations and professionals must consider the advantages and limitations of each type of digital technology for practice, particularly at this time where the discussion is only at its outset. As technology inevitably continues to diffuse throughout mental health care, these considerations alongside patient feedback will be instrumental to ensure that digital tools are, and continue to be, aligned with provider and patient needs. Ultimately, compassion and the integration of digital technology in mental health care should be seen as vital and complementary aspects of obtaining the best patient outcomes, as mediums to accentuate meaningful human connections rather than inanimate products of modern innovation.

Acknowledgments

This work was made possible through funding from the AMS Healthcare Phoenix Fellowship program. The authors would like to thank Alicia Morgan for supporting the screening of articles for this review, and to the various stakeholders at Western University, University of Toronto, Waypoint Centre for Mental Health Care, and the Centre for Addiction and Mental Health for their ideas, thoughts, and lively discussion regarding the findings of this review. Additional data access may be granted through reasonable request to the corresponding author.

Authors' Contributions

This work was first conceived by GS and DW. Article identification and screening was led by TZ, and data extraction was led by JK. TZ and JK co-led manuscript writing with significant contribution from all authors in editing and revisions, particularly from SS, AC, and DW in the Results and Discussion.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Review and Meta-Analysis—Scoping Review Checklist for Scoping Review.
[DOCX File , 15 KB - jmir_v22i3e16263_app1.docx]

References

1. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
2. O'Hanlon P, Aref-Adib G, Fonseca A, Lloyd-Evans B, Osborn D, Johnson S. Tomorrow's world: current developments in the therapeutic use of technology for psychosis. BJPsych Adv 2016 Sep;22(5):301-310. [doi: 10.1192/apt.bp.115.014654]
3. Torous J, Hsin H. Empowering the digital therapeutic relationship: virtual clinics for digital health interventions. NPJ Digit Med 2018;1:16 [FREE Full text] [doi: 10.1038/s41746-018-0028-2] [Medline: 31304301]
4. Torous J, Walker R. Leveraging digital health and machine learning toward reducing suicide-from panacea to practical tool. JAMA Psychiatry 2019 Jul 10:-. [doi: 10.1001/jamapsychiatry.2019.1231] [Medline: 31290952]
5. Tal A, Torous J. The digital mental health revolution: opportunities and risks. Psychiatr Rehabil J 2017 Sep;40(3):263-265. [doi: 10.1037/prj0000285] [Medline: 28891658]
6. McGinty K, Saeed S, Simmons S, Yildirim Y. Telepsychiatry and e-mental health services: potential for improving access to mental health care. Psychiatr Q 2006;77(4):335-342. [doi: 10.1007/s11126-006-9019-6] [Medline: 16927161]
7. CMA PolicyBase.: Canadian Medical Association; 2010. Health Care Transformation in Canada URL: https://policybase.cma.ca/documents/policypdf/PD10-05.PDF [accessed 2020-01-03]
8. Sharp S, McAllister M, Broadbent M. The vital blend of clinical competence and compassion: how patients experience person-centred care. Contemp Nurse 2016;52(2-3):300-312. [doi: 10.1080/10376178.2015.1020981] [Medline: 26077823]
9. McClelland LE, Gabriel AS, DePuccio MJ. Compassion practices, nurse well-being, and ambulatory patient experience ratings. Med Care 2018 Jan;56(1):4-10. [doi: 10.1097/MLR.0000000000000834] [Medline: 29112049]
10. Lown BA, Rosen J, Marttila J. An agenda for improving compassionate care: a survey shows about half of patients say such care is missing. Health Aff (Millwood) 2011 Sep;30(9):1772-1778. [doi: 10.1377/hlthaff.2011.0539] [Medline: 21900669]
11. Zamanzadeh V, Valizadeh L, Rahmani A, van der Cingel M, Ghafourifard M. Factors facilitating nurses to deliver compassionate care: a qualitative study. Scand J Caring Sci 2018 Mar;32(1):92-97. [doi: 10.1111/scs.12434] [Medline: 28156018]
12. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
13. Post SG, Ng LE, Fischel JE, Bennett M, Bily L, Chandran L, et al. Routine, empathic and compassionate patient care: definitions, development, obstacles, education and beneficiaries. J Eval Clin Pract 2014 Dec;20(6):872-880. [doi: 10.1111/jep.12243] [Medline: 25266564]
14. Swanson KM. Nursing as informed caring for the well-being of others. Image J Nurs Sch 1993;25(4):352-357. [doi: 10.1111/j.1547-5069.1993.tb00271.x] [Medline: 8288305]
15. Seppa¨la¨ EM, Simon-Thomas E, Brown SL, Worline MC, Cameron CD, Doty JR, editors. The Oxford Handbook of Compassion Science. New York: Oxford University Press; 2017.
16. Fotaki M. Why and how is compassion necessary to provide good quality healthcare? Int J Health Policy Manag 2015 Apr;4(4):199-201 [FREE Full text] [doi: 10.15171/ijhpm.2015.66] [Medline: 25844380]
17. Pfaff K, Markaki A. Compassionate collaborative care: an integrative review of quality indicators in end-of-life care. BMC Palliat Care 2017 Dec 1;16(1):65 [FREE Full text] [doi: 10.1186/s12904-017-0246-4] [Medline: 29191185]
18. Frampton SB, Guastello S, Lepore M. Compassion as the foundation of patient-centered care: the importance of compassion in action. J Comp Eff Res 2013 Sep;2(5):443-455. [doi: 10.2217/cer.13.54] [Medline: 24236742]
19. Hanford T. Finding words: the use of email in psychotherapy with a disorganised and dissociating client. In: Cundy L, editor. Love in the Age of the Internet: Attachment in the Digital Era. London: Routledge; 2014.
20. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 2;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]

21. Stevenson AC. Compassion and patient centred care. Aust Fam Physician 2002 Dec;31(12):1103-1106. [Medline: 12516513]
22. Jakimowicz S, Perry L, Lewis J. Insights on compassion and patient-centred nursing in intensive care: a constructivist grounded theory. J Clin Nurs 2018 Apr;27(7-8):1599-1611. [doi: 10.1111/jocn.14231] [Medline: 29266484]
23. Newman D, O'Reilly P, Lee SH, Kennedy C. Mental health service users' experiences of mental health care: an integrative literature review. J Psychiatr Ment Health Nurs 2015 Apr;22(3):171-182. [doi: 10.1111/jpm.12202] [Medline: 25707898]
24. Denneson LM, Pisciotta M, Hooker ER, Trevino A, Dobscha SK. 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]
25. Stevens S. Hearing voices: the experience of associate degree nursing students to an auditory hallucinations simulation. Int J Nurs Clin Pract 2018;5(1):289. [doi: 10.15344/2394-4978/2018/289]
26. Barak A, Grohol JM. Current and future trends in Internet-supported mental health interventions. J Technol Hum Serv 2011;29(3):155-196. [doi: 10.1080/15228835.2011.616939]
27. Mental Health Commission of Canada. Ottawa: Mental Health Commission of Canada; 2014. E-Mental Health in Canada: Transforming the Mental Health System Using Technology URL: https://www.mentalhealthcommission.ca/sites/default/files/MHCC_E-Mental_Health-Briefing_Document_ENG_0.pdf [accessed 2020-01-03]
28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
29. 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]
30. 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]
31. 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]
32. Strudwick G, Zhang T, Inglis F, Sockalingam S, Munnery M, Lo B, et al. Delivery of compassionate mental health care in a digital technology-driven age: protocol for a scoping review. BMJ Open 2019 Jul 23;9(7):e027989 [FREE Full text] [doi: 10.1136/bmjopen-2018-027989] [Medline: 31340962]
33. Covidence Systematic Review Software. 2016. URL: https://www.covidence.org/ [accessed 2020-01-08]
34. World Health Organization. 2018. Classification of Digital Health Interventions V1.0: A Shared Language to Describe the Uses of Digital Technology for Health URL: http://www.who.int/iris/handle/10665/260480 [accessed 2020-01-03]
35. Adal J. Bringing Play Therapy Into the New Millennium: Incorporating Video Game Technology Into Mental Health Treatment. Los Angeles: California School of Professional Psychology; 2013.
36. Frueh BC, Monnier J, Grubaugh AL, Elhai JD, Yim E, Knapp R. Therapist adherence and competence with manualized cognitive-behavioral therapy for PTSD delivered via videoconferencing technology. Behav Modif 2007 Nov;31(6):856-866. [doi: 10.1177/0145445507302125] [Medline: 17932240]
37. Aboujaoude E, Salame W, Naim L. Telemental health: a status update. World Psychiatry 2015 Jun;14(2):223-230 [FREE Full text] [doi: 10.1002/wps.20218] [Medline: 26043340]
38. Arndt R. Prison inmates access mental health services through telepsychiatry. Mod Healthc 2018;48(2):18-21 [FREE Full text]
39. Bhuvaneswar CG, Gutheil TG. E-mail and psychiatry: some psychotherapeutic and psychoanalytic perspectives. Am J Psychother 2008;62(3):241-261. [doi: 10.1176/appi.psychotherapy.2008.62.3.241] [Medline: 18846971]
40. Goldstein F, Glueck D. Developing rapport and therapeutic alliance during telemental health sessions with children and adolescents. J Child Adolesc Psychopharmacol 2016 Apr;26(3):204-211. [doi: 10.1089/cap.2015.0022] [Medline: 26491890]
41. Denneson LM, Cromer R, Williams HB, Pisciotta M, Dobscha SK. A qualitative analysis of how online access to mental health notes is changing clinician perceptions of power and the therapeutic relationship. J Med Internet Res 2017 Jun 14;19(6):e208 [FREE Full text] [doi: 10.2196/jmir.6915] [Medline: 28615152]
42. Helton D. Online therapeutic social service provision (Therap-pc): a state of the art review. J Technol Hum Serv 2003;21(4):17-36 [FREE Full text] [doi: 10.1300/J017v21n04_02]
43. Wiarda NR, McMinn MR, Peterson MA, Gregor JA. Use of technology for note taking and therapeutic alliance. Psychotherapy (Chic) 2014 Sep;51(3):443-446. [doi: 10.1037/a0035075] [Medline: 24377404]
44. Yellowlees P, Chan SR, Parish MB. The hybrid doctor-patient relationship in the age of technology - telepsychiatry consultations and the use of virtual space. Int Rev Psychiatry 2015;27(6):476-489. [doi: 10.3109/09540261.2015.1082987] [Medline: 26493089]
45. Bickmore T, Gruber A. Relational agents in clinical psychiatry. Harv Rev Psychiatry 2010;18(2):119-130. [doi: 10.3109/10673221003707538] [Medline: 20235777]
46. Colón Y, Stern S, LCSW-C. Chapter Ten - Counseling groups online: theory and framework. In: Kraus R, Stricker G, Speyer C, editors. Online Counseling: A Handbook for Mental Health Professionals. Second Edition. San Diego: Academic Press; 2011:183-202.

47. Imel ZE, Caperton DD, Tanana M, Atkins DC. Technology-enhanced human interaction in psychotherapy. J Couns Psychol 2017 Jul;64(4):385-393 [FREE Full text] [doi: 10.1037/cou0000213] [Medline: 28318277]
48. Parish MB, Fazio S, Chan S, Yellowlees PM. Managing psychiatrist-patient relationships in the digital age: a summary review of the impact of technology-enabled care on clinical processes and rapport. Curr Psychiatry Rep 2017 Oct 27;19(11):90. [doi: 10.1007/s11920-017-0839-x] [Medline: 29075951]
49. Wood AE, Prins A, Bush NE, Hsia JF, Bourn LE, Earley MD, et al. Reduction of burnout in mental health care providers using the provider resilience mobile application. Community Ment Health J 2017 May;53(4):452-459. [doi: 10.1007/s10597-016-0076-5] [Medline: 28070775]
50. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]
51. Berry N, Bucci S, Lobban F. Use of the internet and mobile phones for self-management of severe mental health problems: qualitative study of staff views. JMIR Ment Health 2017 Nov 1;4(4):e52 [FREE Full text] [doi: 10.2196/mental.8311] [Medline: 29092809]
52. Finlay-Jones A, Kane R, Rees C. Self-compassion online: a pilot study of an internet-based self-compassion cultivation program for psychology trainees. J Clin Psychol 2017 Jul;73(7):797-816. [doi: 10.1002/jclp.22375] [Medline: 27787877]
53. Ozelie R, Panfil P, Swiderski N, Walz E. Hearing voices simulation: impact on occupational therapy students. Open J Occup Ther 2018;6(4):1-11 [FREE Full text] [doi: 10.15453/2168-6408.1452]
54. Henry BW, Block DE, Ciesla JR, McGowan BA, Vozenilek JA. Clinician behaviors in telehealth care delivery: a systematic review. Adv Health Sci Educ Theory Pract 2017 Oct;22(4):869-888. [doi: 10.1007/s10459-016-9717-2] [Medline: 27696102]
55. Turner J, Brown J, Carpenter D. Telephone-based CBT and the therapeutic relationship: the views and experiences of IAPT practitioners in a low-intensity service. J Psychiatr Ment Health Nurs 2018 Jun;25(5-6):285-296. [doi: 10.1111/jpm.12440] [Medline: 29117458]
56. Yamaguchi H, Maki Y, Takahashi K. Rehabilitation for dementia using enjoyable video-sports games. Int Psychogeriatr 2011 May;23(4):674-676. [doi: 10.1017/S1041610210001912] [Medline: 20880429]
57. Berry K, Salter A, Morris R, James S, Bucci S. Assessing therapeutic alliance in the context of mhealth interventions for mental health problems: development of the mobile agnew relationship measure (mARM) questionnaire. J Med Internet Res 2018 Apr 19;20(4):e90 [FREE Full text] [doi: 10.2196/jmir.8252] [Medline: 29674307]
58. Gega L, Smith J, Reynolds S. Cognitive behaviour therapy (CBT) for depression by computer vs therapist: patient experiences and therapeutic processes. Psychother Res 2013;23(2):218-231. [doi: 10.1080/10503307.2013.766941] [Medline: 23390994]
59. van Rijn B, Cooper M, Jackson A, Wild C. Avatar-based therapy within prison settings: pilot evaluation. Br J Guid Counc 2017;45(3):268-283 [FREE Full text] [doi: 10.1080/03069885.2015.1068273]
60. Ball S, Bluteau P, Clouder D, Adefila A, Graham S. MyShoes: An Immersive Simulation of Dementia. In: Proceedings of the International Conference on e-Learning.: Academic Conferences Limited; 2015 Presented at: ICEL'15; June 25-26, 2015; Nassau, Bahamas p. 16-23.
61. Formosa NJ, Morrison BW, Hill G, Stone D. Testing the efficacy of a virtual reality-based simulation in enhancing users' knowledge, attitudes, and empathy relating to psychosis. Aust J Psychol 2018 Mar;70(1):57-65 [FREE Full text] [doi: 10.1111/ajpy.12167]
62. Simpson S, Reid C. Therapeutic alliance in videoconferencing psychotherapy: a review. Aust J Rural Health 2014 Dec;22(6):280-299. [doi: 10.1111/ajr.12149] [Medline: 25495622]
63. Capner M. Videoconferencing in the provision of psychological services at a distance. J Telemed Telecare 2000;6(6):311-319. [doi: 10.1258/1357633001935969] [Medline: 11265099]
64. 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]
65. Talbot F, Thériault J, French DJ. Self-Compassion: Evaluation of a Psychoeducational Website - CORRIGENDUM. Behav Cogn Psychother 2017 Mar;45(2):204. [doi: 10.1017/S1352465816000370] [Medline: 27480792]
66. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
67. Glueck D. Establishing therapeutic rapport in telemental health. In: Myers K, Turvey CL, editors. Telemental Health: Clinical, Technical, and Administrative Foundations for Evidence-Based Practice. London: Elsevier; 2013:29-46.
68. The Centre for Addiction and Mental Health. Toronto: Centre for Addiction and Mental Health; 2019. Mental Illness & Addiction Index URL: https://www.camh.ca/en/health-info/mental-illness-and-addiction-index [accessed 2020-01-08]
69. Sudore RL, Landefeld CS, Pérez-Stable EJ, Bibbins-Domingo K, Williams BA, Schillinger D. Unraveling the relationship between literacy, language proficiency, and patient-physician communication. Patient Educ Couns 2009 Jun;75(3):398-402 [FREE Full text] [doi: 10.1016/j.pec.2009.02.019] [Medline: 19442478]
70. Wiljer D, Strudwick G, Crawford A. Caring in a Digital Age: Exploring the interface of humans and machines in the provision of high quality healthcare. In: Hodges B, editor. Without Compassion…there is no healthcare. Toronto: Associated Medical Services; Forthcoming 2020.

71. The Centre for Addiction and Mental Health. Toronto: Centre for Addiction and Mental Health; 2019. Cognitive Behavioural Therapy URL: https://www.camh.ca/en/health-info/mental-illness-and-addiction-index/cognitive-behavioural-therapy [accessed 2019-07-18]
72. Marino CK, Child B, Krasinski VC. Sharing Experience Learned Firsthand (SELF): Self-disclosure of lived experience in mental health services and supports. Psychiatr Rehabil J 2016 Jun;39(2):154-160. [doi: 10.1037/prj0000171] [Medline: 26653777]

Abbreviations

AMS: Associated Medical Services
CBT: cognitive behavioral therapy
EHR: electronic health record
MEDLINE: Medical Literature Analysis and Retrieval System Online
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
RQ: research question
WHO: World Health Organization

Edited by G Eysenbach; submitted 15.09.19; peer-reviewed by T Risling, I Mircheva, A Younas; comments to author 14.11.19; revised version received 19.11.19; accepted 14.12.19; published 06.03.20.

Please cite as:
Kemp J, Zhang T, Inglis F, Wiljer D, Sockalingam S, Crawford A, Lo B, Charow R, Munnery M, Singh Takhar S, Strudwick G
Delivery of Compassionate Mental Health Care in a Digital Technology–Driven Age: Scoping Review
J Med Internet Res 2020;22(3):e16263
URL: https://www.jmir.org/2020/3/e16263
doi:10.2196/16263
PMID:32141833

Review

The Application of Internet-Based Sources for Public Health Surveillance (Infoveillance): Systematic Review

Joana M Barros[1,2], MSc; Jim Duggan[2], PhD; Dietrich Rebholz-Schuhmann[3], PhD

[1]Insight Centre for Data Analytics, National University of Ireland Galway, Galway, Ireland
[2]School of Computer Science, National University of Ireland Galway, Galway, Ireland
[3]ZB MED - Information Centre for Life Sciences, University Cologne, Cologne, Germany

Corresponding Author:
Joana M Barros, MSc
Insight Centre for Data Analytics
National University of Ireland Galway
The DERI Building
IDA Business Park, Lower Dangan
Galway, H91 AEX4
Ireland
Phone: 353 0838327458
Email: joana.barros@insight-centre.org

Abstract

Background: Public health surveillance is based on the continuous and systematic collection, analysis, and interpretation of data. This informs the development of early warning systems to monitor epidemics and documents the impact of intervention measures. The introduction of digital data sources, and specifically sources available on the internet, has impacted the field of public health surveillance. New opportunities enabled by the underlying availability and scale of internet-based sources (IBSs) have paved the way for novel approaches for disease surveillance, exploration of health communities, and the study of epidemic dynamics. This field and approach is also known as infodemiology or infoveillance.

Objective: This review aimed to assess research findings regarding the application of IBSs for public health surveillance (infodemiology or infoveillance). To achieve this, we have presented a comprehensive systematic literature review with a focus on these sources and their limitations, the diseases targeted, and commonly applied methods.

Methods: A systematic literature review was conducted targeting publications between 2012 and 2018 that leveraged IBSs for public health surveillance, outbreak forecasting, disease characterization, diagnosis prediction, content analysis, and health-topic identification. The search results were filtered according to previously defined inclusion and exclusion criteria.

Results: Spanning a total of 162 publications, we determined infectious diseases to be the preferred case study (108/162, 66.7%). Of the eight categories of IBSs (search queries, social media, news, discussion forums, websites, web encyclopedia, and online obituaries), search queries and social media were applied in 95.1% (154/162) of the reviewed publications. We also identified limitations in representativeness and biased user age groups, as well as high susceptibility to media events by search queries, social media, and web encyclopedias.

Conclusions: IBSs are a valuable proxy to study illnesses affecting the general population; however, it is important to characterize which diseases are best suited for the available sources; the literature shows that the level of engagement among online platforms can be a potential indicator. There is a necessity to understand the population's online behavior; in addition, the exploration of health information dissemination and its content is significantly unexplored. With this information, we can understand how the population communicates about illnesses online and, in the process, benefit public health.

(J Med Internet Res 2020;22(3):e13680) doi:10.2196/13680

KEYWORDS

medical informatics; public health informatics; public health; infectious diseases; chronic diseases; infodemiology; infoveillance

Introduction

Background

Public health is “the art and science of preventing disease, prolonging life and promoting health through the organized efforts of society” [1]. As a research and political field, it is focused on improving the quality of life of the population by identifying, suggesting, and applying prevention measures (eg, through the promotion of healthy behaviors) and health-related treatments [2]. Monitoring health is one important contribution to public health measures and involves the systematic collection, analysis, and interpretation of large amounts of health-related data. The key aim of public health surveillance is to design and guide interventions; in particular, (1) it serves as an early warning system for health emergencies (*epidemics*, ie, acute events), (2) it documents public health interventions and tracks their progress (ie, *monitoring health*), and (3) it monitors and clarifies the epidemiology of health problems, enabling the prioritization of information necessary for the formulation of health policy (ie, targeting chronic events) [3].

In the past, surveillance has been based on reports from health care workers constituting an active surveillance system when consistent and standardized reporting is in place [3,4]. However, this architecture is costly to maintain and involves significant delays between the moment of data capture to the time point of the first diagnosis, thus hampering any rapid or even immediate detection of outbreaks [5]. Instead of attempting to gather surveillance data from a network of health facilities and laboratories, health entities can employ a passive surveillance system in which hospitals, clinics, or other similar sources submit their respective health reports. This system provides an inexpensive way to monitor the community’s health; however, data quality is an issue owing to nonuniform reporting standards, and timeliness remains difficult to achieve [4]. To further complement these systems, syndromic surveillance was created to deal with the timeliness issue by using clinical (eg, emergency department admissions) and nonclinical sources (eg, over-the-counter drug sales), which are available before a diagnosis is confirmed [4]. This type of surveillance is based on the assumption that an outbreak would manifest itself as an anomaly in normal behavior [5]. In line with syndromic surveillance and with the growth of the internet, new opportunities for the detection of health-related information have arisen, with the potential to capture the patient’s input directly from the source. This leads to the ambitious endeavor of being able to monitor the health of a significant portion of the population at any point in time and at any geographical location, with the ultimate goal of monitoring public health.

The abovementioned technological advancements have enabled unofficial informal sources to currently provide more than 60% of epidemic reports [6]. Data analytics based on these data sources can provide near real-time outbreak information in various formats (independently from official governmental output) and have been successfully tested for health-related purposes. Furthermore, these sources offer the unique advantage of providing firsthand evidence for occurrences of health-related events (eg, through social media channels) and real-time informal reports (eg, news), which can be immediately investigated. Any analysis can be focused only on continuous monitoring, or by contrast to the identification of specific events (ie, single disease focus). In the latter case, it can be targeted to identify isolated hints (eg, mentions of flu) or to determine significant changes in public reporting; it can be further extended to consider the location of the population at risk or to monitor the distribution or extension of an epidemic (eg, influenced by the travelling population). The potential of data analytics applied to public data for health-related developments is ever more far-reaching in our increasingly digitally equipped society; thus, these approaches have an important role in the improvement of timeliness and sensitivity (ie, rapidly and correctly identifying health mentions) in public health surveillance [7].

Internet-Based Sources for Public Health Surveillance

IBSs are characterized by providing unstructured information from multiple origins and have proven to detect the first evidence of an outbreak, which is particularly beneficial for locations with a limited capacity for public health surveillance. The use of these sources for public health is also known as infodemiology or infoveillance. With the evidence provided by these sources, health agents are capable of mobilizing rapid response, reducing morbidity and mortality [8,9]. Some examples of IBSs include search queries, web encyclopedias, microblogs, and other social media.

Infectious diseases became the initial case study for the application of IBSs for disease surveillance. These continue to be a major cause of death in low-income countries [10], with research initially focusing on dengue, and are responsible for recurrent threats in the rest of the world (eg, *swine flu* and *bird flu*). Furthermore, these diseases are continuously monitored by official sources through laboratory tests or sentinel systems over many years and such information now forms the ground-truth data used to validate the findings from IBSs [11].

ProMED-mail is one of the first applications of such sources. This system is currently used for communication, via email and reports, among the infectious disease community [12]. Other systems include aggregators such as Global Public Health Intelligence Network, BioCaster, and HealthMap. These initially targeted a variety of sources including emails, Really Simple Syndication feeds, and PDF documents to extract information referring to an increased number of clusters of infected people at a specified time, period, or location, which could indicate a threat. The aggregator systems still in operation also include additional sources such as social media [7,13]. Moving to other sources, influenza-like illnesses (ILIs) served as a prototypical case study owing to being seasonal, worldwide, and well-reported diseases and initiated the monitoring of web-based queries. In particular, one of the first studies utilized Google search volumes to estimate the percentage of ILI-related physician visits [14]. This source was further adapted to the surveillance of other diseases such as dengue [15], gastroenteritis, and chickenpox [16]. This initial success led Google to develop targeted tools for the monitoring of influenza (Google Flu Trends) and dengue (Google Dengue Trends) in 2008 and 2011, respectively, which were later discontinued. Research continued and aimed to identify the most appropriate search terms to utilize as well as other search services (eg, Yahoo [17]) and other languages (eg, Vårdguiden [18]). Following search queries, microblogs, in particular, Twitter, showed to be another source of health information characterized by providing more descriptive information and potentially containing semistructured metadata (eg, location and gender) [19,20]. By filtering messages containing disease-related keywords, the frequency of disease mentions can be tracked and outbreaks can be identified as unusual spikes in the message frequency [21]. Similar to search queries, subsequent research focused on the identification of adequate keywords, as well as the identification of personal health messages, ie, containing a keyword relevant to the disease and describing a first-person infection case, among others [22,23]. With the use of more descriptive albeit semantically ambiguous data, the focus shifted to detecting true signals, ie, first-person occurrences of diseases. The application of IBSs continued to grow [24] in tandem with the addition of new sources such as Facebook [25], Instagram [26], and discussion forums [27]. Noncommunicable diseases (NCDs) are the cause of 71% of deaths globally, ranging from 37% in low-income countries to 88% in high-income countries, hence, internet-based surveillance focus has begun to also include NCDs [10]. In this case, emphasis was given to the online behavior of affected individuals, as well as to the content of the information present in the sources [28], with the goal to establish or improve health practices and support the dissemination of health information to address the needs of the population [29,30].

Owing to the unstructured nature and to the large volumes of data provided by these IBSs, tailoring of solutions, applications, and even tools for retrieving and filtering the content becomes vital for success. Subsequent automatic use of these methods then becomes the key step to monitor the internet sources continuously, and eventually to identify potential public health risks or, even better, risks to individual patients [31]. However, disease surveillance based on online sources must be used with caution. Automatic identification of disease events from web-based data streams has to cope with inherent biases, ie, false-positive events, introduced through geographic or cultural variability in language and reporting when compared with reliable traditional surveillance methods [32]. Furthermore, traditional epidemiological parameters (eg, attack rate) are often not available as a gold standard and thus limit the proper assessment of the applied methods [31].

Objectives

Our objective was to provide a discriminative assessment of the applications of IBSs for disease surveillance and their use as ground truth for future research. To achieve this, we have presented a comprehensive systematic literature review with a focus on IBSs and their limitations, the diseases targeted, and methods commonly applied for disease surveillance. Our research questions (RQs) were as follows:

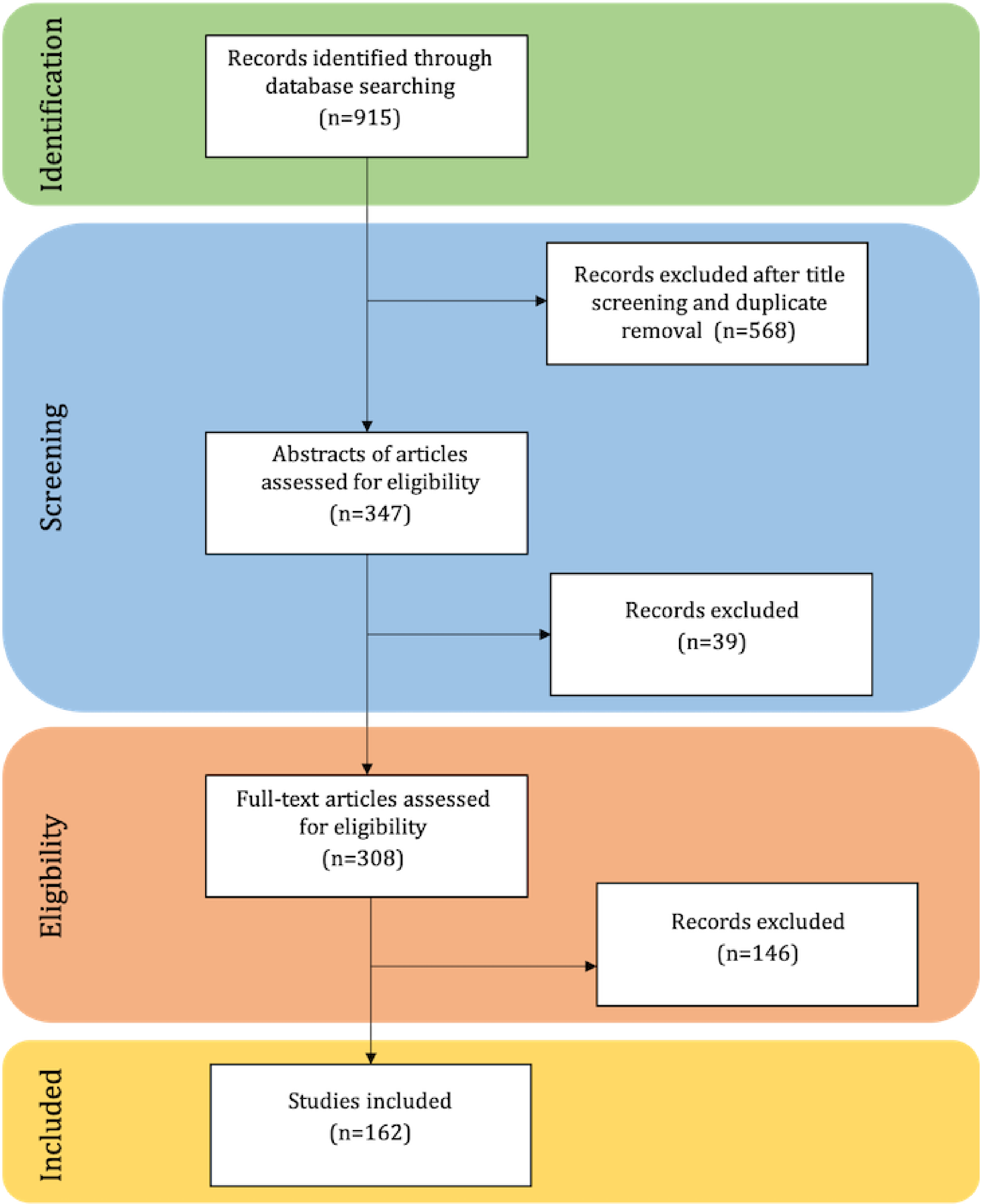
1. What internet-based sources are utilized for infoveillance and infodemiology?
2. What is the aim of the research conducted using these sources?
3. How are internet-based sources applied to generate knowledge?
4. Which sources have shown a preference for studying communicable and noncommunicable diseases?
5. What are the common limitations of internet-based sources?

Methods

Search Strategy

This review was conducted through several stages based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses process (Figure 1). To be the most inclusive, eight mutually exclusive research libraries that contain a variety of journals in the fields of Informatics and Biomedical Sciences were selected. The libraries were Europe PubMed Central, Institute of Electrical and Electronics Engineers Xplore Digital Library, Association for Computing Machinery Digital Library, SpringerLink, EBSCO Host, PubMed, Scopus, and Web of Science. Keyword generation was focused on IBSs of public health data, infoveillance, infodemiology, and disease outbreak and surveillance. We considered all conference and journal articles identified in this process.

Figure 1. Flowchart applied for the literature search.



Article Selection

The keywords were generated taking into consideration a preliminary assessment of the literature through a manual screening of relevant studies to ensure the list was complete. The complete list of these keywords can be found in Multimedia Appendix 1. The literature search was initially performed from October 10 to October 31, 2017, on the abovementioned repositories, focusing on the publication period of 2012 to 2017. This literature search was later augmented to include additional search terms and to extend the publication period until 2018. In total, the literature search had a duration of 2 months, excluding the screening and eligibility steps. Our review focused on the literature published after 2012 to cover a wider variety of sources, given the time lag between their popularity peaks; furthermore, by analyzing the literature published a few years after the first studies (eg, 2009 for search queries [14] and 2010 for social media [19]), we focused on research with finer-grained and adapted methodologies (eg, improved keyword selection and relevancy filters). To select relevant articles, a multiphase process was implemented. First, the article title was screened for relevance and duplicates were removed; subsequently the abstract was screened, and only articles that passed both steps were considered for the eligibility phase. The inclusion and exclusion criteria for the articles were decided according to a modified PICOTS. The criteria are specified in Multimedia Appendix 1. The first author performed the screening process and retrieved the data. When doubts were raised regarding the inclusion of certain publications, the remaining authors were consulted.

Quality Assessment

To address the quality of the studies, we implemented a set of criteria to evaluate the publications retrieved. This assessment was based on a set of questions with regard to the purpose of research, contextualization, methodology, study design, the results obtained, and findings. The quality criteria are based on the work by Kofod-Petersen [33] and are present in Multimedia Appendix 1.

Data Extraction

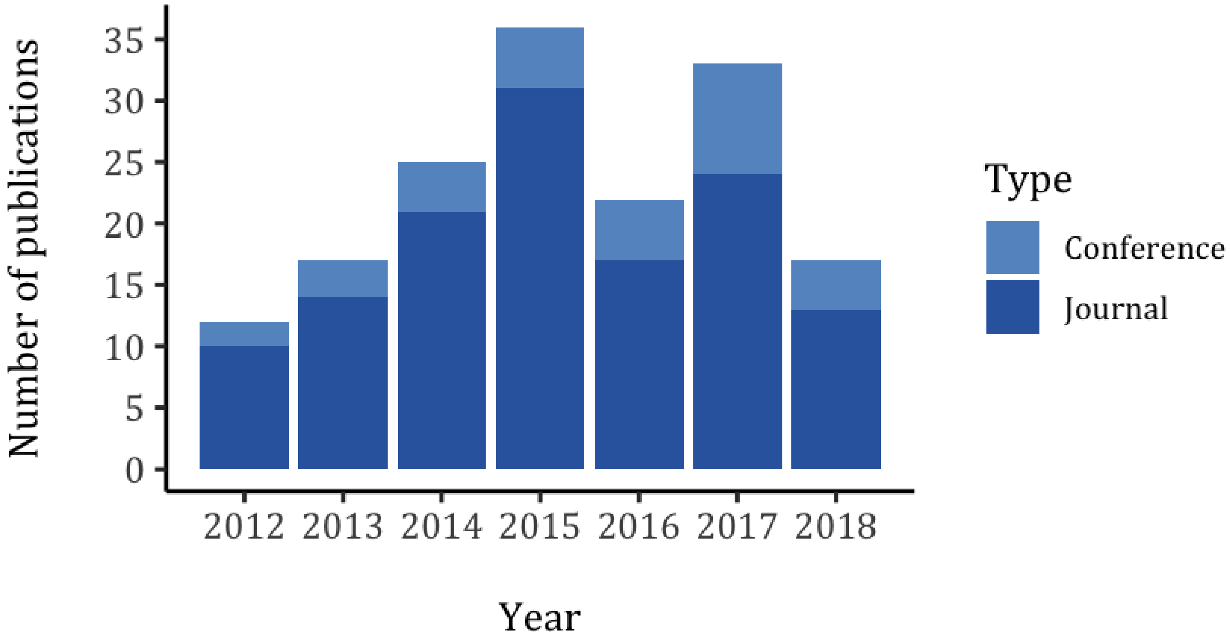
We also developed an extraction form to gather information about the studies allowing us to understand how the issues related to the proposed RQs have been addressed. This step was performed using the NVivo version 11 qualitative software database (QSR International Pty Ltd), nested *cases* were used to annotate each item of the extraction form. The extraction guidelines are available in Multimedia Appendix 1. For each checklist item in the guidelines, we created a classification that has been detailed in the following sections. Each paper was classified as *journal* or *conference*, in accordance with the inclusion criteria. Regarding the targeted diseases, we divided this into *chronic*, *infectious*, *medical conditions*, and *health topics*. The first three categories have been further specified. For the Goal/Objective item, a paper was classified as *outbreak forecasting* if it explicitly stated that the research was aimed at forecasting; else, it was assigned to *surveillance* (ie, when the purpose is only to identify the degree of correlation with ground-truth data and there is no mention of forecasting); *disease characterisation* was assigned when the aim was to determine identifiable characteristics related to a disease, eg, patient search behavior and commonly mentioned treatments, or when the aim was to classify a text as related to a disease; *content analysis* was assigned to the study of the sources' content (eg, news presence and expressed sentiment) referring to a disease or medical condition; *personal health mention classification* focused on separating general mentions of a disease or medical conditions (eg, news) from first-person mentions; and *diagnosis prediction* was assigned when the purpose was to attribute a disease or medical condition to a text and its creator by proxy. The *Internet-based data source* can be classified into search queries, social media (including microblogs), websites, news, discussion forums, web encyclopedia, and media monitoring systems. We also considered the use of data sources external to the IBSs, which can be classified as demographic, socioeconomic, and climate statistics, as well as data from governmental and laboratory sources, among others. To address the study design/methodology, we devised the following criteria: *topic analysis* corresponds to when topic modelling and similar approaches are used; *regression models* encompass all regression and autoregression models (eg, linear regression and autoregressive moving average); *statistical models* was assigned to more complex models (eg, Hidden Markov Chain); *correlation analysis* was used when correlation scores are calculated (eg, Pearson); *rule-based techniques* and *ranking techniques* are self-explanatory; *manual analysis* was assigned when no specific techniques are used other than a manual assessment; *epidemiology theory* refers to the use of techniques and measures commonly used in epidemiology (eg, Susceptible, Exposed, Infectious, and Recovered models); *linguistic analysis* was assigned when sentiment analysis and lexicons, among others, were used; and finally, we split *machine learning* and *deep learning*. Finally, we did not add a classification to the findings and limitations; we chose to keep this as an open field and manually analyzed the outcomes.

Results

Overview

The results from the search strategy are shown in Figure 1; in total, 162 papers were considered for this systematic literature review. The summary of the review results according to the data extraction guidelines is presented in Multimedia Appendix 1 [34-188]. The year with the highest number of publications is 2015 (n=36), followed by 2017 (n=33), 2014 (n=25), 2016 (n=22), 2013 (n=17), 2018 (n=17), and 2012 (n=12; Figure 2). Journal articles accounted for 130 of all publications and the remaining 32, for conferences. The remaining results were split into subsections correspondent to the extraction guidelines followed. The complete summary of the literature analysis is provided in Multimedia Appendix 1.

Figure 2. Distribution of the selected literature per year and type.



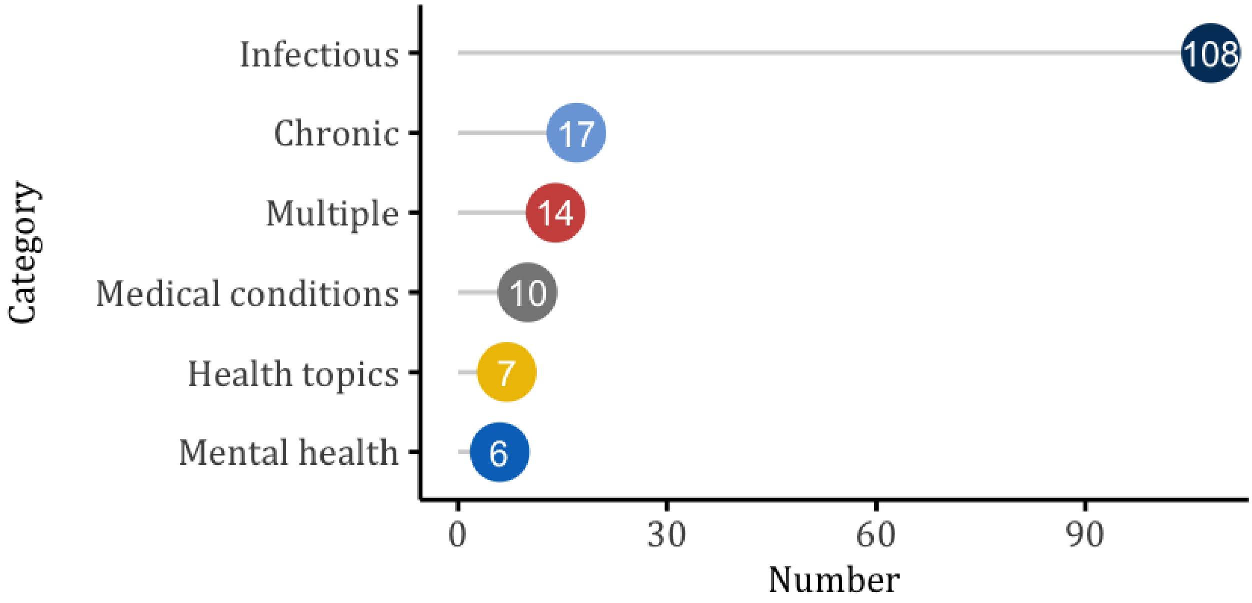
Goal

The analyzed papers mostly focused on surveillance (n=90), content analysis (n=46), and outbreak forecasting (n=45); other goals included personal health mention classification (n=10), disease characterization (n=5), and diagnosis prediction (n=4), with 36 papers having multiple targets.

Diseases, Medical Conditions, and Health Topics

Infectious diseases are markedly the most researched cases, with a total of 108 papers assigned. Chronic diseases are the focus of 17 publications, followed by medical conditions with 10 publications, health topics with 7 mentions, and mental health with 6 assigned articles (Figure 3). A set of 14 publications target multiple diseases from all the previously mentioned categories. Among the infectious diseases, ILIs, dengue, and infectious intestinal diseases are the top choices with 57, 7, and 7 assigned publications, respectively. In terms of chronic illnesses, cancer is the most researched disease (n=3). Excluding the publications focusing on multiple cases, 78% of the diseases appear in less than two articles.

Figure 3. Distribution of the case studies in the literature.



Internet-Based and External Data Sources

The 162 analyzed papers can be classified into eight distinct categories: search queries, social media, news, discussion forums, websites, media monitoring systems, web encyclopedia, and online obituaries. Social media (n=80) and search queries (n=79) are the most utilized IBSs, followed by web encyclopedias (n=13) that, in the selected papers, corresponded solely to Wikipedia. The remaining are utilized in the following decreasing order, forums (n=9), news (n=8), media monitoring systems (n=2), online obituaries (n=2), and websites not related to newspapers (n=1). A total of 29 papers utilize combinations of these sources, with the majority (n=11) combining search queries with social media. For social media, Twitter is mostly used (n=71) with the remaining sources marginally appearing. For search queries, the same behavior is seen with Google Trends; it is present in 42 publications, and when aggregating with Google Flu Trends and Google Dengue Trends, this value rises to 61.

Regarding sources external to IBSs, governmental, or laboratory surveillance statistics are the most utilized and used as ground-truth data (n=107), the second most used external source is hospital and emergency department visits (n=16), which are also used as ground-truth data. Climate or temperature statistics are applied in 8 papers, and socioeconomic statistics in 5 publications followed by health records in 4 publications, demographic or population statistics in 3 papers, and pharmaceutical sales in 2 publications. Scientific search engines, Flu Near You [189], and telephone triage are used individually in only 1 publication. In total, 45 publications do not use any external data source, and 26 publications share multiple external sources.

Study Design

Regarding the methodologies used, correlation analysis (n=59) was predominant and closely followed by regression models (n=46). Machine learning was used in 27 of the analyzed articles, statistical models are preferred in 20 publications, manual analysis was used in 18 of the articles, topic analysis is used in 12 publications, and deep learning and linguistic analysis were used in 10 articles each. Regarding the remaining methodology, rule-based techniques (n=7), epidemiology theory (n=6), surveys (n=3), and ranking techniques (n=1) were used in less than 10 papers.

Findings and Limitations

Qualitatively, the studies reported positive results (n=125), mentioning high or improved correlations with ground-truth data, as well as the outbreak predictive power, and high accuracy when the goal was surveillance, outbreak forecasting, personal health mention classification, disease characterization, and diagnosis prediction. The studies by Olson et al [55], Alicino et al [85], and Yom-tov [109] report negative results caused by questionable reliability with search query data and media influence affecting social networks and web encyclopedias. For the publications solely focused on content analysis (n=30), the findings were reported without a negative or positive association. A total of four publications [54,91,115,130] mention positive and negative results related to a large variation in the correlation score with social network data, surveillance inaccuracies for different age groups and the lack of specificity for search query data, and media influence when applying both social network and search query data.

In terms of limitations, these can be divided into gold standard (n=22); representativeness (n=76); general bias, eg, change in search behavior, symptom variability (n=9), and media effect (n=17); dataset size (n=7); methodology, eg, computational cost, keywords, and spelling errors (n=63); language (n=11); geographical restriction (n=33); and timeframe restriction (n=20).

Discussion

In this study, we aimed to provide a discriminative assessment on the application of IBSs in public health. To achieve this, we focused on the literature published in the last 6 years and applied systematic selection criteria to determine the appropriate studies to include. As a result, we proposed a taxonomy and identified the gaps to be addressed in future research, represented by the identified limitations of IBSs. Hence, this section addresses each RQ stated in the objectives.

Research Question 1: What Internet-Based Sources Are Utilized for Infoveillance and Infodemiology?

As reported in the Results, search queries, social media, discussion forums, news, web encyclopedia, online obituaries, media monitoring systems, and websites constitute the general categories of the IBSs present in the analyzed literature.

For search queries and social media, there is a large variation in the sources, which is mostly caused by geographical differences. The sources include platforms that are only available to certain countries. In the case of search queries, this potentially brings benefits in representativeness as it is possible to estimate the country-wide disease surveillance data from online search behavior. We argue that using a worldwide search engine, cultural differences that shape online search behavior could be diluted further complicating the surveillance task. Google Trends and its variations are the most common and widely represented; Bing also has an extensive geographical representation but a lower market share [190]. Also included are Baidu, Naver, Yandex, Vårdguiden and Websök, and Sapo, which cater to different countries, namely China, South Korea, Russia, Sweden, and Portugal, respectively. Nonetheless, the use of country-specific search engines can be limited by their market share, as is the case for Sapo [191], and their fine-grained geographical representation, eg, Vårdguiden is mostly used by people in Stockholm [68]. These limitations are further discussed in the following subsections.

Regarding social media, the sources differ in content richness. For example, while Twitter is a microblogging service, Weibo incorporates functionalities that can also be found on Facebook [192]. Nevertheless, the same reasoning applies, country-specific platforms can potentially bring benefits in representativeness and more closely estimate the country-wide health-related statistics.

The remaining sources, web encyclopedia and online obituaries, are used without a defined geographical restriction, and only English data were considered. Discussion forums and websites are an exception as they were utilized in different language-specific scenarios. Media monitoring systems also work on data from multiple sources and languages.

Research Question 2: What Is the Aim of the Research Conducted Using These Sources?

With IBSs of health information, the approaches are mostly based on monitoring the internet search and information-sharing behavior; the underlying assumption is that people actively seek and share information on diseases they develop.

In terms of surveillance and outbreak forecasting, estimates of disease activity within a community can be expressed by monitoring the frequency of related internet searches, disease mentions on social and news media, and page views in a web encyclopedia, among others. In addition, these sources also provide complementary information to the ground truth, eg, by targeting sick people who might not go to the hospital. When dealing with outbreak detection, an early and fast response is essential. Traditional surveillance is slower to transmit information across its different channels; therefore, IBSs complement the traditional mechanism when dealing with outbreaks [5,193,194].

Sources that go beyond single keywords pose a challenge as the occurrence of disease mentions does not correspond to the assumption that the text/health report in consideration is referring to the user suffering from the mentioned disease. For example, the microblog “Don’t forget to get your flu shot” is not as valuable as the microblog “I have the flu”; the latter corresponds to a personal health mention that has the potential to more closely correlate with gold standard data. Hence, personal health mention classification is based on the application of classification techniques that aim to filter false-positives, ie, a text containing a disease mention but not stating the user is carrying the disease, from true-positives, ie, a first person mentions of a disease by the affected user [37,72]. This is an important step that has been introduced when dealing with microblogs and online forums, as it has shown improvements for surveillance and outbreak forecasting.

Diagnosis prediction was not a common aim of the analyzed studies as it is difficult to validate owing to the lack of available gold standard data and owing to privacy concerns. The studies by De Choudhury et al [47] and Bodnar et al [64] include a prior user selection process for whom the authors have diagnosis information. In these cases, the source utilized was social media as it provides more contextual information and the potential for sentiment analysis, which is particularly valuable for mental health infodemiology studies. In contrast, the work by Karmen et al [97] targets the diagnosis of a health report (in the form of a forum post) and not the user itself (as not all information for the user is available) utilizing a similar methodology. Yom-tov et al [113], in their study, identify risk markers that correlate with a set of diseases based on the search behavior of assumed affected users.

When considering long-term patients, they also seek the internet for health information but also for online support through the connection with other patients [36,53,86,134]. This corresponds to the task of content analysis and disease characterization. IBSs are not only useful to perform disease surveillance but also to understand the information that is being shared online, which directly relates to public health tasks. The literature also points to the preference of sources for particular user groups, namely, users who seek support groups or connection with other patients and who suffer from chronic illnesses. In this situation, forums and social media, namely, microblogs, provide a suitable medium for the discussion of examination results, symptoms, treatments, and support, offering insights into how diseases are discussed online [36,41,72,86].

Research Question 3: How Are Internet-Based Sources Applied to Generate Knowledge?

As most of the publications aim to perform disease surveillance and outbreak forecasting, correlation analysis is regularly applied to determine the relationship between the IBSs and ground truth. Surveillance data are also commonly incorporated into surveillance and forecasting using regression models, which can also include autoregression, ie, past values of the ground-truth data. However, these methods make several assumptions regarding the distribution of the data, which might not be correct and overly simplistic.

Studies that utilize multiple sources of external data tend to apply more complex statistical methods which attempt to address the assumptions made by regression and autoregression models in trade of higher complexity.

The techniques mentioned earlier are applied to time-series data that can be obtained from the search query volume, page views from a web encyclopedia, and message/health report frequency in social media. In the latter case, to obtain the frequency, machine learning is commonly used to filter messages that are considered nonrelevant for the disease or medical condition in question. Thus, most of the machine learning approaches focus on social media and are reliant on annotated datasets, ie, a set of messages previously labelled as relevant or nonrelevant, which carry an added cost as this is necessary to train the models, as well as the lack of generalization as the labelled dataset targets a specific disease/medical condition.

Deep learning approaches improve generalization, they are capable of modelling complex nonlinear relationships, and do not impose restrictions on the data, eg, distribution; however, they have much higher complexity and can act as a black box owing to the high number of *tunable* hyperparameters. Furthermore, they require large sets of data that might not be available for diseases with a lower prevalence.

Topic analysis is mostly used for content analysis and it provides added benefits to manual analysis and surveys as it is unsupervised, ie, it does not require human input to perform the analysis. However, the topics identified might not be clearly related to a subject, which can lead to subjective interpretations; furthermore, it also carries high computational costs.

Linguistic analysis, in particular, sentiment analysis, can provide insights regarding negative and positive word use, among others, and how it associates with diseases. For example, this is used for mental health research as the sentiment expressed in words can be fundamental to detect the mental state of a user [82,157]. In the same category, named entity recognition aids in the detection of locations and disease names, among others [144]. Furthermore, the use of lexical and syntactic features and the use of lexicons have shown to improve classification tasks, eg, self-mentions of disease and disease-related categories [81,162].

The use of the epidemiology theory is not common as it can require data that are not available through the use of IBSs owing to its limits in terms of user information (eg, age and location); however, some studies have implemented various epidemiological models [96,121,140,170], as well as epidemiological parameters [34,119].

Rule-based techniques are manually created and are specific to the disease/medical condition studied; hence, they suffer from lack of generalization.

Ranking techniques were only used in 1 of the analyzed papers [195], and it was used to rank the topics generated from a topic modelling approach, suggesting that these can be used to facilitate the interpretation of the topic analysis results.

Research Question 4: Which Sources Have Shown a Preference for Studying Communicable and Noncommunicable Diseases, Health Topics, and Mental Health?

The nature of infectious diseases, ie, fast moving and with easier measurable effects, makes these a preferential case study for outbreak detection and surveillance. In tandem, the sources commonly applied for these tasks are search queries and social media, both combined and with other sources. These are the preferred sources as their output can be transformed into time-series data and compared with a gold standard for evaluation. With regard to search queries, a variety has been used to provide the most representative search behavior for the countries and languages targeted. A similar methodology was applied with social media, although mostly restricted to microblogs. Another important task in studying communicable illnesses is to explore what type of information is being shared and when; this is vital to identify the spread of misinformation and to analyze how far-reaching the counteractions are from health care agencies. Hence, content analysis is also performed by mainly utilizing social media as it provides more contextual information than search queries. Forums and news are utilized for the same reason; the higher contextual value allows for more insights into the study of information dissemination.

When discussing NCDs, monitoring and content analysis are the major approaches taken on the papers analyzed. Collecting epidemiological data for NCDs is a labor-intensive process [57,196]; hence, monitoring through digital sources aims primarily to estimate the number of affected individuals, given that official statistics are released with a significant delay [47,57,58,95,136]. To perform such a task, the commonly used sources are social media, search queries, and online obituaries. Content analysis focuses on determining the behavior and characteristics of users who actively mention a disease (eg, through a forum or social media), and the content and dissemination of health information. This is relevant as it allows to explore the information that is spread within these communities, such as personal medical advice. Additionally, past research has shown that online communities can provide a more convenient environment, for some patients, than traditional face-to-face interactions with health providers [36,41,43,82]. For content analysis, the sources applied are social media, forums, and web encyclopedia. As stated earlier, data with greater contextual content could provide more detailed information regarding online behavior and information exchange. With regard to forecasting, Gu et al [95] and Zhang et al [138] mention the task of predicting erythromelalgia-related hospital visits one week ahead and detecting early signals of diabetes, both through the use of search query data.

The research on mental health, medical conditions, and health topics focuses on content analysis; thus, the sources utilized in the analyzed papers refer to social media, in particular, microblogs, and forums.

Overall, the choice of the source of data is significantly related to the health topics, aim of the study, and the data available for evaluation. Infectious diseases have a large incidence variation; hence, they tend to have surveillance data available and most of the approaches focus on surveillance and outbreak forecasting which in itself requires sources that immediately show changes in the online behavior of users. Thus, search queries and microblogs are preferred for an analysis requiring a timely response. Regarding the remaining health topics, the population affected does not fluctuate as is the case with infectious diseases; hence, the focus is on the discussion that occurs online. It is more valuable to determine the information being disseminated, the questions raised online, and the needs of the patients so that the health agencies can cater to this segment of the population.

Research Question 5: What Are the Common Limitations of Internet-Based Sources?

Most of the studies analyzed report on positive outcomes when utilizing IBSs for public health applications; however, some limitations are frequently cited and only a few authors have given these greater importance. Although recent statistics illustrate the growth in search for health information online [197], internet access is neither equally distributed among countries nor equally penetrating in all regions within a country, which significantly affects the application of IBSs of health information [129,169]. In all the sources, common limitations refer to the lack of representativeness and bias caused by internet penetration and access, and preference to certain user age groups. For example, in the case of Twitter, 62% of its 330 million users are aged between 18 and 49 years [198]; around 53% of American internet users look up information on Wikipedia, with these users being mostly highly educated (69%) and under the age of 30 years (62%) [199]. This type of information elucidates on the potential bias caused by the nongeneralized use of these sources, in particular, when a given age group is more susceptible to a disease (eg, elderly and children).

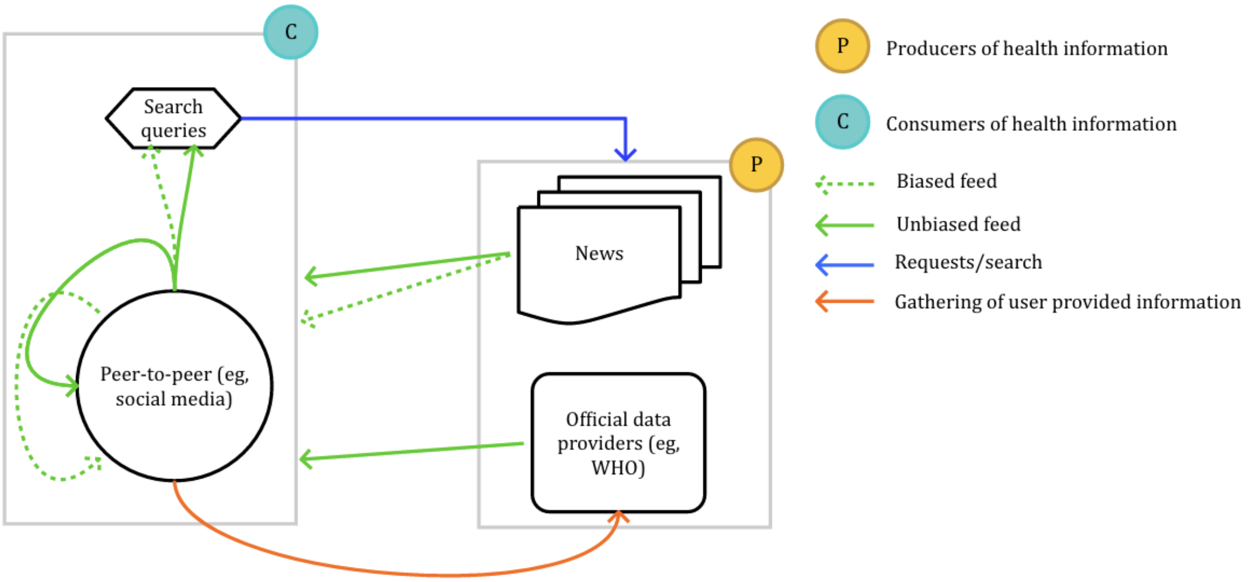
Another common limitation is related to precision issues caused by the inherent nature of diseases. These tend to share symptoms and treatments that are commonly used as keywords to detect relevant health reports, furthermore, the use of unspecific health-related terminology is also common. The layman language used in IBSs is a challenge, given that most approaches in the field rely on the use of keywords selected from specialized medical vocabulary, as summarized by Dai et al [142]. However, the evolution of learning-based, lexicon-based, and embedding approaches has started to mitigate the language specificity effect.

When ground-truth data are available, some studies question their quality as they can be updated after the initial publication; other mentioned limitations concern the amount of data as well as their timespan, and geographical coverage. Language restriction is also a common limitation of the studies, as most are performed only in English.

In particular to search queries and social media, the lack or limited geographical resolution is also cited as a limitation. Using Google Trends, various studies refer to the lack of transparency on how the search volumes are obtained, especially since forecasting systems based on Google Trends (ie, Google Flu Trends and Google Dengue Trends) have shown significant algorithmic problems that led to their termination [200,201]. The need for costly, manually annotated datasets is a common issue when the goal is to perform classification, and it mainly occurs with social media data.

In addition, a common limitation to search queries, social media, and web encyclopedia is the effect caused by media events. A media event is an event or activity conducted for media publicity. In this definition, we include examples such as panic-inducing news [109] and celebrities being diagnosed with medical conditions [173]. Media events have shown to significantly affect the reliability of these sources. The results of studies by Yom-tov [109] and Mollema et al [100] demonstrate how these sources show a higher correlation with media events than actual surveillance data. The study by Alicino et al [85] also reveals that the presence of news strongly influences the search volume in locations where an outbreak is not occurring. In light of this, we present an interaction schema in Figure 4; the sources of public health–related data predominantly comprise search engine queries, which target public sources, and peer-to-peer (P2P) social media networks; we can further distinguish primary consumers of health information, eg, members of P2P networks, from producers, eg, biased and unbiased news and unbiased official data providers (eg, governmental sources). News and official data providers deliver biased and unbiased information to the consumers. Consumers receive this information and spread it, affecting their search and share behavior, namely, in search queries, social media, and web encyclopedias.

Figure 4. Data sources interaction cycle. WHO: World Health Organization.



Conclusions

IBSs of health information are a valuable proxy to study illnesses affecting the population. Their benefits and applications are far-reaching and continue to evolve as a potential asset to public health. The knowledge gathered from this review suggests that search queries and social media provide useful data to monitor infectious diseases. With regard to studying chronic illnesses, discussion forums and social media are preferred.

The methods used to select relevant keywords or messages target specific illnesses, thus requiring constant updates to reflect the population's changing search behavior as well as emerging trends. Here, we identify the first research gap; disease outbreaks outside of the targeted disease names will not be identified, and new terminologies crucial for the detection of previously targeted diseases will be missed. Future approaches must focus on the ever-changing nature of diseases. For example, new related keywords could be identified through services such as Google Trends' *related topics*. To identify emerging illnesses, more emphasis must be given to the structure and syntax of messages describing a first-person mention of a disease, as this could be applicable to other illnesses.

The strong susceptibility to media events and the absence of approaches dealing with this issue constitute the second research gap. As shown in Figure 4, the interactions between the different sources and the type of information (biased and unbiased) reach the consumers and affect their search and share behavior, namely, in search queries, social media, and web encyclopedias. Such an effect must be mitigated to ensure improved reliability when utilizing IBSs.

The third research gap relates to the absence of consistent training and test periods, which impedes the appropriate comparison among the different methodologies. To address this, we suggest the creation of standard datasets, allowing to quantify the improvements of distinct methodologies. We also found a significant lack of interaction with public health officials, which would be the entities receiving the information from these models.

As a final recommendation, we suggest the use of an alternative strategy to better harness the information provided by IBSs. Namely, a proactive approach where the users are asked to report on their health state requesting the user to anonymously publish this information while avoiding the inclination to only publish positive messages. Such implementations can potentially make IBSs more comprehensible and a more valuable asset for disease monitoring.

Systematic Literature Review Limitations

This study makes use of eight databases, aiming to achieve a high coverage of the scientific literature. However, these databases do not guarantee full coverage and, hence, the inclusion of all relevant publications in our systematic methodology. In addition, we only considered articles in English as it is the predominant language of the scientific literature; thus, some contributions were potentially missed. The publication period is restricted to the last 6 years to allow for a focus on recent trends; earlier studies were referenced in the Introduction; however, not in an exhaustive way. We included a variety of keywords for the literature search although we understand that these might not cover all relevant publications.

Given that the authors followed a rigorous and systematic methodology when including and excluding publications for this literature review, selection bias was minimized. However, we cannot guarantee the absence of a bias when qualitatively presenting the findings; some categories and articles might be over- or under-represented.

Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2, cofunded by the European Regional Development Fund.

Authors' Contributions

JB originated the study, collected and analyzed data, and drafted the paper. JD and DR aided in the conceptualization of the study, article selection, and drafting of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supporting information for the systematic literature review (keywords, inclusion and exclusion criteria, quality assessment, data extraction checklist, literature synthesis/summary).
[DOCX File , 63 KB - jmir_v22i3e13680_app1.docx]

References

1. Acheson D. Public Health in England: The Report of the Committee of Inquiry into the Future Development to the Public Health Function. London: London Her Majesty's Stationery Office; 1988.
2. World Health Organization. 2018. Public Health Services URL: http://www.euro.who.int/en/health-topics/Health-systems/public-health-services [accessed 2018-08-27]
3. World Health Organization. 2018. Public Health Surveillance URL: http://www.who.int/topics/public_health_surveillance/en/ [accessed 2018-08-27]
4. Nsubuga P, White ME, Thacker SB, Anderson MA, Blount SB, Broome CV, et al. Public health surveillance: a tool for targeting and monitoring interventions. In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DV, et al, editors. Disease Control Priorities in Developing Countries. Second Edition. New York: Oxford University Press; 2006:997-1018.
5. Hope K, Durrheim DN, d'Espaignet ET, Dalton C. Syndromic Surveillance: is it a useful tool for local outbreak detection? J Epidemiol Community Health 2006 May;60(5):374-375 [FREE Full text] [doi: 10.1136/jech.2005.035337] [Medline: 16680907]
6. World Health Organization. 2018. Epidemic intelligence - systematic event detection URL: https://www.who.int/csr/alertresponse/epidemicintelligence/en/ [accessed 2018-08-27]
7. O'Shea J. Digital disease detection: a systematic review of event-based internet biosurveillance systems. Int J Med Inform 2017 May;101:15-22. [doi: 10.1016/j.ijmedinf.2017.01.019] [Medline: 28347443]

8. Wilson K, Brownstein JS. Early detection of disease outbreaks using the internet. Can Med Assoc J 2009 Apr 14;180(8):829-831 [FREE Full text] [doi: 10.1503/cmaj.090215] [Medline: 19364791]
9. Center for Disease Control and Prevention. 2003. Syndromic Surveillance: Reports from a National Conference URL: https://www.cdc.gov/mmwr/preview/su5301toc.htm [accessed 2020-01-23]
10. World Health Organization. 2018 May 24. The Top 10 Causes of Death URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2019-08-20]
11. Achrekar H, Gandhe A, Lazarus R, Yu S, Liu B, Ssu-Hsin Y, et al. Predicting Flu Trends using Twitter Data. In: Proceedings of the 2011 IEEE Conference on Computer Communications Workshops. 2011 Presented at: INFOCOM WKSHPS'11; April 10-15, 2011; Shanghai, China p. 702-707. [doi: 10.1109/infcomw.2011.5928903]
12. Madoff LC. ProMED-mail: an early warning system for emerging diseases. Clin Infect Dis 2004 Jul 15;39(2):227-232. [doi: 10.1086/422003] [Medline: 15307032]
13. Hartley DM, Nelson NP, Arthur RR, Barboza P, Collier N, Lightfoot N, et al. An overview of internet biosurveillance. Clin Microbiol Infect 2013 Nov;19(11):1006-1013 [FREE Full text] [doi: 10.1111/1469-0691.12273] [Medline: 23789639]
14. 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]
15. Runge-Ranzinger S, McCall PJ, Kroeger A, Horstick O. Dengue disease surveillance: an updated systematic literature review. Trop Med Int Health 2014 Sep;19(9):1116-1160 [FREE Full text] [doi: 10.1111/tmi.12333] [Medline: 24889501]
16. Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron AJ. More diseases tracked by using Google Trends. Emerg Infect Dis 2009 Aug;15(8):1327-1328 [FREE Full text] [doi: 10.3201/eid1508.090299] [Medline: 19751610]
17. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
18. Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. PLoS One 2009;4(2):e4378 [FREE Full text] [doi: 10.1371/journal.pone.0004378] [Medline: 19197389]
19. Culotta A. Towards Detecting Influenza Epidemics by Analyzing Twitter Messages. In: Proceedings of the First Workshop on Social Media Analytics. USA: ACM; 2010 Presented at: SOMA'10; July 25 - 28, 2010; Washington DC p. 115-122. [doi: 10.1145/1964858.1964874]
20. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013 Jul 18;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
21. Kostkova P, Szomszor M, St Louis C. #swineflu: the use of Twitter as an early warning and risk communication tool in the 2009 swine flu pandemic. ACM Trans Manage Inf Syst 2014;5(2):1-25. [doi: 10.1145/2597892]
22. Aramaki E, Maskawa S, Morita M. Twitter Catches the Flu: Detecting Influenza Epidemics Using Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: EMNLP'11; July 27–31, 2011; Edinburgh, United Kingdom p. 1568-1576.
23. Lamb A, Paul MJ, Dredze M. Separating Fact from Fear: Tracking Flu Infections on Twitter. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013 Presented at: NAACL'13; June 9-14, 2013; Atlanta, Georgia p. 789-795.
24. Gianfredi V, Bragazzi NL, Nucci D, Martini M, Rosselli R, Minelli L, et al. Harnessing big data for communicable tropical and sub-tropical disorders: implications from a systematic review of the literature. Front Public Health 2018;6:90 [FREE Full text] [doi: 10.3389/fpubh.2018.00090] [Medline: 29619364]
25. Gittelman S, Lange V, Crawford CA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. J Med Internet Res 2015 Apr 20;17(4):e98 [FREE Full text] [doi: 10.2196/jmir.3970] [Medline: 25895907]
26. Seltzer EK, Jean NS, Kramer-Golinkoff E, Asch DA, Merchant RM. 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]
27. Marcus MA, Westra HA, Eastwood JD, Barnes KL, Mobilizing Minds Research Group. What are young adults saying about mental health? An analysis of Internet blogs. J Med Internet Res 2012 Jan 30;14(1):e17 [FREE Full text] [doi: 10.2196/jmir.1868] [Medline: 22569642]
28. Paul MJ, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: ICWSM'11; July 17-21, 2011; Barcelona, Spain. [doi: 10.1145/2405716.2405728]
29. Roccetti M, Casari A, Marfia G. Inside Chronic Autoimmune Disease Communities: A Social Networks Perspective to Crohn's Patient Behavior and Medical Information. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. 2015 Presented at: ASONAM'15; August 25 - 28, 2015; Paris, France p. 1089-1096. [doi: 10.1145/2808797.2808813]
30. Sciascia S, Radin M. What can Google and Wikipedia can tell us about a disease? Big Data trends analysis in Systemic Lupus Erythematosus. Int J Med Inform 2017 Nov;107:65-69. [doi: 10.1016/j.ijmedinf.2017.09.002] [Medline: 29029693]

31. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Early Alerting Reporting Project Of The Global Health Security Initiative. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. PLoS One 2013;8(3):e57252 [FREE Full text] [doi: 10.1371/journal.pone.0057252] [Medline: 23472077]
32. Yang S, Santillana M, Brownstein JS, Gray J, Richardson S, Kou SC. Using electronic health records and internet search information for accurate influenza forecasting. BMC Infect Dis 2017 May 8;17(1):332 [FREE Full text] [doi: 10.1186/s12879-017-2424-7] [Medline: 28482810]
33. Kofod-Petersen A. Research Gate. 2014. How to Do a Structured Literature Review in Computer Science URL: https://www.researchgate.net/profile/Anders_Kofod-Petersen/publication/265158913_How_to_do_a_Structured_Literature_Review_in_computer_science/links/599a00350f7e9b3edb17cda2/How-to-do-a-Structured-Literature-Review-in-computer-science.pdf [accessed 2020-01-23]
34. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 2012 Jan;86(1):39-45 [FREE Full text] [doi: 10.4269/ajtmh.2012.11-0597] [Medline: 22232449]
35. 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]
36. Gold KJ, Boggs ME, Mugisha E, Palladino CL. Internet message boards for pregnancy loss: who's on-line and why? Womens Health Issues 2012;22(1):e67-e72 [FREE Full text] [doi: 10.1016/j.whi.2011.07.006] [Medline: 21907592]
37. Khan MA, Iwai M, Sezaki K. A Robust and Scalable Framework for Detecting Self-reported Illness from Twitter. In: 2012 IEEE 14th International Conference on e-Health Networking, Applications and Services. 2012 Presented at: Healthcom'12; October 10-13, 2012; Beijing, China p. 303-308. [doi: 10.1109/healthcom.2012.6379425]
38. Patwardhan A, Bilkovski R. Comparison: flu prescription sales data from a retail pharmacy in the US with Google Flu trends and US ILINet (CDC) data as flu activity indicator. PLoS One 2012;7(8):e43611 [FREE Full text] [doi: 10.1371/journal.pone.0043611] [Medline: 22952719]
39. Pervaiz F, Pervaiz M, Rehman NA, Saif U. FluBreaks: early epidemic detection from Google flu trends. J Med Internet Res 2012 Oct 4;14(5):e125 [FREE Full text] [doi: 10.2196/jmir.2102] [Medline: 23037553]
40. Samaras L, García-Barriocanal E, Sicilia M. Syndromic surveillance models using web data: the case of scarlet fever in the UK. Inform Health Soc Care 2012 Mar;37(2):106-124. [doi: 10.3109/17538157.2011.647934] [Medline: 22360741]
41. Sugawara Y, Narimatsu H, Hozawa A, Shao L, Otani K, Fukao A. Cancer patients on Twitter: a novel patient community on social media. BMC Res Notes 2012 Dec 27;5:699 [FREE Full text] [doi: 10.1186/1756-0500-5-699] [Medline: 23270426]
42. van Velsen L, van Gemert-Pijnen JE, Beaujean DJ, Wentzel J, van Steenbergen JE. Should health organizations use web 2.0 media in times of an infectious disease crisis? An in-depth qualitative study of citizens' information behavior during an EHEC outbreak. J Med Internet Res 2012 Dec 20;14(6):e181 [FREE Full text] [doi: 10.2196/jmir.2123] [Medline: 23257066]
43. Bosley JC, Zhao NW, Hill S, Shofer FS, Asch DA, Becker LB, et al. Decoding twitter: surveillance and trends for cardiac arrest and resuscitation communication. Resuscitation 2013 Feb;84(2):206-212 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.10.017] [Medline: 23108239]
44. Bragazzi NL. Infodemiology and infoveillance of multiple sclerosis in Italy. Mult Scler Int 2013;2013:924029 [FREE Full text] [doi: 10.1155/2013/924029] [Medline: 24027636]
45. 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]
46. Cho S, Sohn CH, Jo MW, Shin S, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. PLoS One 2013;8(12):e81422 [FREE Full text] [doi: 10.1371/journal.pone.0081422] [Medline: 24339927]
47. de Choudhury M, Counts S, Horvitz E. Social Media as a Measurement Tool of Depression in Populations. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: WebSci'13; May 2 - 4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]
48. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. PLoS One 2013;8(2):e56176 [FREE Full text] [doi: 10.1371/journal.pone.0056176] [Medline: 23457520]
49. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-like illness surveillance on Twitter through automated learning of naïve language. PLoS One 2013;8(12):e82489 [FREE Full text] [doi: 10.1371/journal.pone.0082489] [Medline: 24324799]
50. Ghosh DD, Guha R. What are we 'tweeting' about obesity? Mapping tweets with Topic Modeling and Geographic Information System. Cartogr Geogr Inf Sci 2013;40(2):90-102 [FREE Full text] [doi: 10.1080/15230406.2013.776210] [Medline: 25126022]
51. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]

52. Kim E, Seok JH, Oh JS, Lee HW, Kim KH. Use of Hangeul Twitter to track and predict human influenza infection. PLoS One 2013;8(7):e69305 [FREE Full text] [doi: 10.1371/journal.pone.0069305] [Medline: 23894447]
53. Lu 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]
54. Nagel AC, Tsou M, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. The complex relationship of realspace events and messages in cyberspace: case study of influenza and pertussis using tweets. J Med Internet Res 2013 Oct 24;15(10):e237 [FREE Full text] [doi: 10.2196/jmir.2705] [Medline: 24158773]
55. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol 2013;9(10):e1003256 [FREE Full text] [doi: 10.1371/journal.pcbi.1003256] [Medline: 24146603]
56. Parker J, Wei Y, Yates A, Frieder O, Goharian N. A Framework for Detecting Public Health Trends with Twitter. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2013 Presented at: ASONAM'13; August 25 - 28, 2013; Niagara Falls, ON, Canada p. 556-563. [doi: 10.1145/2492517.2492544]
57. Willard SD, Nguyen MM. Internet search trends analysis tools can provide real-time data on kidney stone disease in the United States. Urology 2013 Jan;81(1):37-42. [doi: 10.1016/j.urology.2011.04.024] [Medline: 21676450]
58. Won HH, Myung W, Song GY, Lee W, Kim JW, Carroll BJ, et al. Predicting national suicide numbers with social media data. PLoS One 2013;8(4):e61809 [FREE Full text] [doi: 10.1371/journal.pone.0061809] [Medline: 23630615]
59. Zheluk A, Quinn C, Hercz D, Gillespie JA. Internet search patterns of human immunodeficiency virus and the digital divide in the Russian Federation: infoveillance study. J Med Internet Res 2013 Nov 12;15(11):e256 [FREE Full text] [doi: 10.2196/jmir.2936] [Medline: 24220250]
60. Zhou X, Li Q, Zhu Z, Zhao H, Tang H, Feng Y. Monitoring epidemic alert levels by analyzing internet search volume. IEEE Trans Biomed Eng 2013 Feb;60(2):446-452. [doi: 10.1109/TBME.2012.2228264] [Medline: 23192470]
61. Andersson T, Bjelkmar P, Hulth A, Lindh J, Stenmark S, Widerström M. Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries and over-the-counter pharmacy sales. Epidemiol Infect 2014 Feb;142(2):303-313 [FREE Full text] [doi: 10.1017/S0950268813001088] [Medline: 23672877]
62. Araz OM, Bentley D, Muelleman RL. Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska. Am J Emerg Med 2014 Sep;32(9):1016-1023. [doi: 10.1016/j.ajem.2014.05.052] [Medline: 25037278]
63. 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 [FREE Full text] [doi: 10.2196/jmir.3532] [Medline: 25406040]
64. Bodnar T, Barclay VC, Ram N, Tucker CS, Salathé M. On the Ground Validation of Online Diagnosis With Twitter and Medical Records. In: Proceedings of the 23rd International Conference on World Wide Web. 2014 Presented at: WWW'14; April 7 - 11, 2014; Seoul, Korea p. 651-656. [doi: 10.1145/2567948.2579272]
65. Chen L, Hossain KS, Butler P, Ramakrishnan N, Prakash BA. Flu Gone Viral: Syndromic Surveillance of Flu on Twitter Using Temporal Topic Models. In: 2014 IEEE International Conference on Data Mining. 2014 Presented at: ICDM'14; December 14-17. 2014; Shenzhen, China p. 755-760. [doi: 10.1109/icdm.2014.137]
66. Culotta A. Estimating County Health Statistics with Twitter. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 26 - May 1, 2014; Toronto, Ontario p. 1335-1344. [doi: 10.1145/2556288.2557139]
67. Dubey D, Amritphale A, Sawhney A, Dubey D, Srivastav N. Analysis of YouTube as a source of information for West Nile virus infection. Clin Med Res 2014 Dec;12(3-4):129-132 [FREE Full text] [doi: 10.3121/cmr.2013.1194] [Medline: 24573700]
68. Edelstein M, Wallensten A, Zetterqvist I, Hulth A. Detecting the norovirus season in Sweden using search engine data--meeting the needs of hospital infection control teams. PLoS One 2014;9(6):e100309 [FREE Full text] [doi: 10.1371/journal.pone.0100309] [Medline: 24955857]
69. Generous N, Fairchild G, Deshpande A, del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. PLoS Comput Biol 2014 Nov;10(11):e1003892 [FREE Full text] [doi: 10.1371/journal.pcbi.1003892] [Medline: 25392913]
70. Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of internet-based dengue query data: Google Dengue Trends. PLoS Negl Trop Dis 2014 Feb;8(2):e2713 [FREE Full text] [doi: 10.1371/journal.pntd.0002713] [Medline: 24587465]
71. 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]
72. Ku Y, Chiu C, Zhang Y, Chen H, Su H. Text mining self-disclosing health information for public health service. J Assoc Inf Sci Technol 2014;65(5):928-947. [doi: 10.1002/asi.23025]

73. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol 2014 Apr;10(4):e1003581 [FREE Full text] [doi: 10.1371/journal.pcbi.1003581] [Medline: 24743682]
74. Milinovich GJ, Avril SM, Clements AC, Brownstein JS, Tong S, Hu W. Using internet search queries for infectious disease surveillance: screening diseases for suitability. BMC Infect Dis 2014 Dec 31;14:690 [FREE Full text] [doi: 10.1186/s12879-014-0690-1] [Medline: 25551277]
75. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J Med Internet Res 2014 Oct 20;16(10):e236 [FREE Full text] [doi: 10.2196/jmir.3416] [Medline: 25331122]
76. Seo D, Jo M, Sohn CH, Shin S, Lee J, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. J Med Internet Res 2014 Dec 16;16(12):e289 [FREE Full text] [doi: 10.2196/jmir.3680] [Medline: 25517353]
77. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
78. Prieto VM, Matos S, Álvarez M, Cacheda F, Oliveira JL. Twitter: a good place to detect health conditions. PLoS One 2014;9(1):e86191 [FREE Full text] [doi: 10.1371/journal.pone.0086191] [Medline: 24489699]
79. Santos JC, Matos S. Analysing Twitter and web queries for flu trend prediction. Theor Biol Med Model 2014 May 7;11(Suppl 1):S6 [FREE Full text] [doi: 10.1186/1742-4682-11-S1-S6] [Medline: 25077431]
80. Thompson LH, Malik MT, Gumel A, Strome T, Mahmud SM. Emergency department and 'Google flu trends' data as syndromic surveillance indicators for seasonal influenza. Epidemiol Infect 2014 Nov;142(11):2397-2405. [doi: 10.1017/S0950268813003464] [Medline: 24480399]
81. Velardi P, Stilo G, Tozzi AE, Gesualdo F. Twitter mining for fine-grained syndromic surveillance. Artif Intell Med 2014 Jul;61(3):153-163. [doi: 10.1016/j.artmed.2014.01.002] [Medline: 24613716]
82. Wilson ML, Ali S, Valstar MF. Finding Information about Mental Health in Microblogging Platforms: A Case Study of Depression. In: Proceedings of the 5th Information Interaction in Context Symposium. 2014 Presented at: IIiX'14; August 26 - 30, 2014; Regensburg, Germany p. 8-17. [doi: 10.1145/2637002.2637006]
83. Yom-Tov E, Borsa D, Cox IJ, McKendry RA. Detecting disease outbreaks in mass gatherings using internet data. J Med Internet Res 2014 Jun 18;16(6):e154 [FREE Full text] [doi: 10.2196/jmir.3156] [Medline: 24943128]
84. 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;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
85. 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 [FREE Full text] [doi: 10.1186/s40249-015-0090-9] [Medline: 26654247]
86. Betts D, Dahlen HG, Smith CA. A search for hope and understanding: an analysis of threatened miscarriage internet forums. Midwifery 2014 Jun;30(6):650-656. [doi: 10.1016/j.midw.2013.12.011] [Medline: 24439850]
87. Broniatowski DA, Dredze M, Paul MJ, Dugas A. Using social media to perform local influenza surveillance in an inner-city hospital: a retrospective observational study. JMIR Public Health Surveill 2015;1(1):e5 [FREE Full text] [doi: 10.2196/publichealth.4472] [Medline: 27014744]
88. Cleaton JM, Viboud C, Simonsen L, Hurtado AM, Chowell G. Characterizing Ebola transmission patterns based on internet news reports. Clin Infect Dis 2016 Jan 1;62(1):24-31 [FREE Full text] [doi: 10.1093/cid/civ748] [Medline: 26338786]
89. 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;19(7):1125-1132. [doi: 10.1007/s00779-015-0877-5]
90. Davidson MW, Haim DA, Radin JM. Using networks to combine 'big data' and traditional surveillance to improve influenza predictions. Sci Rep 2015 Jan 29;5:8154 [FREE Full text] [doi: 10.1038/srep08154] [Medline: 25634021]
91. Domnich A, Panatto D, Signori A, Lai PL, Gasparini R, Amicizia D. Age-related differences in the accuracy of web query-based predictions of influenza-like illness. PLoS One 2015;10(5):e0127754 [FREE Full text] [doi: 10.1371/journal.pone.0127754] [Medline: 26011418]
92. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
93. Fung IC, Hao Y, Cai J, Ying Y, Schaible BJ, Yu CM, et al. Chinese social media reaction to information about 42 notifiable infectious diseases. PLoS One 2015;10(5):e0126092 [FREE Full text] [doi: 10.1371/journal.pone.0126092] [Medline: 25946020]
94. Gesualdo F, Stilo G, D'Ambrosio A, Carloni E, Pandolfi E, Velardi P, et al. Can Twitter be a source of information on allergy? Correlation of pollen counts with tweets reporting symptoms of allergic rhinoconjunctivitis and names of antihistamine drugs. PLoS One 2015;10(7):e0133706 [FREE Full text] [doi: 10.1371/journal.pone.0133706] [Medline: 26197474]
95. Gu Y, Chen F, Liu T, Lv X, Shao Z, Lin H, et al. Early detection of an epidemic erythromelalgia outbreak using Baidu search data. Sci Rep 2015 Jul 28;5:12649 [FREE Full text] [doi: 10.1038/srep12649] [Medline: 26218589]

96. Hickmann KS, Fairchild G, Priedhorsky R, Generous N, Hyman JM, Deshpande A, et al. Forecasting the 2013-2014 influenza season using Wikipedia. PLoS Comput Biol 2015 May;11(5):e1004239 [FREE Full text] [doi: 10.1371/journal.pcbi.1004239] [Medline: 25974758]
97. Karmen C, Hsiung RC, Wetter T. Screening internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. Comput Methods Programs Biomed 2015 Jun;120(1):27-36. [doi: 10.1016/j.cmpb.2015.03.008] [Medline: 25891366]
98. Koopman B, Karimi S, Nguyen A, McGuire R, Muscatello D, Kemp M, et al. Automatic classification of diseases from free-text death certificates for real-time surveillance. BMC Med Inform Decis Mak 2015 Jul 15;15:53 [FREE Full text] [doi: 10.1186/s12911-015-0174-2] [Medline: 26174442]
99. Lee K, Agrawal A, Choudhary A. Mining Social Media Streams to Improve Public Health Allergy Surveillance. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. 2015 Presented at: ASONAM'15; August 25 - 28, 2015; Paris, France p. 815-822. [doi: 10.1145/2808797.2808896]
100. Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, de Melker H, Paulussen T, et al. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013. J Med Internet Res 2015 May 26;17(5):e128 [FREE Full text] [doi: 10.2196/jmir.3863] [Medline: 26013683]
101. 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]
102. Parker J, Yates A, Goharian N, Frieder O. Health-related hypothesis generation using social media data. Soc Netw Anal Min 2015;5:7. [doi: 10.1007/s13278-014-0239-8]
103. Pathak R, Poudel DR, Karmacharya P, Pathak A, Aryal MR, Mahmood M, et al. YouTube as a source of information on Ebola virus disease. N Am J Med Sci 2015 Jul;7(7):306-309 [FREE Full text] [doi: 10.4103/1947-2714.161244] [Medline: 26258077]
104. Pollett S, Wood N, Boscardin WJ, Bengtsson H, Schwarcz S, Harriman K, et al. Validating the use of Google Trends to enhance pertussis surveillance in California. PLoS Curr 2015 Oct 19;7 [FREE Full text] [doi: 10.1371/currents.outbreaks.7119696b3e7523faa4543faac87c56c2] [Medline: 26543674]
105. Ram S, Zhang W, Williams M, Pengetnze Y. Predicting asthma-related emergency department visits using big data. IEEE J Biomed Health Inform 2015 Jul;19(4):1216-1223. [doi: 10.1109/JBHI.2015.2404829] [Medline: 25706935]
106. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 2015 Oct;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
107. Sueki H. The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan. J Affect Disord 2015 Jan 1;170:155-160. [doi: 10.1016/j.jad.2014.08.047] [Medline: 25240843]
108. Thangarajan N, Green N, Gupta A, Little S, Weibel N. Analyzing Social Media to Characterize Local HIV At-Risk Populations. In: Proceedings of the conference on Wireless Health. 2015 Presented at: WH'15; October 14 - 16, 2015; Bethesda, Maryland p. 1-8. [doi: 10.1145/2811780.2811923]
109. Yom-tov E. Ebola data from the Internet: An Opportunity for Syndromic Surveillance or a News Event? In: Proceedings of the 5th International Conference on Digital Health 2015. 2015 Presented at: DH'15; May 18 - 20, 2015; Florence, Italy p. 115-119. [doi: 10.1145/2750511.2750512]
110. Wang H, Chen D, Yu H, Chen Y. Forecasting the incidence of dementia and dementia-related outpatient visits with Google Trends: evidence from Taiwan. J Med Internet Res 2015 Nov 19;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516] [Medline: 26586281]
111. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci U S A 2015 Nov 24;112(47):14473-14478 [FREE Full text] [doi: 10.1073/pnas.1515373112] [Medline: 26553980]
112. Yin Z, Fabbri D, Rosenbloom ST, Malin B. A scalable framework to detect personal health mentions on Twitter. J Med Internet Res 2015 Jun 5;17(6):e138 [FREE Full text] [doi: 10.2196/jmir.4305] [Medline: 26048075]
113. Yom-Tov E, Borsa D, Hayward AC, McKendry RA, Cox IJ. Automatic identification of web-based risk markers for health events. J Med Internet Res 2015 Jan 27;17(1):e29 [FREE Full text] [doi: 10.2196/jmir.4082] [Medline: 25626480]
114. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. PLoS One 2013;8(5):e64323 [FREE Full text] [doi: 10.1371/journal.pone.0064323] [Medline: 23750192]
115. Zaldumbide J, Sinnott RO. Identification and Validation of Real-Time Health Events through Social Media. In: Proceedings of the 2015 IEEE International Conference on Data Science and Data Intensive Systems. 2015 Presented at: DSDIS'15; December 11-13, 2015; Sydney, Australia. [doi: 10.1109/dsdis.2015.27]
116. Zhang EX, Yang Y, di Shang R, Simons JJ, Quek BK, Yin XF, et al. Leveraging social networking sites for disease surveillance and public sensing: the case of the 2013 avian influenza A(H7N9) outbreak in China. Western Pac Surveill Response J 2015;6(2):66-72 [FREE Full text] [doi: 10.5365/WPSAR.2015.6.1.013] [Medline: 26306219]
117. Zuccon G, Khanna S, Nguyen A, Boyle J, Hamlet M, Cameron M. Automatic detection of tweets reporting cases of influenza like illnesses in Australia. Health Inf Sci Syst 2015;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S4 [FREE Full text] [doi: 10.1186/2047-2501-3-S1-S4] [Medline: 25870759]

118. Allen C, Tsou M, Aslam A, Nagel A, Gawron J. Applying GIS and machine learning methods to Twitter data for multiscale surveillance of influenza. PLoS One 2016;11(7):e0157734 [FREE Full text] [doi: 10.1371/journal.pone.0157734] [Medline: 27455108]
119. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. Proc Natl Acad Sci U S A 2016 Jun 14;113(24):6689-6694 [FREE Full text] [doi: 10.1073/pnas.1523941113] [Medline: 27247405]
120. Byrd K, Mansurov A, Baysal O. Mining Twitter Data for Influenza Detection and Surveillance. In: Proceedings of the International Workshop on Software Engineering in Healthcare Systems. 2016 Presented at: SEHS'16; May 14 - 22, 2016; Austin, Texas p. 43-49. [doi: 10.1145/2897683.2897693]
121. Chen L, Hossain KS, Butler P, Ramakrishnan N, Prakash BA. Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models. Data Min Knowl Discov 2016;30(3):681-710. [doi: 10.1007/s10618-015-0434-x]
122. Deiner MS, Lietman TM, McLeod SD, Chodosh J, Porco TC. Surveillance tools emerging from search engines and social media data for determining eye disease patterns. JAMA Ophthalmol 2016 Sep 1;134(9):1024-1030 [FREE Full text] [doi: 10.1001/jamaophthalmol.2016.2267] [Medline: 27416554]
123. Gosh S, Chakraborty P, Cohn E, Brownstein JS, Ramakrishnan N. Characterizing Diseases from Unstructured Text: A Vocabulary Driven Word2vec Approach. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. 2016 Presented at: CIKM'16; October 24 - 28, 2016; Indianapolis, Indiana p. 1129-1138. [doi: 10.1145/2983323.2983362]
124. Ji X, Chun SA, Geller J. Knowledge-based tweet classification for disease sentiment monitoring. In: Sentiment Analysis and Ontology Engineering. Cham: Springer; 2016:425-454.
125. 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]
126. Ling R, Lee J. Disease monitoring and health campaign evaluation using Google search activities for HIV and AIDS, stroke, colorectal cancer, and marijuana use in Canada: a retrospective observational study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]
127. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015-2016 Colombian Zika virus disease outbreak. JMIR Public Health Surveill 2016 Jun 1;2(1):e30 [FREE Full text] [doi: 10.2196/publichealth.5814] [Medline: 27251981]
128. Martin LJ, Lee BE, Yasui Y. Google Flu Trends in Canada: a comparison of digital disease surveillance data with physician consultations and respiratory virus surveillance data, 2010-2014. Epidemiol Infect 2016 Jan;144(2):325-332. [doi: 10.1017/S0950268815001478] [Medline: 26135239]
129. Moss R, Zarebski A, Dawson P, McCaw JM. Forecasting influenza outbreak dynamics in Melbourne from internet search query surveillance data. Influenza Other Respir Viruses 2016 Jul;10(4):314-323 [FREE Full text] [doi: 10.1111/irv.12376] [Medline: 26859411]
130. Pollett S, Boscardin WJ, Azziz-Baumgartner E, Tinoco YO, Soto G, Romero C, et al. Evaluating Google Flu Trends in Latin America: important lessons for the next phase of digital disease detection. Clin Infect Dis 2017 Jan 1;64(1):34-41 [FREE Full text] [doi: 10.1093/cid/ciw657] [Medline: 27678084]
131. Priest C, Knopf A, Groves D, Carpenter JS, Furrey C, Krishnan A, et al. Finding the patient's voice using Big Data: analysis of users' health-related concerns in the ChaCha Question-and-Answer Service (2009-2012). J Med Internet Res 2016 Mar 9;18(3):e44 [FREE Full text] [doi: 10.2196/jmir.5033] [Medline: 26960745]
132. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in Twitter: a quantitative analysis. JMIR Public Health Surveill 2016;2(1):e1 [FREE Full text] [doi: 10.2196/publichealth.5059] [Medline: 27227144]
133. Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: a comparative analysis. JMIR Public Health Surveill 2016 Oct 20;2(2):e161 [FREE Full text] [doi: 10.2196/publichealth.5901] [Medline: 27765731]
134. Shin SY, Kim T, Seo DW, Sohn CH, Kim S, Ryoo SM, et al. Correlation between National Influenza Surveillance Data and Search Queries from Mobile Devices and Desktops in South Korea. PLoS One 2016;11(7):e0158539 [FREE Full text] [doi: 10.1371/journal.pone.0158539] [Medline: 27391028]
135. Sidana S, Amer-Yahia S, Clausel M, Rebai M, Mai S, Amini M. Health Monitoring on Social Media over Time. IEEE Trans Knowl Data Eng 2018 Aug;30(8):1467-1480. [doi: 10.1109/TKDE.2018.2795606]
136. Tourassi G, Yoon HJ, Xu S. A novel web informatics approach for automated surveillance of cancer mortality trends. J Biomed Inform 2016 Jun;61:110-118 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.027] [Medline: 27044930]
137. 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 4;18(7):e177 [FREE Full text] [doi: 10.2196/jmir.4955] [Medline: 27377323]

138. Zhang W, Ram S, Burkart M, Pengetnze Y. Extracting Signals from Social Media for Chronic Disease Surveillance. In: Proceedings of the 6th International Conference on Digital Health Conference. 2016 Presented at: DH'16; April 11 - 13, 2016; Montreal, Canada p. 79-83. [doi: 10.1145/2896338.2897728]
139. Zou B, Gorton R, Cox IJ, Lampos V. On Infectious Intestinal Disease Surveillance using Social Media Content. In: Proceedings of the 6th International Conference on Digital Health Conference. 2016 Presented at: DH'16; April 11 - 13, 2016; Montreal, Canada p. 157-161. [doi: 10.1145/2896338.2896372]
140. Albinati J, Meira W, Pappa GL, Teixeira M, Marques-toledo C. Enhancement of Epidemiological Models for Dengue Fever Based on Twitter Data. In: Proceedings of the 2017 International Conference on Digital Health. 2017 Presented at: DH'17; July 2 - 5, 2017; London, United Kingdom p. 109-118. [doi: 10.1145/3079452.3079464]
141. Chiu APY, Lin Q, He D. News trends and web search query of HIV/AIDS in Hong Kong. PLoS One 2017;12(9):e0185004 [FREE Full text] [doi: 10.1371/journal.pone.0185004] [Medline: 28922376]
142. Dai X, Bikdash M, Meyer B. From Social Media to Public Health Surveillance: Word Embedding Based Clustering Method for Twitter Classification. In: Proceedings of the IEEE SoutheastCon 2017. 2017 Presented at: SoutheastCon'17; March 30 - April 2, 2017; Charlotte, NC, USA. [doi: 10.1109/secon.2017.7925400]
143. Du Z, Xu L, Zhang W, Zhang D, Yu S, Hao Y. Predicting the hand, foot, and mouth disease incidence using search engine query data and climate variables: an ecological study in Guangdong, China. BMJ Open 2017 Oct 6;7(10):e016263 [FREE Full text] [doi: 10.1136/bmjopen-2017-016263] [Medline: 28988169]
144. Fairchild G, del Valle SY, de Silva L, Segre A. Eliciting disease data from Wikipedia articles. Proc Int AAAI Conf Weblogs Soc Media 2015 May;2015:26-33 [FREE Full text] [doi: 10.5210/ojphi.v8i1.6526] [Medline: 28721308]
145. Fung IC, Zeng J, Chan C, Liang H, Yin J, Liu Z, et al. Twitter and Middle East respiratory syndrome, South Korea, 2015: A multi-lingual study. Infect Dis Health 2018 Mar;23(1):10-16. [doi: 10.1016/j.idh.2017.08.005] [Medline: 30479298]
146. Ghosh S, Chakraborty P, Nsoesie EO, Cohn E, Mekaru SR, Brownstein JS, et al. Temporal topic modeling to assess associations between news trends and infectious disease outbreaks. Sci Rep 2017 Jan 19;7:40841 [FREE Full text] [doi: 10.1038/srep40841] [Medline: 28102319]
147. Guo P, Zhang J, Wang L, Yang S, Luo G, Deng C, et al. Monitoring seasonal influenza epidemics by using internet search data with an ensemble penalized regression model. Sci Rep 2017 Apr 19;7:46469 [FREE Full text] [doi: 10.1038/srep46469] [Medline: 28422149]
148. Haghighi PD, Kang Y, Buchbinder R, Burstein F, Whittle S. Investigating subjective experience and the influence of weather among individuals with fibromyalgia: a content analysis of Twitter. JMIR Public Health Surveill 2017 Jan 19;3(1):e4 [FREE Full text] [doi: 10.2196/publichealth.6344] [Medline: 28104577]
149. Hartley DM, Giannini CM, Wilson S, Frieder O, Margolis PA, Kotagal UR, et al. Coughing, sneezing, and aching online: Twitter and the volume of influenza-like illness in a pediatric hospital. PLoS One 2017;12(7):e0182008 [FREE Full text] [doi: 10.1371/journal.pone.0182008] [Medline: 28753678]
150. Kandula S, Hsu D, Shaman J. Subregional nowcasts of seasonal influenza using search trends. J Med Internet Res 2017 Nov 6;19(11):e370 [FREE Full text] [doi: 10.2196/jmir.7486] [Medline: 29109069]
151. Lampos V, Zou B, Cox I. Enhancing Feature Selection Using Word Embeddings: The Case of Flu Surveillance. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3 - 7, 2017; Perth, Australia p. 695-704. [doi: 10.1145/3038912.3052622]
152. Lee K, Agrawal A, Choudhary A. Forecasting Influenza Levels using Real-Time Social Media Streams. In: 2017 IEEE International Conference on Healthcare Informatics. 2017 Presented at: ICHI'17; August 23-26, 2017; Park City, UT, USA. [doi: 10.1109/ichi.2017.68]
153. Li Z, Liu T, Zhu G, Lin H, Zhang Y, He J, et al. Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: a case study in Guangzhou, China. PLoS Negl Trop Dis 2017 Mar;11(3):e0005354 [FREE Full text] [doi: 10.1371/journal.pntd.0005354] [Medline: 28263988]
154. Liu K, Huang S, Miao Z, Chen B, Jiang T, Cai G, et al. Identifying potential norovirus epidemics in China via internet surveillance. J Med Internet Res 2017 Aug 8;19(8):e282 [FREE Full text] [doi: 10.2196/jmir.7855] [Medline: 28790023]
155. 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]
156. Morsy S, Dang TN, Kamel MG, Zayan AH, Makram OM, Elhady M, et al. Prediction of Zika-confirmed cases in Brazil and Colombia using Google Trends. Epidemiol Infect 2018 Oct;146(13):1625-1627. [doi: 10.1017/S0950268818002078] [Medline: 30056812]
157. Mowery D, Smith H, Cheney T, Stoddard G, Coppersmith G, Bryan C, et al. Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. J Med Internet Res 2017 Feb 28;19(2):e48 [FREE Full text] [doi: 10.2196/jmir.6895] [Medline: 28246066]
158. Noll-Hussong M. Whiplash syndrome reloaded: digital echoes of whiplash syndrome in the European internet search engine context. JMIR Public Health Surveill 2017 Mar 27;3(1):e15 [FREE Full text] [doi: 10.2196/publichealth.7054] [Medline: 28347974]

159. Priedhorsky R, Osthus D, Daughton A, Moran K, Generous N, Fairchild G, et al. Measuring Global Disease with Wikipedia: Success, Failure, and a Research Agenda. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW'17; February 25 - March 1, 2017; Portland, Oregon, USA p. 1812-1834 URL: http://europepmc.org/abstract/MED/28782059 [doi: 10.1145/2998181.2998183]
160. Reichert JR, Kristensen KL, Mukkamala RR, Vatrapu R. A Supervised Machine Learning Study of Online Discussion Forums about Type-2 Diabetes. In: 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services. 2017 Presented at: Healthcom'17; October 12-15, 2017; Dalian, China. [doi: 10.1109/healthcom.2017.8210815]
161. Roccetti M, Marfia G, Salomoni P, Prandi C, Zagari RM, Kengni FL, et al. Attitudes of Crohn's disease patients: infodemiology case study and sentiment analysis of Facebook and Twitter posts. JMIR Public Health Surveill 2017 Aug 9;3(3):e51 [FREE Full text] [doi: 10.2196/publichealth.7004] [Medline: 28793981]
162. Rudra K, Imran M, Sharma A, Ganguly N. Classifying Information from Microblogs during Epidemics. In: Proceedings of the 2017 International Conference on Digital Health. 2017 Presented at: DH'17; July 2 - 5, 2017; London, United Kingdom p. 104-108. [doi: 10.1145/3079452.3079491]
163. Samaras L, García-Barriocanal E, Sicilia M. Syndromic surveillance models using web data: the case of influenza in Greece and Italy using Google Trends. JMIR Public Health Surveill 2017 Nov 20;3(4):e90 [FREE Full text] [doi: 10.2196/publichealth.8015] [Medline: 29158208]
164. Strauss RA, Castro JS, Reintjes R, Torres JR. Google dengue trends: an indicator of epidemic behavior. The Venezuelan Case. Int J Med Inform 2017 Aug;104:26-30. [doi: 10.1016/j.ijmedinf.2017.05.003] [Medline: 28599813]
165. Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, et al. Dynamic Forecasting of Zika Epidemics Using Google Trends. PLoS One 2017;12(1):e0165085 [FREE Full text] [doi: 10.1371/journal.pone.0165085] [Medline: 28060809]
166. Tkachenko N, Chotvijit S, Gupta N, Bradley E, Gilks C, Guo W, et al. Google Trends can improve surveillance of Type 2 diabetes. Sci Rep 2017 Jul 10;7(1):4993 [FREE Full text] [doi: 10.1038/s41598-017-05091-9] [Medline: 28694479]
167. Marques-Toledo CD, Degener CM, Vinhal L, Coelho G, Meira W, Codeço CT, et al. Dengue prediction by the web: tweets are a useful tool for estimating and forecasting Dengue at country and city level. PLoS Negl Trop Dis 2017 Jul;11(7):e0005729 [FREE Full text] [doi: 10.1371/journal.pntd.0005729] [Medline: 28719659]
168. Xu Q, Gel YR, Ramirez LL, Nezafati K, Zhang Q, Tsui K. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. PLoS One 2017;12(5):e0176690 [FREE Full text] [doi: 10.1371/journal.pone.0176690] [Medline: 28464015]
169. Zhang K, Arablouei R, Jurdak R. Predicting Prevalence of Influenza-Like Illness From Geo-Tagged Tweets. In: Proceedings of the 26th International Conference on World Wide Web Companion. 2017 Presented at: WWW'17; April 3 - 7, 2017; Perth, Australia p. 1327-1334. [doi: 10.1145/3041021.3051150]
170. Zhang Q, Perra N, Perrotta D, Tizzoni M, Paolotti D, Vespignani A. Forecasting Seasonal Influenza Fusing Digital Indicators and a Mechanistic Disease Model. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3 - 7, 2017; Perth, Australia p. 311-319. [doi: 10.1145/3038912.3052678]
171. Zhang Y, Milinovich G, Xu Z, Bambrick H, Mengersen K, Tong S, et al. Monitoring pertussis infections using internet search queries. Sci Rep 2017 Sep 5;7(1):10437 [FREE Full text] [doi: 10.1038/s41598-017-11195-z] [Medline: 28874880]
172. Barros JM, Duggan J, Rebholz-Schuhmann D. Disease mentions in airport and hospital geolocations expose dominance of news events for disease concerns. J Biomed Semantics 2018 Jun 12;9(1):18 [FREE Full text] [doi: 10.1186/s13326-018-0186-9] [Medline: 29895320]
173. Brigo F, Lattanzi S, Bragazzi N, Nardone R, Moccia M, Lavorgna L. Why do people search Wikipedia for information on multiple sclerosis? Mult Scler Relat Disord 2018 Feb;20:210-214. [doi: 10.1016/j.msard.2018.02.001] [Medline: 29428464]
174. Chen TH, Chen YC, Chen JL, Chang FC. Flu Trend Prediction Based on Massive Data Analysis. In: Proceedings of the 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis. 2018 Presented at: ICCCBDA'18; April 20-22, 2018; Chengdu, China. [doi: 10.1109/icccbda.2018.8386532]
175. Chen B, Shao J, Liu K, Cai G, Jiang Z, Huang Y, et al. Does eating chicken feet with pickled peppers cause avian influenza? Observational case study on Chinese social media during the avian influenza A (H7N9) outbreak. JMIR Public Health Surveill 2018 Mar 29;4(1):e32 [FREE Full text] [doi: 10.2196/publichealth.8198] [Medline: 29599109]
176. Gianfredi V, Bragazzi NL, Mahamid M, Bisharat B, Mahroum N, Amital H, et al. Monitoring public interest toward pertussis outbreaks: an extensive Google Trends-based analysis. Public Health 2018 Dec;165:9-15. [doi: 10.1016/j.puhe.2018.09.001] [Medline: 30342281]
177. Ho HT, Carvajal TM, Bautista JR, Capistrano JD, Viacrusis KM, Hernandez LF, et al. Using Google Trends to examine the spatio-temporal incidence and behavioral patterns of dengue disease: a case study in Metropolitan Manila, Philippines. Trop Med Infect Dis 2018 Nov 11;3(4):pii: E118 [FREE Full text] [doi: 10.3390/tropicalmed3040118] [Medline: 30423898]
178. Kagashe I, Yan Z, Suheryani I. Enhancing seasonal influenza surveillance: topic analysis of widely used medicinal drugs using Twitter data. J Med Internet Res 2017 Sep 12;19(9):e315 [FREE Full text] [doi: 10.2196/jmir.7393] [Medline: 28899847]
179. Karisani P, Agichtein E. Did You Really Just Have a Heart Attack?: Towards Robust Detection of Personal Health Mentions in Social Media. In: Proceedings of the 2018 World Wide Web Conference. 2018 Presented at: WWW'18; April 23 - 27, 2018; Lyon, France p. 137-146. [doi: 10.1145/3178876.3186055]

180. Livelo ED, Cheng C. Intelligent Dengue Infoveillance Using Gated Recurrent Neural Learning and Cross-Label Frequencies. In: Proceedings of the 2018 IEEE International Conference on Agents. 2018 Presented at: ICA'18; July 28-31, 2018; Singapore. [doi: 10.1109/agents.2018.8459963]
181. Mavragani A, Ochoa G. Infoveillance of infectious diseases in USA: STDs, tuberculosis, and hepatitis. J Big Data 2018;5:30. [doi: 10.1186/s40537-018-0140-9]
182. Nan Y, Gao Y. A machine learning method to monitor China's AIDS epidemics with data from Baidu trends. PLoS One 2018;13(7):e0199697 [FREE Full text] [doi: 10.1371/journal.pone.0199697] [Medline: 29995920]
183. Phillips CA, Leahy AB, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship between state-level Google online search volume and cancer incidence in the United States: retrospective study. J Med Internet Res 2018 Jan 8;20(1):e6 [FREE Full text] [doi: 10.2196/jmir.8870] [Medline: 29311051]
184. Shah MP, Lopman BA, Tate JE, Harris J, Esparza-Aguilar M, Sanchez-Uribe E, et al. Use of internet search data to monitor rotavirus vaccine impact in the United States, United Kingdom, and Mexico. J Pediatric Infect Dis Soc 2018 Feb 19;7(1):56-63 [FREE Full text] [doi: 10.1093/jpids/pix004] [Medline: 28369477]
185. Verma M, Kishore K, Kumar M, Sondh AR, Aggarwal G, Kathirvel S. Google search trends predicting disease outbreaks: an analysis from India. Healthc Inform Res 2018 Oct;24(4):300-308 [FREE Full text] [doi: 10.4258/hir.2018.24.4.300] [Medline: 30443418]
186. Wakamiya S, Kawai Y, Aramaki E. Twitter-based influenza detection after flu peak via tweets with indirect information: text mining study. JMIR Public Health Surveill 2018 Sep 25;4(3):e65 [FREE Full text] [doi: 10.2196/publichealth.8627] [Medline: 30274968]
187. Young SD, Torrone EA, Urata J, Aral SO. Using search engine data as a tool to predict syphilis. Epidemiology 2018 Jul;29(4):574-578 [FREE Full text] [doi: 10.1097/EDE.0000000000000836] [Medline: 29864105]
188. Zou B, Lampos V, Cox I. Multi-Task Learning Improves Disease Models from Web Search. In: Proceedings of the 2018 World Wide Web Conference. 2018 Presented at: WWW'18; April 23 - 27, 2018; Lyon, France p. 87-96. [doi: 10.1145/3178876.3186050]
189. Flu Near You. URL: https://flunearyou.org [accessed 2019-09-16]
190. Statista. 2018. Worldwide Desktop Market Share of Leading Search Engines From January 2010 to July 2019 URL: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [accessed 2018-08-28]
191. Statcounter. Desktop, Mobile & Console Search Engine Market Share Portugal, 2009-2019 URL: https://gs.statcounter.com/search-engine-market-share/desktop-mobile-console/portugal/#yearly-2009-2019-bar [accessed 2019-09-04]
192. Chen S, Zhang H, Lin M, Lv S. Comparision of Microblogging Service Between Sina Weibo and Twitter. In: Proceedings of 2011 International Conference on Computer Science and Network Technology. 2011 Presented at: ICCSNT'11; December 24-26, 2011; Harbin, China. [doi: 10.1109/iccsnt.2011.6182424]
193. 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]
194. Kass-Hout TA, Alhinnawi H. Social media in public health. Br Med Bull 2013;108:5-24. [doi: 10.1093/bmb/ldt028] [Medline: 24103335]
195. Diaz-Aviles E, Stewart A, Velasco E, Denecke K, Nejdl W. arXiv e-Print archive. 2012. Epidemic Intelligence for the Crowd, by the Crowd (Full Version) URL: http://arxiv.org/abs/1203.1378 [accessed 2020-01-23]
196. Kroll M, Phalkey RK, Kraas F. Challenges to the surveillance of non-communicable diseases--a review of selected approaches. BMC Public Health 2015 Dec 16;15:1243 [FREE Full text] [doi: 10.1186/s12889-015-2570-z] [Medline: 26672992]
197. Pew Research Center. 2013 Feb 12. The Internet and Health URL: http://www.pewinternet.org/2013/02/12/the-internet-and-health/ [accessed 2018-08-28]
198. Omnicore. 2018. Twitter by the Numbers: Stats, Demographics & Fun Facts URL: https://www.omnicoreagency.com/twitter-statistics/ [accessed 2018-08-28]
199. Zickuhr K, Rainie L. Pew Research Center. 2011 Jan 13. Wikipedia, Past and Present: A Snapshot of Current Wikipedia Users URL: http://www.pewinternet.org/2011/01/13/wikipedia-past-and-present/ [accessed 2020-01-23]
200. Butler D. When Google got flu wrong. Nature 2013 Feb 14;494(7436):155-156. [doi: 10.1038/494155a] [Medline: 23407515]
201. 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]

Abbreviations

IBS: internet-based source
ILI: influenza-like illness
NCD: noncommunicable disease
P2P: peer-to-peer
RQ: research question
SFI: Science Foundation Ireland

Edited by G Eysenbach; submitted 24.02.19; peer-reviewed by V Gianfredi, D Paolotti, D Gunasekeran, N Bragazzi; comments to author 18.05.19; revised version received 18.09.19; accepted 26.11.19; published 13.03.20.

Please cite as:
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;22(3):e13680
URL: http://www.jmir.org/2020/3/e13680/
doi:10.2196/13680
PMID:32167477

Review

Mapping the Evidence on the Effectiveness of Telemedicine Interventions in Diabetes, Dyslipidemia, and Hypertension: An Umbrella Review of Systematic Reviews and Meta-Analyses

Patrick Timpel[1*], MSc, MSc, Dipl Soz arb/Soz päd; Sarah Oswald[2], BA; Peter E H Schwarz[1,3,4], MD, MBA, PhD; Lorenz Harst[5*], BA, MA

[1] Department for Prevention and Care of Diabetes, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[2] Master Program Health Sciences / Public Health at the Institute and Policlinic of Occupational and Social Medicine, Faculty of Medicine at the University Clinic Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[3] Paul Langerhans Institute Dresden, Helmholtz Center Munich, University Hospital and Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

[4] German Center for Diabetes Research (DZD e V), Neuherberg, Germany

[5] Research Association Public Health Saxony / Center for Evidence-Based Healthcare, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[*] these authors contributed equally

Corresponding Author:
Patrick Timpel, MSc, MSc, Dipl Soz arb/Soz päd
Department for Prevention and Care of Diabetes
Faculty of Medicine Carl Gustav Carus
Technische Universität Dresden
Fetscherstraße 74
Dresden, 01307
Germany
Phone: 49 3513177209
Email: patrick.timpel@tu-dresden.de

Abstract

Background: Telemedicine is defined by three characteristics: (1) using information and communication technologies, (2) covering a geographical distance, and (3) involving professionals who deliver care directly to a patient or a group of patients. It is said to improve chronic care management and self-management in patients with chronic diseases. However, currently available guidelines for the care of patients with diabetes, hypertension, or dyslipidemia do not include evidence-based guidance on which components of telemedicine are most effective for which patient populations.

Objective: The primary aim of this study was to identify, synthesize, and critically appraise evidence on the effectiveness of telemedicine solutions and their components on clinical outcomes in patients with diabetes, hypertension, or dyslipidemia.

Methods: We conducted an umbrella review of high-level evidence, including systematic reviews and meta-analyses of randomized controlled trials. On the basis of predefined eligibility criteria, extensive automated and manual searches of the databases PubMed, EMBASE, and Cochrane Library were conducted. Two authors independently screened the studies, extracted data, and carried out the quality assessments. Extracted data were presented according to intervention components and patient characteristics using defined thresholds of clinical relevance. Overall certainty of outcomes was assessed using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) tool.

Results: Overall, 3564 references were identified, of which 46 records were included after applying eligibility criteria. The majority of included studies were published after 2015. Significant and clinically relevant reduction rates for glycated hemoglobin (HbA_{1c}; ≤−0.5%) were found in patients with diabetes. Higher reduction rates were found for recently diagnosed patients and those with higher baseline HbA_{1c} (>8%). Telemedicine was not found to have a significant and clinically meaningful impact on blood pressure. Only reviews or meta-analyses reporting lipid outcomes in patients with diabetes were found. GRADE assessment revealed that the overall quality of the evidence was low to very low.

Conclusions: The results of this umbrella review indicate that telemedicine has the potential to improve clinical outcomes in patients with diabetes. Although subgroup-specific effectiveness rates favoring certain intervention and population characteristics were found, the low GRADE ratings indicate that evidence can be considered as limited. Future updates of clinical care and practice guidelines should carefully assess the methodological quality of studies and the overall certainty of subgroup-specific outcomes before recommending telemedicine interventions for certain patient populations.

(J Med Internet Res 2020;22(3):e16791) doi:10.2196/16791

KEYWORDS

telemedicine; diabetes mellitus; hypertension; dyslipidemia; review; GRADE approach; treatment outcome

Introduction

Background

Diabetes is affecting 463 million people worldwide (aged between 20 and 79 years) [1]. Hypertension and hyperlipidemia (or hypercholesterolemia) are common comorbidities in patients with type 2 diabetes (T2D), and also show an increasing coprevalence [2,3]. The risk of diabetes-associated complications can be reduced by continuous control of blood glucose [4], blood pressure (BP) lowering [5-8], and blood lipid profile [9,10]. Current guidelines of the American Diabetes Association (ADA) recommend that most adults with diabetes achieve glycated hemoglobin (HbA_{1c}) <7.0%, BP<140/90 mmHg (<130/90 for patients with increased cardiovascular [CV] risk), and low-density lipoprotein cholesterol (LDL-c) <100 mg/dL [11]. Diabetes self-management education and support, defined as an interactive and continuous process intended to increase knowledge, skills, and abilities required for successful self-management of diabetes interventions [12], has proven to be effective [13,14]. Similarly, hypertensive patients may benefit from the combination of self-monitoring with education or counseling in terms of increased medication adherence and improved BP control [15].

The application of information and communication technologies (ICTs) in health care has been rapidly increasing worldwide. Telemedicine is defined by three characteristics: (1) using ICTs, (2) covering a geographical distance, and (3) involving professionals who deliver care directly to a patient or a group of patients [16,17]. Owing to the need for individualized and continuous monitoring and self-management support for patients, chronic diseases are considered the ideal target conditions for the development and implementation of telemedicine approaches [18,19].

However, detailed guidance is still lacking on how to choose and integrate tools for specific target groups in diabetes care [20,21]. Earlier systematic reviews of high-quality review articles already uncovered key elements for technology-enabled self-management, such as (1) communication between a health care provider and patient, (2) patient-generated health data, (3) education, and (4) feedback [22], or they simply underlined the promising nature of telemedicine [23]. However, the available overviews mainly focus on a specific target group, do not take into account the heterogeneity of telemedicine applications, or focus on a specific tool [24]. Heterogeneous applications of the term telemedicine [16] further limit the external validity of single studies. Owing to the differentiated phenotypes of applied telemedicine solutions, their components, and settings, as well as missing analyses of the quality of studies (and certainty of effects), evidence-based guidance on the best available digital intervention is challenging [25-27].

Objective

Therefore, the primary objective of this umbrella review is to identify, synthesize, and critically appraise the evidence on the effectiveness of telemedicine solutions and their components on clinical outcomes—HbA_{1c}, high-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol (TC), triglycerides (TGC), systolic BP (SBP), diastolic BP (DBP)—in patients with diabetes (type 1 diabetes [T1D] and T2D), hypertension, or dyslipidemia. Owing to the increasing number of available reviews and meta-analysis as well as the potential of addressing three prevalent chronic conditions with multiple digital interventions, the analysis was conducted as an umbrella review [28,29].

The research question is based on the Population, Intervention, Control, Outcome, and Time (PICOT) criteria: *In patients with diabetes, hypertension or dyslipidemia, what is the evidence for the effectiveness of telemedicine-supported chronic care on disease-specific clinical outcomes?*

Methods

Search Strategy and Eligibility Criteria

We conducted an umbrella review using extensive automated and manual searches of the databases PubMed, EMBASE, and the Cochrane Library to identify relevant evidence on the effectiveness of telemedicine interventions on the three target diseases. Umbrella reviews summarize and contrast evidence from existing systematic reviews and meta-analyses by looking at specific outcomes across included records [28].

The search was carried out in October 2018. PICOT-criteria (Table 1) for “population,” “intervention,” and “study design” were combined to develop the search strings (Multimedia Appendix 1). No time limitation was applied.

Table 1. Population, Intervention, Control, Outcome, and Time and eligibility criteria.

Population, Intervention, Control, Outcome, and Time criteria	Eligibility	
	Inclusion	Exclusion
Population	Humans; only studies addressing at least one of the predetermined target diseases within their initial search	Studies addressing chronic diseases in general, other than the three diseases defined, or not addressing any disease at all; specific populations (pregnant women and ethnical minorities); and animals
Intervention	Primary studies applying telemedicine intervention specified as (1) using ICTs[a], (2) covering distance, and (3) involving a health care provider for delivering care to the patient	Studies focusing solely on monitoring or data storage and exchange tools (such as electronic health records)
Control	Usual care	No control group available or not specified
Outcome	Effectiveness analyses allowing for quantitative comparisons between groups using clinical parameters (primary outcome HbA$_{1c}$[b], SBP[c], DBP[d], HDL-c[e], LDL-c[f], TC[g], and TGC[h])	Studies primarily investigating mortality, costs or cost-effectiveness, or feasibility; or efficacy
Time	Follow-up time of at least three months	No or shorter follow-up periods described
Study design	Study design being either a systematic review or meta-analysis of randomized controlled trials	Other, including a systematic review or meta-analysis of observational studies

[a] ICT: information and communication technology.

[b] HbA$_{1c}$: glycated hemoglobin.

[c] SBP: systolic blood pressure.

[d] DBP: diastolic blood pressure.

[e] HDL-c: high-density lipoprotein cholesterol.

[f] LDL-c: low-density lipoprotein cholesterol.

[g] TC: total cholesterol.

[h] TGC: triglycerides.

Records that fulfilled the following eligibility criteria were included (Table 1): systematic reviews or meta-analyses of randomized controlled trials (RCTs; as this is regarded as highest level of evidence) [30] evaluating the effectiveness of telemedicine in at least one of the target diseases (diabetes, hypertension, and/or dyslipidemia) in adults. No restrictions were made with respect to the kind of participating medical providers. We included all eligible articles in English language and with full text available.

Relevant reviews or meta-analyses were excluded if their primary studies mainly assessed mortality, utilization of health services, the usability of the technology studied, or patients' acceptance of or satisfaction with the telemedicine tools, or if no quantitative comparison based on clinical outcomes was reported. Studies evaluating interventions using automated feedback without involving a professional or those providing only monitoring of relevant parameters (without feedback) were excluded. In addition, studies evaluating telemedicine use of medical providers only or those in which the components of the intervention were not transparently described were excluded. Eligible records had to report a change in one of the specified clinical outcomes after a follow-up time of at least three months, as this period is in line with current treatment guidelines [15,31,32].

Conference abstracts or protocols were excluded as well. Research was excluded if it focused on specific countries or regions or targeted specified populations (eg, minorities and pregnant women with diabetes). We excluded those studies for which updates of the evidence—indicated by the same group of authors and/or application of identical search string—were available.

We further searched the reference lists of all relevant publications by hand, to identify any additional studies. After carrying out the title-abstract screening, we conducted a hand search in Google Scholar and the three most relevant journals in the field of digital health, as indicated by the highest number of potentially relevant publications (Multimedia Appendix 2).

Data Extraction and Quality Assessment

Two authors (PT and LH) independently screened the records, extracted data, and carried out the quality assessments. The quality assessment of records was done using the Oxford Quality Assessment Questionnaire (OQAQ) to eliminate records of low quality before data extraction [33]. Any disagreement over the suitability of certain records was discussed among the raters and resolved by consensus.

As the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) is the established tool for assessing the overall certainty of evidence by analyzing its risk of bias, imprecision, inconsistency, indirectness, and publication bias, it was used to assess the quality of included records [34]. This assessment was performed by three independent researchers (PT, SO, and LH), using independent pairwise ratings. Disagreements were again resolved by discussion or, where not possible, by consulting the independent third coder [35].

The results of the included records were extracted using a piloted, standardized data extraction form. According to the methodological considerations for conducting umbrella or meta-reviews, the results were reported descriptively and in tabular form [28,29].

Data Analysis

The presentation of data is descriptive; however, the results of meta-analyses and subgroup analyses were specifically analyzed to find effective components or modes of delivery (intensity and frequency) in subgroups or settings. In light of previous trials, a clinically relevant reduction of –0.5% in HbA_{1c} is considered a suitable threshold (Table 2) [36,37]. The definition of clinically relevant reduction rates (direction of arrows) and the statistical significance (green) were used to compare interventions’ effectiveness (Tables 3-5).

Table 2. Definition of clinically relevant differences in glycated hemoglobin.

Reduction rate in glycated hemoglobin (%)	*P* value	Guidance
≤−0.5	>.05	[a]
>−0.5, <0	>.05	[b]
>0	>.05	[c]
>−0.5, <0	<.05	[d]
≤−0.5	<.05	[e]

[a] non-significant but clinically relevant change.

[b] non-significant and not clinically relevant change.

[c] non-significant and not clinically relevant change.

[d] significant but not clinically relevant change.

[e] significant and clinically relevant change.

In terms of BP control, a −10 mmHg reduction in SBP or a −5 mmHg reduction in DBP is considered as clinically relevant [38]. No exact clinical relevance margins for lipid profiles could be prespecified, as European guidelines recommend a risk-based approach with regard to the presence of CV risk or established CV disease [32].

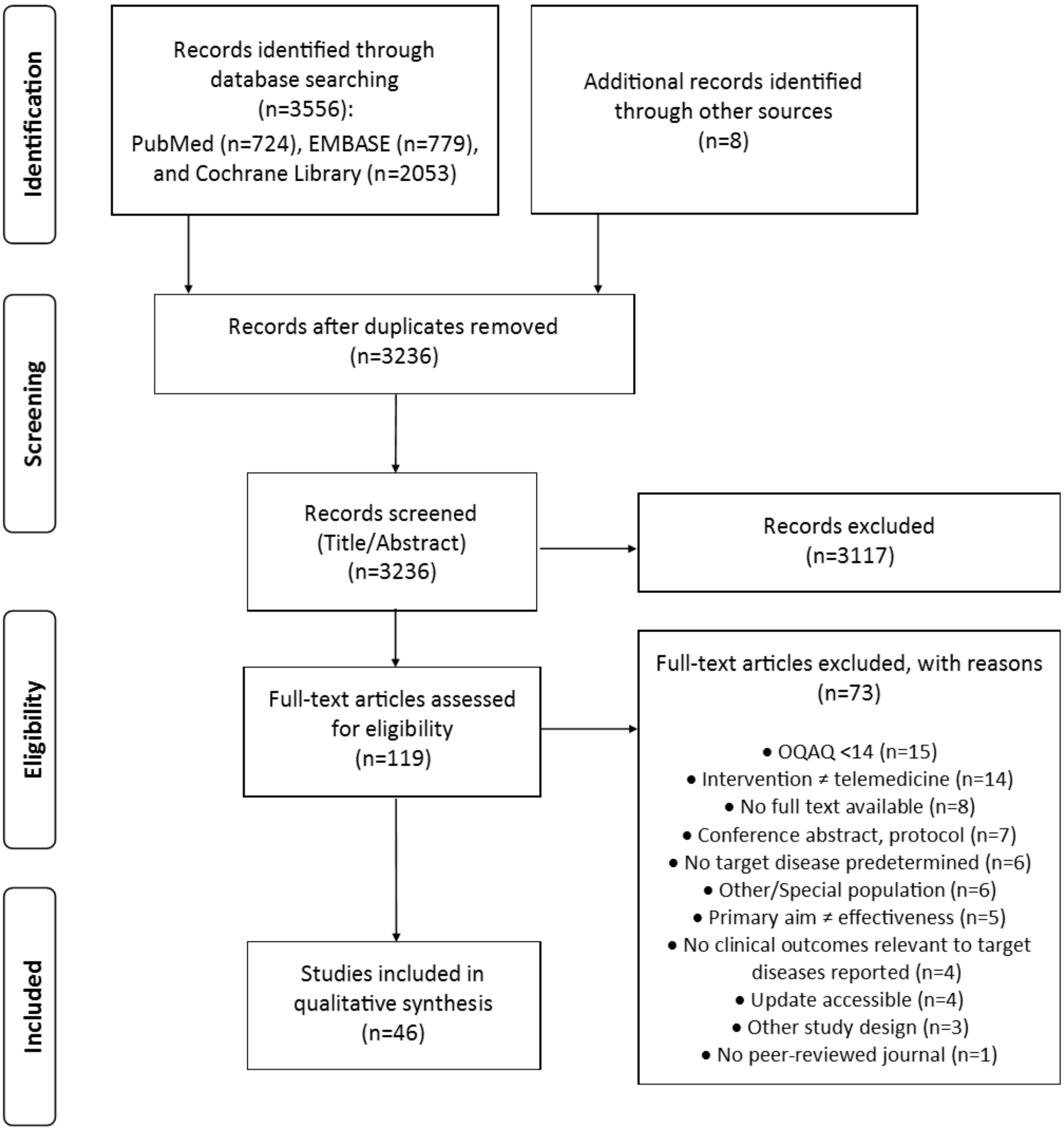
To compare overall treatment effects between baseline and follow-up, meta-analyses reporting treatment effects as mean differences (MD), standardized mean difference (SMD), Cohen *d*, and Hedge *g* were included. For heterogeneity testing, results of I^2 statistics (indicating variation across studies that is not due to chance) were used. A value of <40% indicates a low, 30%-60% a moderate, and >75% a substantial-to-high level of heterogeneity [39].

Results

Review Characteristics

Overall, 3564 references were identified. After title-abstract screening, 119 records remained for further full-text analysis. Details of the extracted evidence are provided in the Multimedia Appendices 3-9. The most important reasons for exclusion were low quality (n=15) and applied interventions not matching the prespecified telemedicine definition (n=14; annex section V). Overall, 46 studies were included in this umbrella review (Figure 1). In Figure 1, the Preferred Reporting Items for Systematic Reviews and Meta-analyses flowchart shows the study selection process, covering the single steps of identification via a 2-step screening (title and abstract and full-text base) for eligibility and inclusion into the qualitative synthesis of this review.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses flowchart of the study selection process. OQAQ: Overview Quality Assessment Questionnaire.



Study Characteristics

Study designs included 16 systematic reviews [40-55], 7 meta-analyses [56-62], 19 records conducting both a systematic review and meta-analysis [63-81], three systematic reviews and meta-analyses with meta-regression [82-84], and one systematic review and network meta-analysis [85]. The included meta-studies were published between 2009 and 2018, the majority was published after 2015 (Multimedia Appendix 5) and focused on diabetes. No high-quality reviews or meta-analyses were found targeting patients with dyslipidemia.

An analysis of primary studies revealed significant overlaps among the 26 meta-analyses (Multimedia Appendix 5). The majority of primary studies were published after 2010 (Multimedia Appendix 5).

On a scale of 0 to 18, the median OQAQ score of the 46 included studies was 16 (IQR 1), indicating that they were good quality systematic reviews and meta-analyses.

Results of Included Systematic Reviews

Data from 16 systematic reviews were extracted (Multimedia Appendix 6) [40-54]. Diabetes was the chronic disease covered most often by the included reviews. A total of 5 systematic reviews dealt with T2D [41,43,46,49]; however, only one systematic review dealt with T1D [53]. Both types were studied together for a total of 4 times [42,45,48,51], four other systematic reviews did not specify which type of diabetes they focused on [44,50,52,54]. Among the other diseases studied, hypertension was the most common [40,44,50,55]. The results of the included systematic reviews were mixed, presenting a tendency for positive effects of telemedicine, in general, and digitally supported self-management using phones or apps on HbA_{1c} [42,44,54] and SBP/DBP [40,55]. In contrast, the majority of studies evaluating telemonitoring and self-monitoring interventions found no significant improvements in HbA_{1c}, fasting plasma glucose, or BP [49-51].

Results of Meta-Analyses

Effectiveness of Telemedicine in Patients With Diabetes

Digital self-management in diabetes (T1D and T2D) was analyzed by 13 meta-analyses, of which 4 meta-analyses evaluated the effectiveness of mobile health (mHealth) [63,65,84] and one meta-analysis evaluated the use of social network services (SNS) [78]. In digital self-management interventions, those including prescription (−0.75%, 95% CI −1.05 to −0.43; *P*=.013), teleconsultation (−0.62%; *P*<.001), and health information technologies on top of usual care (mostly based on tele-education; −0.57%, 95% CI −0.71 to −0.43; *P*<.005) showed significant clinically relevant mean reductions in HbA_{1c} [60,72,80]. Digital self-management interventions using mHealth showed significant clinically relevant reductions in HbA_{1c} if they provided remote access to usual care (−0.55%, 95% CI −0.72 to −0.38; *P*<.001), used one or two features (−0.52%, 95% CI −0.76 to −0.28; *P*<.001), used SMS-based feedback (−0.64%, 95% CI −1.09 to −0.19; *P*=.005), included a potential risk intervention (−0.61%, 95% CI −0.95 to −0.27; *P*<.001), included a structured display (−0.69%, 95% CI −0.32 to −1.06; *P*=.008), provided medication management (−0.56%, 95% CI −0.99 to −0.13; *P*<.001), and provided lifestyle modification management (−0.52%, 95% CI −0.84 to −0.20; *P*<.001) [63,65,80]. SNS applied in diabetes self-management interventions proved to be effective if they were Web-based (−0.51%, 95% CI −0.68 to −0.34; *P*<.001) or combined Web-based SNS with mobile technologies (−0.54%, 95% CI −0.72 to −0.37; *P*<.001) [78].

Overall mean reductions in HbA_{1c} of telemedicine interventions in patients with T1D ranged between −0.12% and −0.86% [60,61,63,70,72,78,82,84]. Overall mean reductions were mostly not significant. Highest mean reductions were observed for the combination of tele-education with teleconsultation (−0.91%, 95% CI −1.21 to −0.61), although data on statistical significance were not provided [70]. No significant clinically relevant reductions for population characteristics such as baseline HbA_{1c} or age were identified in patients with T1D.

Although varying in range (−0.01% to −1.13%), telemedicine significantly reduced HbA_{1c} in patients with T2D [59,60,64,66,68,72,73,75,76,78,79,81,85].

Effectiveness According to Intervention Duration

Table 3 summarizes the effectiveness of telemedicine in patients with diabetes by comparing intervention durations.

Table 3. Effectiveness of telemedicine on glycated hemoglobin in patients with diabetes, according to intervention duration.

Application category and type of diabetes	Intervention duration	Trials, n	Patients, n	Outcome	MD[a] (95% CI) of percent change in HbA_{1c}[b]	*P* value	I^2 (%)	Grading of Recommendations, Assessment, Development, and Evaluation
Digital health education [56]								
T1D[c]/T2D[d]	3 months	3	203	☒ e	−0.71 (−1.0 to −0.43)	.90	0	☒
T1D/T2D	6 months	2	562	☒ e	−0.52 (−0.75 to −0.29)	.65	0	☒
T1D/T2D	12 months	6	1153	☒ e,f	−0.55 (−0.7 to −0.39)	<.001	78	☒
Telemedicine [70,85]								
T1D	<6 months	7	NS[g]	☒ e	0.07 (−0.16 to 0.31)	NS	NS	☒
T1D	≥6 months	21	NS	☒ e	−0.24 (−0.41 to −0.07)	NS	NS	☒
T2D	≤3 months	17	1377	☒ e	−0.67 (−0.93 to −0.41)	NS	NS	☒
T2D	4-6 months	36	4538	☒ e	−0.41 (−0.84 to 0.02)	NS	NS	☒
T2D	7-11 months	4	659	☒ e	−0.66 (−1.18 to −0.15)	NS	NS	☒
T2D	≥12 months	36	10,237	☒ e	−0.26 (−0.40 to −0.12)	NS	NS	☒
Digital self-management [59,60,72,83]								
T2D	≤3 months	10	NS	☒ e,f	−0.51 (−0.71 to −0.31)	<.001	41.8	☒
T2D	>3 and ≤6 months	10	NS	☒ e,f	−0.48 (−0.68 to −0.28)	<.001	34.5	☒
T2D	3-4 months	11	1613	☒ e,f	−0.30 (−0.50 to −0.11)	<.001	89.1	☒
T2D	>6 months	15	NS	☒ e,f	−0.35 (−0.53 to −0.18)	<.001	70.5	☒
T2D	6-8 months	14	2389	☒ e,f	−0.59 (−0.78 to −0.39)	<.001	84.8	☒
T2D	9-12 months	7	1272	☒ e	−0.21 (−0.35 to −0.075)	.131	39.1	☒
T1D/T2D	≤ 6 months	30	NS	☒ e,f	−0.56 (NS)	<.001	30	☒
T1D/T2D	6 months	6	741	☒ e	−0.57 (−0.85 to −0.30)	.099	NS	☒
T1D/T2D	>6 months	25	NS	☒ e,f	−0.40 (NS)	<.001	25	☒
T1D/T2D	12 months	7	3466	☒ e	−0.30 (−0.48 to −0.11)	.099	NS	☒
Digital self-management (SMS) [75]								
T2D	<6 months	6	NS	☒ e,f	−0.60 (−0.80 to −0.40)	<.001	NS	☒
T2D	≥6 months	4	NS	☒ e,f	−0.40 (−0.56 to −0.24)	<.001	NS	☒
Digital self-management (social network service) [78]								
T1D/T2D	≤3 months	13	799	☒ e,f	−0.54 (−0.80 to −0.28)	<.001	23	☒
T1D/T2D	3-12 months	11	1465	☒ e,f	−0.41 (−0.63 to −0.19)	<.001	25	☒
T1D/T2D	>12 months	10	2713	☒ e,f	−0.36 (−0.59 to −0.14)	<.002	90	☒

[a] MD: mean difference.

[b] HbA_{1c}: glycated haemoglobin

[c] T1D: type 1 diabetes.

[d] T2D: type 2 diabetes.

[e] The direction of the arrows indicates potential clinically relevant reduction rates (see Table 2).

[f] Green arrows show statistical significance.

[g] NS: not specified—cases in which no data were provided. Missing data on statistical significance were handled as nonsignificant.

Significant and clinically relevant reductions were found for short (≤3 months), middle (4-8 months), and long (>12 months) intervention durations. Digital health education, analyzed in the meta-analysis by Angeles et al [56], on average, reduced HbA_{1c} above the predefined clinical relevance margin (HbA_{1c} ≤−0.5%; Table 2). However, only the effects of interventions with a long-term study duration (12 months) were statistically significant (−0.55%, 95% CI −0.7 to −0.39; *P*<.001). Although three meta-analyses observed a tendency for higher reduction rates in shorter intervention durations [59,75,85], no general significant differences in reduction rates among intervention durations were found.

Short-term intervention durations (≤6 months) of digital self-management showed greater mean reductions (−0.56%; *P*<.001) [60] compared with mid- and long-term durations (>6 months) [60,72]. Clinically relevant mean reductions in SNS were significant for short-term intervention durations (≤3 months) as well [78]. Using Web-based SNS for digital self-management proved to be significantly effective in the three pooled follow-up measurements. Again, the greatest mean reductions were apparent during the short-term (≤3 months) follow-up (−0.54%, 95% CI −0.80 to −0.28; *P*<.001) [78].

Effectiveness According to Feedback Mode, Frequency, and Intensity

Although telemedicine interventions using feedback functions significantly reduced HbA_{1c} in several studies [56,60-63,66,67,72,80], the highest reduction rates were found when no personalized feedback was provided (−0.61%, 95% CI −1.40 to 0.19; *P*=.001) [63]. No difference in HbA_{1c} change was found for the type of health care professionals providing the feedback (eg, nurses or physicians) [72].

In addition, feedback, provided either via human telephone calls (−1.13%, 95% CI −1.51 to −0.75; *P*<.05) or via the internet (−0.62%, 95% CI −0.82 to −0.42; *P*<.001), significantly reduced HbA_{1c} to a clinically relevant extent (≤−0.5 change) [68,81]. Higher frequency of provider feedback also showed greater reductions in HbA_{1c} (−1.12%, 95% CI −1.32 to −0-91; *P*<.001) when compared with mean reduction rates of interventions utilizing low frequency rates (−0.33%, 95% CI −0.59 to −0.07; *P*<.01) [82] (Table 4).

Table 4. Effectiveness of telemedicine on glycated hemoglobin in patients with diabetes, according to feedback mode, frequency, and intensity.

Application category and type of diabetes	Feedback characteristics	Trials, n	Patients, n	Outcome	MD[a] (95% CI) of percent change in HbA_{1c}	*P* value	I^2 (%)	Grading of Recommendations, Assessment, Development, and Evaluation
Telemedicine [70,82]								
T1D[b]	App based	5	336	☒ [c]	−0.37 (−0.94 to 0.20)	.20	81.74	☒
T1D	High intensity[d]	13	NS	☒ [c]	−0.24 (−0.49 to 0.01)	NS[e]	NS	☒
T1D	≠High intensity	14	NS	☒ [c]	−0.09 (−0.23 to 0.06)	NS	NS	☒
T1D	Audit + feedback	24	NS	☒ [c]	−0.22 (−0.38 to −0.06)	NS	NS	☒
T1D	No audit + feedback	4	NS	☒ [c]	0.01 (−0.27 to −0.30)	NS	NS	☒
Digital self-management [59,68,72,81,83]								
T2D[f]	Human call/telephone	5	NS	☒ [c,g]	−1.13 (−1.51 to −0.75)	<.05	38	☒
T2D	Human call/telephone	12	NS	☒ [c,g]	−0.53 (−0.81 to −0.26)	<.001	76.35	☒
T2D	Manual	6	1180	☒ [c,g]	−0.44 (−0.74 to −0.15)	.04	NS	☒
T2D	Manual	22	NS	☒ [c,g]	−0.50 (−0.65 to −0.34)	<.001	67.2	☒
T2D	Automated	5	NS	☒ [c,g]	−0.50 (−0.69 to −0.32)	<.001	0	☒
T2D	Automated calls	2	NS	☒ [c]	−0.01 (−0.32 to 0.29)	.94	0	☒
T2D	Automated text	9	NS	☒ [c]	−0.36 (−0.47 to −0.24)	NS	0	☒
T2D	Text message	3	380	☒ [c,g]	−0.52 (−1.04 to 0.00)	<.05	73.5	☒
T2D	Web-based	13	2405	☒ [c,g]	−0.41 (−0.55 to −0.27)	<.05	79.6	☒
T2D	Web-based	19	NS	☒ [c,g]	−0.62 (−0.82 to −0.42)	<.001	77.57	☒
Digital self-management (mobile health) [63,82,84]								
T2D	Low frequency	7	440	☒ [c,g]	−0.33 (−0.59 to −0.07)	.01	47.35	☒
T2D	High frequency	5	326	☒ [c,g]	−1.12 (−1.32 to −0.91)	<.001	0	☒
T1D/T2D	Personalized feedback	8	NS	☒ [c,g]	−0.43 (−0.74 to −0.12)	<.001	75	☒
T1D/T2D	≠Personalized feedback	4	NS	☒ [c,g]	−0.61 (−1.40 to 0.19)	.001	81	☒
T1D/T2D	Frequency (daily)	15	NS	☒ [c]	−0.6 (−0.9 to −0.4)	.27	NS	☒
T1D/T2D	Frequency (weekly)	3	NS	☒ [c]	−0.2 (−0.6 to 0.2)	.27	NS	☒
T1D/T2D	Frequency (not specified)	4	NS	☒ [c]	−0.4 (−0.5 to −0.2)	.27	NS	☒

---

[a] MD: mean difference.

[b] T1D: type 1 diabetes.

[c] The direction of the arrows indicates potential clinically relevant reduction rates (see Table 2).

[d] Direct contact at least once a week.

[e] NS: not specified—cases in which no data were provided. Missing data on statistical significance were handled as nonsignificant.

[f] T2D: type 2 diabetes.

[g] Green arrows show statistical significance.

The meta-regression carried out by Huang et al [68] also revealed that factors we previously disregarded, such as study location, sample size, and feedback methods, were associated significantly with changes in HbA_{1c}. Their combination in multivariate meta-regression analyses explained almost 100% of the variance among studies.

Effectiveness According to Population Characteristics

Subgroup analyses on the effectiveness of telemedicine in certain patient populations (Table 5) were carried out by 12 meta-analyses [60-62,66,68,70,72,75,79,83-85].

Although differences were not always significant, those subgroups with higher baseline HbA_{1c} (>7.5% or >8.0%) showed increased reductions rates [62,68,70,72,79,83,85]. Only for interventions categorized as digital self-management using SMS, the reduction rates were higher (−0.71%, 95% CI −0.93 to −0.48; *P*<.001) in patients with lower baseline HbA_{1c} (<8%) when compared with those with higher (≥8%) baseline HbA_{1c} (−0.38%, 95% CI −0.53 to −0.24; *P*<.001) [75]. Using meta-regression methods, Kebede et al [83] found significant reduction rates in HbA_{1c} for baseline HbA_{1c} >7.5% (beta=−.44, 95% CI −0.81 to −0.06; *P*=.031), self-monitoring of behavioral outcomes, such as diets and physical activity (beta=−1.21, 95% CI −1.95 to −0.46; *P*=.009), and for support in problem solving (beta=−1.30, 95% CI −2.05 to −0.54; *P*=.007).

Significant differences for age groups were sparse, as only three meta-analysis found significant reduction rates in patients with T2D [75] and both types combined [60,61]. The meta-analysis by Saffari et al [75] on SMS-based digital self-management found significantly greater mean reductions (*P*=.006) in HbA_{1c} for patients younger than 55 years (−0.65%, 95% CI −0.88 to −0.41; *P*<.001) when compared with the older age group (−0.42%, 95% CI −0.56 to −0.27; *P*<.001) [75]. The greatest significant mean reductions were observed for patients with diabetes aged 41 to 50 years (−1.83%, 95% CI −3.17 to −0.48; *P*<.001) and those over 50 years (−1.05%, 95% CI −1.50 to −0.60; *P*<.001) [60,61].

For digital self-management, a shorter time since diagnosis (<8.5 years) was associated with significantly greater mean reduction in HbA_{1c} (−0.83%, 95% CI −1.10 to −0.56; *P*=.007) when compared with patients being diagnosed more than 8.5 years ago (−0.22%, 95% CI −0.44 to 0.01; *P*=.007) [79]. Similarly, patients diagnosed less than 7 years ago showed higher mean reductions (−0.61%, 95% CI −0.79 to −0.42) compared with their counterparts (−0.37%, 95% CI −0.61 to −0.13; *P*=.03) after using SMS-based digital self-management [75].

Table 5. Effectiveness of telemedicine on glycated hemoglobin in patients with diabetes, according to population characteristics.

Category of application and type of diabetes	Population characteristics	Trials, n	Patients, n	Outcome	MD[a] (95% CI) of percent change in HbA_{1c}[b]	*P* value	I^2 (%)	Grading of Recommendations, Assessment, Development, and Evaluation
Telemedicine [70,85]								
T1D[c]	Adults	15	1256	☒[d,e]	−0.26 (−0.47 to −0.05)	<.01	79.7	☒
T1D	Children and adolescents	11	796	☒	−0.12 (−0.30 to 0.05)	.70	0	☒
T1D	Baseline HbA_{1c} <9.0%	16	NS	☒	−0.06 (−0.02 to 0.09)	NS[f]	NS	☒
T1D	Baseline HbA_{1c} ≥9.0%	12	NS	☒	−0.34 (−0.57 to −0.11)	NS	NS	☒
T2D[g]	Baseline HbA_{1c} <8.0%	48	5720	☒	−0.22 (−0.25 to −0.19)	NS	NS	☒
T2D	Baseline HbA_{1c} ≥8.0%	45	8100	☒	−0.60 (−0.61 to −0.60)	NS	NS	☒
Digital self-management [60-62,68,72,79,83]								
T2D	Age <55 years	7	701	☒	−0.67 (−1.15 to −0.20)	.52	75	☒
T2D	Age ≥55 years	8	541	☒	−0.41 (−0.62 to −0.21)	.52	0	☒
T2D	Age undetermined	2	289	☒	−0.72 (−1.60 to 0.16)	.52	47	☒
T2D	Diagnosis[h] <8.5 years ago	7	549	☒	−0.83 (−1.10 to 0.56)	.007	0	☒
T2D	Diagnosis[h] ≥8.5 years ago	4	394	☒	−0.22 (−0.44 to 0.01)	.007	0	☒
T2D	Diagnosis time[h] undetermined	6	588	☒	−0.43 (−0.71 to −0.30)	.007	55	☒
T2D	Baseline HbA_{1c} ≤8.0%	6	590	☒	−0.49 (−0.71 to −0.27)	.69	0	☒
T2D	Baseline HbA_{1c} ≤8.0%	7	NS	☒	−0.33 (−0.53 to −0.13)	<.05	46	☒
T2D	Baseline HbA_{1c} >7.0%	11	1707	☒	−0.33 (−0.48 to −0.18)	<.001	77.8	☒
T2D	Baseline HbA_{1c} >7.5%	10	1921	☒	−0.45 (−0.70 to −0.21)	<.001	80.4	☒
T2D	Baseline HbA_{1c} >8.0%	11	941	☒	−0.57 (−0.93 to −0.22)	.69	65	☒
T2D	Baseline HbA_{1c} >8.0%	11	NS	☒	−0.70 (−1.03 to −0.36)	<.05	81	☒
T2D	Baseline BMI <30 kg/m^2	5	359	☒	−0.64 (−0.91 to −0.36)	.49	0	☒
T2D	Baseline BMI ≥30 kg/m^2	10	966	☒	−0.43 (−0.68 to −0.17)	.49	35	☒
T2D	Baseline BMI undetermined	2	206	☒	−0.96 (−2.76 to 0.85)	.49	91	☒
T1D/T2D	Age <40 years	14	NS	☒	−0.32	.02	NS	☒
T1D/T2D	Age <40 years	11	NS	☒	−0.85 (−1.79 to 0.10)	.07	98	☒

Category of application and type of diabetes	Population characteristics	Trials, n	Patients, n	Outcome	MD[a] (95% CI) of percent change in HbA_{1c}[b]	*P* value	I^2 (%)	Grading of Recommendations, Assessment, Development, and Evaluation
T1D/T2D	Age ≥40 years	40	NS	☒	−0.53	<.001	NS	☒
T1D/T2D	Age 41-50 years	8	NS	☒	−1.83 (−3.17 to −0.48)	<.001	96.2	☒
T1D/T2D	Age >50 years	17	NS	☒	−1.05 (−1.50 to −0.60)	<.001	97	☒
T1D/T2D	Baseline HbA_{1c} <8.0%	6	NS	☒	−0.26 (−0.43 to −0.10)	.03	NS	☒
T1D/T2D	Baseline HbA_{1c} ≥ 8.0%	8	NS	☒	−0.64 (−0.93 to −0.35)	.03	NS	☒
T1D/T2D	Baseline HbA_{1c} <9.0%	NS	NS	☒	−0.35	NS	NS	☒
T1D/T2D	Baseline HbA_{1c} ≥9.0%	NS	NS	☒	−1.22	NS	NS	☒
Digital self-management (mobile health) [66,84]								
T2D	Baseline HbA_{1c} <8.0%	4	696	☒	−0.33 (−0.59 to −0.06)	.02	70	☒
T1D/T2D	Average age <25 years	5	NS	☒	−0.5 (−0.8 to −0.1)	.54	NS	☒
T1D/T2D	Average age ≥25 years	17	NS	☒	−0.5 (−0.7 to −0.3)	.54	NS	☒
T1D/T2D	BMI ≥25 kg/m^2	7	NS	☒	−0.8 (−1.1 to −0.5)	.93	NS	☒
T1D/T2D	24 kg/m^2≤ BMI <25 kg/m^2	3	NS	☒	−0.8 (−1.7 to 0.1)	.93	NS	☒
T1D/T2D	BMI unspecified	12	NS	☒	−0.3 (−0.5 to −0.1)	.93	NS	☒
Digital self-management (SMS) [75]								
T2D	Age <55 years	5	NS	☒	−0.65 (−0.88 to −0.41)	<.001	NS	☒
T2D	Age ≥55 years	5	NS	☒	−0.42 (−0.56 to −0.27)	.006	NS	☒
T2D	Diagnosis[h] <7 years ago	4	NS	☒	−0.61 (−0.79 to −0.42)	.001	NS	☒
T2D	Diagnosis[h] ≥7 years ago	3	NS	☒	−0.37 (−0.62 to −0.13)	.031	NS	☒
T2D	Baseline HbA_{1c} <8.0%	5	NS	☒	−0.71 (−0.93 to −0.48)	<.001	NS	☒
T2D	Baseline HbA_{1c} ≥8.0%	5	NS	☒	−0.38 (−0.53 to −0.24)	<.001	NS	☒

[a] MD: mean difference.

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

[c] T1D: type 1 diabetes.

[d] The direction of the arrows indicates potential clinically relevant reduction rates (see Table 2).

[e] Green arrows show statistical significance.

[f] NS: not specified—cases in which no data were provided. Missing data on statistical significance were handled as nonsignificant.

[g] T2D: type 2 diabetes.

[h] Diagnosis time: time since first diagnosis of diabetes.

Effect of Telemedicine on Blood Pressure in Patients With Diabetes

Mean reductions of both SBP and DBP were also found in T2D patients. Toma et al [78] found highly significant mean reductions in patients with both T1D and T2D for SBP (−3.47 mmHg, 95% CI −5.01 to −1.94; *P*<.001) and DBP (−1.84 mmHg, 95% CI −2.98 to −0.70; *P*=.002) because of Web- and mobile-based SNS interventions. Evaluating the effect of digitally supported dietary interventions in patients with chronic diseases, Kelly et al [69] also reported significant mean reductions in SBP (−5.91 mmHg, 95% CI −11.14 to −0.68; *P*=.003) in the diabetes subgroup (although showing high heterogeneity between the two studies; I^2=69%). Although no information on statistical significance was provided, Lee et al [85] showed greatest mean reductions in SBP for the telemedicine subgroups focusing on tele-education (−4.05 mmHg, 95% CI −5.64 to −1.10), as well as those combining tele-education and telemonitoring (−3.91 mmHg, 95% CI −10.07 to 2.25). Analyzing the data of four studies, Cui et al [66] found nonsignificant reductions for both DBP (−1.76 mmHg, 95% CI −3.6 to 0.07; *P*=.06) and SBP (−2.62 mmHg, 95% CI −5.6 to 0.36; *P*=.08). Digitally supported dietary interventions in patients with diabetes resulted in significant mean reductions in SBP (−5.91 mmHg, 95% CI −11.14 to −0.68; *P*=.003) [69]. However, none of the presented reduction rates reached the predefined clinical relevance margin.

Effect of Telemedicine on Lipid Profiles in Patients With Diabetes

Only 8 of the included studies reported on lipid profiles; 4 in T1D/T2D patients [56,65,72,78], 2 in T2D patients [73,85], 1 in T1D patients [70], and 1 in several chronic diseases [69]. On the basis of these studies, evidence on the effectiveness of telemedicine in lowering LDL-c or TGC or increasing high-density lipoprotein cholesterol (HDL-c) in patients with diabetes is scarce and heterogeneous. Marcolino et al [72] found evidence that digital self-management applications for both diabetes types can reduce LDL-c levels; however, although significant, the effect was small (−6.6 mg/dL, 95% CI −8.3 to −4.9; *P*<.001; I^2=24%) [72]. In addition, for both types of diabetes, Toma et al [78] found evidence for a significant improvement in TC (−5.74 mg/dL, 95% CI −9.71 to −1.78; *P*<.005; I^2=53%), HDL (1.90 mg/dL, 95% CI 0.24 to 3.57; *P*=.02; I^2=19%), and TGC (−11.05 mg/dL, 95% CI −20.92 to −1.18; *P*<.03; I^2=0%). Reductions in LDL (−1.15 mg/dL, 95% CI −5.19 to 2.88; *P*=.58; I^2=47%) were not significant. Again, for patients with T2D, the pooled analysis of Lee et al [85] found little and rather inconsistent effects, be it for LDL-c, HDL-c, TC, and TGC.

Effects of Telemedicine in Patients With Hypertension

A total of 3 of the included meta-analyses focused on patients with hypertension [57,71]. Although the two analyses of Omboni et al [57,58] focused on home BP monitoring, Liu et al [71] examined the effect of internet-based interventions. Liu et al [71] reported a significant overall mean reduction in SBP (−3.8 mmHg, 95% CI −5.63 to −2.06; *P*=.001) and DBP (−2.1 mmHg, 95% CI −3.51 to −0.65; *P*<.05). Owing to the identified heterogeneity for SBP (I^2=61%), the authors carried out a subgroup analysis, revealing that mean change in SBP was greater in long-term interventions (6-12 months; −5.8 mmHg, 95% CI −4.3 to −4.1) when compared with short-term interventions (<6 months; −3.47 mmHg, 95% CI −5.2 to −1.7). However, data on statistical significance were not provided [71]. The results of Omboni et al [57] show significant mean reductions in SBP when using ambulatory measurement (−2.28 mmHg, 95% CI −4.32 to −0.24; *P*<.05). In their more recent analysis, they included studies evaluating additional features such as combined data transmission to physician, feedback, advice, and medication regulation. This time, they observed significant mean reductions (−3.48 mmHg, 95% CI −5.31 to −1.64; *P*<.001) [58].

Grading of Evidence

The quality assessment of outcomes using the GRADE framework revealed the following levels of certainty (Multimedia Appendix 8). Of the 219 HbA_{1c} outcomes assessed overall, 170 (77.63%) outcomes were rated as very low evidence and 42 (19.18%) outcomes were rated as low evidence. All of the 42 outcomes measuring SBP or DBP resulted in very low ratings of overall certainty (Table 6).

Table 6. Grading of Recommendations, Assessment, Development, and Evaluation assessment of certainty of glycated hemoglobin and systolic blood pressure/diastolic blood pressure outcomes.

GRADE[a]	HbA_{1c}[b], n (%)	SBP[c]/DBP[d], n (%)
	—[e]	—
	2 (0.92)	—
	42 (19.8)	—
	170 (77.63)	42 (100)

[a] GRADE: Grading of Recommendations, Assessment, Development, and Evaluation.

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

[c] SBP: systolic blood pressure.

[d] DBP: diastolic blood pressure.

[e] Not applicable.

The main reasons for low-quality assessment results in both outcome categories were as follows:

- Unclear or high-risk of bias: Missing allocation concealment, missing blinding of patients, study personnel and outcome assessors, high risk of selection bias and reporting bias (intention-to-treat analysis), and high or unclear losses to follow-up.
- Inconsistency: High heterogeneity in subgroup analysis, inconsistent confidence intervals crossing the mark for no effect.
- Indirectness: Differences in populations (type of diabetes, baseline HbA_{1c}, age, duration of diabetes, and gender), differences in interventions (devices used, components combined, feedback intensity and frequency, and professional or professionals involved), and differences in settings (community, hospital, and primary care) in the pooled subgroups.
- Imprecision: Large confidence intervals and small effect sizes mostly because of small sample sizes.
- Publication bias: Visual and statistical or missing publication bias assessment; the reasons for the increased risk of publication bias mostly referred to the overrepresentation of smaller studies with higher effect sizes (favoring telemedicine). Furthermore, one reason is the paucity of data on mid- and long-term effects (6-12 months).
- Underreporting of relevant information: Reporting of study duration, dropouts/missing data, and follow-up time. Guidance on this matter was further complicated as some authors did not make a clear distinction between study duration and follow-up [61].

Only for two outcomes (0.92%) measuring HbA_{1c}, overall certainty was judged as moderate (Tables 5 and 6). In 6 (5 in HbA_{1c} and 1 in DBP) cases, the outcomes of subgroup analyses were not assessed using GRADE, as results of only one trial were used by the authors of meta-analyses to pool data.

As the initial search did not identify records primarily targeting patients with dyslipidemia and subgroup analyses on HDL, LDL, TC, and TGC were sparse, no grading of lipid outcomes was performed.

Discussion

Principal Findings

High-level evidence from the 46 included meta-analyses and systematic reviews suggests that telemedicine interventions can be effective in improving clinical outcomes in patients with diabetes. Observed reduction rates are comparable with those of nonpharmacological eg, nutrition intervention [86] or increased physical activity [87]) and some pharmacological interventions (−0.5% to −1.25%) [88]. The observed reduction rates are encouraging, bearing in mind that the United Kingdom Prospective Diabetes Study (UKPDS) revealed that a 0.9% decrease in HbA_{1c} was associated with a 25% reduction in microvascular complications, a 10% decrease in diabetes-related mortality, and a 6% reduction in all-cause mortality [89].

In patients with diabetes, significant differences between telemedicine interventions and for certain population characteristics were identified. Telemedicine interventions embedded in frequent and intense patient-provider interactions and interventions with short durations (≤6 months) showed greater benefits. In addition, higher reduction rates were found for recently diagnosed patients and those with higher baseline HbA_{1c}. However, quality assessment using GRADE revealed that overall and subgroup-specific certainty of evidence is low to very low. Therefore, the identified reduction rates have to be dealt with caution when translating them into evidence-based recommendations for treatment guidelines.

Telemedicine was not found to have a significant and clinically meaningful impact on BP. Assessing the certainty of SBP and DBP outcomes, GRADE only revealed very low ratings. No records primarily targeting patients with dyslipidemia were found.

According to the recent consensus report of the ADA and European Association for the Study of Diabetes, the application of telemedicine in diabetes is associated with a modest improvement in glycemic control [31]. The European Society of Cardiology/European Society of Hypertension (ESC/ESH) guidelines for the management of arterial hypertension also report that telemonitoring and mobile phone apps may lead to improved outcomes for patients with hypertension [15]. Our umbrella review updates this assessment of the effectiveness of telemedicine with special regard to intervention components, population characteristics, and it provides an in-depth assessment of the certainty of evidence. A brief summary of the study results can be found in Textbox 1.

Textbox 1. Brief summary of the study results. HbA1c: glycated hemoglobin; GRADE: Grading of Recommendations, Assessment, Development, and Evaluation.

- Telemedicine has the potential to improve clinical outcomes in patients with diabetes. Mixed results were found for patients with hypertension, none for those with dyslipidemia.
- Specific characteristics of the intervention (eg, high frequency and intensity of feedback/interaction and short treatment duration) and the patient (age <55 years, high baseline HbA_{1c}, and recent diagnosis) seem to be associated with increased benefits in patients with diabetes.
- An assessment of the overall certainty using GRADE resulted in low and very low ratings, indicating that effects have to be dealt with caution.

Intervention Components

Looking at the characteristics of the telemedicine applications analyzed by the included meta-analyses, those encompassing frequent and intense patient-provider communication interactions showed greater benefit in HbA_{1c} reduction. This was especially true for the combination of tele-case management with either teleconsultation (−1.20%, 95% CI −2.30 to −0.10; *P*<.001) or telemonitoring (−0.54%, 95% CI −2.44 to −0.06) in patients with T2D [85]. Similarly, analogue disease self-management education interventions are known to be more effective in terms of HbA_{1c} reduction when they offer additional support (eg, structured dietary or exercise interventions) [37,90]. On the basis of the analysis by Kastner et al [91], the combination of case management and self-management in addition to education provides potential for reduced HbA_{1c} levels when compared with education and plain care coordination. Therefore, continuous and frequent communication, either via intensive feedback [68,81] or psychological support [92], seems most promising.

With a longer duration of follow-ups, the quality of evidence steadily declines because of considerable risk of bias and heterogeneity of study populations and interventions included. As for digital self-management, the evidence base is larger yet more diverse, as SMS (1 meta-analysis), social networks (1 meta-analysis), and mHealth apps (4 meta-analysis) can be used. However, the quality of evidence for digital self-management is low to very low, irrespective of the basal technology or the type of diabetes.

In our analysis, some application types were found to reduce BP, for example, in SBP after tele-education (−4.05 mmHg, 95% CI −5.64 to −1.10), as well as strategies combining tele-education and telemonitoring (−3.91 mmHg) [85]. In patients with diabetes, Web- and mobile-based SNS interventions significantly reduced DBP (−3.47 mmHg) [78], and digitally supported dietary interventions led to significant mean reductions in SBP (−5.91 mmHg) [69]. Although these reduction rates did not reach clinical relevance of ≥10 mmHg in SBP or ≥5 mmHg in DBP, they are similar to the expected reduction rates of nonpharmacologic interventions in patients with hypertension. Our results support the identified potential of telemonitoring and mobile phone apps in home BP self-monitoring, articulated in the current ESC/ESH guideline [15] because of the additional advantages in memorizing, reviewing, and transmitting BP measurements [58,93].

On the basis of the identified potential of telemedicine to provide individual self-management support, it is likely that embedded or additional components may have an additive and/or sustained impact on clinical outcomes. As such, recent evidence identified social media [94,95], gamification [96], and machine learning models [97,98] as successful strategies to improve clinical outcomes and prevent disease-related complications.

Population Characteristics

According to the included meta-analyses, telemedicine interventions are more effective for patients with T2D, higher baseline HbA_{1c}, and a more recent diagnosis of diabetes. The increased potential for newly diagnosed patients was also identified by systematic reviews [99,100] and landmark trials such as the UKPDS [5]. As for hypertension, the results did not allow for population-specific analyses, which might be because of the rather passive interventions studies, such as telemonitoring.

With the exception of a baseline BMI <30 kg/m^2 (considered in one meta-analysis), all population-specific subgroup analyses were of low or very low evidence, the latter being more prevalent. This is also true for differences among age groups, for which no significant evidence was found. However, there was a tendency for higher reduction rates of HbA_{1c} in younger patient cohorts with diabetes [60,75]. Owing to the increased risk of elevated BP levels (> 130/80) and long-term risk of CV events, the current ESC/ESH guideline suggests treatment in younger adults (<50 years) [15]. In terms of age-specific BP control, ADA suggests focusing on DBP in patients under 50 years [101].

Overall, as the results concerning population characteristics are diverse and of low to very low quality, our analysis did not find enough high-level evidence to recommend telemedicine for the treatment of patients with both hypertension and diabetes.

Only reviews or meta-analyses reporting lipid outcomes in patients with diabetes were found. The extracted results on lipid outcomes are sparse and too heterogeneous to draw a conclusion on the effectiveness of telemedicine on these outcomes [41,44,46,49,53,55]. With special regard to the effects of statins, as the first-line agents used to decrease cholesterol in the management of dyslipidemia and hypertension, the extracted effects of telemedicine on lipid profiles can be considered as minor [15,32]. However, recent evidence underlines the promising potential of mobile phone-based self-monitoring interventions in patients with lipid metabolism disorders [102], because of the combination of therapy and lifestyle behavior changes.

Limitations

Robust systematic reviewing methods were used to generate an overview of high-quality evidence on the effects of telemedicine in three prevalent chronic conditions. The protocol of this umbrella review was presented to the research community [103]. However, this study has several limitations, starting with the initial search and inclusion process. Although a comprehensive and piloted search strategy has been used, it is possible that some relevant studies were missed, if the exact search terms were used neither by the authors nor by the databases examined (Multimedia Appendix 10). The search within three different databases, complemented by a comprehensive hand search within the most important journals in the field of telemedicine, the use of MeSH terms, and a broad set of keywords, may have limited this risk of selection bias.

In addition, some full-text articles were excluded because of their definition and application of the term “telemedicine,” which did not comply with standardized definitions, such as the one provided by Sood et al [17]. Although the technology applied to deliver telemedicine has made tremendous advances during the past 10 years, our thorough application of the telemedicine definition and subgroup analyses using the GRADE assessment ensures comparability of intervention types. Intensive full-text assessment was applied to limit the bias of falsely including/excluding systematic reviews and meta-analyses because of mislabeling and inadequate delimitations of efficacy and effectiveness, as studies focusing on efficacy were excluded. As telemedicine is mostly embedded in low-risk interventions, mortality as an outcome was not considered. Although internationally recommended to be reported in addition to changes in HbA_{1c} [104], parameters such as the time below, in, or above range, the number of hypoglycemic episodes, and quality of life were only reported by a few study authors and therefore did not allow for evidence-based guidance on this matter. A reason may be the publication date (median=2011) of the primary studies (Multimedia Appendix 6), which is before these recommendations were made.

We also included different types of statistical analyses, including meta-analysis, network-meta-analysis, and meta-regression. Although the majority reported MD, there was a considerable methodological heterogeneity. This was because of the application of fixed- and random-effects models, as well as the reporting of SMD, Hedge *g*, or Cohen *d* instead of MD. Comparing the aggregated results of those statistical values without considering their weight (on the basis of the number of studies or number of patients per subgroup analysis) may have biased our analysis. However, this process was impeded by inconsistent reporting of baseline data such as the number of trials and participants in subgroup analysis. In addition, it is likely that reporting bias within the included systematic reviews and meta-analysis also affected our analysis. When studying the funnel plots, we also observed a tendency toward overrepresented smaller studies with higher effect sizes (favoring telemedicine), thereby increasing the risk of publication bias within some of the included analyses.

Further Methodological Considerations

Owing to the multimodal and individualized nature of digital interventions, the low GRADE results, especially the increase I^2, are not surprising. In addition, we found significant overlaps among the primary studies of the included records (Multimedia Appendix 6). The results of the subgroup analysis therefore need to be considered with exceptional care before recommending certain intervention components for certain populations. However, as GRADE is the established procedure to evaluate the certainty of evidence when developing or updating guidelines, new quality assessment tools appropriate for the tailored and hybrid design of digital interventions should be developed [105]. Along with the need for rather adaptive study designs, there is growing criticism on the suitability of RCTs for evaluating the effectiveness of digital interventions. In light of current efforts to support the clinical effectiveness, quality, and economic value of new technologies by using new assessment frameworks [106-110], our analysis underlines the challenges in this endeavor. In addition, future assessments on the clinical effectiveness should also include consolidated core outcome sets and patient-reported outcomes [111,112]. However, as stated by the included records, longer study durations and more rigorously designed studies are needed for these future research needs.

Conclusions

The results of this umbrella review indicate that telemedicine has the potential to improve clinical outcomes in patients with diabetes. Evidence extracted from systematic reviews and meta-analyses of RCTs showed subgroup-specific effectiveness rates favoring certain intervention and population characteristics. However, as indicated by the low GRADE ratings, evidence on the effectiveness of telemedicine in the three chronic conditions can be considered as limited.

Future updates of clinical care and practice guidelines should carefully assess the methodological quality of studies and assess the overall certainty of subgroup-specific outcomes before recommending telemedicine interventions for certain patient populations.

Acknowledgments

The authors wish to thank Jochen Schmitt, Hendrikje Lantzsch, and Kristin Kemple for their valuable input to the conduct of the umbrella review and the design of the manuscript. The work on this review was partly funded by the European Social Fund and the Free State of Saxony (Grant number: 100310385).

Parts of this manuscript were presented on a poster during a German Conference (Diabetes Kongress 2019) on May 30, 2019 and during a talk on October 10, 2019, (18 Deutscher Kongress für Versorgungsforschung) in Berlin.

Authors' Contributions

PT and LH designed the study and also conducted the search. PT and LH were responsible for critical evaluation, analysis, and presentation of the results. PT, LH, and SO conducted the GRADE Assessment. PT and LH drafted the manuscript. PT, LH, SO, and PS critically evaluated the article and gave their final approval before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Population, Intervention, Control, Outcome, and Time criteria and principles of data extraction.
[DOC File , 70 KB - jmir_v22i3e16791_app1.doc]

Multimedia Appendix 2
Number of manuscripts per journal after title/abstract screening.
[DOC File , 79 KB - jmir_v22i3e16791_app2.doc]

Multimedia Appendix 3
Quality assessment for study inclusion.
[DOC File , 167 KB - jmir_v22i3e16791_app3.doc]

Multimedia Appendix 4
List of excluded studies with reasons.
[DOC File , 46 KB - jmir_v22i3e16791_app4.doc]

Multimedia Appendix 5
Characteristics of included records.
[DOC File , 185 KB - jmir_v22i3e16791_app5.doc]

Multimedia Appendix 6
Results of included systematic reviews.
[DOC File , 78 KB - jmir_v22i3e16791_app6.doc]

Multimedia Appendix 7
Results of included meta-analyses.
[DOC File , 382 KB - jmir_v22i3e16791_app7.doc]

Multimedia Appendix 8
Grading of Recommendations Assessment, Development and Evaluation of glycated haemoglobin and diastolic blood pressure/systolic blood pressure outcomes.
[DOC File , 542 KB - jmir_v22i3e16791_app8.doc]

Multimedia Appendix 9
References of multimedia appendices.
[DOC File , 86 KB - jmir_v22i3e16791_app9.doc]

Multimedia Appendix 10
Electronic database search strategy.
[DOC File , 67 KB - jmir_v22i3e16791_app10.doc]

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9 edition. Diabetes Res Clin Pract 2019 Nov;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
2. Iglay K, Hannachi H, Joseph Howie P, Xu J, Li X, Engel SS, et al. Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus. Curr Med Res Opin 2016 Jul;32(7):1243-1252. [doi: 10.1185/03007995.2016.1168291] [Medline: 26986190]
3. Song Y, Liu X, Zhu X, Zhao B, Hu B, Sheng X, et al. Increasing trend of diabetes combined with hypertension or hypercholesterolemia: NHANES data analysis 1999-2012. Sci Rep 2016 Nov 2;6:36093 [FREE Full text] [doi: 10.1038/srep36093] [Medline: 27805013]
4. Schwarz PE, Timpel P, Harst L, Greaves CJ, Ali MK, Lambert J, et al. Blood sugar regulation for cardiovascular health promotion and disease prevention: JACC health promotion series. J Am Coll Cardiol 2018 Oct 9;72(15):1829-1844 [FREE Full text] [doi: 10.1016/j.jacc.2018.07.081] [Medline: 30286928]

5. -. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet 1998 Sep 12;352(9131):837-853. [Medline: 9742976]
6. -. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. Br Med J 1998 Sep 12;317(7160):703-713 [FREE Full text] [Medline: 9732337]
7. Turnbull F, Neal B, Algert C, Chalmers J, Chapman N, Cutler J, Blood Pressure Lowering Treatment Trialists' Collaboration. Effects of different blood pressure-lowering regimens on major cardiovascular events in individuals with and without diabetes mellitus: results of prospectively designed overviews of randomized trials. Arch Intern Med 2005 Jun 27;165(12):1410-1419. [doi: 10.1001/archinte.165.12.1410] [Medline: 15983291]
8. Thomopoulos C, Parati G, Zanchetti A. Effects of blood-pressure-lowering treatment on outcome incidence in hypertension: 10 - Should blood pressure management differ in hypertensive patients with and without diabetes mellitus? Overview and meta-analyses of randomized trials. J Hypertens 2017 May;35(5):922-944. [doi: 10.1097/HJH.0000000000001276] [Medline: 28141660]
9. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, Cholesterol Treatment Trialists' (CTT) Collaborators. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 2005 Oct 8;366(9493):1267-1278. [doi: 10.1016/S0140-6736(05)67394-1] [Medline: 16214597]
10. ADVANCE Collaborative Group, Patel A, MacMahon S, Chalmers J, Neal B, Billot L, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2560-2572. [doi: 10.1056/NEJMoa0802987] [Medline: 18539916]
11. American Diabetes Association. 9. Cardiovascular Disease and Risk Management: Standards of Medical Care in Diabetes-2018. Diabetes Care 2018 Jan;41(Suppl 1):S86-104. [doi: 10.2337/dc18-S009] [Medline: 29222380]
12. 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]
13. He X, Li J, Wang B, Yao Q, Li L, Song R, et al. Diabetes self-management education reduces risk of all-cause mortality in type 2 diabetes patients: a systematic review and meta-analysis. Endocrine 2017 Mar;55(3):712-731. [doi: 10.1007/s12020-016-1168-2] [Medline: 27837440]
14. Sherifali D, Bai J, Kenny M, Warren R, Ali MU. Diabetes self-management programmes in older adults: a systematic review and meta-analysis. Diabet Med 2015 Nov;32(11):1404-1414. [doi: 10.1111/dme.12780] [Medline: 25865179]
15. Williams B, Mancia G, Spiering W, Rosei EA, Azizi M, Burnier M, ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J 2018 Sep 1;39(33):3021-3104. [doi: 10.1093/eurheartj/ehy339] [Medline: 30165516]
16. Otto L, Harst L, Schlieter H, Wollschlaeger B, Richter P, Timpel P. Towards a Unified Understanding of eHealth and Related Terms – Proposal of a Consolidated Terminological Basis. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies. Health Informatics Europe; 2018; Madeira: HEALTHINF; 2018 Presented at: BIOSTEC'18; Jan 19-21, 2018; Funchal, Madeira, Portugal p. 533-539 URL: http://www.scitepress.org/DigitalLibrary/Link.aspx? [doi: 10.5220/0006651005330539]
17. 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]
18. Kruse CS, Soma M, Pulluri D, Nemali NT, Brooks M. The effectiveness of telemedicine in the management of chronic heart disease - a systematic review. JRSM Open 2017 Mar;8(3):2054270416681747 [FREE Full text] [doi: 10.1177/2054270416681747] [Medline: 28321319]
19. 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]
20. American Diabetes Association. 5. Prevention or Delay of Type 2 Diabetes: Standards of Medical Care in Diabetes-2018. Diabetes Care 2018 Jan;41(Suppl 1):S51-S54. [doi: 10.2337/dc18-S005] [Medline: 29222376]
21. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes-2018. Diabetes Care 2018 Jan;41(Suppl 1):S55-S64. [doi: 10.2337/dc18-S006] [Medline: 29222377]
22. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
23. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
24. 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]

25. Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: a systematic review of reviews. Int J Med Inform 2012 Jan;81(1):1-11. [doi: 10.1016/j.ijmedinf.2011.10.009] [Medline: 22104370]
26. Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K, et al. Personalized Telehealth in the future: a global research agenda. J Med Internet Res 2016 Mar 1;18(3):e53 [FREE Full text] [doi: 10.2196/jmir.5257] [Medline: 26932229]
27. Kirley K, Sachdev N. Digital health-supported lifestyle change programs to prevent Type 2 Diabetes. Diabetes Spectr 2018 Nov;31(4):303-309 [FREE Full text] [doi: 10.2337/ds18-0019] [Medline: 30510384]
28. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015 Sep;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
29. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: planning a relevant research question and objective for an overview. Syst Rev 2018 Mar 1;7(1):39 [FREE Full text] [doi: 10.1186/s13643-018-0695-8] [Medline: 29490699]
30. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ, GRADE Working Group. What is 'quality of evidence' and why is it important to clinicians? Br Med J 2008 May 3;336(7651):995-998 [FREE Full text] [doi: 10.1136/bmj.39490.551019.BE] [Medline: 18456631]
31. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2018 Dec;41(12):2669-2701 [FREE Full text] [doi: 10.2337/dci18-0033] [Medline: 30291106]
32. Catapano AL, Graham I, de Backer G, Wiklund O, Chapman MJ, Drexel H, ESC Scientific Document Group. 2016 ESC/EAS Guidelines for the Management of Dyslipidaemias. Eur Heart J 2016 Oct 14;37(39):2999-3058. [doi: 10.1093/eurheartj/ehw272] [Medline: 27567407]
33. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991;44(11):1271-1278. [doi: 10.1016/0895-4356(91)90160-b] [Medline: 1834807]
34. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
35. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, United States: Wiley; 2019.
36. Little RR, Rohlfing CL, Sacks DB, National Glycohemoglobin Standardization Program (NGSP) Steering Committee. Status of hemoglobin A1c measurement and goals for improvement: from chaos to order for improving diabetes care. Clin Chem 2011 Feb;57(2):205-214 [FREE Full text] [doi: 10.1373/clinchem.2010.148841] [Medline: 21148304]
37. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016 Jun;99(6):926-943 [FREE Full text] [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]
38. Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. Br Med J 2009 May 19;338:b1665 [FREE Full text] [doi: 10.1136/bmj.b1665] [Medline: 19454737]
39. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADEpro GDT. 2013. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations URL: https://gdt.gradepro.org/app/handbook/handbook.html [accessed 2020-01-16]
40. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. JMIR Mhealth Uhealth 2018 Jul 23;6(7):e10723 [FREE Full text] [doi: 10.2196/10723] [Medline: 30037787]
41. Aspry KE, Furman R, Karalis DG, Jacobson TA, Zhang AM, Liptak GS, et al. Effect of health information technology interventions on lipid management in clinical practice: a systematic review of randomized controlled trials. J Clin Lipidol 2013;7(6):546-560. [doi: 10.1016/j.jacl.2013.10.004] [Medline: 24314354]
42. Baron J, McBain H, Newman S. The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review. J Diabetes Sci Technol 2012 Sep 1;6(5):1185-1196 [FREE Full text] [doi: 10.1177/193229681200600524] [Medline: 23063046]
43. Connelly J, Kirk A, Masthoff J, MacRury S. The use of technology to promote physical activity in Type 2 diabetes management: a systematic review. Diabet Med 2013 Dec;30(12):1420-1432. [doi: 10.1111/dme.12289] [Medline: 23870009]
44. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12:CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
45. 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]
46. 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 2016 May;33(5):565-579. [doi: 10.1111/dme.12987] [Medline: 26470750]

47. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: a systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
48. Holmen H, Wahl AK, Småstuen MC, Ribu L. Tailored communication within mobile apps for diabetes self-management: a systematic review. J Med Internet Res 2017 Jun 23;19(6):e227 [FREE Full text] [doi: 10.2196/jmir.7045] [Medline: 28645890]
49. Mushcab H, Kernohan WG, Wallace J, Martin S. Web-based remote monitoring systems for self-managing Type 2 Diabetes: a systematic review. Diabetes Technol Ther 2015 Jul;17(7):498-509. [doi: 10.1089/dia.2014.0296] [Medline: 25830528]
50. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010 Jun 16;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
51. Porter J, Huggins CE, Truby H, Collins J. The effect of using mobile technology-based methods that record food or nutrient intake on diabetes control and nutrition outcomes: a systematic review. Nutrients 2016 Dec 17;8(12):pii: E815. [doi: 10.3390/nu8120815] [Medline: 27999302]
52. Rush KL, Hatt L, Janke R, Burton L, Ferrier M, Tetrault M. The efficacy of telehealth delivered educational approaches for patients with chronic diseases: A systematic review. Patient Educ Couns 2018 Aug;101(8):1310-1321. [doi: 10.1016/j.pec.2018.02.006] [Medline: 29486994]
53. Russell-Minda E, Jutai J, Speechley M, Bradley K, Chudyk A, Petrella R. Health technologies for monitoring and managing diabetes: a systematic review. J Diabetes Sci Technol 2009 Nov 1;3(6):1460-1471 [FREE Full text] [doi: 10.1177/193229680900300628] [Medline: 20144402]
54. Sun C, Malcolm JC, Wong B, Shorr R, Doyle M. Improving glycemic control in adults and children with Type 1 Diabetes with the use of smartphone-based mobile applications: a systematic review. Can J Diabetes 2019 Feb;43(1):51-8.e3. [doi: 10.1016/j.jcjd.2018.03.010] [Medline: 30026048]
55. Vargas G, Cajita MI, Whitehouse E, Han H. Use of short messaging service for hypertension management: a systematic review. J Cardiovasc Nurs 2017;32(3):260-270 [FREE Full text] [doi: 10.1097/JCN.0000000000000336] [Medline: 27111819]
56. Angeles RN, Howard MI, Dolovich L. The effectiveness of web-based tools for improving blood glucose control in patients with diabetes Mellitus: a meta-analysis. Can J Diabetes 2011;35(4):344-352. [doi: 10.1016/S1499-2671(11)54011-0]
57. Omboni S, Guarda A. Impact of home blood pressure telemonitoring and blood pressure control: a meta-analysis of randomized controlled studies. Am J Hypertens 2011 Sep;24(9):989-998. [doi: 10.1038/ajh.2011.100] [Medline: 21654858]
58. Omboni S, Gazzola T, Carabelli G, Parati G. Clinical usefulness and cost effectiveness of home blood pressure telemonitoring: meta-analysis of randomized controlled studies. J Hypertens 2013 Mar;31(3):455-67; discussion 467. [doi: 10.1097/HJH.0b013e32835ca8dd] [Medline: 23299557]
59. Shen Y, Wang F, Zhang X, Zhu X, Sun Q, Fisher E, et al. Effectiveness of internet-based interventions on glycemic control in patients with Type 2 Diabetes: meta-analysis of randomized controlled trials. J Med Internet Res 2018 May 7;20(5):e172 [FREE Full text] [doi: 10.2196/jmir.9133] [Medline: 29735475]
60. Su D, Zhou J, Kelley MS, Michaud TL, Siahpush M, Kim J, et al. Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials. Diabetes Res Clin Pract 2016 Jun;116:136-148. [doi: 10.1016/j.diabres.2016.04.019] [Medline: 27321329]
61. Tchero H, Kangambega P, Briatte C, Brunet-Houdard S, Retali G, Rusch E. Clinical effectiveness of Telemedicine in Diabetes Mellitus: a meta-analysis of 42 randomized controlled trials. Telemed J E Health 2019 Jul;25(7):569-583. [doi: 10.1089/tmj.2018.0128] [Medline: 30124394]
62. Wu C, Wu Z, Yang L, Zhu W, Zhang M, Zhu Q, et al. Evaluation of the clinical outcomes of telehealth for managing diabetes: a PRISMA-compliant meta-analysis. Medicine (Baltimore) 2018 Oct;97(43):e12962 [FREE Full text] [doi: 10.1097/MD.0000000000012962] [Medline: 30412116]
63. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
64. Alharbi NS, Alsubki N, Jones S, Khunti K, Munro N, de Lusignan S. Impact of Information Technology-based interventions for Type 2 Diabetes Mellitus on glycemic control: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 25;18(11):e310 [FREE Full text] [doi: 10.2196/jmir.5778] [Medline: 27888169]
65. 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 1;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
66. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
67. Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, Alberta Kidney Disease Network. Effect of telemedicine on glycated hemoglobin in diabetes: a systematic review and meta-analysis of randomized trials. Can Med Assoc J 2017 Mar 6;189(9):E341-E364 [FREE Full text] [doi: 10.1503/cmaj.150885] [Medline: 27799615]

68. Huang Z, Tao H, Meng Q, Jing L. Management of endocrine disease. Effects of telecare intervention on glycemic control in type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. Eur J Endocrinol 2015 Mar;172(3):R93-101. [doi: 10.1530/EJE-14-0441] [Medline: 25227131]
69. Kelly JT, Reidlinger DP, Hoffmann TC, Campbell KL. Telehealth methods to deliver dietary interventions in adults with chronic disease: a systematic review and meta-analysis. Am J Clin Nutr 2016 Dec;104(6):1693-1702. [doi: 10.3945/ajcn.116.136333] [Medline: 27935523]
70. Lee SW, Ooi L, Lai YK. Telemedicine for the management of glycemic control and clinical outcomes of Type 1 Diabetes Mellitus: a systematic review and meta-analysis of randomized controlled studies. Front Pharmacol 2017;8:330 [FREE Full text] [doi: 10.3389/fphar.2017.00330] [Medline: 28611672]
71. Liu S, Dunford SD, Leung YW, Brooks D, Thomas SG, Eysenbach G, et al. Reducing blood pressure with internet-based interventions: a meta-analysis. Can J Cardiol 2013 May;29(5):613-621. [doi: 10.1016/j.cjca.2013.02.007] [Medline: 23618507]
72. Marcolino MS, Maia JX, Alkmim MB, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]
73. 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]
74. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-930. [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]
75. 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]
76. Spencer-Bonilla G, Ponce OJ, Rodriguez-Gutierrez R, Alvarez-Villalobos N, Erwin PJ, Larrea-Mantilla L, et al. A systematic review and meta-analysis of trials of social network interventions in type 2 diabetes. BMJ Open 2017 Aug 21;7(8):e016506 [FREE Full text] [doi: 10.1136/bmjopen-2017-016506] [Medline: 28827256]
77. Suksomboon N, Poolsup N, Nge YL. Impact of phone call intervention on glycemic control in diabetes patients: a systematic review and meta-analysis of randomized, controlled trials. PLoS One 2014;9(2):e89207 [FREE Full text] [doi: 10.1371/journal.pone.0089207] [Medline: 24586596]
78. Toma T, Athanasiou T, Harling L, Darzi A, Ashrafian H. Online social networking services in the management of patients with diabetes mellitus: systematic review and meta-analysis of randomised controlled trials. Diabetes Res Clin Pract 2014 Nov;106(2):200-211. [doi: 10.1016/j.diabres.2014.06.008] [Medline: 25043399]
79. Wu IX, Kee JC, Threapleton DE, Ma RC, Lam VC, Lee EK, et al. Effectiveness of smartphone technologies on glycaemic control in patients with type 2 diabetes: systematic review with meta-analysis of 17 trials. Obes Rev 2018 Jun;19(6):825-838. [doi: 10.1111/obr.12669] [Medline: 29345109]
80. 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]
81. Zhai Y, Zhu W, Cai Y, Sun D, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014 Dec;93(28):e312 [FREE Full text] [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]
82. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018 Aug;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
83. Kebede MM, Zeeb H, Peters M, Heise TL, Pischke CR. Effectiveness of digital interventions for improving glycemic control in persons with poorly controlled type 2 diabetes: a systematic review, meta-analysis, and meta-regression analysis. Diabetes Technol Ther 2018 Nov;20(11):767-782. [doi: 10.1089/dia.2018.0216] [Medline: 30257102]
84. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
85. Lee SW, Chan CK, Chua SS, Chaiyakunapruk N. Comparative effectiveness of telemedicine strategies on type 2 diabetes management: a systematic review and network meta-analysis. Sci Rep 2017 Oct 4;7(1):12680 [FREE Full text] [doi: 10.1038/s41598-017-12987-z] [Medline: 28978949]
86. Snorgaard O, Poulsen GM, Andersen HK, Astrup A. Systematic review and meta-analysis of dietary carbohydrate restriction in patients with type 2 diabetes. BMJ Open Diabetes Res Care 2017;5(1):e000354 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000354] [Medline: 28316796]

87. Pai L, Li T, Hwu Y, Chang S, Chen L, Chang P. The effectiveness of regular leisure-time physical activities on long-term glycemic control in people with type 2 diabetes: a systematic review and meta-analysis. Diabetes Res Clin Pract 2016 Mar;113:77-85 [FREE Full text] [doi: 10.1016/j.diabres.2016.01.011] [Medline: 26822261]
88. Sherifali D, Nerenberg K, Pullenayegum E, Cheng JE, Gerstein HC. The effect of oral antidiabetic agents on A1C levels: a systematic review and meta-analysis. Diabetes Care 2010 Aug;33(8):1859-1864 [FREE Full text] [doi: 10.2337/dc09-1727] [Medline: 20484130]
89. -. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. Lancet 1998 Sep 12;352(9131):854-865. [Medline: 9742977]
90. Pillay J, Armstrong MJ, Butalia S, Donovan LE, Sigal RJ, Vandermeer B, et al. Behavioral programs for type 2 diabetes mellitus: a systematic review and network meta-analysis. Ann Intern Med 2015 Dec 1;163(11):848-860. [doi: 10.7326/M15-1400] [Medline: 26414227]
91. Kastner M, Cardoso R, Lai Y, Treister V, Hamid JS, Hayden L, et al. Effectiveness of interventions for managing multiple high-burden chronic diseases in older adults: a systematic review and meta-analysis. Can Med Assoc J 2018 Aug 27;190(34):E1004-E1012 [FREE Full text] [doi: 10.1503/cmaj.171391] [Medline: 30150242]
92. Viana LV, Gomes MB, Zajdenverg L, Pavin EJ, Azevedo MJ, Brazilian Type 1 Diabetes Study Group. Interventions to improve patients' compliance with therapies aimed at lowering glycated hemoglobin (HbA1c) in type 1 diabetes: systematic review and meta-analyses of randomized controlled clinical trials of psychological, telecare, and educational interventions. Trials 2016 Feb 17;17:94 [FREE Full text] [doi: 10.1186/s13063-016-1207-6] [Medline: 26888087]
93. Parati G, Omboni S. Role of home blood pressure telemonitoring in hypertension management: an update. Blood Press Monit 2010 Dec;15(6):285-295. [doi: 10.1097/MBP.0b013e328340c5e4] [Medline: 21084882]
94. Gabarron E, Årsand E, Wynn R. Social media use in interventions for diabetes: rapid evidence-based review. J Med Internet Res 2018 Aug 10;20(8):e10303 [FREE Full text] [doi: 10.2196/10303] [Medline: 30097421]
95. Litchman ML, Edelman LS, Donaldson GW. Effect of diabetes online community engagement on health indicators: cross-sectional study. JMIR Diabetes 2018 Apr 24;3(2):e8 [FREE Full text] [doi: 10.2196/diabetes.8603] [Medline: 30291079]
96. Theng Y, Lee JW, Patinadan PV, Foo SS. The use of videogames, gamification, and virtual environments in the self-management of diabetes: a systematic review of evidence. Games Health J 2015 Oct;4(5):352-361. [doi: 10.1089/g4h.2014.0114] [Medline: 26287926]
97. Faruqui SH, Du Y, Meka R, Alaeddini A, Li C, Shirinkam S, et al. Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2019 Nov 1;7(11):e14452 [FREE Full text] [doi: 10.2196/14452] [Medline: 31682586]
98. Jahmunah V, Oh SL, Wei JK, Ciaccio EJ, Chua K, San TR, et al. Computer-aided diagnosis of congestive heart failure using ECG signals - a review. Phys Med 2019 Jun;62:95-104. [doi: 10.1016/j.ejmp.2019.05.004] [Medline: 31153403]
99. Weaver RG, Hemmelgarn BR, Rabi DM, Sargious PM, Edwards AL, Manns BJ, et al. Association between participation in a brief diabetes education programme and glycaemic control in adults with newly diagnosed diabetes. Diabet Med 2014 Dec;31(12):1610-1614. [doi: 10.1111/dme.12513] [Medline: 24890340]
100. Bongaerts BW, Müssig K, Wens J, Lang C, Schwarz P, Roden M, et al. Effectiveness of chronic care models for the management of type 2 diabetes mellitus in Europe: a systematic review and meta-analysis. BMJ Open 2017 Mar 20;7(3):e013076 [FREE Full text] [doi: 10.1136/bmjopen-2016-013076] [Medline: 28320788]
101. de Boer IH, Bangalore S, Benetos A, Davis AM, Michos ED, Muntner P, et al. Diabetes and hypertension: a position statement by The American Diabetes Association. Diabetes Care 2017 Sep;40(9):1273-1284. [doi: 10.2337/dci17-0026] [Medline: 28830958]
102. Steinert A, Eicher C, Haesner M, Steinhagen-Thiessen E. Effects of a long-term smartphone-based self-monitoring intervention in patients with lipid metabolism disorders. Assist Technol 2018 Jun 26:1-8. [doi: 10.1080/10400435.2018.1493710] [Medline: 29944463]
103. Timpel P, Harst L. German Medical Science.: GMS Publishing House; 2018 Oct 12. Effectiveness of Telemedicine Interventions for the Most Prevalent Chronic Diseases in German Primary Care – A Protocol for an Umbrella Review URL: https://www.egms.de/static/en/meetings/dkvf2018/18dkvf200.shtml [accessed 2020-01-16]
104. Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care 2019 Aug;42(8):1593-1603. [doi: 10.2337/dci19-0028] [Medline: 31177185]
105. Law LM, Wason JM. Design of telehealth trials--introducing adaptive approaches. Int J Med Inform 2014 Dec;83(12):870-880 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.09.002] [Medline: 25293533]
106. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Br Med J 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
107. Kidholm K, Clemensen J, Caffery LJ, Smith AC. The Model for Assessment of Telemedicine (MAST): A scoping review of empirical studies. J Telemed Telecare 2017 Oct;23(9):803-813. [doi: 10.1177/1357633X17721815] [Medline: 28758525]

108. Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of eHealth interventions: systematic literature review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
109. National Institute for Health and Care Excellence. National Institute for Health and Care Excellence. 2019. Evidence Standards Framework for Digital Health Technologies (March 2019) URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2020-01-16]
110. 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]
111. Clarke M, Williamson PR. Core outcome sets and systematic reviews. Syst Rev 2016 Jan 20;5:11 [FREE Full text] [doi: 10.1186/s13643-016-0188-6] [Medline: 26792080]
112. Anker SD, Agewall S, Borggrefe M, Calvert M, Caro JJ, Cowie MR, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. Eur Heart J 2014 Aug 7;35(30):2001-2009. [doi: 10.1093/eurheartj/ehu205] [Medline: 24904027]

Abbreviations

ADA: American Diabetes Association
BP: blood pressure
CV: cardiovascular
DBP: diastolic blood pressure
ESC/ESH: European Society of Cardiology/European Society of Hypertension
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
HbA$_{1c}$: glycated hemoglobin
HDL: high-density lipoprotein
HDL-c: high-density lipoprotein cholesterol
ICT: information and communication technology
LDL: low-density lipoprotein
LDL-c: low-density lipoprotein cholesterol
MD: mean difference
mHealth: mobile health
OQAQ: Oxford Quality Assessment Questionnaire
PICOT: Population, Intervention, Control, Outcome, and Time
RCT: randomized controlled trial
SBP: systolic blood pressure
SMD: standardized mean difference
SNS: social network services
T1D: type 1 diabetes
T2D: type 2 diabetes
TC: total cholesterol
TGC: triglycerides
UKPDS: United Kingdom Prospective Diabetes Study

Edited by G Eysenbach; submitted 28.10.19; peer-reviewed by J Apolinário-Hagen, D Vogel; comments to author 15.11.19; revised version received 26.11.19; accepted 15.12.19; published 18.03.20.

Please cite as:
Timpel P, Oswald S, Schwarz PEH, Harst L
Mapping the Evidence on the Effectiveness of Telemedicine Interventions in Diabetes, Dyslipidemia, and Hypertension: An Umbrella Review of Systematic Reviews and Meta-Analyses
J Med Internet Res 2020;22(3):e16791
URL: http://www.jmir.org/2020/3/e16791/
doi:10.2196/16791
PMID:32186516

Review

An Analysis of the Learning Health System in Its First Decade in Practice: Scoping Review

Jodyn E Platt[1*], MPH, PhD; Minakshi Raj[2*], MPH; Matthias Wienroth[3*], PhD

[1]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, United States
[2]Department of Health Management and Policy, University of Michigan School of Public Health, Ann Arbor, MI, United States
[3]School of Geography, Politics & Sociology, Newcastle University, Newcastle upon Tyne, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Minakshi Raj, MPH
Department of Health Management and Policy
University of Michigan School of Public Health
1420 Washington Heights
Ann Arbor, MI, 48109
United States
Phone: 1 3155593112
Email: miraj@umich.edu

Abstract

Background: In the past decade, Lynn Etheredge presented a vision for the Learning Health System (LHS) as an opportunity for increasing the value of health care via rapid learning from data and immediate translation to practice and policy. An LHS is defined in the literature as a system that seeks to continuously generate and apply evidence, innovation, quality, and value in health care.

Objective: This review aimed to examine themes in the literature and rhetoric on the LHS in the past decade to understand efforts to realize the LHS in practice and to identify gaps and opportunities to continue to take the LHS forward.

Methods: We conducted a thematic analysis in 2018 to analyze progress and opportunities over time as compared with the initial *Knowledge Gaps and Uncertainties* proposed in 2007.

Results: We found that the literature on the LHS has increased over the past decade, with most articles focused on theory and implementation; articles have been increasingly concerned with policy.

Conclusions: There is a need for attention to understanding the ethical and social implications of the LHS and for exploring opportunities to ensure that these implications are salient in implementation, practice, and policy efforts.

(J Med Internet Res 2020;22(3):e17026) doi:10.2196/17026

KEYWORDS

learning health system; review; knowledge management; bioethics; health information exchange

Introduction

Background

In 2007, Lynn Etheredge [1] envisioned the Learning Health System (LHS) in which he described a system aimed at increasing the value of health care without *draconian* cost cutting. He encouraged facilitating what he called *rapid learning* from new evidence for practice and policy. The publication of this first article to explicitly use the language of the LHS and urge the further consideration of such a system coincided with the adoption of electronic health records (EHRs) in clinical settings aimed at integrating clinical, financial, and administrative data [1]. Etheredge [1] described several opportunities for answering key questions about population health and health care delivery. Addressing these questions would demand a change in institutional (eg, Medicaid) and organizational (eg, in hospitals) leadership and in funding structures to advance the use of health information to improve health [1]. Etheredge's proposals envisioned competitive markets led by health plans and providers who use EHRs; payment linked to evidence-based protocols; and Medicaid and the State Children's Health Insurance Program as national leaders in EHR adoption and in the use of EHR research databases. National computer-searchable clinical trial databases and national assessments of new technologies would support these efforts. Here, we considered Etheredge's article [1] as an initial conceptualization of gaps and opportunities for future examination and development of the LHS. Many subsequent articles on the LHS topic have cited the guiding principles presented in Etheredge's article [1,2].

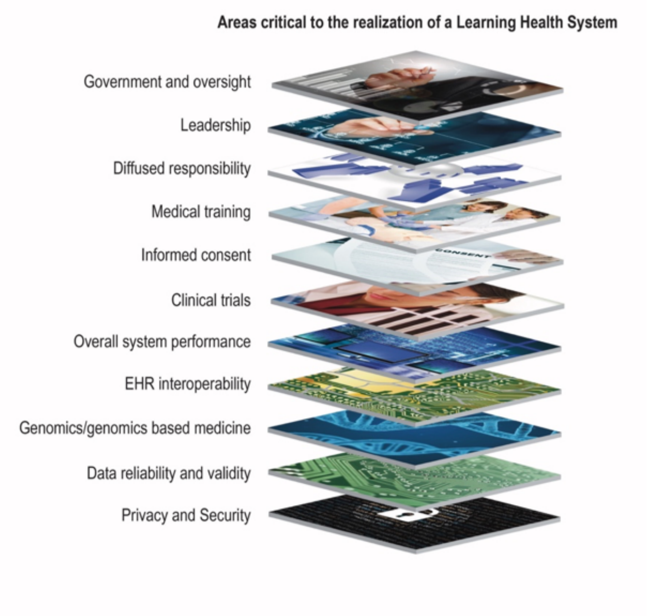
Over a decade has passed since this early vision. Health care provision generates significant amounts of patient and experiential data, and health records, laboratory results, population health surveillance, and patient-generated data, that can be agglomerated and analyzed within health systems. These activities are the result of (1) an increase in health data availability within growing information technology systems via the widening use of patients' EHRs; (2) efforts to increase the volume of clinical research with patients undertaken at health care facilities; and (3) considerable research and data generation by government-funded and commercial enterprises [3-8]. Bringing together diverse actors in the health and life sciences context, LHSs aim to gather and analyze differently sourced data to create useful knowledge that is disseminated to all stakeholders, put into practice, and then evaluated [9].

The LHS framework marks a departure from data practices that are governed by the intended use for data, be it research, quality improvement (QI), clinical care, or public health. The aim of LHS is to enable continuously and rapidly operating virtuous cycles of study, feedback, and practice change, regardless of the original intention for data collection; its vision has come to shape the goals of initiatives in the United States, the United Kingdom, and globally [1,2]. Examples include the American Society of Clinical Oncology's CancerLinQ initiative, which offers an emerging large-scale database for an oncology learning community to improve the quality of cancer care [10], the CommonWell Health Alliance, which enables querying of treatment data for over 17 million unique individuals [11], and the National Institutes of Health *All of Us* research initiative, which seeks to connect genomic, health, and social media data from over 1 million individuals [12]. Similar avenues toward large-scale agglomeration, analysis, and sharing of patient health information exist in the United Kingdom, such as the National Health Service–supported data platform, Lambeth DataNet, which provides general physicians and academic researchers with access to data from approximately 350,000 patients [13] and the Connected Health Cities initiative [14], which is building data stores to support direct care across the North of England region.

Objectives

In this review, we aimed to examine the literature and rhetoric on the LHS, identify trends and themes in published efforts undertaken to moving this idea from concept to reality, and assess gaps in the literature that suggest areas for future research and policy considerations that are necessary for moving the LHS forward (Figure 1).

Figure 1. Areas critical to the realization of a learning health system.



Methods

Search Strategy

We conducted a scoping review using search terms "learning health system(s)," "learning health care system(s)," and "learning healthcare system(s)" on PubMed, Web of Science, and Scopus databases to identify peer-reviewed publications. The search was limited to peer-reviewed articles published in the English language between January 2007 and December 2017.

We identified and included articles published in a variety of clinical contexts, with most articles focusing on the United States and written by US-based authors and some focused on the United Kingdom, India, Sweden, Kenya, and China. Articles describing or examining the LHS from clinical, technical, and ethical perspectives were included. Our initial search yielded 542 articles for review, including USA Institute of Medicine (now USA National Academy of Medicine) proceedings and chapters that informed our search but were not included in the final full analyses.

Data Extraction and Article Selection

Following a title and abstract review, we excluded 222 articles. Articles were excluded if they were duplicates, conference proceedings or posters, or not peer reviewed. Results were compared among members of the study team before exclusion. We created a Microsoft Excel spreadsheet charting information from the 320 remaining articles, including (1) article title; (2) year; (3) country; (4) category; (5) concern or focus; (6) field; and (7) number of papers citing the article of interest (Table 1).

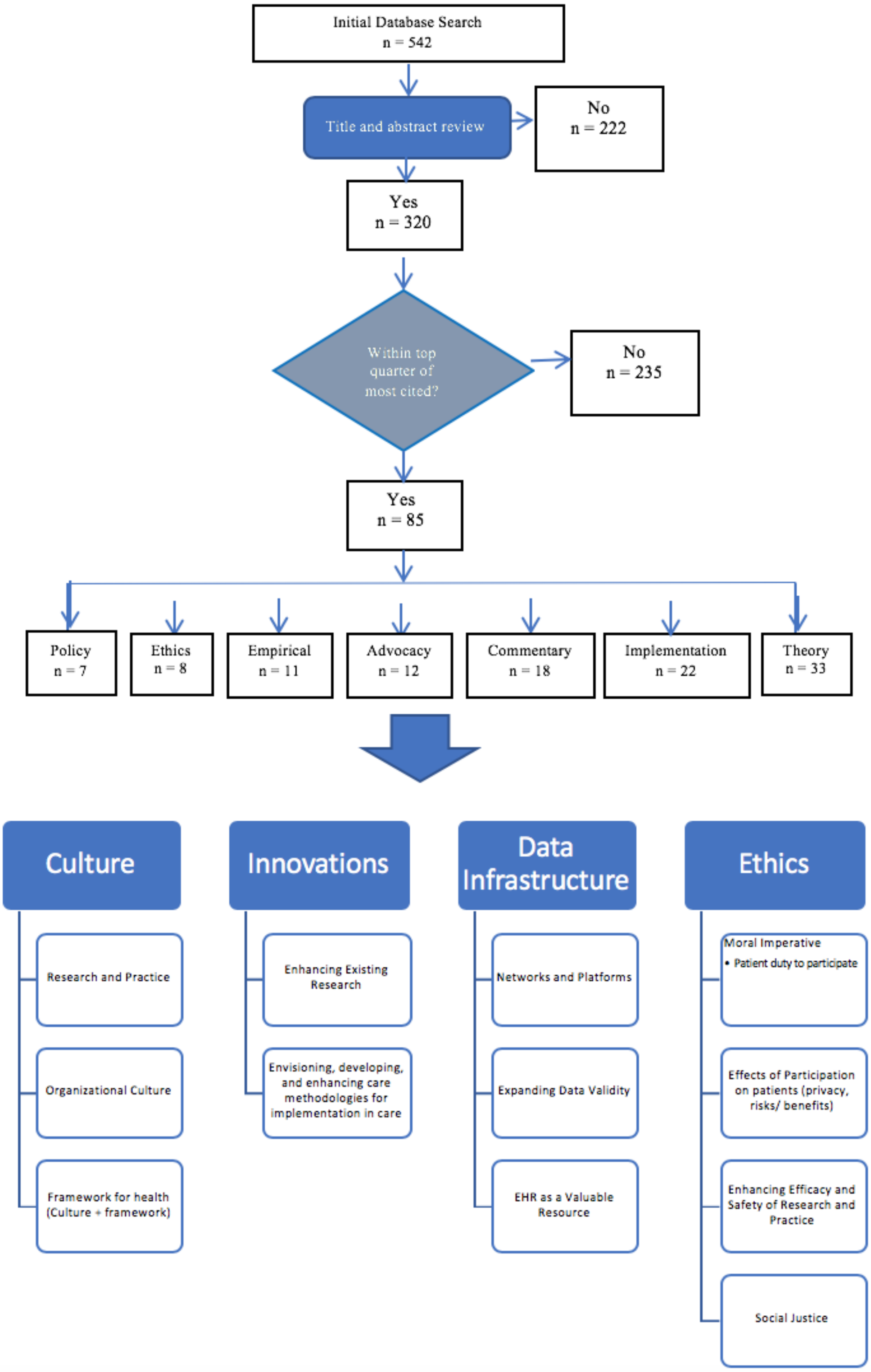
Next, we sorted articles by year, identified the top quarter of most cited papers within each year. We used citation data provided by Web of Science and Scopus to determine the number of citations for each article, sorted all included articles in the sample to identify the top quarter, and then compiled a list of these articles. We chose to focus on the most cited papers as representatives of those with the most salient themes to the discourse and scholarship on the LHS. This sorting resulted in 85 articles for our final review, which represents just over 15.7% (85/542) of the total number of articles identified in our initial search (see Figure 2). The 85 articles were read in full to conduct a more thorough discourse (thematic mapping) analysis that could be compared with the vision laid out in Etheredge's [1] manuscript in 2007, which is described in the section, *Analytical Strategy*.

Table 1. Summary of abstracted information.

Information documented	Description
Article title	Title of publication
Year	Year of publication
Country	Global context of publication
Category	Preset categories for thematic understanding, including the following: • Policy (eg, relating to health reform) • Advocacy (eg, encouraging implementation into organizations) • Theory (ie, generating or enhancing LHS[a] frameworks) • Empirical (eg, studies testing infrastructures or hypotheses) • Implementation (ie, evaluating implementation into clinical contexts) • Ethics (ie, introducing or describing ethical perspectives or limitations of the LHS) • General commentary
Concern and/or Focus	Information on the primary focus or concern of the article, such as the following: • Quality improvement • Personalized health care • Evidence-based medicine • Electronic health record • Ethical oversight
Field	Information on perspective through which article discusses the LHS, such as the following: • Clinical context (eg, oncology and surgery) • Professional context (eg, nursing) • Discourse context (eg, medical informatics, research, and ethics) • Health care system (ie, articles discussing the LHS across a larger scope)
Number of citations	Number of times each article has been cited according to Web of Science or Scopus

[a] LHS: Learning Health System.

Figure 2. Summary of search strategy and themes.



Analytical Strategy

The analysis presented in this paper draws on both quantitative and qualitative approaches. First, we identified historical trends in the LHS discourse by quantitatively assessing the number of articles published on the LHS over time. We then evaluated article categories over time and assessed the variety of clinical domains that have integrated the LHS discourse for the 85 articles selected for review.

Next, we performed a thematic mapping analysis to qualitatively examine the variety of ways in which articles define the LHS. This is a method for extracting, analyzing, and reporting themes in data, such as from interviews or texts [15]. Themes describe textual data that are grouped around a main issue [16] and are recurrent and systematic [17]. Moreover, themes emerge from the data and can be refined into levels or subthemes that reflect on the theme [18]. In this review, we initially identified broad themes that emerged from the articles. These themes were then refined into subthemes. Finally, we compared the qualitative findings from our literature review with the vision developed by Etheredge [1] in 2007 to assess the extent to which these initiatives have been undertaken and discussed according to the literature over the past decade and to identify areas for continued effort. One member of the study team coded the 85 articles using the MAXQDA software and shared codes and code relation matrices with the other members to ensure agreement. The study team held meetings every 2 weeks to discuss emerging themes and subthemes, thus arriving at the final list iteratively and collaboratively.

Results

Defining the Learning Health System

Articles in our final sample most commonly draw upon the Institute of Medicine's (now the National Academy for Medicine) ambitious and encompassing definition describing LHSs as those that: *generate and apply the best evidence for the collaborative health carechoices of each patient and provider;…drive the process of discovery as a natural outgrowth of patient care; and ensure innovation, quality, safety, and value in health care* [19].

These articles describe the LHS as an avenue for delivering more targeted, safe, and effective health care by using information from the experiences and treatment of patients to inform decision making and subsequent care in real time [20,21]. A central ambition, as well as a vehicle for the aspirations of the LHS, is the notion of *learning*. Importantly, learning itself is considered the transfer of knowledge through formal curricula (eg, during medical training) and a transfer of culture, attitudes, and beliefs in ways that can be implemented in research as well as practice when combined with QI methods [21,22]. For example, Faden et al [23] describe learning as a process including research, information gleaned from QI efforts, and comparative effectiveness research, culminating in improved practice [24]. In this way, QI is described as a critical stepping stone to learning, which depends on research and also has the potential to bridge the traditional gap between research and practice to inform the goals of an LHS [24].

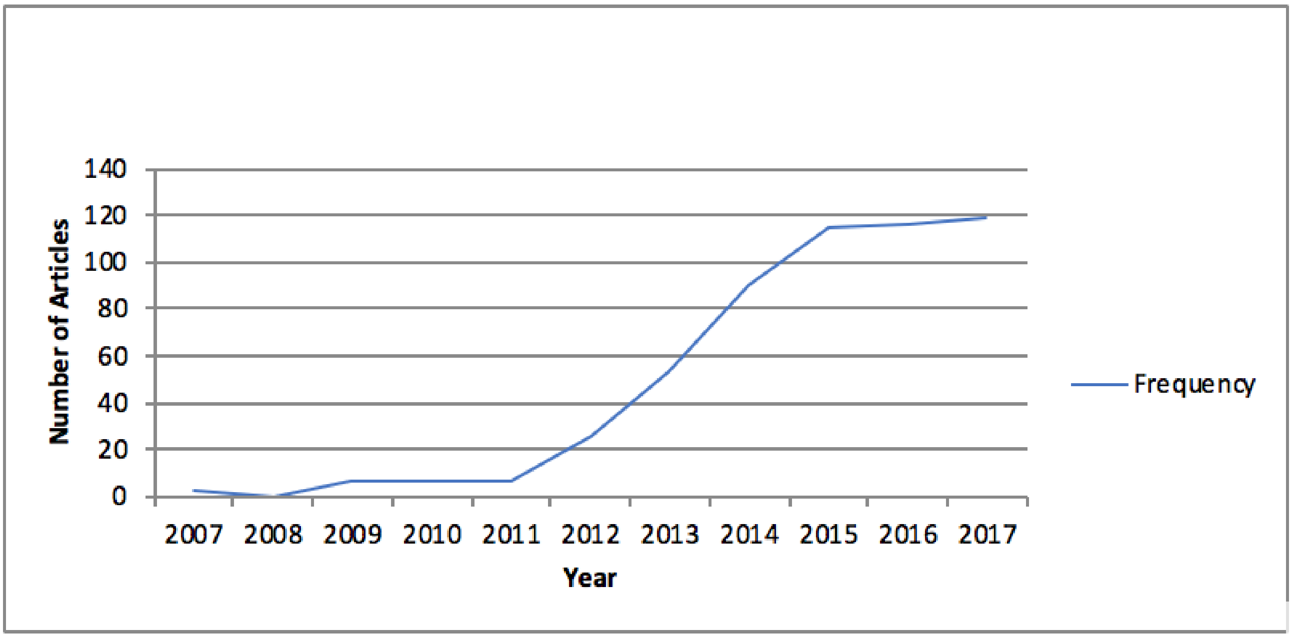
Some articles refer to the LHS more specifically, as a *rapid learning health system* to emphasize the celerity—eg, *real-time* delivery—with which evidence-based medicine can be used to determine health care decisions [1,25]. A system, in this context, is characterized as requiring a coherent, flexible organizational structure, with data maintained in a repository until needed for a particular purpose [26]. In addition, mechanisms are in place to ensure that data are usable by all entities within the system and that they are transferred appropriately, safely, and ethically. Data originate from clinical practice, research, participation, and inquiry provided by organizational leaders, physicians, researchers, patients, and research participants [26]. Although some publications in our sample offer a discussion of the LHS as a large-scale system spanning borders, others implement, evaluate, or present challenges on the system at a local or regional scale.

Quantitative Trends in the Learning Health System Discourse

Our review indicated historical trends in the LHS discourse, with a particular surge in articles published between 2013 and 2017. This trend, we noticed in the broader literature search, is consistent with the trends in articles we selected for review (Figure 3).

Our review resulted in the selection of 85 articles published between 2007 and 2017. Most of the 85 articles included in our review were published in 2014, and the articles cited most were published in 2010 (132 citations); 2007 (116 citations); and 2013 (102 citations).

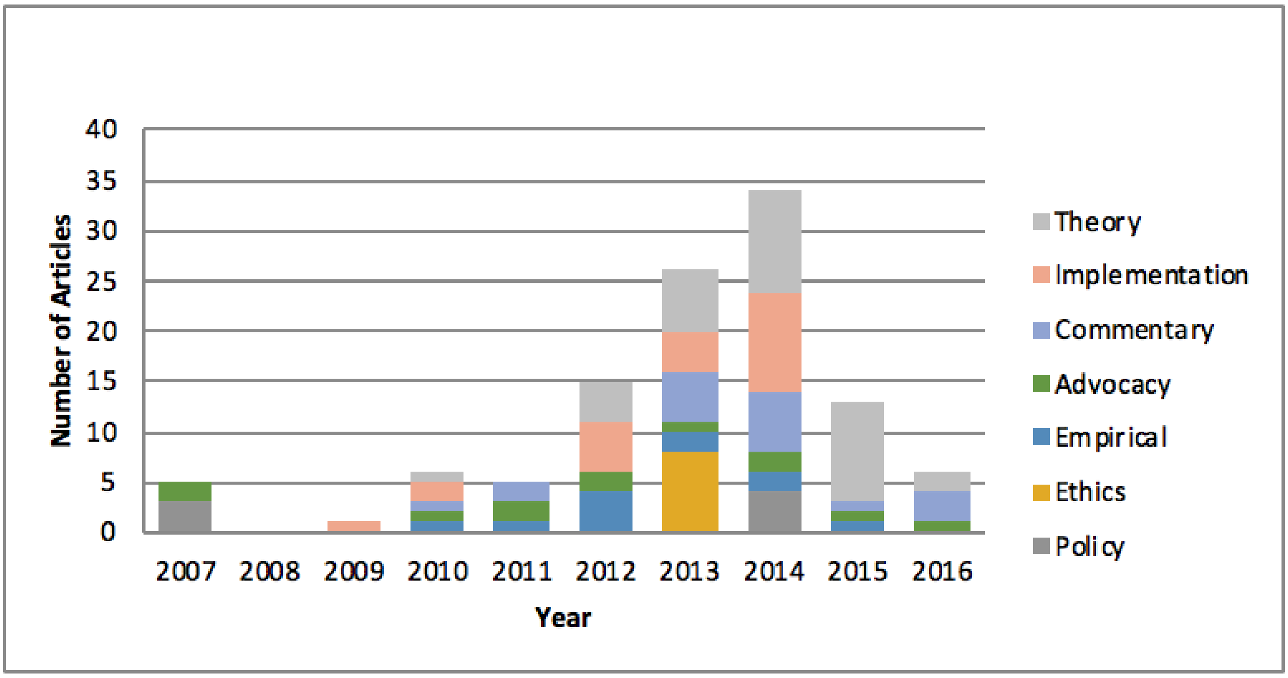
Figure 3. Frequency of articles published by year (n=542).



Trends in Categories of Articles

Following the categorization of each paper according to its primary concerns, we found that a striking majority of articles are concerned with theory (n=33), followed by implementation (n=23), commentary (n=18), advocacy, ie, promoting the idea or driving the demand for an LHS (n=12), empirical data (n=11), ethics (n=8), and policy (n=7). Articles published in 2007 were exclusively theory and/or advocacy; in 2010, articles were published across a broader variety of categories. However, articles predominantly discussing ethics in our sample were only published in 2013; similarly, policy articles were only published in 2007 and 2014 (Figure 4).

Figure 4. Trends in article categories among frequently cited articles based on review of frequently cited articles (n=85).



Most of the implementation and empirical articles discuss the LHS in the context of oncology (n=10) [27-36]; however, other contexts include primary care (n=3) [37,38], surgery (n=3) [20,39], and pediatrics (n=7) as well as others. Some articles, such as commentary articles, exemplify clinical contexts such as pediatrics (n=3) and oncology (n=3) but are not exclusively concerned with a particular clinical context. In Table 2, we summarize the clinical contexts of articles included in our review, including the initiative of interest in each article. Articles were classified based on the objective of the literature—ie, (1) whether the article discusses or is concerned with the use of the initiative for QI, (2) whether the article discusses or is concerned with the use of the initiative for research, and/or (3) whether the article specifically evaluates the initiative with regard to improving quality or research. Most of the articles are classified as being concerned with QI, eg, around EHR point-of-care diagnostic capabilities or in clinical guidelines and subsequent process measures via networks through which information is shared. Some of these articles frame research as a stepping stone for improving point-of-care quality; however, their ultimate objective is not to improve research itself. Some studies evaluate the efficacy of QI initiatives. Fewer articles (n=9) are concerned with initiatives seeking to improve research, eg, through the development of clinical trials, or by considering the limitations in the processes through which data are collected in research used to inform practice (eg, informed consent).

Table 2. Clinical context of articles.

Article type, clinical context, source	Name or focus of initiative	Classification
Implementation [a]		
Oncology		
Buetow and Niederhuber [29]	Cancer Biomedical Informatics Grid	QI[b]
Abernathy et al [28]	Rapid learning health care	QI
Elson et al [27]	Athena Breast Health Network	QI
Sledge et al [30]	CancerLinQ	QI
Schilsky et al [31]	CancerLinQ	QI
Abernathy et al [32]	Innovation in oncology	QI
Primary care		
Delaney et al [37]	Electronic Primary Care Research Network	E[c]
Peterson et al [38]	Practice-based Research Networks	E
Surgery		
Kwon et al [39]	Surgical Care and Outcomes Assessment Program	QI/E
Flum et al [20]	Comparative Effectiveness Research Translation Network	QI/E
Pediatrics		
Forrest et al [40]	PEDSnet consortium of 8 children’s hospitals	QI
Forrest et al [41]	ICN[d]	QI/E
Endocrinology		
Fiore et al [42]	Point-of-care clinical trial	R[e]/E
Cardiology		
Califf and Sugarman [43]	Pragmatic clinical trials	R
Maddox et al [44]	Veterans Administration Clinical Assessment, Reporting, and Tracking program	QI/E
Intensive Care Unit		
Warner et al [45]	Temporal phenotype data via an EHR[f]	QI/E
Gastro-enterology		
Forrest et al [41]	ICN	QI/E
Public Health Departments		
Klann et al [46]	Query Health	QI/E
Health systems		
Greene et al [47]	Rapid learning health system	QI
Harper [48]	Clinical Demand Index	QI/E
Weng et al [49]	Integrated Model for Patient Care and Clinical Trials	R/E
Forrest et al [40]	PEDSnet consortium of 8 children’s hospitals	QI
McGlynn et al [50]	Patient Outcomes Research to Advance Learning network	R
Mandl et al [51]	Scalable Collaborative Infrastructure for a Learning Healthcare System	R/E
Empirical		
Pediatrics		
Lannon and Peterson [52]	Pediatric Collaborative Improvement Networks	QI/E
Kelley et al [53]	Examining attitudes toward research	R

Article type, clinical context, source	Name or focus of initiative	Classification
Oncology		
Spinks et al [33]	Not applicable	Not applicable
Surgery		
Pingleton et al [21]	Evaluating quality and patient safety curricula	QI/E
Nephrology		
Kelley et al [53]	Not applicable	Not applicable
Primary care and/or internal medicine		
Pingleton et al [21]	Evaluating quality and patient safety curricula	QI/E
General		
Pediatrics		
Clancy et al [22]	Collaborative Improvement Networks	QI/R
Gardner and Kelleher [54]	LHS[g] for Pediatrics	QI
Kahn et al [55]	Common Pediatric Research Terminology	R/E
Oncology		
Feeley et al [34]	Health Information Technology to improve quality of cancer care	QI
Shaikh et al [35]	Collaborative Biomedicine	QI
Shah et al [36]	CancerLinQ	R/E

[a] Italicized categories describe concern or focus of article.

[b] QI: quality improvement.

[c] E: evaluation.

[d] ICN: ImproveCareNow.

[e] R: research.

[f] EHR: electronic health record.

[g] LHS: Learning Health System

Themes

Our thematic analysis revealed five broad themes across the 85 articles we reviewed: (1) culture, ie, the environment or environmental change required to support the LHS; (2) innovations, including new tools or ideas needed or being developed to realize the vision of an LHS; (3) data infrastructure; and (4) ethical considerations generally framed as moral imperatives or in terms of principles such as privacy and efficiency [15]. Although some articles present evidence or discussion on multiple themes, others tend to focus exclusively on one theme, eg, articles commenting on ethical aspects. Furthermore, articles focusing on technical or research aspects rarely comment on social or ethical considerations of the LHS.

Culture

Most articles (n=81) set clear expectations and priorities for advancing the LHS and achieving its goals. Culture, describing the way in which research and clinical care are considered and how they contribute to the vision or mission of the LHS, is central to enhancing the system. For example, academic health centers are discussed for balancing high-quality teaching with attention to increasing customer service, productivity, and research missions and reducing knowledge gaps for evidence-based medicine [56]. In the reviewed papers, aspects of culture included (1) organizational culture; (2) research and practice; and (3) establishing a framework for health, ie, establishing the boundaries of research, practice, and QI while ensuring appropriate oversight of risks and benefits. These aspects of culture align with the initial priorities set by Etheredge in 2007, including and extending beyond leadership and collaboration, to include actionable steps such as professional education [1].

Innovations, Tools, and Ideas

Most of the reviewed articles (n=77) propose ideas and tools or evaluate innovations that are anticipated to contribute to achieving an LHS by (1) enhancing existing research and (2) envisioning, developing, and enhancing care methodologies to ultimately be implemented in care. For example, the Patient Outcomes Research to Advance Learning network is a research network studying the effectiveness of various approaches to diagnosis, treatment, and management, which could create cohorts of patients with common diagnoses to conduct large-scale comparative effectiveness research to accomplish its goals of assisting patients, caregivers, and physicians in making informed decisions [50]. Other tools, in contrast, are concerned with enhancing methodologies for clinical practice. For example, ImproveCareNow (ICN) is a network aiming to transform the health of children with Crohn disease and ulcerative colitis through a collaboration of pediatric gastroenterology practices working together to develop and enhance care methodologies [57]. The innovations, tools, and ideas that have been developed since 2007 seem particularly responsive to Etheredge's envisioning of moving toward an LHS via the EHR, predictive modeling, and software development [1].

Data Infrastructure

Data infrastructure emerges as a key theme in articles (n=67), with authors discussing a variety of technical avenues for pursuing and/or achieving the LHS. Given that the LHS framework is heavily reliant on an emerging technical infrastructure, the significance of this theme is not surprising. Publications are concerned with (1) networks and platforms; (2) expanding study and data reliability and/or validity; and (3), particularly, the EHR as a valuable resource. In 2007, Etheredge [1] discussed the Cancer Research Network, the Vaccine Safety Datalink network, and the American Medical Group Association as exemplars of data infrastructure progressing toward an LHS. A decade later, the literature is discussing myriad networks and platforms such as the National Patient-Centered Clinical Research Network and the data-sharing platform, PopMedNet [46,50].

The literature argues that these networks and platforms rely on robust participation to build reliability and validity; however, reported participation is limited owing to inadequate sampling, limited availability of clinical information in datasets, lack of rigorous inclusion and/or exclusion criteria resulting in confounders, and resulting issues around generalizability of findings [58-60]. Furthermore, the quality of data suffers because of issues such as inconsistent terminology and the use of statistical techniques that are not advanced enough for complex observational data aggregated from the EHR [37,61]. However, articles regard the EHR as a valuable resource with the capacity to capture, communicate, aggregate, store, and analyze large pools of data for real-time clinician decision support and, broadly, for establishing a nationwide LHS [26,30,48]. Indeed, articles address limitations of the EHR spanning from the practice of individual teams building their own data repositories to the need for manual data aggregation [25,39]. However, articles urge enhancing data interoperability and tailoring the technology to specific clinical contexts such as oncology to ensure effective use of the EHR [28].

Ethics

According to Faden et al [23], the development of an LHS relies heavily on altruism and the understanding that although participating in research may not offer personal therapeutic or curative benefits, what is inherent to clinical research is the potential for large-scale benefits to society by filling knowledge gaps and enhancing care methods. Only eight papers focus on ethics, and most of these are part of the same special issue of the Hastings Center Report in 2013 [24,60,62]. Yet, many other papers—just under half of the reviewed papers (n=39)—discuss at least some ethical and/or social aspects of the LHS, including what is often posited as *challenges* in combining research and clinical frameworks and cultures. Specifically, articles address (1) efficacy and patient safety in research and practice, (2) the moral imperative on patients to participate in research, (3) effects of participation, eg, on privacy and the lack of guaranteed therapeutic benefits in research studies, and (4) questions of social justice (ie, fair subject selection and a just distribution of research benefits and burdens) [23,24,43,63]. For example, patients who are actively undergoing treatment may be more willing to participate in research as a result of their current experience of benefiting from treatment and scientific knowledge. However, as participants in an LHS, EHRs and other data sources will be used during, and after, specific instances of being a patient. Notions of benefit and risk are likely to change over time, both for a specific individual and within the context of a given institution. Ongoing conversations about the risks and benefits of participation, and what the idiom *research* refers to, will be critical to achieving adequate engagement and ensuring that willingness to participate is not confounded with a need for care [64,65]. Furthermore, potential users of health care services may anticipate receiving further and/or advanced treatment for long-term and rare diseases. For example, in the context of the United States, health care is not free of charge and features some of the highest costs globally for health services and long-term care. A question of social justice would be how data from patients are used for the health system to ultimately give back to patients who are charged for services and whose health and associated data may be turned into monetizable assets by health care providers.

It is a positive sign that such a significant number of articles in the sample engage with ethical and social aspects of the LHS. However, this engagement still focuses primarily on efficacy and safety and on espousing a moral duty for physicians, clinicians, and patients to participate in the LHS. In his 2007 article, Etheredge [1] discusses the ethical implications of an LHS, particularly in terms of responsibility, gaps in our understanding of comparative benefits and risks of clinical research, and prescriptions across minority and special needs groups, and raises issues of patient confidentiality. The literature, to date, concerned with ethical implications of an LHS, though sparse, continues to prioritize these gaps. Furthermore, it urges the consideration of participation and recognizes some of these issues as relating to social justice more broadly.

Progress and Opportunities Over the Past Decade

Over the past decade, there has been tremendous movement with regard to the technical side of LHS infrastructure including the EHR, which is discussed in much of the literature as a valuable resource for research, clinical care, and the translation from one to the other [66,67]. Much of the literature in our review focuses on using EHRs for QI, signaling progress in the use and impact of health information technology on moving toward the vision of the LHS. However, our findings suggested a lack of attention to ethical considerations, eg, social justice issues, to intersections of different types of data that might enter potentially large-scale data systems such as the LHS (health, welfare, education, criminal justice, etc), and to the role, recruitment, and retention of participants in research, and equitable distribution of benefits. These aspects of health care and research have been discussed in other bodies of health literature but do not seem to have been adapted widely for and by the LHS discourse. There is also a need for further consideration of leadership, collaboration, and responsibilities of various entities within the health care system. Etheredge [1] cites Medicaid as a key leader of the LHS; for leaders to emerge, it is essential to identify, understand, and map the responsibilities of LHS leaders and identify how other stakeholders such as providers, patients, and hospitals conceptualize the LHS as a sociotechnical system.

Discussion

Principal Findings

Since Etheredge's paper [1] and the US Institute of Medicine's roundtable report in 2007 [2], more than 500 academic papers have been published discussing the LHS, with most of these emerging since 2014. This work predominantly cites empirical research describing technical developments and analytical capacity. However, LHS infrastructure concepts navigate complex systems of policy, ethics, networks, and processes in local clinical care and research settings as well as in the global consortia. Although the literature is expanding with the rate of technological innovation, an empirically based ethical analysis of large health data systems, including the deliberation of their ethical management and sociotechnical engagements, further requires considerable attention.

Articles in our review generally suggest an engagement with the LHS discourse by exploring or evaluating tools and innovations to bring the LHS framework to fruition in research and practice; by examining technical infrastructures to test the process of aggregating data from research to implement findings into practice; and by evaluating challenges and subsequent priorities to facilitate the LHS across teams and institutions. Our sample suggests that the LHS framework is considered suitable for, and is in parts already utilized across, a variety of institutions and clinical contexts, albeit not anywhere near the capacity of what the LHS framework suggests might be possible. The publication numbers themselves are indicative of the desire and effort to present, develop, and apply the LHS in action as research, technical, and clinical teams collaborate and propose new models. However, while some articles test the LHS framework as an avenue to pursuing QI, others consider the LHS a framework for implementable research. There is a need to examine whether, in applying an LHS model, a boundary exists between QI and research, and, further, the terms and implications of such a boundary. At present, the expectations and requirements for informed consent, notification, return of research results, risk assessment, privacy, and confidentiality are vastly different for research vs QI.

Future studies should examine stakeholder perceptions of the boundaries between QI and research, engaging patients, healthy participants, researchers, clinicians, and technical designers involved in LHS teams to understand the extent to which they distinguish between the two. Importantly, the ways that the LHS might be utilized need to be clarified to develop appropriate and effective governance frameworks.

Our review revealed a number of trends and themes in the LHS discourse that may inform avenues for further deliberation to contribute to a more robust discussion of the LHS. For example, the relatively small number of articles comprehensively engaging with its social, ethical, and governance aspects is surprising. This may suggest limited discussion of the LHS as a concept in the social sciences and humanities and an opportunity for further examination of aspects and implications of the LHS from these perspectives—ongoing discussions in bodies of social science scholarship around health and data would provide fruitful platforms for further discussion. The LHS relies on the collaboration of various stakeholders in research, clinical, and technical arenas to utilize technology for the improvement of patient health, in fact forming an ecosystem of interrelated and linked activities. One might expect that these interpersonal interactions as well as interactions between stakeholders and technologies would prompt further discussion of the social and ethical concerns that might arise or of the social and societal enablers, drivers, and impacts of the LHS. For example, the intersection between the patient duty to participate and the technical challenge of maintaining privacy, including data security, could be examined further to understand whether there are nuances to when patients may be more or less willing to share data across a network and to understand what the processes for decision making and data sharing need to look like to be compatible with data subjects' needs [23].

To date, the LHS has primarily focused on the technical access between different users (via EHRs) and across local networks; this is reflected in the articles in our sample as well, suggesting a need for tests of implementation at a larger scale. ICN is an example of a regional-level network; there is the possibility of testing similar networks at a global level. It is possible that owing to our criteria of including only articles written in English, our sample excluded global-level examples of LHSs; however, it is unclear whether there are current collaborations between global settings and whether this is a part of the current LHS vision. We do know that aspects of types of LHSs—such as data governance in health care research—are being discussed in other bodies of the literature but that scholarship either does not refer to the LHS or does not take a systemic approach to health research and care.

Implications for Future Research, Practice, and Policy

Our review suggests that although some knowledge gaps and uncertainties presented by Etheredge [1] have received consideration over the past decade, significant gaps persist, presenting opportunities for future research and policy (Table 3).

Table 3. Progress over the past decade in addressing knowledge gaps and uncertainties.

Knowledge gaps and uncertainties (Etheredge, 2007) [1]	Perspectives from the literature (2007-2017)	Areas for considering the way forward
Diffused responsibility	Robust discussion of the need for leadership and coordination across various stakeholders, research, policy, and practice. Tools exist for development of new technologies; however, they appear to be largely segregated into those for enhancing research vs practice, leading to disparate regulatory oversight	Further examination is needed of stakeholder perspectives of funding and responsibility for supporting research. Need for a better understanding of the roles and responsibilities of those involved in research from collection to sharing and implementation
Concerns about clinical trials	The past decade has seen a variety of innovations for enhancing research as well as clinical care, along with the emergence of networks and platforms for trials and studies across various clinical contexts. The literature sees a continued need for clinical trials to support precision medicine but asserts the need for ensuring that participants are well-informed about risks and benefits of participating, and that sampling and distribution of benefits are equitable	Continued examination of public needs, attitudes, beliefs, and knowledge about participating in research and clinical trials, especially at a systemic level of health research and care data agglomeration and learning cycles
Consequences of underinvestment	The literature suggests that the emergence of networks and platforms, as well as the increasing use of the EHR[a] has facilitated investment in evidence-based research. However, there is a need for ensuring that data are interoperable and meaningful, that studies and data are reliable and valid, and that the public is better able to envision its role in research	Need for an enhanced design of the EHR to increase interoperability, standardization, and quality of data to ensure that findings are translatable and generalizable. Ensuring adequate regulation of research and informed consent for participants
Genetics and genomics-based medicine	Data quality, reliability, and validity are critical to genetics and genomics-based medicine. The literature primarily discusses a need for ensuring the just and equitable distribution of benefits and minimization of risks	Sampling must include patient populations beyond the socioeconomic environment of a hospital or region, and rigorous inclusion and exclusion criteria
Overall system performance	The literature extensively discusses the EHR as a resource for improving quality; several articles urge for a consideration of research vs quality improvement. In addition, there is a need for improving and developing curricula in academic medical centers on research and clinical safety and quality	Moving the LHS[b] forward requires further theoretical consideration of research vs quality improvement. The literature is mixed and defining boundaries will be beneficial to examining the policy and practice implications of rapid learning

[a] EHR: electronic health record.

[b] LHS: Learning Health System.

Several authors have theorized and advocated the need and potential for the LHS to enhance patient safety and social justice. However, few empirical studies have actually examined the effects of immersing in an LHS culture on the patient-doctor relationship, on professional relationships in health systems, or on public accountability and governance. Generally, only a small subset of publications is acutely engaged with key ethical concerns such as trust, solidarity, equity, or privacy. Those that are engaged with these ethical concerns primarily do so from the perspective of calling attention to these concepts as potential rather than actual concerns. Given the ideal of a technical infrastructure that supports and is supported by a social and cultural learning system, it is important for studies to examine whether and how, eg, relationships between providers and administrative leaders may be different in institutions that have implemented an LHS-based technical framework. As such, we suggest that rather than considering concepts such as trust as a challenge, burden, or barrier necessary to master or overcome, the LHS discourse requires studies on how trust affects and is affected by the LHS context to understand its potential to enhance the LHS, if strengthened.

Many cited challenges and needs in our review align with governance concerns—eg, limitations of data safety and reliability, validity of data, or validity of the algorithmic outputs of data in an LHS. While this literature appears to focus on *overcoming these hurdles* through consideration of the moral obligation of researchers and participants, there was little indication of future research specifically examining the ethical and social aspects of health care and clinical research, eg, through the lens of dignity, equity, equality, and social justice. We suggest that future research needs to examine recruitment and participation in research with a particular focus on the moral imperative, to better understand implications of various approaches to enhancing participation in research on the LHS. Furthermore, the discourse around the LHS needs to attend more to the ethical implications of such systems on patients, participants, and their data and those who may share patient or participant data from both technical and ethical perspectives (eg, physicians, hospitals, public health departments, and payers). For example, it would be important to clearly understand and present practical (eg, economic and therapeutic), social, and ethical rationales for patient participation in clinical research at a systemic level of data agglomeration and analysis, outputs of which may be subject to commercial interests.

The EHR is considered a valuable resource in publications in our review, a number of articles also point specifically to this resource as requiring improvements to effectively and ethically carry through the LHS vision. Notably, EHRs are financially incentivized in the United States, making the local context ideal for implementing the LHS nationally. At the time Etheredge [1] made his call for the LHS, the Health Information Technology for Economic and Clinical Health (HITECH) Act and the US $35 billion investment in the adoption and meaningful use of EHRs had yet to materialize [68]. Since the HITECH Act, adoption and use of the EHR has continued to increase with evidence of better health outcomes after systems have had time to mature [67,69]. However, studies have also suggested consequences of EHR use on user (patient and provider) satisfaction, calling to attention the need for considering and understanding implications of the LHS on patients, and the ways in which their embeddedness in the LHS is influenced by the systemic pressures driving health information technology [70-72]. In addition, there is a need to improve ethical governance, and that continues to include necessary progress on enabling data security. Importantly, improving security does not necessarily mean restricting data from leaving the environment of a screen in an office room at a particular institution; however, there is a need to better understand the boundaries of data as understood by stakeholders including patients, clinicians, and researchers, and their rights and views on how data are retained and used. Furthermore, the information that is stored on the EHR is not always in a standardized format—eg, some data may be in the form of narrative text, whereas others may be more easily coded. For information to be usable, it is necessary to identify ways to make EHRs interoperable and, further, the information on them interpretable in a consistent and standardized way, eg, by using natural language processing techniques [73,74].

Although our review primarily focused on papers based in the United States, there is a need to examine the role and implications of an LHS in other health care system contexts. For example, the expectations and incentives for EHR use may vary across contexts. The procedures for recruitment and enrollment in research may also vary, and the principles governing use and sharing of data and health information may be different. In the US context, historical injustices in the research domain have brought to the forefront regulatory principles such as the Belmont Report describing the ethical foundations of research [75]. In addition, the influx of new actors within the LHS and big data practices expands questions related to who is doing research and innovation, including sources of funding and commercial interests. These evolutions are likely to guide the direction of work related to developing, implementing, and attending to the ethical and social considerations of an LHS [76]. As other countries and health care contexts envision and implement the LHS, it will be important to study how the histories and policies of these countries influence the LHS and how key stakeholders perceive and have expectations for what an LHS can deliver.

Finally, our sample of publications is indicative of the successes or anticipated challenges with the LHS; however, learning also results from an understanding of what goes wrong. Attempts at implementing an LHS along with evaluations of the process should be disseminated as well, for a more robust understanding of what might be optimal conditions for an LHS and what contexts may need more malleability to adopt an LHS framework and culture. For example, future studies may seek to identify the conditions under which the use of emerging technologies and delivery models achieve the goals of an LHS. For example, are the adoption of telemedicine and the use of patient-generated data via health apps and wearable devices contributing to systems that are also engaged in learning and/or feedback? Furthermore, how are the technical, social, and ethical implications of learning from clinical care delivered through telemedicine different from learning and knowledge generation that occurs through traditional, face-to-face care delivery?

Conclusions

The past decade has observed tremendous progress in the technical infrastructure and innovative spirit of advancing the framework and implementation of the LHS. However, there are opportunities to continue to examine and develop policies ensuring the cultural and ethical imperatives for an LHS. Since Etheredge [1] presented an initial vision of the LHS in 2007, the growing implementation of the EHR and emergence of networks and platforms have contributed to the initial stages of understanding the LHS framework. However, to continue to define and realize the aims of an LHS, there is a need to consider the boundaries of research, practice, and QI. Furthermore, there is a need for engaging with various stakeholders including patients, participants, providers, and organizational leaders in clinical, technical, administrative, and research domains to understand the broader implications of an LHS.

Acknowledgments

The authors thank M Grace Trinidad for developing Figure 1 that illustrates the areas critical to the realization of an LHS.

Conflicts of Interest

None declared.

References

1. Etheredge LM. A rapid-learning health system. Health Aff (Millwood) 2007;26(2):w107-w118. [doi: 10.1377/hlthaff.26.2.w107] [Medline: 17259191]
2. Olsen L, Aisner D, McGinnis JM, Roundtable on Evidence-Based Medicine, Institute of Medicine. The Learning Healthcare System: Workshop Summary. Washington, DC: National Academies Press; 2007.

3. Dean BB, Lam J, Natoli JL, Butler Q, Aguilar D, Nordyke RJ. Review: use of electronic medical records for health outcomes research: a literature review. Med Care Res Rev 2009 Dec;66(6):611-638. [doi: 10.1177/1077558709332440] [Medline: 19279318]
4. Treweek S, Mitchell E, Pitkethly M, Cook J, Kjeldstrøm M, Taskila T, et al. Strategies to improve recruitment to randomised controlled trials. Cochrane Database Syst Rev 2010 Jan 20(1):MR000013. [doi: 10.1002/14651858.MR000013.pub4] [Medline: 20091668]
5. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
6. Bourne PE, Lorsch JR, Green ED. Perspective: sustaining the big-data ecosystem. Nature 2015 Nov 5;527(7576):S16-S17. [doi: 10.1038/527S16a] [Medline: 26536219]
7. Hafen E, Kossmann D, Brand A. Health data cooperatives - citizen empowerment. Methods Inf Med 2014;53(2):82-86. [doi: 10.3414/ME13-02-0051] [Medline: 24514946]
8. Platt J, Kardia S. Public trust in health information sharing: implications for biobanking and electronic health record systems. J Pers Med 2015 Feb 3;5(1):3-21 [FREE Full text] [doi: 10.3390/jpm5010003] [Medline: 25654300]
9. Flynn AJ, Friedman CP, Boisvert P, Landis-Lewis Z, Lagoze C. The Knowledge Object Reference Ontology (KORO): a formalism to support management and sharing of computable biomedical knowledge for learning health systems. Learn Health Syst 2018 Apr;2(2):e10054 [FREE Full text] [doi: 10.1002/lrh2.10054] [Medline: 31245583]
10. ASCO CancerLinQ. URL: https://www.cancerlinq.org/ [accessed 2020-01-23]
11. Asnaani J. HealthIT. 2015 Aug. CommonWell Health Alliance: Interoperability for the Common Good URL: https://www.healthit.gov/sites/default/files/facas/ITF_CommonWell_POV_2015-08-14_v2.pdf [accessed 2020-01-23]
12. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795 [FREE Full text] [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
13. NHS Lambeth CCG. Lambeth DataNet URL: https://www.lambethccg.nhs.uk/your-health/Information-for-patients/lambeth-datanet/Pages/default.aspx [accessed 2020-01-23]
14. Connected Health Cities. CHC HUB URL: https://www.connectedhealthcities.org/chc-hub/ [accessed 2020-01-23]
15. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: Sage; 1998.
16. Brink P, Wood MJ, Marilynn JW, Pamela JB. Advanced Design in Nursing Research. Thousand Oaks, CA: Sage; 1997.
17. Polit DF, Hungler BP, Beck CT. Essentials of Nursing Research: Methods, Appraisal, and Utilization. Philadelphia: Lippincott Williams & Wilkins; 1994.
18. DeSantis L, Ugarriza DN. The concept of theme as used in qualitative nursing research. West J Nurs Res 2000 Apr;22(3):351-372. [doi: 10.1177/019394590002200308] [Medline: 10804897]
19. Olsen L, Saunders RS, McGinnis JM, Institute of Medicine, Roundtable on Value & Science-Driven Health Care, The Learning Health System Series. Patients Charting the Course: Citizen Engagement and the Learning Health System: Workshop Summary. Washington, DC: National Academies Press; 2011.
20. Flum DR, Alfonso-Cristancho R, Devine EB, Devlin A, Farrokhi E, Tarczy-Hornoch P, CERTAIN Collaborative. Implementation of a 'real-world' learning health care system: Washington State's Comparative Effectiveness Research Translation Network (CERTAIN). Surgery 2014 May;155(5):860-866. [doi: 10.1016/j.surg.2014.01.004] [Medline: 24787113]
21. Pingleton SK, Davis DA, Dickler RM. Characteristics of quality and patient safety curricula in major teaching hospitals. Am J Med Qual 2010;25(4):305-311. [doi: 10.1177/1062860610367677] [Medline: 20606210]
22. Clancy CM, Margolis PA, Miller M. Collaborative networks for both improvement and research. Pediatrics 2013 Jun;131(Suppl 4):S210-S214. [doi: 10.1542/peds.2012-3786H] [Medline: 23729762]
23. Faden RR, Beauchamp T, Kass N. Learning health care systems and justice. Hastings Cent Rep 2011;41(4):3. [doi: 10.1002/j.1552-146x.2011.tb00105.x] [Medline: 21845906]
24. Faden RR, Kass NE, Goodman SN, Pronovost P, Tunis S, Beauchamp TL. An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics. Hastings Cent Rep 2013;43(s1):S16-S27. [doi: 10.1002/hast.134] [Medline: 23315888]
25. Etheredge LM. Rapid learning: a breakthrough agenda. Health Aff (Millwood) 2014 Jul;33(7):1155-1162. [doi: 10.1377/hlthaff.2014.0043] [Medline: 25006141]
26. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med 2010 Nov 10;2(57):57cm29. [doi: 10.1126/scitranslmed.3001456] [Medline: 21068440]
27. Elson SL, Hiatt RA, Anton-Culver H, Howell LP, Naeim A, Parker BA, Athena Breast Health Network. The Athena Breast Health Network: developing a rapid learning system in breast cancer prevention, screening, treatment, and care. Breast Cancer Res Treat 2013 Jul;140(2):417-425. [doi: 10.1007/s10549-013-2612-0] [Medline: 23887672]
28. Abernethy AP, Etheredge LM, Ganz PA, Wallace P, German RR, Neti C, et al. Rapid-learning system for cancer care. J Clin Oncol 2010 Sep 20;28(27):4268-4274 [FREE Full text] [doi: 10.1200/JCO.2010.28.5478] [Medline: 20585094]

29. Buetow KH, Niederhuber J. Infrastructure for a learning health care system: CaBIG. Health Aff (Millwood) 2009;28(3):923-4; author reply 924. [doi: 10.1377/hlthaff.28.3.923-a] [Medline: 19414908]
30. Sledge GW, Hudis CA, Swain SM, Yu PM, Mann JT, Hauser RS, et al. ASCO's approach to a learning health care system in oncology. J Oncol Pract 2013 May;9(3):145-148 [FREE Full text] [doi: 10.1200/JOP.2013.000957] [Medline: 23942494]
31. Schilsky RL, Michels DL, Kearbey AH, Yu PP, Hudis CA. Building a rapid learning health care system for oncology: the regulatory framework of CancerLinQ. J Clin Oncol 2014 Aug 1;32(22):2373-2379. [doi: 10.1200/JCO.2014.56.2124] [Medline: 24912897]
32. Abernethy A, Abrahams E, Barker A, Buetow K, Burkholder R, Dalton WS, et al. Turning the tide against cancer through sustained medical innovation: the pathway to progress. Clin Cancer Res 2014 Mar 1;20(5):1081-1086 [FREE Full text] [doi: 10.1158/1078-0432.CCR-13-3430] [Medline: 24523437]
33. Spinks TE, Ganz PA, Sledge GW, Levit L, Hayman JA, Eberlein TJ, et al. Delivering high-quality cancer care: the critical role of quality measurement. Healthc (Amst) 2014 Mar 1;2(1):53-62 [FREE Full text] [doi: 10.1016/j.hjdsi.2013.11.003] [Medline: 24839592]
34. Feeley TW, Sledge GW, Levit L, Ganz PA. Improving the quality of cancer care in America through health information technology. J Am Med Inform Assoc 2014;21(5):772-775 [FREE Full text] [doi: 10.1136/amiajnl-2013-002346] [Medline: 24352553]
35. 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 7;16(4):e101 [FREE Full text] [doi: 10.2196/jmir.2496] [Medline: 24711045]
36. Shah A, Stewart AK, Kolacevski A, Michels D, Miller R. Building a rapid learning health care system for oncology: why CancerLinQ collects identifiable health information to achieve its vision. J Clin Oncol 2016 Mar 1;34(7):756-763. [doi: 10.1200/JCO.2015.65.0598] [Medline: 26755519]
37. Delaney BC, Peterson KA, Speedie S, Taweel A, Arvanitis TN, Hobbs FD. Envisioning a learning health care system: the electronic primary care research network, a case study. Ann Fam Med 2012;10(1):54-59 [FREE Full text] [doi: 10.1370/afm.1313] [Medline: 22230831]
38. Peterson KA, Lipman PD, Lange CJ, Cohen RA, Durako S. Supporting better science in primary care: a description of practice-based research networks (PBRNs) in 2011. J Am Board Fam Med 2012;25(5):565-571 [FREE Full text] [doi: 10.3122/jabfm.2012.05.120100] [Medline: 22956691]
39. SCOAP Collaborative, Writing Group for the SCOAP Collaborative, Kwon S, Florence M, Grigas P, Horton M, Horvath K, et al. Creating a learning healthcare system in surgery: Washington State's Surgical Care and Outcomes Assessment Program (SCOAP) at 5 years. Surgery 2012 Feb;151(2):146-152 [FREE Full text] [doi: 10.1016/j.surg.2011.08.015] [Medline: 22129638]
40. Forrest CB, Margolis PA, Bailey LC, Marsolo K, del Beccaro MA, Finkelstein JA, et al. PEDSnet: a national pediatric learning health system. J Am Med Inform Assoc 2014;21(4):602-606 [FREE Full text] [doi: 10.1136/amiajnl-2014-002743] [Medline: 24821737]
41. Forrest CB, Crandall WV, Bailey LC, Zhang P, Joffe MM, Colletti RB, et al. Effectiveness of anti-TNFα for Crohn disease: research in a pediatric learning health system. Pediatrics 2014 Jul;134(1):37-44 [FREE Full text] [doi: 10.1542/peds.2013-4103] [Medline: 24935993]
42. Fiore LD, Brophy M, Ferguson RE, D'Avolio L, Hermos JA, Lew RA, et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. Clin Trials 2011 Apr;8(2):183-195 [FREE Full text] [doi: 10.1177/1740774511398368] [Medline: 21478329]
43. Califf RM, Sugarman J. Exploring the ethical and regulatory issues in pragmatic clinical trials. Clin Trials 2015 Oct;12(5):436-441 [FREE Full text] [doi: 10.1177/1740774515598334] [Medline: 26374676]
44. Maddox TM, Plomondon ME, Petrich M, Tsai TT, Gethoffer H, Noonan G, et al. A national clinical quality program for Veterans Affairs catheterization laboratories (from the Veterans Affairs clinical assessment, reporting, and tracking program). Am J Cardiol 2014 Dec 1;114(11):1750-1757. [doi: 10.1016/j.amjcard.2014.08.045] [Medline: 25439452]
45. Warner JL, Zollanvari A, Ding Q, Zhang P, Snyder GM, Alterovitz G. Temporal phenome analysis of a large electronic health record cohort enables identification of hospital-acquired complications. J Am Med Inform Assoc 2013 Dec;20(e2):e281-e287 [FREE Full text] [doi: 10.1136/amiajnl-2013-001861] [Medline: 23907284]
46. Klann JG, Buck MD, Brown J, Hadley M, Elmore R, Weber GM, et al. Query Health: standards-based, cross-platform population health surveillance. J Am Med Inform Assoc 2014;21(4):650-656 [FREE Full text] [doi: 10.1136/amiajnl-2014-002707] [Medline: 24699371]
47. Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. Ann Intern Med 2012 Aug 7;157(3):207-210. [doi: 10.7326/0003-4819-157-3-201208070-00012] [Medline: 22868839]
48. Harper EM. The economic value of health care data. Nurs Adm Q 2013;37(2):105-108. [doi: 10.1097/NAQ.0b013e318286db0d] [Medline: 23454988]
49. Weng C, Li Y, Berhe S, Boland MR, Gao J, Hruby GW, et al. An Integrated Model for Patient Care and Clinical Trials (IMPACT) to support clinical research visit scheduling workflow for future learning health systems. J Biomed Inform 2013 Aug;46(4):642-652 [FREE Full text] [doi: 10.1016/j.jbi.2013.05.001] [Medline: 23684593]

50. McGlynn EA, Lieu TA, Durham ML, Bauck A, Laws R, Go AS, et al. Developing a data infrastructure for a learning health system: the PORTAL network. J Am Med Inform Assoc 2014;21(4):596-601 [FREE Full text] [doi: 10.1136/amiajnl-2014-002746] [Medline: 24821738]
51. Mandl KD, Kohane IS, McFadden D, Weber GM, Natter M, Mandel J, et al. Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): architecture. J Am Med Inform Assoc 2014;21(4):615-620 [FREE Full text] [doi: 10.1136/amiajnl-2014-002727] [Medline: 24821734]
52. Lannon CM, Peterson LE. Pediatric collaborative improvement networks: background and overview. Pediatrics 2013 Jun;131(Suppl 4):S189-S195. [doi: 10.1542/peds.2012-3786E] [Medline: 23729759]
53. Kelley M, James C, Kraft SA, Korngiebel D, Wijangco I, Rosenthal E, et al. Patient perspectives on the learning health system: the importance of trust and shared decision making. Am J Bioeth 2015;15(9):4-17 [FREE Full text] [doi: 10.1080/15265161.2015.1062163] [Medline: 26305741]
54. Gardner W, Kelleher K. A learning health care system for pediatrics. JAMA Pediatr 2014 Apr;168(4):303-304. [doi: 10.1001/jamapediatrics.2013.5163] [Medline: 24493388]
55. Kahn MG, Bailey LC, Forrest CB, Padula MA, Hirschfeld S. Building a common pediatric research terminology for accelerating child health research. Pediatrics 2014 Mar;133(3):516-525 [FREE Full text] [doi: 10.1542/peds.2013-1504] [Medline: 24534404]
56. Grumbach K, Lucey CR, Johnston SC. Transforming from centers of learning to learning health systems: the challenge for academic health centers. J Am Med Assoc 2014 Mar 19;311(11):1109-1110. [doi: 10.1001/jama.2014.705] [Medline: 24643597]
57. Seid M, Margolis PA, Opipari-Arrigan L. Engagement, peer production, and the learning healthcare system. JAMA Pediatr 2014 Mar;168(3):201-202 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.5063] [Medline: 24446048]
58. Katzan IL, Rudick RA. Time to integrate clinical and research informatics. Sci Transl Med 2012 Nov 28;4(162):162fs41. [doi: 10.1126/scitranslmed.3004583] [Medline: 23197569]
59. Mullins CD, Vandigo J, Zheng Z, Wicks P. Patient-centeredness in the design of clinical trials. Value Health 2014 Jun;17(4):471-475 [FREE Full text] [doi: 10.1016/j.jval.2014.02.012] [Medline: 24969009]
60. Brody H, Miller FG. The research-clinical practice distinction, learning health systems, and relationships. Hastings Cent Rep 2013;43(5):41-47. [doi: 10.1002/hast.199] [Medline: 24092591]
61. Wallace PJ, Shah ND, Dennen T, Bleicher PA, Bleicher PD, Crown WH. Optum Labs: building a novel node in the learning health care system. Health Aff (Millwood) 2014 Jul;33(7):1187-1194. [doi: 10.1377/hlthaff.2014.0038] [Medline: 25006145]
62. Largent EA, Miller FG, Joffe S. A prescription for ethical learning. Commentary. Hastings Cent Rep 2013;43(s1):S28-S29. [doi: 10.1002/hast.135] [Medline: 23315890]
63. Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. Nature 2015 Oct 15;526(7573):336-342 [FREE Full text] [doi: 10.1038/nature15816] [Medline: 26469044]
64. Wienroth M, Caffrey L, Wolfe C, McKevitt C. Patient-initiated recruitment for clinical research: evaluation of an outpatient letter research statement. Health Expect 2018 Apr;21(2):494-500 [FREE Full text] [doi: 10.1111/hex.12642] [Medline: 29164743]
65. Wienroth M, Pearce C, McKevitt C. Research campaigns in the UK National Health Service: patient recruitment and questions of valuation. Sociol Health Illn 2019 Sep;41(7):1444-1461. [doi: 10.1111/1467-9566.12957] [Medline: 31119753]
66. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]
67. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic health record adoption in US hospitals: Progress continues, but challenges persist. Health Aff (Millwood) 2015 Dec;34(12):2174-2180. [doi: 10.1377/hlthaff.2015.0992] [Medline: 26561387]
68. Thune J, Alexander L, Roberts P, Burr R, Enzi M. Health Affairs. 2015 Mar 4. Where Is HITECH's $35 Billion Dollar Investment Going? URL: https://www.healthaffairs.org/do/10.1377/hblog20150304.045199/full/ [accessed 2020-01-23]
69. Lin SC, Jha AK, Adler-Milstein J. Electronic health records associated with lower hospital mortality after systems have time to mature. Health Aff (Millwood) 2018 Jul;37(7):1128-1135. [doi: 10.1377/hlthaff.2017.1658] [Medline: 29985687]
70. Meyerhoefer CD, Sherer S, Deily M, Chou SY, Guo X, Chen J, et al. Provider and patient satisfaction with the integration of ambulatory and hospital EHR systems. J Am Med Inform Assoc 2018 Aug 1;25(8):1054-1063. [doi: 10.1093/jamia/ocy048] [Medline: 29788287]
71. Rose D, Richter LT, Kapustin J. Patient experiences with electronic medical records: lessons learned. J Am Assoc Nurse Pract 2014 Dec;26(12):674-680 [FREE Full text] [doi: 10.1002/2327-6924.12170] [Medline: 25234112]
72. Street RL, Liu L, Farber NJ, Chen Y, Calvitti A, Zuest D, et al. Provider interaction with the electronic health record: the effects on patient-centered communication in medical encounters. Patient Educ Couns 2014 Sep;96(3):315-319 [FREE Full text] [doi: 10.1016/j.pec.2014.05.004] [Medline: 24882086]
73. Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff (Millwood) 2014 Jul;33(7):1163-1170 [FREE Full text] [doi: 10.1377/hlthaff.2014.0053] [Medline: 25006142]

74. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. N Engl J Med 2018 Oct 11;379(15):1452-1462 [FREE Full text] [doi: 10.1056/NEJMra1615014] [Medline: 30304648]
75. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The US Department of Health and Human Services. 1979. The Belmont Report URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html [accessed 2020-01-23]
76. Sharon T. The Googlization of health research: from disruptive innovation to disruptive ethics. Per Med 2016 Nov;13(6):563-574. [doi: 10.2217/pme-2016-0057] [Medline: 29754546]

Abbreviations

EHR: electronic health record
HITECH: Health Information Technology for Economic and Clinical Health
ICN: ImproveCareNow
LHS: Learning Health System
QI: quality improvement

Edited by G Eysenbach; submitted 12.11.19; peer-reviewed by D Wyatt, S Choi; comments to author 21.12.19; revised version received 30.12.19; accepted 31.12.19; published 19.03.20.

Please cite as:
Platt JE, Raj M, Wienroth M
An Analysis of the Learning Health System in Its First Decade in Practice: Scoping Review
J Med Internet Res 2020;22(3):e17026
URL: http://www.jmir.org/2020/3/e17026/
doi:10.2196/17026
PMID:32191214

Proposal

The Value of Data: Applying a Public Value Model to the English National Health Service

James Wilson[1], BA, PhD; Daniel Herron[2], BSc, PhD; Parashkev Nachev[3], BM BCh, PhD; Nick McNally[2], BA, PhD; Bryan Williams[4], MBBS, PhD; Geraint Rees[5,6,7], BSc, MB BChir, PhD

[1]Department of Philosophy, University College London, London, United Kingdom
[2]Research & Development, UCL Hospitals/UCL Biomedical Research Centre, London, United Kingdom
[3]Institute of Neurology, University College London, London, United Kingdom
[4]UCL Institute of Cardiovascular Sciences, University College London, London, United Kingdom
[5]Institute of Cognitive Neuroscience, University College London, London, United Kingdom
[6]Faculty of Life Sciences, University College London, London, United Kingdom
[7]Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom

Corresponding Author:
James Wilson, BA, PhD
Department of Philosophy
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 207 679 0213
Email: james.wilson@ucl.ac.uk

Abstract

Research and innovation in biomedicine and health care increasingly depend on electronic data. The emergence of data-driven technologies and associated digital transformations has focused attention on the value of such data. Despite the broad consensus of the value of health data, there is less consensus on the basis for that value; thus, the nature and extent of health data value remain unclear. Much of the existing literature presupposes that the value of data is to be understood primarily in financial terms, and assumes that a single financial value can be assigned. We here argue that the value of a dataset is instead relational; that is, the value depends on who wants to use it and for what purposes. Moreover, data are valued for both nonfinancial and financial reasons. Thus, it may be more accurate to discuss the values (plural) of a dataset rather than the singular value. This plurality of values opens up an important set of questions about how health data should be valued for the purposes of public policy. We argue that public value models provide a useful approach in this regard. According to public value theory, public value is created, or captured, to the extent that public sector institutions further their democratically established goals, and their impact on improving the lives of citizens. This article outlines how adopting such an approach might be operationalized within existing health care systems such as the English National Health Service, with particular focus on actionable conclusions.

(J Med Internet Res 2020;22(3):e15816) doi:10.2196/15816

KEYWORDS

health policy; innovation; public value; intellectual property; NHS Constitution

Background

Research and innovation in biomedicine and health care increasingly depend on electronic data. Health care systems are data-rich environments, within which data are collected, used, and shared. The emergence of data-driven technologies and associated digital transformations has brought the value of such data into focus; in particular, the "value proposition" focuses on better patient outcomes relative to cost [1,2]. Despite the broad consensus on the value of data, the basis for that value, and thus its nature and extent, is not clear. It is essential that the conceptual framework within which any valuation of data can be coherently couched is clarified. We here argue that the idea of public value holds much promise for elucidating the value of public sector data. We further highlight how such an approach might be operationalized within health care systems such as England's National Health Service (NHS), with particular focus on reaching actionable conclusions.

Defining Value

To clarify the value of data, it is useful to begin by focusing on the elements of *data* and *value* as separate entities. Most familiar everyday goods such as cars and tables are rival goods: they can only be in one place at a time, and may only be possessed or used by a single user. Moreover, many rival goods in health care research and innovation (eg, tissue or DNA samples) are nondurable such that consumption destroys the good and allows only one user to enjoy it. In contrast, data are a nonrival and durable good: the same data items can simultaneously be used (and subsequently copied for reuse) by many different agents for multifarious purposes without difficulty. For example, the same underlying data can be used to provide direct care to patients, help plan future services, conduct multiple health care research projects, and realize economic benefits through commercialization. Although datasets themselves are nonrival in use and are durable, these same properties may not apply to their value [3]. Data that provide a competitive advantage will lose much of their value or usefulness if they become common knowledge. Data used for location-sensitive advertising while an individual is in a particular shop will quickly depreciate in value. Thus, the fact that data *can* be used for an infinite variety of purposes simultaneously does not necessarily mean that they *will* be, or even that they should be.

There are several ways in which access to goods that are intrinsically nonrival can be governed for reasons of social policy. Commonly used models conceive of data as: (1) a private good, in which a dataset that could be shared much more widely is not, perhaps for ethical or legal reasons, or due to a competitive advantage (as in test data exclusivity in clinical trials); (2) a licensed good, in which access to the data is controlled and managed via approval processes that stipulate factors such as the purposes for which the data can be used; or (3) an open commons, in which the data are openly accessible for anyone to reuse or add value [4]. The optimal governance solution for data in each case will require balancing the ability to generate the data, the opportunity to make use of it, and maintaining the underpinning social license.

The starting point for considering value within the framework of public policy is often economic or market value. This may be driven by an assumption that cost-benefit analysis is the best way to approach public sector decision making, but may also reflect the perceived importance of commercial imperatives within public sector data policy. In any case, this framing naturally excludes valuations for which there is no market, and degrades those where commodification is difficult or impossible. In addition, some aspects of value can be difficult to assess financially. For example, innovation in health care technology from data can create intangible assets that are not physical in nature, such as patents, copyrights, software, goodwill, and brand recognition. Although such intangibles are frequently difficult to value, they nevertheless are considered to represent a significant proportion of the value that can be created from data [5].

It is important to consider values other than financial value as relevant to health care. For example, patients value being treated with respect by health care professionals, and also value being told candidly when something has gone wrong in their care or treatment, but it would be a mistake to consider either of these values as primarily financial. In the context of health care, nonfinancial values are often grouped together under the heading of social value judgments [6].

The value of datasets should be understood in relational terms. That is, the value is not intrinsic to the dataset itself, but rather depends on the varied purposes for which it will be used. Different agents may value the same dataset for different reasons and to different extents. Access to vast datasets of patient-level data will be mission-critical for a company whose main product is health data analytics, but will be merely useful and nonessential for a company with a different focus. Moreover, the same agent may simultaneously value a given dataset in multiple ways. For example, patients may simultaneously place a nonfinancial value on their health data—insofar as they value their privacy, and the assurance of confidentiality within clinical encounters—and a financial value, insofar as they might be willing to sell rights to use the data to researchers or commercial companies. In summary, it may be less misleading to discuss the *values* (plural) of a dataset than to imply that there is a singular value of a dataset, regardless of the context.

The financial value of particular datasets will be determined by a combination of supply-side and demand-side features. On the supply side, these features will include: (1) data quality, (2) the format in which the data are stored and the extent to which this format is machine-readable, (3) the ability to link the data, (4) the type of data (eg, identifiable, anonymized, or aggregate), (5) the quantity of data, and (6) the degree to which the data are actionable [7]. On the demand side, the main factors determining the value will include: (1) the use of the data determined by an organization, and the amount of data required to fulfill this purpose; (2) the wealth and willingness to pay of the organization; and (3) the relative cost of obtaining access to data elsewhere [7]. While much attention is rightly devoted to large-scale datasets that could be viewed as strategic assets, questions about the value of data also need to be addressed within locally implemented systems. Textbox 1 provides a worked example of the value of data in an appointment scheduling system.

Textbox 1. The value of data in an artificial intelligence appointment scheduling system.

An innovative artificial intelligence (AI) solution to health care challenges such as optimizing hospital nonattendance [8] illustrates the intricacies of valuing data. The value of reducing nonattendance is two-fold: first, it accelerates the delivery of care for patients, achieving better outcomes faster; and second, it enhances the productivity of the institution, producing a greater health care impact for the quantity of dedicated resources. The problem is too complex to be solved intuitively, and therefore requires a solution drawn from historical data, such as a predictive model of nonattendance that enables appropriately focused reminding. The higher the predictive fidelity, the more efficient the interventions introduced will be to maximize attendance.

The extent to which a datum contributes to the fidelity of the model will depend on the complexity of the underlying problem: simple problems are data-efficient, requiring only small samples, whereas complex problems such as the present example are often profligate, requiring large, diverse collections of data.

Although the marginal value of a single datum might be considered low, the value of a relatively small increase in predictive performance can be very high. A large hospital trust could easily have around 1 million scheduled events a year; thus, savings of even a few pence per appointment could be cumulatively substantial. Moreover, the social value of even a single instance of death or major disability thereby prevented is obviously far greater than the cost of computational modeling and the increasingly inexpensive digital systems required for its implementation. Accordingly, the marginal contribution of a datum needs to be contextualized by its ultimate impact.

Could we nonetheless assign a value to the data before the modeling work has been done? The answer in this case is clearly no, given that the complexity of the problem is only revealed in the course of analysis once data of the necessary scale have been evaluated. Just as the statistical power of complex models cannot be estimated from small, pilot samples, the maximum achievable predictive performance (ie, the determinant of value in this case) can only be known once the modeling work is complete. Reflection on the multifactorial nature of hospital attendance could allow one to *guess* that models of dimensionality only supported by large-scale data would be optimal, but this can only be *confirmed* in a post hoc context.

Even if the model is demonstrated to be stable over time at the source institution, generalizability to other institutions cannot be assumed. Indeed, a more sophisticated model is likely to be closely tailored to the local population. Therefore, the features of a model that enhance its *local* value may degrade its *global* value beyond the confines of the specific institution. This is not a reason to keep models simple, but rather highlights the need for building a distributed machine-learning infrastructure that allows individual institutions to build their own models rather than relying on generic models drawn from pooled data.

The Financial Value of Data

There is a widespread assumption that the value of public sector data is not being realized to its full capacity, and that this value needs to be optimized [2,9]. Nonetheless, how to measure the value of data for the purposes of public policy is uncertain. It is therefore unclear what a commitment to optimize this value might mean.

From an economic perspective, it might initially be tempting to think that the correct value of data is its price within a market, which will be determined by the operation of supply and demand. Optimizing the value of public sector health data might then be equated with maximizing the resale value of the data. However, such an approach is too narrowly focused on economic value, and completely disregards some of the nonfinancial ways in which data are valued.

Even in narrow financial terms, there will be circumstances in which making the data available for free or on a cost-recovery basis will be a more successful strategy, if doing so stimulates markets to build services on top of the data, which then create employment and tax revenue. For example, the US government's investment in the Global Positioning System, which it makes freely available, created new private markets by adding value to smartphones and to satellite navigation systems, which far outweighs the cost of the initial investment [10].

Thus, there is no reason to think that maximizing the amount that a business pays for access to public sector health data would also maximize the amount of economic value created. Moreover, asymmetric information might dominate health care markets for data, by which the commercial parties involved in such transactions are likely to possess much greater material knowledge of the potential uses of health care data and the likely future products. This could lead to suboptimal pricing of the data, moral hazard, or poor selection of market opportunities.

Governments should be primarily concerned with the economic value created through the use of the data over a life cycle rather than what could be gained from its initial sale. The success of data repositories such as the European Nucleotide Archive [11] illustrates this point vividly, as the biological sequences are made available to researchers free of charge, thereby vastly increasing the efficiency of related research. There is a consensus within the omics community that such curation and repositories are effective ways of creating value, despite (and partly because of) the fact that the sequences are given away free of charge.

Working out the optimal price and commercial strategies for public sector health care data will be complex, even in narrow financial terms. There will rarely be a matter of a single monolithic entity that controls all of the public sector health data in a particular state; rather, there is likely to be fragmentation at a regional or functional basis. This greatly complicates the economic picture. First, in cases with multiple potential sources for data—for example, if a company requires a clinical dataset from 1 million patients to build a tool, but could obtain a dataset from distinct regional bodies—competition between these bodies could significantly reduce the market price, leading to a less than optimal capture of the value by the public sector. This is one reason why the UK government's Office for Life Sciences has proposed that any commercial arrangements made by local NHS organizations should not be allowed to "undermine, inhibit or impact" the ability of the system at a national level to maximize the value of NHS data [12].

Second, to the extent that access to health data is commercialized, there is room for disagreement about where the money should go (ie, individual patients, hospitals, regional bodies, or the health care system as a whole) and the extent to which this income should be used for delivery of care or to further enhance the system’s ability to commercialize data. Third, there may be no guarantee that the value created through the release of a particular dataset for free or on a cost-recovery basis can be captured within its territory of origin. Even if these issues can be addressed, the data under control of health care systems will be derived from both the patients and from the operation of the health care system and its employees. Therefore, determining the ultimate recipient of reimbursement for any value remains to be resolved.

Public Value

Any policy that aims to optimize the value of public sector health data also needs to give due weight to the nonfinancial values that health systems are seeking to promote and protect. Health systems most obviously aim to restore those who are sick back to normal functioning, and to prevent the onset of disease. However, they also embed other values such as confidentiality and respect for autonomy. These values can be realized through role-based access controls that ensure access to patient data is only on a need-to-know basis, and in the ways in which health systems allow patients to choose how their identifiable health data will be shared. These values of confidentiality and respect for autonomy may come into potential conflict with the financial value to be realized through the commercialization of data; thus, any attempt to optimize value will need to take this into account in such cases.

In brief, the kind of value that public sector organizations should be seeking to optimize is not in any straightforward sense a financial value. Because of the interplay of financial and nonfinancial values, concepts such as “value chain,” “adding value,” and “value proposition” need to be translated, and partially transformed, from a business context. The public management theorist Mark Moore labels the kind of value that the public sector should be aiming to create as public value [13]. Public value is created, or captured, to the extent that public sector institutions further their democratically established goals or improve the lives of citizens [14].

Discussion of the factors contributing to enhancement of public value thus requires clarity about the purposes of different parts of the public sector, and how they fit together into an overall conception of the public good. Although realizing this type of clarity will require the public articulation and discussion of value judgments, this is a virtue rather than a vice in any account of value that will guide public policy.

In some cases, a public sector organization will have already articulated a set of values or principles that together express what is needed for the institution to do its job well. England's NHS offers a good example of such an organization. The constitution of the NHS establishes seven core principles: (1) the aim is to provide a comprehensive service, available to all; (2) access to those services will be based on clinical need not ability to pay; (3) care should be delivered with the highest standards of excellence and professionalism; (4) care should be patient-centered; (5) the NHS should be integrated across organizational boundaries; (6) the NHS should be committed to providing good value for tax payers’ money; and (7) the NHS should be accountable to the public, communities, and patients [15].

These principles could be used to help specify a model of public value for the NHS, which could inform questions about the value of data. For example, public value would be created if investing in interoperability and linkage of datasets allows for a system of care that is better integrated across organizational boundaries (principle 5). Similarly, implementing changes that improve value for money would create public value (principle 6).

Crowding Out, Perverse Incentives, and the Optimization of Public Value

There are good reasons for considering that, all else being equal, reducing the costs of a service to citizens while maintaining its quality will increase public value, even in public sector organizations that are not as explicit as the NHS in adopting this as a core principle. However, some or many citizens will not accept that the importance of value for money provides a reason to commercialize access to goods that were previously either not available or available on a noncommercial basis. The introduction of commercial motives may be resisted for many reasons, including that it may crowd out altruistic or solidaristic motivations [16,17]. Other citizens may worry that such commercialization could undermine the system’s core focus on meeting patient needs: in a system that is chronically short of money, managers are likely to take an opportunity to make extra money to cover shortfalls. Over time, the system may change so that rather than being perceived as extra cash, this resource would become part of the core budget. The system could then develop incentives to improve its ability to sell access to the data through employment of skilled professionals.

To the extent that such concerns are widespread—and a public sector organization is committed to a principle similar to the NHS’s seventh principle of accountability to the public, communities, and patients—they will need to be taken seriously. Citizens’ reasonable expectations about how the system should (and does) operate are an important force shaping public value, not least because of the need to maintain a social license [18,19]. These debates will require particular nuance and sensitivity in cases where—as perhaps in the NHS—there is a significant gap between the reality of the commercially funded and supplied nature of pharmaceuticals and medical devices, and the picture that some citizens have of a proudly noncommercial system.

However, completely avoiding commercial entanglements would cause conflict with other aspects of public value. The public sector will rarely have the skills to optimally innovate on its own, and to insist on complete lack of commercial involvement would result in delivering suboptimal care. Of course, the other extreme possibility—adopting a blanket policy of allowing commercial access to data without charge—would be in conflict with further aspects of public value. Although there will be circumstances in which making data available without charge to commercial companies will be reasonable from the perspective of public value, if those data will be used to design devices or services that will then be sold back to the public sector, there is room for concern about the fairness of the exchange.

Conclusion

If implemented without adequate thought, the attempt to realize the financial value of public sector health data through commercialization could destabilize the broader nonfinancial values that public sector health organizations aim to promote. Introducing the idea of public value provides a more clearly expressed overview of the challenge governments face in optimizing the value of health data, but cannot by itself resolve conflicts within competing conceptions of the public good.

Detailed discussion of what a successful public value account will look like is beyond the scope of this article; however, it is important to note some key elements that will be shared by all such accounts. First, public value presupposes a background public political culture, which shapes the conventions within which democratic deliberation takes place [20]. Political cultures, and their animating values, typically shift much more slowly than specific policy choices. The core elements of the public culture of most Western societies that profoundly shape the conceptions of public value include transparency, democratic accountability, due process and the rule of law, human dignity, using resources efficiently, and maintaining public trust.

Second, public value approaches require domain-specific goals and values for different elements of public services, which specify the requirements for the service to do its job well. Here, we have provided examples of the NHS constitution and the AI appointment scheduling system that represent these values at a high level and at a more specific level, respectively. These goals and values gain their legitimacy from their congruence with the background values that structure the public political culture, the agreed direction of government policy, and the results of public deliberation. In cases requiring changes to the balance or specification of domain-specific values (eg, rapidly shifting possibilities as a result of the opportunities created by AI), public deliberation will be needed within the framework set by the background values on how best to specify the approach taken for a particular domain or institution. Public value approaches have been worked out in more detail in some cases outside of the realm of data policy, such as for priority setting in health care [21,22].

Constructing an adequate model of public value for public sector data will require both articulation of the goals and values implicit in different public services, and open public dialogue about how best to specify these goals and values in light of fast-changing circumstances. It is important to place a range of social values on the table to best understand different publics’ attitudes to public sector organizations, along with their perceived opportunities and threats, toward realizing the financial value of health data. Facilitating such conversations in a constructive manner will depend on gaining public trust. Importantly, the limits of what policy makers can do to capture the value of public sector health data while maintaining public trust cannot be determined a priori.

Acknowledgments

This work was funded by the National Institute for Health Research, University College London Hospitals Biomedical Research Centre, and the Wellcome Trust. The funders had no role in the design, implementation, interpretation, and reporting of the study.

Authors' Contributions

JW and DH performed the initial literature review, and wrote the initial draft. JW led the writing of subsequent drafts. PN contributed the appointment scheduling textbox. GR, NM, and BW contributed examples, ideas, and textual clarifications to various drafts.

Conflicts of Interest

GR undertakes paid consultancy work for DeepMind. JW provides advice as a Panel Member to England’s National Data Guardian. The remaining authors declare no conflicts of interest.

References

1. Department of Health and Social Care. Code of conduct for data-driven health and care technology. London: Department of Health & Social Care; 2019 Jul 18. URL: https://tinyurl.com/y9yxs65q [accessed 2019-08-08]
2. Her Majesty's Treasury. The economic value of data: discussion paper. London: Her Majesty's Treasury; 2018 Aug. URL: https://www.gov.uk/government/publications/the-economic-value-of-data-discussion-paper [accessed 2019-08-08]
3. Prainsack B. Data Donation: How to Resist the iLeviathan. In: Krutzinna J, Floridi L, editors. The Ethics of Medical Data Donation. Cham, Switzerland: Springer; 2019:9-22.
4. Benkler Y. Commons and Growth: The Essential Role of Open Commons in Market Economies. The University of Chicago Law Review 2013;80(3):1499-1555 [FREE Full text]
5. Haskel J, Westlake S. Capitalism Without Capital: The Rise Of The Intangible Economy. New Jersey: Princeton University Press; 2017.

6. National Institute for Health and Care Excellence. Social Value Judgements: Principles for the development of NICE guidance. London: National Institute for Health and Care Excellence; 2008.
7. Harwich E, Lasko-Skinnner R. Making NHS data work for everyone. London: Reform; 2018. URL: https://reform.uk/research/making-nhs-data-work-everyone [accessed 2019-08-08]
8. Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. NPJ Digit Med 2019;2:26. [doi: 10.1038/s41746-019-0103-3] [Medline: 31304373]
9. Her Majesty's Treasury. Getting smart about intellectual property and other intangibles in the public sector: Budget 2018. London: Her Majesty's Treasury; 2018 Oct. URL: https://tinyurl.com/t2nok5n [accessed 2019-08-08]
10. Mazzucato M. The Entrepreneurial State: Debunking Public Vs. Private Sector Myths. USA: PublicAffairs; 2015.
11. European Nucleotide Archive. European Nucleotide Archive Browser. URL: https://www.ebi.ac.uk/ena/browser/home [accessed 2020-02-06]
12. Department of Health & Social Care. Creating the right framework to realise the benefits for patients and the NHS where data underpins innovation. London: Department of Health & Social Care; 2019. URL: https://tinyurl.com/y3nnqs5d [accessed 2019-07-15]
13. Moore MH. Creating Public Value: Strategic Management In Government. Cambridge, MA: Harvard University Press; 1995.
14. Moore MH. Public Value Accounting: Establishing the Philosophical Basis. Public Admin Rev 2014 May 28;74(4):465-477. [doi: 10.1111/puar.12198]
15. Department of Health & Social Care. The NHS Constitution for England. London: Department of Health & Social Care; 2015. URL: https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england [accessed 2019-07-15]
16. Frey B, Oberholzer-Gee F. The cost of price incentives: An empirical analysis of motivation crowding-out. Am Econ Rev 1997;87(4):746-755.
17. Sandel M. What Money Can't Buy: The Moral Limits Of Markets. USA: Farrar, Straus And Giroux; 2013.
18. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015 May;41(5):404-409. [doi: 10.1136/medethics-2014-102374] [Medline: 25617016]
19. Taylor M, Wilson J. Reasonable Expectations of Privacy and Disclosure of Health Data. Med Law Rev 2019;27(3):432-460 [FREE Full text] [doi: 10.1093/medlaw/fwz009] [Medline: 31220873]
20. Rawls J. Political Liberalism. New York: Columbia University Press; 2005.
21. Wilson J. Public Value, Maximization and Health Policy: An Examination of Hausman’s Restricted Consequentialism. Public Health Ethics 2016 May 15:phw020. [doi: 10.1093/phe/phw020]
22. Rumbold B, Weale A, Rid A, Wilson J, Littlejohns P. Public Reasoning and Health-Care Priority Setting: The Case of NICE. Kennedy Inst Ethics J 2017;27(1):107-134 [FREE Full text] [doi: 10.1353/ken.2017.0005] [Medline: 28366905]

Abbreviations

AI: artificial intelligence
NHS: National Health Service

Edited by G Eysenbach; submitted 09.08.19; peer-reviewed by B Prainsack, E Khajeh; comments to author 22.10.19; revised version received 14.11.19; accepted 15.12.19; published 27.03.20.

Please cite as:
Wilson J, Herron D, Nachev P, McNally N, Williams B, Rees G
The Value of Data: Applying a Public Value Model to the English National Health Service
J Med Internet Res 2020;22(3):e15816
URL: http://www.jmir.org/2020/3/e15816/
doi:10.2196/15816
PMID:32217501

Viewpoint

Deep Digital Phenotyping and Digital Twins for Precision Health: Time to Dig Deeper

Guy Fagherazzi[1], MSc, PhD

[1] Luxembourg Institute of Health, Department of Population Health, Digital Epidemiology Hub, Strassen, Luxembourg

Corresponding Author:
Guy Fagherazzi, MSc, PhD
Luxembourg Institute of Health
Department of Population Health
Digital Epidemiology Hub
1 A-B Rue Thomas Edison
Strassen, 1445
Luxembourg
Phone: 33 669396334
Email: guy.fagherazzi@lih.lu

Abstract

This viewpoint describes the urgent need for more large-scale, deep digital phenotyping to advance toward precision health. It describes why and how to combine real-world digital data with clinical data and omics features to identify someone's digital twin, and how to finally enter the era of patient-centered care and modify the way we view disease management and prevention.

(J Med Internet Res 2020;22(3):e16770) doi:10.2196/16770

KEYWORDS

digital health; digital epidemiology; deep digital phenotyping; digital orthodoxy; precision medicine; precision health; personalized medicine; digital phenotyping; precision prevention; big data; omics; digitosome; data lake; digital cohort

Introduction

It has been said that there are "a hundred ways of being diabetic," but I could argue that there are much more, which is true for virtually any chronic disease. The more we advance in the understanding of a disease, from a biological, clinical, genetic, epidemiological, sociological, behavioral, or psychological point of view, the more we uncover the complexity of medical conditions that health care systems will then have to prevent, treat, and manage.

Precision Medicine is Much More Than a Matter of Genetic Features

Precision medicine has been defined as [1]:

> *An emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person.*

Recent achievements have been made in the field of precision oncology [2], but so far, only a small proportion of patients can benefit from personalized treatment each year. Currently, the focus is on the use of genetic or molecular markers to stratify diagnoses and corresponding treatment strategies. However, most are still in the discovery stage in mice models or *in silico*, even if some commercial companies are already starting to use them [3].

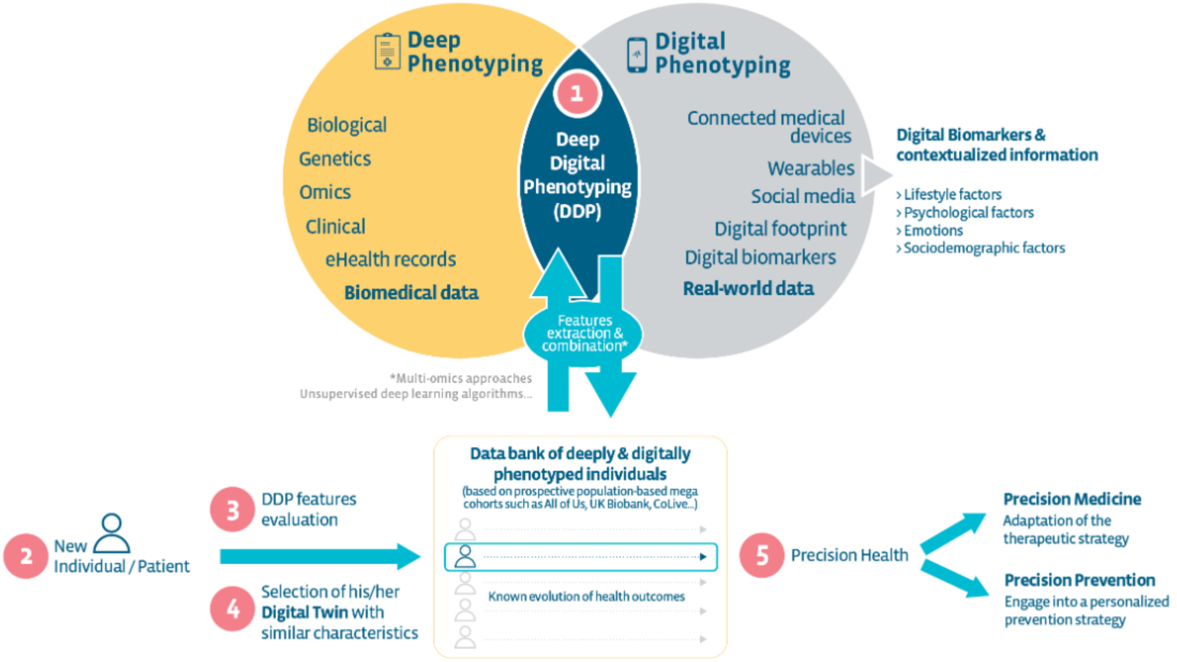
How can precision medicine claim to involve patients in their care [4] if it ignores the data they generate in real life? A personalized therapeutic strategy could theoretically suit someone's biological or genetic phenotype but could fail because of their level of stress, dietary habits, working or living environment, or their cultural background. However, from an economic perspective, we should expect these costly therapeutics (sometimes up to tens of thousands of dollars a month per patient in oncology) to have the best return on investment in terms of compliance and success rate [5].

To achieve the ultimate goal of precision health, which is to match one individual, given their unique profile, with their one, best, medical, therapeutic, and preventive strategy, I argue that we will have to invest in the concept of deep digital phenotyping.

Deep Digital Phenotyping Is the Missing Link

Achieving the true purpose of precision health requires integrating, from scratch, features from the “digitosome” (ie, all data generated digitally by individuals online or by their smartphones or connected devices) [6]. Deep digital phenotyping is the combination of deep phenotyping (defined for almost a decade now as the “precise and comprehensive analysis of phenotypic abnormalities in which the individual components of the phenotype are observed and described”) [7], with digital phenotyping, (defined as the moment-by-moment quantification of the individual-level human phenotype *in situ* using data from personal digital devices [8]). If clinically relevant [9], the power of digital data [10] will give us insights, usually in an automated and objective way, about the lifestyle, psychological state, sociodemographics, and environment of a given individual and thus will help capture the bigger picture and reach the full potential of precision health (see Figure 1). Digital phenotyping has already proven relevant in a few areas, such as psychiatry [11] or cardiovascular diseases [12].

Figure 1. The concept of deep digital phenotyping and digital twin identification for precision health.



Surprisingly, despite the relative ease of acquisition, high volume, and low cost and burden for the individual, these types of data are often neglected and not exploited to characterize the phenotypes of the patient. Do not forget: a person living with diabetes spends roughly 6 hours per year face-to-face with health care professionals and takes more than 600 hours per year to manage (generating data about) their disease(s) by themselves.

Large clinical and epidemiological initiatives can now gather details about disease manifestations, risk factors, and health determinants in a more individualized and detailed way, and use advanced algorithms to integrate data on human behaviors and human-machine interactions through smartphones or connected devices. For example, tracking digital heart rate variability (HRV) as a marker of stress [13] and optimizing drug intake when HRV is at its peak.

Let Us Meet Our Digital Twin

We have known for some time that “one size fits all” strategies are not efficient therapeutically or preventively [14]. To observe a true leap forward in the domain of precision health, I encourage developing large data banks of “digital twins.” The digital twin is a very new concept in health research and comes from the industrial world, where a digital replica of a physical entity is virtually recreated, with similar elements and dynamics, to perform real-time optimization and testing. The use of digital twins can be extended to the medical field, the elements being features from deep digital phenotyping and the dynamics being the evolution of health outcomes over time. Thus, a digital twin would be a virtual patient with similar or close characteristics as a new patient seen during a clinical visit, and for whom the health status, risks of complications, and disease evolutions are known. This new patient will have a digital twin represented by the average characteristics of its closest cluster group, obtained thanks to deep digital phenotyping.

Methodology and Ethical Concerns

It seems this is the right moment to implement this digital twin concept. From multi-omics approaches to unsupervised deep learning algorithms, along with the proper computational power, we now have the appropriate tools to deal with the diversity and quantity of information and move from coarsely stratified groups to refined, small groups of individuals defined by numerous features. Methods such as variational autoencoders, an unsupervised deep learning framework, can be used to learn latent representations to cluster and identify deep digital phenotyping patterns [15], or uniform manifold approximation and projection could also be used, which is a dimensionality reduction technique for machine learning [16]. This is on top of using hierarchical agglomerative cluster analyses or k-means clustering to identify refined subgroups of individuals whose detailed characteristics can be averaged to provide someone's digital twin.

The main challenge to address here is accessing large populations and their detailed information. This will be resolved in the short term with the development of mega cohorts (such as the All of Us Research Program, the UK Biobank, CoLive Diabetes, etc) and other prospective digital health data lakes and big data infrastructure, which will soon provide both deep digital phenotyping of volunteers enrolled in these initiatives and the evolution of their health outcomes.

Going so deep in the phenotyping of populations will also raise ethical and data security concerns. Appropriate clinical and research practices will have to be updated and extended, in parallel with medical and technological evolution, without preventing innovation and ensuring that it will benefit most people. Privacy by design and by default, pseudonymization, traceability, and data portability, key elements from the General Data Protection Regulation [17] or European guidelines on data security for Big Data projects [18], should be the standard of research and included from the beginning in the study and the information technology infrastructure associated with it. Also, conducting transparent research, obtaining informed consent, including participants at every stage of the research, communicating continuously on the different uses of the data collected, and going back to the community and the lay public will ensure trust in deep digital phenotyping methods to combine many sources of data on a large population. In parallel, open data and open source practices should particularly be encouraged in this field.

Let Us Dig Deep!

We are moving from an ancient world where people with diabetes were characterized by only a few measurements of fasting glucose levels or glycated hemoglobin to a world where frontiers between subclinical types of diabetes are being redrawn. Soon, we will also be moving to a future where we will be able to deeply phenotype individuals with thousands of points of combined clinical, biological, genetic, sociological, psychological, and real-world digital parameters, which will profoundly change the way we characterize patients, and how we understand and contextualize the various forms of diseases. This is where modern epidemiology, combined with computer science, data science, and behavioral psychology, will play a significant role in medical research.

Ultimately, we will enter the era of true precision health and patient-centered care and modify the way we consider disease management and prevention, through the identification of someone's digital twins. This will augment the capabilities of health care professionals and empower patients by fine-tuning disease management, treatments, and devices to use, as well as biomarkers to monitor.

Acknowledgments

GF would like to thank Dr Courtney Dow and Dr Laetitia Huiart for the diligent proofreading of this article and colleagues from the Department of Population Health at the Luxembourg Institute of Health for their support.

Conflicts of Interest

None declared.

References

1. Psaty BM, Dekkers OM, Cooper RS. Comparison of 2 Treatment Models: Precision Medicine and Preventive Medicine. JAMA 2018 Aug 28;320(8):751-752. [doi: 10.1001/jama.2018.8377] [Medline: 30054607]
2. Wise HC, Solit DB. Precision Oncology: Three Small Steps Forward. Cancer Cell 2019 Jun 10;35(6):825-826. [doi: 10.1016/j.ccell.2019.05.009] [Medline: 31185208]
3. Muse ED, Topol EJ. Digital orthodoxy of human data collection. The Lancet 2019 Aug;394(10198):556. [doi: 10.1016/s0140-6736(19)31727-1]
4. Wynn RM, Adams KT, Kowalski RL, Shivega WG, Ratwani RM, Miller KE. The Patient in Precision Medicine: A Systematic Review Examining Evaluations of Patient-Facing Materials. J Healthc Eng 2018;2018:9541621 [FREE Full text] [doi: 10.1155/2018/9541621] [Medline: 30250657]
5. Kasztura M, Richard A, Bempong N, Loncar D, Flahault A. Cost-effectiveness of precision medicine: a scoping review. Int J Public Health 2019 Dec;64(9):1261-1271 [FREE Full text] [doi: 10.1007/s00038-019-01298-x] [Medline: 31650223]
6. Fagherazzi G, Ravaud P. Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]

7. Robinson PN. Deep phenotyping for precision medicine. Hum Mutat 2012 May 13;33(5):777-780 [FREE Full text] [doi: 10.1002/humu.22080] [Medline: 22504886]
8. Torous J, Kiang MV, Lorme J, Onnela JP. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
9. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
10. Raballo A. Digital phenotyping: an overarching framework to capture our extended mental states. Lancet Psychiatry 2018 Mar;5(3):194-195. [doi: 10.1016/S2215-0366(18)30054-3] [Medline: 29482758]
11. Lydon-Staley DM, Barnett I, Satterthwaite TD, Bassett DS. Digital phenotyping for psychiatry: Accommodating data and theory with network science methodologies. Curr Opin Biomed Eng 2019 Mar;9:8-13. [doi: 10.1016/j.cobme.2018.12.003] [Medline: 31650093]
12. Teo JX, Davila S, Yang C, Hii AA, Pua CJ, Yap J, et al. Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. Commun Biol 2019;2:361 [FREE Full text] [doi: 10.1038/s42003-019-0605-1] [Medline: 31602410]
13. Kim H, Cheon E, Bai D, Lee YH, Koo B. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
14. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
15. Wang Z, Wang Y. Extracting a biologically latent space of lung cancer epigenetics with variational autoencoders. BMC Bioinformatics 2019 Nov 25;20(Suppl 18):568 [FREE Full text] [doi: 10.1186/s12859-019-3130-9] [Medline: 31760935]
16. Sánchez-Rico M, Alvarado JM. A Machine Learning Approach for Studying the Comorbidities of Complex Diagnoses. Behav Sci (Basel) 2019 Nov 22;9(12):122 [FREE Full text] [doi: 10.3390/bs9120122] [Medline: 31766665]
17. Chico V. The impact of the General Data Protection Regulation on health research. Br Med Bull 2018 Dec 01;128(1):109-118. [doi: 10.1093/bmb/ldy038] [Medline: 30445448]
18. 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 Oct 11;13(10):e0204937 [FREE Full text] [doi: 10.1371/journal.pone.0204937] [Medline: 30308031]

Abbreviations

HRV: heart rate variability

Edited by G Eysenbach; submitted 23.10.19; peer-reviewed by D Pougheon Bertrand, A Honka; comments to author 14.12.19; revised version received 21.12.19; accepted 21.12.19; published 03.03.20.

Please cite as:
Fagherazzi G
Deep Digital Phenotyping and Digital Twins for Precision Health: Time to Dig Deeper
J Med Internet Res 2020;22(3):e16770
URL: https://www.jmir.org/2020/3/e16770
doi:10.2196/16770
PMID:32130138

Viewpoint

Connected Medical Technology and Cybersecurity Informed Consent: A New Paradigm

Jeffrey Tully[1*], MD; Andrea Coravos[2,3,4,5*], BS; Megan Doerr[6*], MS, BS; Christian Dameff[7,8*], MD, MS

[1]Department of Anesthesiology and Pain Medicine, UC Davis Medical Center, Sacramento, CA, United States
[2]Elektra Labs, Boston, MA, United States
[3]Digital Medicine Society, Boston, MA, United States
[4]Harvard-MIT Center for Regulatory Science, Boston, MA, United States
[5]Policy Innovation Lab of Tomorrow, Penn State University, State College, PA, United States
[6]Sage Bionetworks, Seattle, WA, United States
[7]Department of Emergency Medicine, University of California San Diego, La Jolla, CA, United States
[8]Department of Computer Science and Engineering, La Jolla, CA, United States
[*]all authors contributed equally

Corresponding Author:
Jeffrey Tully, MD
Department of Anesthesiology and Pain Medicine
UC Davis Medical Center
4150 V Street
Suite 1200 PSSB
Sacramento, CA, 95817
United States
Phone: 1 9167345031
Email: jefftullymd@gmail.com

Abstract

Background: Connected medical technology is increasingly prevalent and offers both a host of new therapeutic potentials and cybersecurity-related considerations. Current practice largely does not include discussions of cybersecurity issues when clinicians obtain informed consent.

Objective: This paper aims to raise awareness about cybersecurity considerations for connected medical technology as they relate to informed consent discussions between patients and clinicians.

Methods: Clinicians, health care cybersecurity researchers, and informed consent experts propose the concept of a cybersecurity informed consent for connected medical technology.

Results: This viewpoint discusses concepts designed to facilitate further discussion on the need, development, and execution of cybersecurity informed consent.

Conclusions: Cybersecurity informed consent may be a necessary component of informed consent practices, as connected medical technology proliferates in the health care environment.

(J Med Internet Res 2020;22(3):e17612) doi:10.2196/17612

KEYWORDS

connected medical technology; cybersecurity; informed consent; privacy; patient autonomy; digital health; medical devices; ethics

The practice of medicine is built on the foundation of clinician-patient engagement, and consent is a key pillar supporting this essential relationship [1]. As medicine has shifted from a paternalistic, subordinating art to a collaborative effort of shared goal setting and decision making between parties, so too has the understanding of the ethics and acquisition of consent evolved.

It is widely acknowledged that consent must be *informed*. This charge demands that clinicians empower patients in shared decision making through culturally competent, plain language dialogue. In doing so, a patient’s informed consent becomes the embodiment of the principle of autonomy as well as a symbol of their investment as the most important stakeholder in the therapeutic alliance.

In the clinical realm, informed consent precedes care such as new treatment regimens or proposed surgical procedures. The discussion between healer and cared-for has come to possess a distinct anatomy. First, the patient is identified, which must involve not just a name and a record number, but the understanding of their personal story, beliefs, and objectives. Second the clinician discusses the nature of the intervention and presents a tailored list of benefits, risks, and alternatives to the stated plan. The persistent pace of progress in medicine ensures that these latter considerations are closely tied to new frontiers in clinical science.

Connected medical technology occupies one such frontier. From wearable activity trackers and mobile software apps to implantable medical devices and telemedicine platforms, digital tools are assuming an ever-growing role in health care, with an amplified potential depending on the degree these technologies connect to other devices, computers, or networks [2]. The acquisition of large amounts of increasingly granular data and the facilitation of longitudinal and remote clinical interactions are all enabled in part by the continuous connectivity of these devices.

Though the benefits of such connectivity in diagnosing, monitoring, and treating diseases are widely touted, connectivity may introduce additional risks to patients. As connected medical devices are built with the same or similar hardware and software used in mobile technologies and computers, flaws in code, components, or networks can lead to exploitation and disruption of these devices [3]. The analysis of and protection against such attacks constitute a central element of the practice of cybersecurity [4].

Researchers have demonstrated cybersecurity vulnerabilities in medical devices including automated internal cardioverter defibrillators, bedside infusion pumps, and implantable insulin delivery systems [5]. Such flaws, if abused, could lead to a number of consequences ranging from exposure of personal and private health information to the malfunction of devices resulting in physical harm. Though there are not yet any reports of patients directly affected by the exploitation of a medical device's cybersecurity vulnerability, the potential for such events has led to concerted efforts from manufacturers, regulators, and security professionals to advocate for and improve medical cybersecurity practices.

Clinicians are expected to acquire knowledge on various medications, procedures, and therapies to understand and articulate the risks, benefits, and alternatives of such interventions during the informed consent process. As an increasing number of interventions rely solely or in part on connected technologies, the same attendant framework should exist for these tools. We propose that the unique characteristics of connected technologies warrant development of a "cybersecurity-informed consent" to address the cybersecurity implications of planned interventions. Several challenges exist in creating a model for such a consent.

First, the epidemiology of health care cybersecurity is an emerging science. The benefits of connected medical devices have been quickly embraced; however, cybersecurity itself has been a blind spot for many practicing connected medicine. This disinclination may arise from a lack of knowledge or understanding of cybersecurity vulnerabilities, or may relate to the lack of identified real-world incidents of cybersecurity vulnerabilities interfering with clinical care. However, the US Food and Drug Administration, in its authority as primary regulator of medical devices, has issued multiple safety communications and recalls consistent with the maxim that "absence of evidence is not evidence of absence" [6-8].

A second challenge is the risk-benefit ratio of addressing cybersecurity vulnerabilities. One of the most basic remedies for cybersecurity vulnerabilities is the practice of "patching"—updating a device's software with new and improved code to address flaws. This is usually a relatively straightforward exercise, occurring almost continuously in many enterprises that use commercial operating systems. However, patching has added complexity with medical devices, because a small but nonnegligible risk exists when software updates intended to address cybersecurity dangers lead to unintended corruption of the device's normal functioning. Furthermore, while the risk of exploitation of a vulnerability may be unknown, the failure rate of a patch may be well-documented [9]. This risk-to-reward imbalance may result in some clinicians advising patients to forego patching of vulnerabilities entirely.

Third, in contrast to much of medicine, the cybersecurity risk of connected medical interventions may not be fixed—or even consistent—across the life span of a device or app used in an intervention. The same properties in software that allow for updates in functionality and features ensure that the potential for new vulnerabilities exist alongside the promise of additional benefits. A clinician accustomed to largely stable probabilities of, for example, infection with transfusion of blood products may confront a situation with a connected device where current cybersecurity risk is undefined and future danger remains amorphous.

Given the increasing ubiquity of networked functionality in digital devices ranging from the smallest wearables to the largest surgical robots, the simple option of using nonconnected technologies may not be a choice for clinicians or patients concerned about cybersecurity risks. Those looking to benefit from the most sophisticated pacemakers or insulin pumps may face the realization that alternatives to these devices might not exist.

The traditional model of clinician-facilitated informed consent relies on the previously discussed core knowledge that many clinicians currently lack. There are increasing efforts to raise awareness of the potential consequences of cybersecurity vulnerabilities on patient safety; however, there is still no widespread curricula for allied health professional students or concerted continuing medical education for practicing providers. Without at least a basic, conceptual understanding of both the risks and remedies, it is both unfair and impractical to expect the clinician to impart the same information for patient consideration.

Yet patients surely deserve the chance to consider cybersecurity elements within the context of their treatment plans or procedures, and cybersecurity-informed consent creates the environment for the patient to be an active participant in this process. The challenge then becomes incorporating cybersecurity-informed consent into a complex workflow process that already demands significant time and effort, which is largely uncompensated. Several potential models may warrant further evaluation.

The highest yield method for cyber-informed consent adoption may be educational interventions for the providers whom are most likely to interface with connected medical technologies, including electrophysiologists, endocrinologists, and informaticians, in the form of digital modules, tool kits, or simulations.

An argument may be made that, as cybersecurity concepts are not immediately germane to the practice of medicine, there exists no explicit demand for clinician involvement in the consent process. This framework implies that the obligation to cyber-consent belongs to other stakeholders such as the manufacturers of the technology. In such a system, a patient might interface with a particular vendor to be educated about the cybersecurity implications of a device in a similar way that proceduralists rely on device representatives to provide technical guidance during the implantation process.

Informed consent methodology in clinical research has recently evolved in a number of interesting ways that may serve as a good model for cyber-informed consent.

The opportunity to obtain consent remotely through online portals, along with the ability to present a wide variety of information customized to individual capabilities and cultures, has increased inclusivity and interactivity. It has also generated the idea of “independently navigable” consents that patients complete autonomously without requiring direct interaction with the research team [10,11].

It is clear that more research is needed to determine which cybersecurity-informed consent approach best addresses the unique hurdles of digital medicine. Having a clear, concise informed consent process can be challenging in light of clinician inexperience, as well as uncertainty regarding the true level of risk posed to patients by connected technologies (and the potential for these risks to fluctuate over time). As connected medical interventions offer personalization of care and expanded accessibility, so too will an effective cyber-informed consent empower clinicians and patients in goal-affirming health care decision making.

Acknowledgments

The authors wish to thank Adam Conner-Simons for assistance with copyediting this manuscript.

Conflicts of Interest

None declared.

References

1. Grady C. Enduring and emerging challenges of informed consent. N Engl J Med 2015 Feb 26;372(9):855-862. [doi: 10.1056/nejmra1411250]
2. Iqbal MH, Aydin A, Brunckhorst O, Dasgupta P, Ahmed K. A review of wearable technology in medicine. J R Soc Med 2016 Oct 11;109(10):372-380 [FREE Full text] [doi: 10.1177/0141076816663560] [Medline: 27729595]
3. Kruse CS, Frederick B, Jacobson T, Monticone DK. Cybersecurity in healthcare: a systematic review of modern threats and trends. Technol Health Care 2017;25(1):1-10. [doi: 10.3233/THC-161263] [Medline: 27689562]
4. Perakslis ED. Cybersecurity in health care. N Engl J Med 2014 Jul 31;371(5):395-397. [doi: 10.1056/NEJMp1404358] [Medline: 25075831]
5. Woods B, Coravos A, Corman JD. The case for a hippocratic oath for connected medical devices: viewpoint. J Med Internet Res 2019 Mar 19;21(3):e12568 [FREE Full text] [doi: 10.2196/12568] [Medline: 30888323]
6. Klonoff D, Han J. The first recall of a diabetes device because of cybersecurity risks. J Diabetes Sci Technol 2019 Sep 17;13(5):817-820. [doi: 10.1177/1932296819865655] [Medline: 31313589]
7. Kramer DB, Fu K. Cybersecurity concerns and medical devices: lessons from a pacemaker advisory. JAMA 2017 Dec 05;318(21):2077-2078. [doi: 10.1001/jama.2017.15692] [Medline: 29049709]
8. Ronquillo JG, Zuckerman DM. Software-related recalls of health information technology and other medical devices: implications for FDA regulation of digital health. Milbank Q 2017 Sep 12;95(3):535-553 [FREE Full text] [doi: 10.1111/1468-0009.12278] [Medline: 28895231]
9. Baranchuk A, Refaat MM, Patton KK, Chung MK, Krishnan K, Kutyifa V, American College of Cardiology’s Electrophysiology Section Leadership. Cybersecurity for cardiac implantable electronic devices: what should you know? J Am Coll Cardiol 2018 Mar 20;71(11):1284-1288 [FREE Full text] [doi: 10.1016/j.jacc.2018.01.023] [Medline: 29475627]
10. Kraft SA, Doerr M. Engaging populations underrepresented in research through novel approaches to consent. Am J Med Genet C Semin Med Genet 2018 Mar 07;178(1):75-80 [FREE Full text] [doi: 10.1002/ajmg.c.31600] [Medline: 29512940]
11. Moore S, Tassé AM, Thorogood A, Winship I, Zawati M, Doerr M. Consent Processes for Mobile App Mediated Research: Systematic Review. JMIR Mhealth Uhealth 2017 Aug 30;5(8):e126 [FREE Full text] [doi: 10.2196/mhealth.7014] [Medline: 28855147]

Edited by G Eysenbach; submitted 26.12.19; peer-reviewed by R Hendricks-Sturrup, F Abujarad, K Morse, J Bobe; comments to author 22.01.20; revised version received 28.01.20; accepted 28.01.20; published 30.03.20.

Please cite as:
Tully J, Coravos A, Doerr M, Dameff C
Connected Medical Technology and Cybersecurity Informed Consent: A New Paradigm
J Med Internet Res 2020;22(3):e17612
URL: http://www.jmir.org/2020/3/e17612/
doi:10.2196/17612
PMID:32224492

Original Paper

A Values-Tailored Web-Based Intervention for New Mothers to Increase Infant Vaccine Uptake: Development and Qualitative Study

Amanda Dempsey[1*], MD, MPH, PhD; Bethany M Kwan[1*], PhD; Nicole M Wagner[2], PhD; Jennifer Pyrzanowski[1], MSPH; Sarah E Brewer[1], PhD, MPA; Carter Sevick[1], MS; Komal Narwaney[2], PhD; Kenneth Resnicow[3], PhD; Jason Glanz[2], PhD

[1]University of Colorado, Denver, Aurora, CO, United States

[2]Kaiser Permanente Colorado Institute for Health Research, Denver, CO, United States

[3]University of Michigan, Department of Health Behavior & Health Education, School of Public Health, Ann Arbor, MI, United States

[*]these authors contributed equally

Corresponding Author:
Amanda Dempsey, MD, MPH, PhD
University of Colorado, Denver
13199 East Montview Blvd, Suite 300
Aurora, CO, 80045
United States
Phone: 1 3037246679
Email: amanda.dempsey@ucdenver.edu

Abstract

Background: Vaccine hesitancy among parents leads to childhood undervaccination and outbreaks of vaccine-preventable disease. As the reasons for vaccine hesitancy are diverse, there is often not enough time during regular clinical visits for medical providers to adequately address all the concerns that parents have. Providing individually tailored vaccine information via the internet before a clinical visit may be a good mechanism for effectively allaying parents' vaccination concerns while also being time efficient. Including tailoring based on values is a promising, but untested, approach to message creation.

Objective: This study aimed to describe the process by which we developed a Web-based intervention that is being used in an ongoing randomized controlled trial aimed at improving the timeliness of infant vaccination by reducing parental vaccine hesitancy.

Methods: Development of the intervention incorporated evidence-based health behavior theories. A series of interviews, surveys, and feedback sessions were used to iteratively develop the intervention in collaboration with vaccination experts and potential end users.

Results: In all, 41 specific content areas were identified to be included in the intervention. User feedback elucidated preferences for specific design elements to be incorporated throughout the website. The tile-based architecture chosen for the website was perceived as easy to use. Creating messages that were two-sided was generally preferred over other message formats. Quantitative surveys identified associations between specific vaccine values and vaccination beliefs, suggesting that values tailoring should vary, depending on the specific belief being endorsed.

Conclusions: Using health behavior theories, qualitative and quantitative data, and significant expert and end user input, we created a novel, Web-based intervention to improve infant vaccination timeliness. The intervention is based on tailoring messages according to each individual's values and beliefs. This intervention is currently being tested in a controlled randomized clinical trial.

(J Med Internet Res 2020;22(3):e15800) doi:10.2196/15800

KEYWORDS

immunization; parents

Introduction

Vaccine Hesitancy

Vaccination is widely recognized as one of the most effective public health interventions ever [1]. However, despite the well-established safety and effectiveness of vaccines, a growing number of parents are choosing to delay or forgo them for their children because of questions about vaccines' necessity and safety (also sometimes because of firmly held religious or political beliefs). This phenomenon is referred to as *vaccine hesitancy* [2,3]. Vaccine hesitancy often results in undervaccination among children and has led to increasing numbers of vaccine-preventable disease outbreaks in the United States over the last two decades [4,5]. For example, measles, which was considered eradicated from the United States in 2000, caused infections among more than 2000 people between 2014 and 2018 alone, with the majority of cases occurring in individuals under- or unvaccinated against the virus [6]. Vaccine-preventable disease outbreaks are associated with significant cost and morbidity, and in some cases, even death [7]. Therefore, it is a public health priority to find interventions to mitigate this trend of parental vaccine hesitancy and childhood undervaccination [8]. This paper reports on the development of a novel, tailored, Web-based intervention to promote timely vaccination by addressing these issues. A more detailed description of the study design and protocol is available elsewhere [9].

Study Rationale

Most interventions to increase vaccination developed thus far have focused primarily on correcting knowledge deficits, with the hypothesis that correcting these deficits will lead to improved attitudes and behaviors (ie, parents become less vaccine hesitant and thus are more likely to get their children vaccinated). Unfortunately, the majority of interventions based on this concept have not been effective at increasing vaccination rates [10,11]. This is because, as research has elucidated, parents' vaccination decisions are multiply determined—based not *just* on their knowledge about the risk and severity of infectious diseases and the benefits and risks of vaccines but also on trust, emotion, values, past experience, access to health care, and peer influences [12,13].

Given the multiple determinants of vaccination decisions, novel intervention strategies that account for factors beyond knowledge deficits—such as personal values and emotions associated with parents' individual concerns and barriers to vaccination—are needed [14]. One such approach to address these multiple issues is message tailoring. Message tailoring involves providing customized vaccine-promotion messages based on an individual's unique beliefs, experiences, knowledge, and barriers to action [15]. Research on tailored messaging in multiple domains shows that by increasing the personal relevance of the information, people are more receptive to new information that may challenge their beliefs. Regardless of whether their *knowledge* of a given situation is altered by message tailoring, the technique is thought to work by lowering psychological resistance to information or suggested action that may counter what an individual initially thinks or believes [15]. Tailored messaging interventions have been shown to be effective for a number of health behaviors but have not been extensively tested for vaccine promotion [16,17].

In this paper, we describe the development of a Web-based, tailored messaging intervention used in a randomized controlled trial that is ongoing (ClinicalTrials.gov protocol number NCT02665013). This intervention, targeted to new and expectant mothers, was designed to promote vaccination by either reinforcing provaccination decisions among parents who are not hesitant to vaccinate or by decreasing vaccine hesitancy and thus increasing vaccination intention among parents who are vaccine hesitant. To do this, we developed our intervention based on evidence-based health behavior theories and included tailoring in both standard (ie, gender, primary vaccination beliefs, and concerns) and novel (personal values) domains. Although many additional factors could be novel targets for tailoring (trust, emotion, access to health care, etc), values were chosen as the potential new tailoring target for this intervention because, unlike beliefs, experiences, and barriers, which frequently change over time or by situation, values are believed to remain stable over the life course and across contexts [18-21]. Moreover, the behavior of vaccine hesitancy appears rooted in values as well as knowledge, skills, and self-efficacy [13,22-24]. Although there has been minimal research on how to use values to effectively promote vaccination, many researchers in the field recognize its potential importance [23,25-28]. Thus, a secondary goal in developing this values-tailored intervention was to begin to address the knowledge gap regarding the role values may play in creating messages effective for increasing maternal acceptance of vaccines. In addition, although fathers can be an important influence in childhood vaccination decisions, they were not included as the target for this intervention based on our previous research suggesting that mothers were the main vaccine decision makers in the study population of interest [29,30]. In this paper, we describe the overall intervention development process, with particular attention to how the novel values tailoring was developed.

Methods

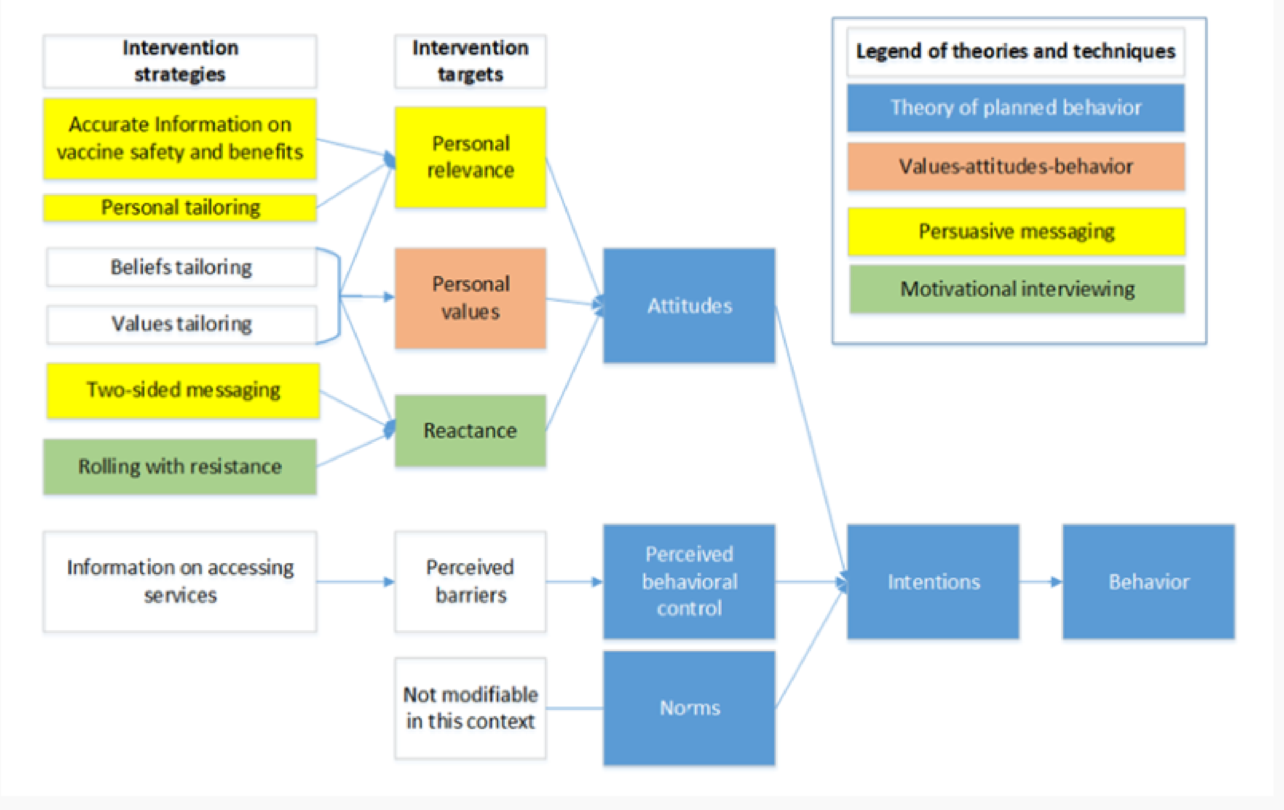
Design Overview

The Web-based, tailored vaccine promotion messaging intervention was developed using a multiphase, iterative, user-centered design process. It was informed by behavioral theory, empirical data from maternal surveys and interviews, and expert and end-user input. When possible, end users' opinions about intervention design were prioritized over those of the research team. For the intervention trial, we planned that mothers would be recruited to the website between the last trimester of their pregnancy and when their child was less than or equal to 2 months of age (the primary time for vaccination decisions to be solidified [30]). Mothers would then receive additional exposures to the website (with retailored information based on updated beliefs) at three additional time points between enrollment and infant age of 15 months. Vaccination timeliness would be assessed at age 200 days (ie, approximately 6 months) and 489 days (ie, approximately 15 months). Two versions of the website were developed for the randomized trial: a version with messages tailored to the individual participant based on personal characteristics (age, baby's gender, and pregnancy status) and mothers' vaccination beliefs and concerns, vaccination values, and intention to vaccinate (described in this paper), and an untailored version that was identical in appearance and content to the tailored version except for the tailored components (described elsewhere) [31].

Methods used to establish the tailored website's architecture, content, and tailoring included the following:

1. A conceptual model reflecting empirically supported theories and intervention strategies for attitude and behavior change (Figure 1).
2. Developing informational content for the intervention using data from a previous intervention [32], the published literature, and end-user input.
3. Assessing the architecture of the intervention by evaluating an untailored prototype with usability testing and one-on-one interviews with potential end users of the website.
4. Iteratively developing and testing different message tailoring approaches using (a) survey data that assessed the relationships between maternal values and vaccination beliefs, (b) structured interviews with potential end users of the intervention on different types of message framing approaches in combination with tailored information, and (c) health communication expert and research team consensus.

Figure 1. Conceptual model for a tailored messaging intervention to promote childhood vaccination.



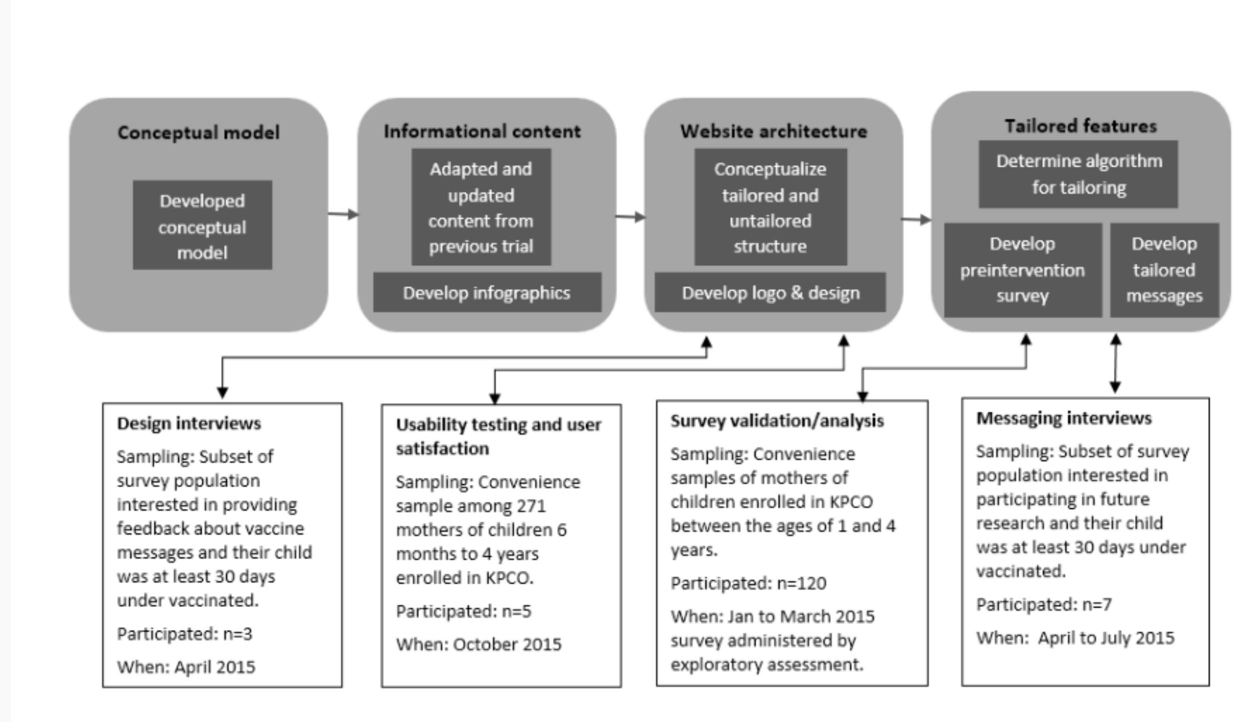
Conceptual Model

The theoretical framework for this intervention (Figure 1) was a hybrid of the Theory of Planned Behavior (TPB) [33] and the Values-Attitudes-Behavior (VAB) model [34]. According to the TPB, behavior (in this case, childhood vaccination) is directly influenced by intentions (intention to vaccinate), which are based on one's attitudes, perceived norms, and perceived behavioral control. In the TPB, perceived behavioral control is a function of control beliefs (eg, beliefs that the decision to vaccinate is in the parents' power to control) and attitudes are a function of behavioral beliefs (eg, beliefs that vaccinating one's child will prevent infectious disease and will not harm the child). The VAB theory goes beyond the TPB to posit that personal values are factors influencing attitudes—this is the basis for our hypothesis that values-tailored messaging would be an effective behavior change target in the intervention. We also used select principles of motivational interviewing (MI) [35] and persuasive messaging [36,37] to inform intervention design, aiming to increase the personal relevance of the materials while minimizing reactance. Specifically for MI, the idea of *rolling with resistance* (not directly counteracting a person's antivaccination attitudes for example) is believed to decrease a person's reactance to receiving information that may be counter to their current beliefs [35,38]. Persuasive messaging in this context refers to making the information more personally relevant by tailoring the information to the person's needs and more trustworthy by providing information that is perceived as accurate and balanced (two-sided messaging) with regard to potential *risks and harms* related to vaccination.

Setting and Participants

The setting for this project was Kaiser Permanente Colorado (KPCO). KPCO is a managed care organization in the Denver metropolitan area that maintains an electronic health record with demographic, medical encounter, and vaccination data on all members. Between January 2015 and October 2015, various convenience samples (described below and in Figure 2) of mothers of young children currently enrolled at KPCO were recruited for the design and tailored message testing interviews, usability testing, and surveys.

Figure 2. Study participant flow and activities.



Design interviews
Sampling: Subset of survey population interested in providing feedback about vaccine messages and their child was at least 30 days under vaccinated.
Participated: n=3
When: April 2015

Usability testing and user satisfaction
Sampling: Convenience sample among 271 mothers of children 6 months to 4 years enrolled in KPCO.
Participated: n=5
When: October 2015

Survey validation/analysis
Sampling: Convenience samples of mothers of children enrolled in KPCO between the ages of 1 and 4 years.
Participated: n=120
When: Jan to March 2015 survey administered by exploratory assessment.

Messaging interviews
Sampling: Subset of survey population interested in participating in future research and their child was at least 30 days under vaccinated.
Participated: n=7
When: April to July 2015

Developing Informational Content

Informational content for the website was developed by editing content adapted from an intervention used in a previous trial to reflect the most current vaccine information and recommendations [32], and developing new content based on emerging vaccination issues identified by vaccine experts and from vaccine questions received from parents in a previous trial [39]. When possible, content was contextualized to the local Colorado environment and health system (ie, reflected vaccines in the KPCO vaccine formulary) and was written at an eighth-grade reading level. All information presented was evidence based and included updated references from peer-reviewed journal articles and materials from the Centers for Disease Control and Prevention.

Assessing and Finalizing the Architecture of the Intervention

Design Interviews, Usability Testing, and User Satisfaction

We solicited feedback from parents of children aged 1 to 2 years on preliminary color schemes, logos, and general architecture of the intervention using printed prototypes. These parents were recruited based on a prior indication of interest in participating in the intervention's development when participating in the values survey described in the following sections.

On the basis of their feedback, we created an interactive, untailored, Web-based prototype to assess usability and user satisfaction. Using the electronic health record, we identified a random sample of 271 English-speaking mothers of children aged 6 months to 4 years currently enrolled in KPCO and not part of the previous study [32]. These individuals were recruited via email. Usability was assessed using the *think aloud* methodology, in which users provided verbal feedback as they completed specific tasks, such as logging on to the website and identifying specific types of information. User satisfaction was then assessed using the System Usability Scale (SUS) that includes 10 questions measured on a 5-point Likert scale [40]. SUS scores above 70 (total possible range 0-100) are considered passable. This feedback was incorporated to create the final version of the tailored (and untailored) websites.

Iterative Development of Tailored Messages Incorporating Values and Framing

To create a personalized Web-based experience for each participant, the intervention was designed such that a preintervention survey would assess individuals' values, beliefs, and vaccine hesitancy. The Web-based tailoring engine would then use this information to present corresponding tailored messages incorporated with evidence-based informational content about vaccination to each user. The process by which we developed and evaluated the values tailoring and message framing strategies is described in the following sections.

Values Survey

We first conducted a survey among KPCO mothers to identify which values were important for vaccination and whether values may have an impact beyond tailoring on beliefs. Described elsewhere [24], these data suggested that values had an influence on vaccination behavior that was separate but additive to the influence from beliefs and should therefore be included as a tailoring target. In this paper, we report our exploratory assessment of the associations between values and different beliefs and concerns as posited that some values would be better matched than others to some vaccination beliefs. Owing to time constraints of the study, we used email to recruit a convenience sample of mothers of children aged between 1 and 4 years who were currently enrolled at KPCO. Surveys were completed on the Web using REDCap (Research Electronic Data Capture, Nashville, TN, USA) and SurveyGizmo (SurveyGizmo, Boulder, CO, USA) survey systems. A US $30 incentive was provided. In addition to vaccination values, the survey also included measures of vaccination attitudes and beliefs and a global values measurement scale called the Schwartz Theory of Basic Human Values [21] to ascertain how values and attitudes relate to one another. Specific survey measures that were included are described in the following sections.

Vaccine Values

Parental values pertaining to childhood vaccination decisions (*vaccine values*) were measured using a scale the study team developed and assessed during this survey, called the Parental Vaccine Value Scale (PVVS). The PVVS is a 20-item scale that assesses 6 domains of values related to vaccination: security—disease prevention (valuing protecting one’s children from the harm of infectious disease; Cronbach alpha=.74), security—vaccine risk (valuing protecting one’s children from perceived harm of vaccines; α=.73), universalism (valuing protecting one’s community as a whole from the harm of infectious disease; α=.86), self-direction (valuing the process of gathering information to make an informed decision; α=.66), conformity (valuing the recommendations of experts and authority; α=.62), and tradition (valuing following the established norm in one’s religion or family; α=.79). These 6 domains were created to reflect a subset of global values in the Schwartz Theory of Basic Human Values, which was also examined during the survey [21,41,42]. The development and validation of the PVVS, including factor structure and alignment with the Schwartz global values, is described in a separate publication [24]. On the basis of the results of this analysis, the PVVS was chosen in favor of the Schwartz Theory of Basic Human Values scale as the tailoring target in the intervention. Values were assessed with a 4-point Likert scale that ranged from strongly agree to strongly disagree.

Vaccine Beliefs and Concerns

Beliefs and concerns about childhood vaccination were measured using a 10-item scale developed in a previous study [39], with three additional questions later added by the study team for the purposes of the project. A 5-point Likert scale (strongly agree to strongly disagree) was used to assess the responses.

Intention to Vaccinate

In all, 2 items that assessed mothers’ intention to vaccinate their newborn during the first year of life were used for message tailoring. These were based on performance of similar questions in a past study [39]. One item assessed how many of the 8 vaccines in the infant series a participant planned to have their infant receive with the following options: *none of the vaccines*, *some of the vaccines*, and *all of the vaccines*. The second item assessed when the mother intended to have their infant vaccinated with the following response options: *all on-time as recommended by my baby’s doctor* and *all or some later than my baby’s doctor recommends* (often referred to as a *delayed scheduled* or an *alternative schedule*). Combining these 2 items, mothers were categorized into 3 groups: refuses all vaccines, uses an alternative schedule (receives some or all vaccines but does not follow recommended timing), and full vaccine acceptor. The first 2 of these categories were grouped together to define *vaccine-hesitant* mothers when assessing vaccination outcomes at the end of the trial.

Analysis of Survey Data

From the survey data, we computed Spearman correlations between PVVS domain scores and the belief items. These results, combined with subsequent research team input, qualitative data, and expert opinion, were used to determine which values were associated with each specific belief or concern about vaccination and would therefore be included in the tailoring algorithm.

Interviews Exploring Message Framing and Values Incorporation

We considered several message framing options for the intervention messages. Using general information marketing approaches [43-45], we evaluated the potential usefulness of combining message framing with values framing approaches for the intervention messages. We considered the following approaches: 1) only values tailoring; (2) 1-sided messages, which present information about vaccines without acknowledgement of antivaccine arguments or negative aspects of vaccination [44]; (3) nonrefutational two-sided messages, which present both provaccination information and potential negatives (such as side effects) to the same extent (50% of information is pro and 50% is con) [44]; (4) *push* messages, which directly refute myths about vaccines and adopt a directive tone; and (5) *pull* messages, which provide information and invite the reader into a conversation or discussion about the topic while trying to encourage the desired action or behavior and encourage central processing or personal engagement in the content [37,43,45].

These various framing options were tested, along with the impact of different values and beliefs combinations, using structured interviews with 7 new mothers. The interviews took place between April and July 2015. Mothers were recruited from participants who took the survey described above and were eligible if they indicated in the survey an interest in providing feedback about vaccine messages, their child was at least 30 days undervaccinated as indicated in the medical record, and the mother had endorsed at least one of the six vaccine values domains in their survey (however, the messages tested were not necessarily matched to mothers’ most highly affirmed values, though this was attempted when possible). Potential participants were recruited using email and phone outreach. The interviews focused on assessing the participant's acceptability of the messages and preferences between different formats. The interviews continued until saturation was reached. The participants received a US $50 gift card for their time.

We designed the architecture of the intervention based on the concept of a *home page* consisting of different *tiles* representing different vaccination issues (Figure 3) as a central navigation point. We planned to have the most highly tailored information for the intervention presented in 3 *Just for You* tiles located prominently on the top of home page (Figure 3). These tiles represented the mothers' three most pressing concerns or questions regarding vaccination. The intent of these tiles was to (1) facilitate the ease with which each parent could locate the content of highest interest, (2) increase mothers' receptivity to information about their vaccine concern(s), (3) succinctly and credibly summarize key information about the concern, and (4) provide interested mothers with access to additional information about their concern. In the message framing interviews, 4 mothers were presented with up to 9 messages, which addressed as many as three featured concerns in the *Just for You* tiles. These messages included up to three messages tailored to value domains and up to three messages with different framing options (ie, a 3 × 3 factorial design; see Table 1 for examples of values tailoring messages). After viewing all the messages for an area of concern, mothers ranked the messages in order of their preferences and described their reasoning around their rankings. When a mother's top choices aligned with their measured values, we considered the message a candidate for the final intervention. When mothers' measured values and top ranked messages conflicted, we revised the messages to better align with the reasons that the mothers provided in the message framing interviews. Revised messages were further tested in subsequent interviews using the same interview format but with 3 different mothers.

Figure 3. Architecture of the Final Tailored Intervention - Vaccines and Your Baby Home Page.

Table 1. Examples of values-tailored messages. (Vaccines and Your Baby: Tailored Messages.)

Value	Topics	
	Alternative/delayed vaccine schedules message	Doing your own research on vaccines message
Security—disease prevention	Like many parents, your main goal is to keep your child healthy. The last thing you want is for your child to get an illness you could have prevented with a simple vaccine.	You're the kind of person who will do everything she can to protect her baby from illnesses.
Self-direction	You're not one to just do what other people tell you to do. You know your child better than anyone, and you have choices to make. You want to do your own research about vaccines. You don't want him/her to get a disease. But you don't want to put him/her at risk by getting vaccines.	You're the kind of person who plays an active role in decisions about her baby's health.
Security—vaccine risk	That's a lot of needles (and a lot of tears)! You want to protect your child. But with so many vaccines at once, you're concerned about exposing him/her to too many unnatural ingredients all at once.	You're the kind of person who will do everything she can to protect her baby from pain or unnecessary medicines.

Tailoring Components

Tailoring on Beliefs, Concerns, Hesitancy, and Demographics

Consistent with many previous tailored interventions and the TPB, we planned a priori to incorporate tailoring based on mothers' beliefs and concerns, intention to vaccinate, and demographic characteristics, as described in the following sections.

Belief Tailoring

Of the 13 vaccination concerns assessed in the preintervention survey, those receiving an average score of 3 or less (some were composite measures, possible range of 1-5, with lower values corresponding to increasing antivaccine views) were considered a *qualifying concern* that could potentially be tailored on. This value was chosen as the cutoff for categorizing a concern as qualifying, given that it would capture participants with less-than-positive (ie, neutral or negative) beliefs about the topic. If more than three concerns met this threshold, mothers were asked to select their top three concerns that formed the basis of the *Just for You* tiles. If only three concerns were identified, they became the 3 *Just for You* tiles. Concerns beyond the top 3 selected by the mother were highlighted in the website's base content (see Figure 3). If there were less than three concerns, up to 3 default *Just for You* tiles were featured in priority order: *The Vaccine Schedule*, *Baby Vaccine Visit*, and *Kaiser Permanente Clinics*. These topics were chosen as they were felt to have the least potential for raising new concerns among mothers without at least three concerns. For mothers with no vaccination concerns, *Just for You* content was tailored (first tile only) by providing positive reinforcement about their decision to vaccinate. The vaccine schedule tile was tailored based on the child's age at the time the website was being viewed.

Intention Tailoring

We tailored content on vaccination intention based on responses to the 2 vaccine intention questions described earlier (how many of the 8 vaccines in the infant series a participant planned to have their infant receive and when the mother intended to have their infant vaccinated). These responses were combined, and mothers were categorized into 1 of 3 mutually exclusive groups for tailoring (as opposed to only 2 groups for the planned final vaccination analysis of the trial) that was incorporated into the content of the *Just for You* tiles: refuses all vaccines, uses an alternative schedule (receives some or all vaccines but does not follow recommended timing), and full vaccine acceptor.

Demographic Tailoring

Tailoring on personal characteristics, including the mother's pregnancy status and child's nickname, was incorporated throughout the website content.

Results

Developing Informational Content

In total, there were 17 general topic areas that the intervention addressed, divided into 41 specific content areas (Table 2). These 41 content areas were further grouped into 6 broad categories corresponding to the 6 *standard* tiles (ie, not the *Just for You* tiles) presented on the intervention's home page (Figure 3). Furthermore, 4 of the 41 content areas were newly developed for the intervention. The remaining 37 were adapted from the previous intervention [39]. In addition, three newly developed, interactive infographics were also included on the following topics: herd immunity (included under the *Community Benefits* topic), antigen counts in the past and current vaccines (included under the *Vaccine Ingredient Types* topic), and disease rates before and after vaccines (included under the *Risk of Diseases* topic). Content areas included in the 6 home page tiles were identical in the tailored and untailored websites.

Table 2. Topic areas (general tops: n=17; specific topics: n=41) covered in the intervention and corresponding tile on the home page (n=6).

General topic area of interest and specific topics covered within each	Name of corresponding tile on home page
Vaccine schedule	
Recommended vaccine schedule	Vaccine schedule and timing
Vaccines	Vaccine schedule and timing
Safety of the schedule	Vaccine schedule and timing
How the schedule is made	Vaccine schedule and timing
Importance of vaccine timing	Vaccine schedule and timing
Alternative schedules	
Reasons why we cannot recommend an alternative schedule	Vaccine schedule and timing
Immunity and timing	
Baby’s developing immune system	Vaccine schedule and timing
Parents’ main concerns about baby vaccines	Vaccine schedule and timing
Vaccine safety research	
How vaccine studies are done	Vaccine safety and side effects
Vaccine Adverse Event Reporting System, Vaccine Safety Datalink, Vaccine Injury Compensation Program—How vaccine side effects are reported	Vaccine safety and side effects
Doing your own research[a]	Vaccine safety and side effects
Vaccine side effects	
Mild/common side effects	Vaccine safety and side effects
Serious/rare side effects	Vaccine safety and side effects
Illnesses/conditions not currently linked to vaccines	Vaccine safety and side effects
Who should be vaccinated?	Vaccine safety and side effects
How vaccines are made	
Vaccine production	How vaccines work
How Pharma works	How vaccines work
Kaiser and Pharma	How vaccines work
Vaccine ingredients	
Why some ingredients are needed for vaccine production	How vaccines work
Vaccine ingredient types	How vaccines work
Individual ingredients	How vaccines work
Vaccines and immunity	
Natural versus vaccine immunity	How vaccines work
Why vaccinate	
Community benefits of vaccination	Reasons to vaccinate
Risk of diseases[a]	Reasons to vaccinate
Worldwide risk of diseases[a]	Reasons to vaccinate
Current outbreaks (measles and pertussis)[a]	Reasons to vaccinate
Diseases prevented	
Vaccine-preventable diseases and vaccines given at Kaiser Permanente Colorado	Reasons to vaccinate
The baby vaccine visit	
Before the vaccine visit	Getting vaccines at Kaiser
During the vaccine visit	Getting vaccines at Kaiser

General topic area of interest and specific topics covered within each	Name of corresponding tile on home page
After the vaccine visit	Getting vaccines at Kaiser
Maternal vaccination	
Vaccines in pregnancy	Getting vaccines at Kaiser
Safety of vaccines in pregnancy	Getting vaccines at Kaiser
Kaiser Permanente clinics	
Transportation to Kaiser clinics	Getting vaccines at Kaiser
Bus routes	Getting vaccines at Kaiser
Kaiser clinic hours	Getting vaccines at Kaiser
Colorado vaccine laws	
Colorado vaccine laws	Learn more
Colorado vaccine exemptions	Learn more
Travel vaccination	
Kaiser travel clinic	Learn more
Vaccines and traveling abroad	Learn more
Adolescent vaccination	
Adolescent vaccination	Learn more
Other FAQs[b]	
Other FAQs	Learn more

[a] Newly developed for the intervention.

[b] FAQ: frequently asked question.

Assessing and Finalizing the Architecture of the Intervention

Design Interviews

Participants included 3, white, non-Hispanic mothers of children aged between 1 and 4 years. All the mothers preferred design features that appeared to be associated with the health system. This included color palettes with blue and green and logos with clean lines, similar to KPCO website pages. On the basis of the user feedback, we selected a tailored (and matching untailored) website design that used an interactive tile-based homepage (Figure 3), menu navigation in multiple locations, layering of information using accordion style grouping, and pop-up information in select locations.

Usability Testing and Satisfaction Survey Results

Of 271 mothers contacted for usability and satisfaction testing sessions, we scheduled a convenience sample of the first 6 respondents for usability testing interviews. Mothers who completed the interviews (n=5) were all female, white, non-Hispanic, with at least some college education, and had income ranging from US $50,000 to US $90,000 or more. On the basis of their input, changes to the prototype architecture included adding submenus to the main page, using words in place of images to assist in identifying the schedule toggle feature, adding new links to content pages, and automatically closing content accordions to assist in reading and navigating long content areas. All but 1 user had SUS scores above the passable score of 70 (range 65-97.5), suggesting that the usability of the site was acceptable. Owing to the similarity in responses among mothers, additional interviews beyond these 5 were not undertaken.

Description of the Architecture of the Final Tailored Intervention

In the final intervention, mothers are first directed to an onboarding page designed to engage the participant in the website so that they are inclined to continue to view the content. It includes a welcome message, references the participant's intention to vaccinate, explains the intent of the website and that the information presented will be tailored based on their survey responses, and visually depicts where they can find the tailored content.

Following this, mothers are taken to the main home page of the intervention (Figure 3). On this page, the bulk of message tailoring is received via the 3 featured *Just for You* tiles. Any additional topics of concern beyond the top three concerns are highlighted within the 6 interactive tiles on this page. All the tiles lead to additional content that is tailored on the participant's personal characteristics.

Iterative Development of Tailored Messages Incorporating Values and Framing

Values Survey Results

Table 3 shows the results of correlations between belief items and values. These results informed team discussions about which values might be most appropriate for framing each concern topic. A combination of these data, research team consensus, and results from the parent interviews (described in the following sections) were used to determine which values were available as tailoring targets for each of the vaccine beliefs and concerns.

Table 3. Correlations between values and beliefs corresponding to website topic areas. Correlations (r) are Spearman correlation coefficients.

Beliefs[a]	Associations with provaccine beliefs			Associations with antivaccine beliefs		
	Universalism, r	Conformity, r	Security (disease prevention), r	Security (vaccine risk), r	Tradition, r	Self-direction, r
Enough research	−0.249[b]	−0.202[c]	−0.126	0.339 [d]	0.247 [b]	0.261 [b]
Disease risk and benefit	−0.423[d]	−0.257[b]	−0.311[b]	0.202 [c]	0.199 [c]	0.080
Too many, too soon	−0.139	−0.364[d]	-0.199 [c]	0.411 [d]	0.257 [b]	0.365 [d]
Natural immunity	−0.315[b]	−0.226[c]	−0.204[c]	0.259 [b]	0.324 [d]	0.106
Vaccine safety	−0.123	−0.173	−0.078	0.352 [d]	0.360 [d]	0.180
Do own research	−0.173	−0.180[c]	−0.221[c]	0.249 [b]	0.178	0.505 [d]
Vaccine ingredients	−0.251[b]	−0.347[d]	−0.117	0.239 [b]	0.240 [b]	0.156
Autism	−0.205[c]	−0.231[c]	−0.092	0.338 [d]	0.358 [d]	0.164
Vaccine risk versus benefit	−0.186[c]	−0.115	−0.343[d]	0.474 [d]	0.344 [d]	0.333 [d]
Combined risk/benefit	−0.319[d]	−0.176	−0.351[d]	0.412 [d]	0.301 [d]	0.257 [d]

[a] Three additional vaccine topics were added based on expert feedback after this analysis was completed for a total of 13 belief topics.

[b] P<.01.

[c] P<.05.

[d] P<.001.

Interviews Exploring Message Framing and Values Incorporation

Of the 39 mothers contacted for interviews, 7 participated. All of them were mothers of children aged between 1 and 2 years.

Across topic areas, qualitative interviews assessing the general approach for framing messages revealed that two-sided messages were preferred by mothers compared with 1-sided, *push*, or *pull* messages. The two-sided message on side effects, presented as a table, was particularly well received. All mothers responded favorably to the intervention layout and reported that presenting both mild and rare side effects seemed *honest*. On the topic of alternative schedules (delaying or skipping certain vaccines), mothers generally (n=5) preferred the two-sided messages, largely because of the detail provided. The 2 mothers who followed an alternative schedule disliked the *black-and-white* messaging against this practice, believing this approach could sway mothers away from vaccinating at all. Finally, when 1- and two-sided messages were presented without values framing, mothers often (n=5) mentioned that values framing would improve the likeability of the message.

Among the values-framed messages, most (n=4) mothers preferred statements from *self-direction* values-tailored versions that acknowledged their *right to choose*. Mothers also responded favorably to the tailored versions of *security* that specifically addressed mothers' motivation for keeping their child safe. Values-framed messages around conformity were the least preferred, with only 1 mother in the sample endorsing this value strongly. Values were not favored by any of the mothers for messages on the topic of side effects. These findings guided the team to adopt a two-sided messaging approach that incorporated values tailoring for topics of concern identified for mothers.

Tailoring Components

Final Values Tailoring Algorithm

To build the final values tailoring algorithm, we used the results from the survey and interviews to identify the values or set of values that appeared impactful and relevant for each belief item assessed (Table 4). Any values exceeding the threshold of a 2.5 score (a score well above the *neutral* value of 2 on the 4-point scale used to assess values) that were also deemed impactful for that belief based on the survey and interview results were considered *relevant values*. We built an algorithm to randomly select one of these values and incorporate it into the corresponding *Just for You* tile. If the value has already been used in a previous topic area, another available value is selected at random from the available pool of values for that topic. If no additional values remain, all previously available values are made available for random selection. If the participant has no values that meet the threshold, a message without values framing for that vaccination concern is displayed. Values are generally incorporated into these tiles as *wrap-around* introductory sentences allowing for the core informational content on that belief to remain similar for each user. The same home page structure consisting of 9 tiles (Figure 2) was used for each retailoring.

Table 4. Values available for tailoring according to topic areas of the Just for You Tiles.

Title of the *Just for You* tile (content of corresponding general or specific topic areas potentially linked to tile)[a]	Applicable values					
	Security—disease risk	Security—vaccine risk	Self-direction	Conformity	Universalism	Tradition
Vaccine Safety Research (vaccine safety research)	x[b]	x	x	x	—[c]	—
Vaccine-Preventable Diseases (why vaccinate and diseases prevented)	x	x	x	x	x	—
Number and Timing of Vaccines (how the schedule is made, reasons why we cannot recommend an alternative schedule, immunity and timing, and vaccines and immunity)	x	x	x	x	—	—
Vaccine Ingredients (vaccine safety research and vaccine ingredients)	x	x	x	x	—	—
Vaccine Side Effects (vaccine safety research and vaccine side effects)	x	x	x	—	—	—
Doing Your Own Research on Vaccines (vaccine safety research and doing your own research)	x	x	x	x	x	—
The Immune System and Vaccines (immunity and timing, vaccines and immunity, and why vaccinate)	x	x	—	x	x	—
Vaccines and Autism (side effects and vaccine ingredients)	x	x	x	—	—	—
Vaccination Risks and Benefits (vaccine side effects, who should not be vaccinated, and why vaccinate)	x	x	x	—	x	—
Vaccines During Pregnancy (maternal vaccination)	x	x	x	x	—	—
The Role of Pharmaceutical Companies (how vaccines are made)	—	x	x	—	—	—
Alternative/Delayed Vaccine Schedules (reasons why we cannot recommend an alternative schedule, immunity, and timing)	x	x	x			—
Tips for Vaccinating at Kaiser (recommended vaccine schedule, baby vaccine Visit, and Kaiser Permanente clinics)	No values tailoring	No values tailoring	No values tailoring	No values tailoring	No values tailoring	No values tailoring

[a] Multiple General or Specific Topics related to several of the Just for You tiles. These tiles could include information on 1 or more general or specific topics depending on the user's input.

[b] An 'x' in the table denotes that a given value is available to incorporate into the *Just for You* tile content.

[c] A '—' in the table denotes that the given value is not available to incorporate into the *Just for You* tile content.

Discussion

Principal Findings

In this paper, we described in detail how we developed a Web-based, tailored messaging intervention to address maternal vaccine hesitancy using an iterative development process and a mixed method approach. This intervention, which was developed using validated health behavior theories, is expected to be effective, engaging, and easy to use based on end-user feedback and pilot testing. Ultimately, the intervention included common tailoring elements such as demographics and beliefs as well as a novel tailoring target, personal values. In addition to examining tailoring targets, we also used user feedback to assess the potential impact of different message framing strategies in combination with the tailored elements. This was necessary as there is ongoing debate about the optimal messages framing strategy with regard to vaccine hesitancy and immunization [46-51].

Application of Theory

Of the 4 health behavior theories and techniques used to inform the development of the intervention, the TPB has the most evidence for its applicability to vaccine decision making. Owing to this, we planned a priori to incorporate elements from this theory into our intervention and therefore did not focus any of our data collection described in this paper on these elements. Specifically, the intervention was tailored based on TPB constructs of attitudes, and when possible, elements of social norms were incorporated into the messages. The other remaining theories and techniques—the VAB theory and elements of persuasive messaging and MI—did not have a strong evidence base with regard to vaccine decision making at the time the intervention was being developed. However, the data collected in this study suggest that each is relevant to the vaccination decision. Specifically, results from our quantitative survey of values and from the qualitative interviews on message framing and values tailoring both suggest that certain values are more important than others with regard to vaccine decision making, and that some values are better incorporated with certain beliefs than others. A subsequent study done by our group examining the role that values play in mothers' vaccine decision making further supports this notion [24]. Support for incorporating techniques from persuasive messaging into the intervention was also derived from the qualitative message framing and values tailoring interviews. The respondents clearly indicated that vaccine-hesitant parents strongly preferred messages that were perceived as *balanced*, where both positive and negative information (ie, two-sided messages) about vaccine safety or benefits was presented as these messages were deemed more trustworthy than messages that only conveyed one side of these issues. This point supports the concept in persuasive messaging that messages deemed as trustworthy are more likely to be reflected upon and more persuasive. Also supporting the importance of persuasive messaging was the finding from the message framing and values tailoring interviews that showed that messages tailored to the user were deemed more personally relevant. The design feedback interviews showing that mothers uniformly liked intervention architecture that clearly allowed a user to choose the specific information to view also support this concept. Evidence from our study for incorporating techniques from MI into the intervention is somewhat indirect in that mothers participating in the message framing and values interviews who were following an alternative vaccination schedule did not like *black-and-white* messages that argued against this practice. Such messages could be considered to counter the MI tenet of *rolling with resistance*. Another important MI tenet is that of using intrinsic motivation to effect behavior change. Evidence in support of this concept also comes from the message framing and values tailoring interviews showing that mothers generally preferred values-tailored messages to those not tailored to values. Further work by our group and others that occurred after our intervention was developed lends additional support to the important role that MI likely plays in motivating parents to vaccinate [52,53].

Future Work

The next step in our study is to assess the efficacy of this intervention in a randomized controlled trial of KPCO expectant and new mothers. This trial (ClinicalTrials.gov protocol number NCT02665013) will examine the relative effect of the tailored versus untailored websites for their effect on timely infant vaccine utilization during the first 15 months of life. In addition, a number of secondary outcomes will also be assessed including whether the intervention modified maternal vaccination beliefs and concerns or vaccine hesitancy, and how any changes in these outcomes relate to vaccination values. This will allow for a more thorough investigation of the VAB model used as a basis for this study.

Potential Importance of Values

Tailoring on values represents a novel, and potentially important, innovation incorporated into this intervention. A large body of literature in the social sciences provides compelling evidence that aligning educational messages with personal values can make information more salient and actionable [54-57]. The somewhat more limited number of studies examining values tailoring in the realm of health behavior change specifically further support this concept [58-61]. Recently, researchers interested in improving immunization delivery have begun to recognize the potential role of values tailoring in promoting vaccination [23,27,62]. Although there has been limited prior research examining the impact of values tailoring on vaccine acceptance, our own study [24] and that of others [63], which were completed after the study presented in this paper, have shown that differences in personal values are associated with variation in the acceptance of recommended vaccines. Taken together, values appear to be a potentially untapped tailoring resource that warrants further exploration. We believe results from the randomized controlled trial that will evaluate this intervention will add important and novel information to this growing body of research.

Limitations

This paper should be considered in the context of some important limitations. First, when developing the various aspects of the intervention, we generally received input from only a small number of mothers, most of whom were white and non-Hispanic. Maternal input was not designed to be comprehensive. Rather, we opted to obtain in-depth information from a handful of mothers to provide more nuanced insight for optimizing the intervention and making it more relevant to potential end users, and we generally solicited mothers' input until thematic saturation was reached (although we did not do a formal qualitative analysis). In addition, we focused much of the intervention development on the incorporation of values as a tailoring element, an approach to tailoring that is in need of further study. Moreover, although the study population providing input into the interventions' design was diverse with regard to demographic characteristic such as race, ethnicity, education, and income, the population was relatively homogeneous with regard to insurance status (all had access to care) and primary language (we only gathered data from English-speaking mothers) and did not include fathers. As such, any impact of the intervention found in the upcoming clinical trial may not be generalizable to other populations that differ in these respects, and the intervention may need further refinement based on these characteristics. In addition, there are several items that previous research has shown as potentially important variables in the vaccination decision that could represent potential tailoring variables (education level, exposure to scientific controversy, degree of social networking, etc) [64,65] and were not included in our intervention. This was a purposeful decision to be able to isolate any potential impacts of values tailoring on vaccine uptake. However, these variables may need to be considered as tailoring targets in future iterations of the intervention. A notable strength of the project was the use of multiple methods to collect data (qualitative and quantitative), which may increase the validity of our findings.

Conclusions

We used both qualitative and quantitative approaches and significant end-user input to develop a Web-based, theory-driven, tailored messaging intervention designed to address maternal vaccine hesitancy and subsequently improve uptake of infant vaccines. Our results suggest that elements from each of the 4 models and techniques incorporated into our conceptual model for the intervention were important. Specifically, results of qualitative user design and message framing interviews supported using several techniques from persuasive messaging and MI in the intervention. A quantitative survey on parents' vaccination values, combined with qualitative data from the message framing interviews, supported the importance of the VAB model as a foundation of the intervention and supported the idea of using values as a novel tailoring variable. Elements from the TPB were not assessed directly in this study as there was already an evidence base supporting the importance of this theory in parents' vaccine decision making. By incorporating elements from these four theories and techniques into the intervention, we believe it will be highly effective in changing mothers' vaccination attitudes and behaviors. This hypothesis will be tested in a subsequent randomized controlled trial.

Acknowledgments

This project was funded by the National Institutes of Health Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD079457).

Conflicts of Interest

AD serves on the advisory boards for Merck, Sanofi, and Pfizer and has worked as a consultant for Pfizer. None of these companies played a role in this research, and she does not receive research funding from these groups. All other authors have no conflicts to declare.

References

1. Roush SW, Murphy TV, Vaccine-Preventable Disease Table Working Group. Historical comparisons of morbidity and mortality for vaccine-preventable diseases in the United States. JAMA 2007 Nov 14;298(18):2155-2163. [doi: 10.1001/jama.298.18.2155] [Medline: 18000199]
2. Omer SB, Pan WK, Halsey NA, Stokley S, Moulton LH, Navar AM, et al. Nonmedical exemptions to school immunization requirements: secular trends and association of state policies with pertussis incidence. JAMA 2006 Oct 11;296(14):1757-1763. [doi: 10.1001/jama.296.14.1757] [Medline: 17032989]
3. Glanz JM, Newcomer SR, Narwaney KJ, Hambidge SJ, Daley MF, Wagner NM, et al. A population-based cohort study of undervaccination in 8 managed care organizations across the United States. JAMA Pediatr 2013 Mar 01;167(3):274-281. [doi: 10.1001/jamapediatrics.2013.502] [Medline: 23338829]
4. Omer SB, Salmon DA, Orenstein WA, deHart MP, Halsey N. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. N Engl J Med 2009 May 7;360(19):1981-1988. [doi: 10.1056/NEJMsa0806477] [Medline: 19420367]
5. Siddiqui M, Salmon DA, Omer SB. Epidemiology of vaccine hesitancy in the United States. Hum Vaccin Immunother 2013 Dec;9(12):2643-2648 [FREE Full text] [doi: 10.4161/hv.27243] [Medline: 24247148]
6. Sundaram ME, Guterman LB, Omer SB. The true cost of measles outbreaks during the postelimination era. JAMA 2019 Mar 26;321(12):1155-1156. [doi: 10.1001/jama.2019.1506] [Medline: 30844023]
7. Salmon DA, Dudley MZ, Glanz JM, Omer SB. Vaccine hesitancy: causes, consequences, and a call to action. Vaccine 2015 Nov 27;33(Suppl 4):D66-D71. [doi: 10.1016/j.vaccine.2015.09.035] [Medline: 26615171]
8. N/A. Assessing the State of Vaccine Confidence in the United States: Recommendations from the National Vaccine Advisory Committee: Approved by the National Vaccine Advisory Committee on June 9, 2015 [corrected]. Public Health Rep 2015;130(6):573-595 [FREE Full text] [doi: 10.1177/003335491513000606] [Medline: 26556929]
9. Dempsey AF, Wagner N, Narwaney K, Pyrzanowski J, Kwan BM, Kraus C, et al. 'Reducing Delays In Vaccination' (REDIVAC) trial: a protocol for a randomised controlled trial of a web-based, individually tailored, educational intervention to improve timeliness of infant vaccination. BMJ Open 2019 May 22;9(5):e027968 [FREE Full text] [doi: 10.1136/bmjopen-2018-027968] [Medline: 31122997]
10. Trivedi D. Cochrane review summary: face-to-face interventions for informing or educating parents about early childhood vaccination. Prim Health Care Res Dev 2014 Oct;15(4):339-341. [doi: 10.1017/S1463423614000322] [Medline: 25146256]
11. Sadaf A, Richards JL, Glanz J, Salmon DA, Omer SB. A systematic review of interventions for reducing parental vaccine refusal and vaccine hesitancy. Vaccine 2013 Sep 13;31(40):4293-4304. [doi: 10.1016/j.vaccine.2013.07.013] [Medline: 23859839]
12. Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert Rev Vaccines 2015 Jan;14(1):99-117. [doi: 10.1586/14760584.2015.964212] [Medline: 25373435]
13. MacDonald NE. Vaccine hesitancy: definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]

14. MacDonald NE, Butler R, Dubé E. Addressing barriers to vaccine acceptance: an overview. Hum Vaccin Immunother 2018 Jan 02;14(1):218-224. [doi: 10.1080/21645515.2017.1394533] [Medline: 29048975]
15. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
16. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. Ann Behav Med 1999;21(4):276-283. [Medline: 10721433]
17. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
18. Stockard J, Carpenter G, Kahle LR. Continutity and change in values in midlife: testing the age stability hypothesis. Experimental Aging Research 2014 Mar 13;40(2):224-244. [doi: 10.1080/0361073x.2014.882215]
19. Rokeach M. The Nature Of Human Values. New York: Free Press; 1973.
20. Konty MA, Dunham CC. Differences in value and attitude change over the life course. Sociological Spectrum 2010 Jul 30;17(2):177-197. [doi: 10.1080/02732173.1997.9982159]
21. Schwartz S. Are there universal aspects in the structure and contents of human values? J Soc Issu 1994;50:19-45. [doi: 10.1111/j.1540-4560.1994.tb01196.x]
22. Bedford H, Attwell K, Danchin M, Marshall H, Corben P, Leask J. Vaccine hesitancy, refusal and access barriers: the need for clarity in terminology. Vaccine 2018 Oct 22;36(44):6556-6558. [doi: 10.1016/j.vaccine.2017.08.004] [Medline: 28830694]
23. Leask J, Macartney K. Parental decisions about vaccination: collective values are important. J Paediatr Child Health 2008 Oct;44(10):534-535. [doi: 10.1111/j.1440-1754.2008.01381.x] [Medline: 19012625]
24. Cataldi JR, Sevick C, Pyrzanowski J, Wagner N, Brewer SE, Narwaney KJ, et al. Addressing personal parental values in decisions about childhood vaccination: measure development. Vaccine 2019 Sep;37(38):5688-5697. [doi: 10.1016/j.vaccine.2019.08.009]
25. Dore RA, Stone ER, Buchanan CM. A social values analysis of parental decision making. J Psychol 2014;148(4):477-504. [doi: 10.1080/00223980.2013.808603] [Medline: 24946390]
26. Salmon DA, Omer SB. Individual freedoms versus collective responsibility: immunization decision-making in the face of occasionally competing values. Emerg Themes Epidemiol 2006 Sep 27;3:13 [FREE Full text] [doi: 10.1186/1742-7622-3-13] [Medline: 17005041]
27. Witteman HO. Addressing vaccine hesitancy with values. Pediatrics 2015 Aug;136(2):215-217 [FREE Full text] [doi: 10.1542/peds.2015-0949] [Medline: 26195531]
28. Witteman HO, Chipenda Dansokho S, Exe N, Dupuis A, Provencher T, Zikmund-Fisher BJ. Risk communication, values clarification, and vaccination decisions. Risk Anal 2015 Oct;35(10):1801-1819. [doi: 10.1111/risa.12418] [Medline: 25996456]
29. Glanz JM, Wagner NM, Narwaney KJ, Shoup JA, McClure DL, McCormick EV, et al. A mixed methods study of parental vaccine decision making and parent-provider trust. Acad Pediatr 2013;13(5):481-488 [FREE Full text] [doi: 10.1016/j.acap.2013.05.030] [Medline: 24011751]
30. Glanz JM, Kraus CR, Daley MF. Addressing parental vaccine concerns: engagement, balance, and timing. PLoS Biol 2015 Aug;13(8):e1002227 [FREE Full text] [doi: 10.1371/journal.pbio.1002227] [Medline: 26252770]
31. Dempsey AF, Wagner N, Narwaney K, Pyrzanowski J, Kwan BM, Kraus C, et al. 'Reducing Delays In Vaccination' (REDIVAC) trial: a protocol for a randomised controlled trial of a web-based, individually tailored, educational intervention to improve timeliness of infant vaccination. BMJ Open 2019 May 22;9(5):e027968 [FREE Full text] [doi: 10.1136/bmjopen-2018-027968] [Medline: 31122997]
32. Glanz JM, Wagner NM, Narwaney KJ, Kraus CR, Shoup JA, Xu S, et al. Web-based social media intervention to increase vaccine acceptance: a randomized controlled trial. Pediatrics 2017 Dec;140(6):- [FREE Full text] [doi: 10.1542/peds.2017-1117] [Medline: 29109107]
33. Ajzen I. The theory of planned behavior. Organizational Beh Hum Dec Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
34. Ajzen I. Values, Attitudes, and Behavior. In: Methods, Theories, and Empirical Applications in the Social Sciences. Wiesbaden: VS Verlag für Sozialwissenschaften; 2012:33-38.
35. Miller W, Rollnick S. Motivational interviewing. New York: Guilford Press; 1991.
36. O'Keefe G. Persuasion: Theory and Research. Thousand Oaks: SAGE Publications; 2002.
37. Petty R, Cacioppo JT. The elaboration likelihood model of persuasion. Adv Exp Med Soc Psych 1986;19:123-205. [doi: 10.1016/s0065-2601(08)60214-2]
38. Gagneur A, Gosselin V, Dubé È. Motivational interviewing: a promising tool to address vaccine hesitancy. Vaccine 2018 Oct 22;36(44):6553-6555. [doi: 10.1016/j.vaccine.2017.10.049] [Medline: 30301523]
39. Shoup JA, Wagner NM, Kraus CR, Narwaney KJ, Goddard KS, Glanz JM. Development of an interactive social media tool for parents with concerns about vaccines. Health Educ Behav 2015 Jun;42(3):302-312. [doi: 10.1177/1090198114557129] [Medline: 25413375]
40. Jordan P, Thomas B, Lyall I. SUS: A quick and dirty usability scale. In: Usability Evaluation in Industry. Bristol, PA: Taylor & Francis; 1996.

41. Schwartz SH, Boehnke K. Evaluating the structure of human values with confirmatory factor analysis. J Res Personality 2004 Jun;38(3):230-255. [doi: 10.1016/s0092-6566(03)00069-2]
42. Schwartz SH. An overview of the Schwartz theory of basic values. Online Read Psychol Cult 2012;2(1):- [FREE Full text] [doi: 10.9707/2307-0919.1116]
43. Olver J, Farris PW. Push and pull: a one-two punch for packaged products. MIT Sloan Manag Rev 1989;31(1):53-61.
44. O’Keefe DJ. How to handle opposing arguments in persuasive messages: a meta-analytic review of the effects of one-sided and two-sided messages. AICA 2016 May 18;22(1):209-249. [doi: 10.1080/23808985.1999.11678963]
45. Robertson T. Chron. 2019. Difference between push & pull marketing URL: https://smallbusiness.chron.com/difference-between-push-pull-marketing-31806.html [accessed 2017-05-01]
46. Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997 Jan;121(1):3-19. [Medline: 9000890]
47. Gerend MA, Shepherd JE. Using message framing to promote acceptance of the human papillomavirus vaccine. Health Psychol 2007 Nov;26(6):745-752. [doi: 10.1037/0278-6133.26.6.745] [Medline: 18020847]
48. Abhyankar P, O'Connor DB, Lawton R. The role of message framing in promoting MMR vaccination: evidence of a loss-frame advantage. Psychol Health Med 2008 Jan;13(1):1-16. [doi: 10.1080/13548500701235732] [Medline: 18066916]
49. Gainforth HL, Cao W, Latimer-Cheung AE. Message framing and parents' intentions to have their children vaccinated against HPV. Public Health Nurs 2012 Nov;29(6):542-552. [doi: 10.1111/j.1525-1446.2012.01038.x] [Medline: 23078425]
50. Frew PM, Owens LE, Saint-Victor DS, Benedict S, Zhang S, Omer SB. Factors associated with maternal influenza immunization decision-making. Evidence of immunization history and message framing effects. Hum Vaccin Immunother 2014;10(9):2576-2583 [FREE Full text] [doi: 10.4161/hv.32248] [Medline: 25483468]
51. Hendrix KS, Finnell SM, Zimet GD, Sturm LA, Lane KA, Downs SM. Vaccine message framing and parents' intent to immunize their infants for MMR. Pediatrics 2014 Sep;134(3):e675-e683 [FREE Full text] [doi: 10.1542/peds.2013-4077] [Medline: 25136038]
52. Reno JE, O'Leary ST, Pyrzanowski J, Lockhart S, Thomas J, Dempsey AF. Evaluation of the implementation of a multicomponent intervention to improve health care provider communication about human papillomavirus vaccination. Acad Pediatr 2018;18(8):882-888. [doi: 10.1016/j.acap.2018.08.004] [Medline: 30107236]
53. Gagneur A, Lemaître T, Gosselin V, Farrands A, Carrier N, Petit G, et al. A postpartum vaccination promotion intervention using motivational interviewing techniques improves short-term vaccine coverage: PromoVac study. BMC Public Health 2018 Jun 28;18(1):811 [FREE Full text] [doi: 10.1186/s12889-018-5724-y] [Medline: 29954370]
54. Kahan D. Fixing the communications failure. Nature 2010 Jan 20;463(7279):296-297. [doi: 10.1038/463296a]
55. Vinson DE, Scott JE, Lamont LM. The role of personal values in marketing and consumer behavior. J Marketing 2018 Nov 28;41(2):44-50. [doi: 10.1177/002224297704100215]
56. Edwards JR, Cable DM. The value of value congruence. J Appl Psychol 2009 May;94(3):654-677. [doi: 10.1037/a0014891] [Medline: 19450005]
57. McLeod MS. Values, communication behavior, and political participation. Political Commun 2001 Jul;18(3):273-300. [doi: 10.1080/10584600152400347]
58. Sweeney AM, Moyer A. Self-affirmation and responses to health messages: a meta-analysis on intentions and behavior. Health Psychol 2015 Feb;34(2):149-159. [doi: 10.1037/hea0000110] [Medline: 25089345]
59. 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. [Medline: 11392867]
60. Falk EB, O'Donnell MB, Cascio CN, Tinney F, Kang Y, Lieberman MD, et al. Self-affirmation alters the brain's response to health messages and subsequent behavior change. Proc Natl Acad Sci U S A 2015 Feb 17;112(7):1977-1982 [FREE Full text] [doi: 10.1073/pnas.1500247112] [Medline: 25646442]
61. Kelly MP, Heath I, Howick J, Greenhalgh T. The importance of values in evidence-based medicine. BMC Med Ethics 2015 Oct 12;16(1):69 [FREE Full text] [doi: 10.1186/s12910-015-0063-3] [Medline: 26459219]
62. Attwell K, Freeman M. I Immunise: an evaluation of a values-based campaign to change attitudes and beliefs. Vaccine 2015 Nov 17;33(46):6235-6240 [FREE Full text] [doi: 10.1016/j.vaccine.2015.09.092] [Medline: 26458802]
63. Flannigan L, Orenstein WA, Bednarczyk R. Personal values as mediators of receipt of non-influenza childhood vaccines. Open Forum Infect Dis 2017;4(1):S514. [doi: 10.1093/ofid/ofx163.1335]
64. Lane S, MacDonald NE, Marti M, Dumolard L. Vaccine hesitancy around the globe: analysis of three years of WHO/UNICEF Joint Reporting Form data-2015-2017. Vaccine 2018 Jun 18;36(26):3861-3867 [FREE Full text] [doi: 10.1016/j.vaccine.2018.03.063] [Medline: 29605516]
65. Marti M, de Cola M, MacDonald NE, Dumolard L, Duclos P. Assessments of global drivers of vaccine hesitancy in 2014-Looking beyond safety concerns. PLoS One 2017;12(3):e0172310 [FREE Full text] [doi: 10.1371/journal.pone.0172310] [Medline: 28249006]

Abbreviations

KPCO: Kaiser Permanente Colorado

MI: motivational interviewing
PVVS: Parental Vaccine Value Scale
SUS: System Usability Scale
TPB: Theory of Planned Behavior
VAB: Values-Attitudes-Behavior

Edited by G Eysenbach; submitted 07.08.19; peer-reviewed by T Provencher, N Abed Elhadi Shahbari; comments to author 05.09.19; revised version received 30.10.19; accepted 16.12.19; published 05.03.20.

Please cite as:
Dempsey A, Kwan BM, Wagner NM, Pyrzanowski J, Brewer SE, Sevick C, Narwaney K, Resnicow K, Glanz J
A Values-Tailored Web-Based Intervention for New Mothers to Increase Infant Vaccine Uptake: Development and Qualitative Study
J Med Internet Res 2020;22(3):e15800
URL: https://www.jmir.org/2020/3/e15800
doi:10.2196/15800
PMID:32134394

Original Paper

The Interactive Web-Based Program MSmonitor for Self-Management and Multidisciplinary Care in Persons With Multiple Sclerosis: Quasi-Experimental Study of Short-Term Effects on Patient Empowerment

Peter Joseph Jongen[1,2], PhD; Gezien ter Veen[3,4], MNAP; Wim Lemmens[5]; Rogier Donders[5], PhD; Esther van Noort[6], MSc; Esther Zeinstra[4], PhD

[1]Department of Community and Occupational Medicine, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands
[2]MS4 Research Institute, Nijmegen, Netherlands
[3]Zorggroep Noorderboog, Meppel, Netherlands
[4]Isala Hospital, Meppel, Netherlands
[5]Radboud University Medical Center, Department for Health Evidence, Nijmegen, Netherlands
[6]Curavista bv, Geertruidenberg, Netherlands

Corresponding Author:
Peter Joseph Jongen, PhD
MS4 Research Institute
Ubbergseweg 34
Nijmegen, 6522 KJ
Netherlands
Phone: 31 243239146
Email: ms4ri@kpnmail.nl

Abstract

Background: Empowerment helps persons with a chronic disease to self-manage their condition and increase their autonomy and participation. MSmonitor (Curavista bv) is an interactive Web-based program for self-management and multidisciplinary care in multiple sclerosis (MS). It includes, among others, short questionnaires on fatigue (Modified Fatigue Impact Scale-5 [MFIS-5]) and health-related quality of life (HRQoL, Leeds Multiple Sclerosis Quality of Life [LMSQoL]); long questionnaires on disabilities, perception of disabilities (Multiple Sclerosis Impact Profile), and HRQoL (Multiple Sclerosis Quality of Life-54); a Medication and Adherence Inventory and an Activity Diary. The combination MFIS-5, LMSQoL, and Medication and Adherence Inventory constitutes the Quick Scan.

Objective: This study aimed to investigate the short-term effects of MSmonitor on empowerment in patients with MS.

Methods: We conducted a quasi-experimental study in a general hospital. Of the 180 patients with MS, 125 were eligible, 30 used MSmonitor, and 21 participated in the study (mean age 45.4 years, SD 10.2 years). A total of 24 eligible patients who did not use MSmonitor constituted the control group (mean age 49.3 years, SD 11.4 years). At baseline and at 4 months, we assessed self-efficacy (Multiple Sclerosis Self-Efficacy Scale [MSSES]), participation and autonomy (Impact on Participation and Autonomy [IPA] questionnaire), and self-management (Partners In Health [PIH] questionnaire). Differences between time points and groups were tested with paired *t* tests and χ^2 tests.

Results: In the MSmonitor group, follow-up values remained unchanged for MSSES control (*P*=.19), MSSES function (*P*=.62), IPA limitations (*P*=.26), IPA problems (*P*=.40), PIH recognition and management of symptoms (*P*=.52), PIH adherence to treatment (*P*=.80), and PIH coping (*P*=.73), whereas the PIH knowledge score had improved (mean 27.8, SD 1.7 vs mean 28.7, SD 2.0; *P*=.02). The overall utilization rate of the program components was 83% and that of the Quick Scan was 95%. In the control group, all outcomes had remained unchanged.

Conclusions: The results suggest that for first-time users of the MSmonitor program and their health care providers, it may not be justified to expect a short-term improvement in empowerment in terms of self-efficacy, self-management, autonomy, or participation. Furthermore, a lack of effect on empowerment is not because of nonusage of the program components.

(J Med Internet Res 2020;22(3):e14297) doi:10.2196/14297

KEYWORDS

multiple sclerosis; empowerment; self-management; eHealth; internet-based intervention; internet-based communication; personal autonomy; social participation; self-efficacy

Introduction

Background

Multiple Sclerosis

Multiple sclerosis (MS) is a chronic inflammatory and degenerative disease of the central nervous system (CNS). It is the most frequent chronic CNS disease in young adulthood, and the majority of patients experience their first symptoms at the age of 20 to 40 years [1]. Intermittent or continuous disease activity results in a stepwise or slow increase in disabilities over time [1]. The disease course is largely unpredictable, as is the response to disease-modifying drug (DMD) treatment [1,2]. MS is incurable, as the effectiveness of DMD treatment is only partial and limited to the inflammatory component of the disease [1,2].

Patient Empowerment and Web-Based Health Services

Persons with chronic conditions such as MS depend on their own insights to manage daily activities and self-care. To make optimal choices, to evaluate the effects of their choices, and to also contribute to a preventive, personalized, and participatory health care, it is paramount for them to be *empowered* [3]. Empowerment has been defined as a process: *the mechanisms by which people, organizations and communities gain mastery over their lives* [4]. Thus, patient empowerment may be defined as the process by which patients discover and develop the inherent capacity to be responsible for one’s own life [5,6]. Although the concept of patient empowerment is still developing [6,7], empowered patients are generally considered to control their situation, have a critical attitude, and participate and perform tasks in an encouraging environment [3,8,9]. A recent systematic literature review of qualitative studies identified control, coping, knowledge, participation, support, and legitimacy as key aspects of patient empowerment [6]. Notably, interventions that aimed at improving patient empowerment have resulted in higher self-efficacy and self-care competence [10].

Web-based health services use telecommunications and information technology to provide care, education, and monitoring services to patients [11]. In patients with MS, Web-based health services have been shown to result in improved health care because of improved symptom management and treatment adherence [12-15]. E-communication can be defined as communication via Web-based platforms or apps [16-18], and among MS patients, e-communication has high levels of acceptance for exchanging information with health care providers [19]. Information systems with an e-communication function have also been found to be useful in enhancing interdisciplinary communication [20,21]. Consequently, it has been suggested that e-communication should be integrated into electronic health services for patients with MS [19].

MSmonitor

Against this background, we developed MSmonitor, an interactive Web-based program for self-management and multidisciplinary care in persons with MS, that can be used on computers, tablets, and smart phones [22-24]. At the time of the study, MSmonitor included short questionnaires on fatigue (Modified Fatigue Impact Scale-5 [MFIS-5]) [25,26], health-related quality of life (HRQoL, Leeds Multiple Sclerosis Quality of Life [LMSQoL] questionnaire) [27,28], and anxiety and depression (Hospital Anxiety and Depression Scale [HADS]) [29-31]; long questionnaires on disabilities and perception of disabilities (Multiple Sclerosis Impact Profile [MSIP]) [32,33] and HRQoL (Multiple Sclerosis Quality of Life-54 [MSQoL-54] questionnaire) [34]; inventories (Medication and Adherence Inventory, Miction Inventory); and diaries (Activity Diary, Miction Diary) [22-24]. We previously reported that patients who used the combined MFIS-5 and LMSQoL questionnaires at least twice in a period of up to 6 months showed an improved HRQoL and that in these patients, the degree of fatigue improvement correlated with the frequency of Activity Diary usage [24].

Conceivably, MSmonitor usage may lead in various ways to an improvement of empowerment. For example, the Activity Diary and MFIS-5 give insight into factors affecting fatigue and thus facilitate self-management of fatigue and fatigue-related symptoms. The quantified overview of (perceived) disabilities given by the MSIP informs patients about the relative importance of their symptoms and thus facilitates focused self-management. Documentation of missed doses in the Medication and Adherence Inventory may help improve adherence to DMD treatment.

While developing MSmonitor, we did not intend the program to generate short-term effects, as MS is a progressive disease that most patients are afflicted with for decades. However, as we live in an *instant gratification era*, where everything seems to be available immediately via smart phone or the internet [35,36], we became aware that patients might indeed expect early results. Therefore, to obtain knowledge about the short-term effect of MSmonitor on patient empowerment, we conducted a quasi-experimental study.

Objective

The aim of the study was to explore short-term changes in empowerment in persons with MS using MSmonitor.

Methods

MSmonitor

MSmonitor is used by about 1500 patients and their health care providers in 23 hospitals in the Netherlands.

Concept

The concept is based on the autonomy of patients, the multidisciplinary character of MS care, and the collaboration between stakeholders involved in MS care [23,24]. The program was developed gradually by the immaterial and material input of various stakeholders [22-24]. By facilitating self-assessments and self-management, MSmonitor aims to use and increase patients' autonomy. Patients own their personal data generated by the program and decide which health care providers can have access to their data [23,24]. By making self-assessment outcomes available to the multidisciplinary team, MSmonitor helps patients in preventing unnecessary measurements and promotes the use of patient-reported outcomes [23,24].

Content

At the time of the study, the content comprised 9 components in 3 categories: psychometrically validated questionnaires, inventories, and diaries. The characteristics and availability of the various components are presented in Table 1.

Alerts are sent when questionnaires are available, and reminders are sent when scheduled questionnaires are not completed. For all questionnaires, scores are automatically generated and presented in graphs and tables to patients and authorized caregivers (Figure 1), as well as changes over time (Figure 2).

Table 1. Characteristics and availability of MSmonitor components.

Name	Purpose	Structure	Min[a]-max[b]	Validation	Availability
MFIS-5[c] questionnaire	Perceived impact of fatigue on daily activities over past month	5 items (0-4)	0-20 (lower=better)	Fisk et al [25]; NMSS[d] [26]	Monthly
LMSQoL[e] questionnaire	MS[f]-related aspects of QoL[g] over past month	8 items (0-3)	0-24 (higher=better)	Ford et al [27]; Ensari et al [28]	Monthly
MSIP[h] questionnaire	Overview of actual MS-related disabilities (a) and perception of disabilities (b)	36 a-items; 36 b-items; scorings variable	7 domain and 4 symptom scores, 0-variable (lower=better)	Wynia et al [32]; Wynia et al [33]	6 monthly
MSQoL[i]-54 questionnaire	Multidimensional assessment of physical and mental MS-related QoL over 4 weeks	54 items, various scorings	Physical QoL, 0-100; mental QoL, 0-100 (higher=better)	Vickrey et al [34]	Yearly
HADS[j] questionnaire	Anxiety in past week; depression in past week	7 items (0-3); 7 items (0-3)	0-21 (lower=better); 0-21 (lower=better)	Honarmand and Feinstein [30]; Watson et al [31]	On indication
Medication and Adherence Inventory	Medication and DMD[k] adherence in past month	N/A[l]	N/A	N/A	Monthly
Miction Inventory	Actual urological symptoms	N/A	N/A	N/A	On indication
Activity Diary	Activities and rest periods in 24 hours	N/A	N/A	N/A	Daily
Miction Diary	Frequency and quantity of miction and fluid intake in 24 hours	N/A	N/A	N/A	On indication

[a] Min: minimum score.

[b] Max: maximum score.

[c] MFIS-5: Modified Fatigue Impact Scale-5 items.

[d] NMSS: National Multiple Sclerosis Society.

[e] LMSQoL: Leeds Multiple Sclerosis Quality of Life.

[f] MS: multiple sclerosis.

[g] QoL: quality of life.

[h] MSIP: Multiple Sclerosis Impact Profile.

[i] MSQoL-54: Multiple Sclerosis Quality of Life-54 items.

[j] HADS: Hospital Anxiety and Depression Scale.

[k] DMD: disease-modifying drug.

[l] Not applicable.

Figure 1. Screenshot of graphic presentation of Multiple Sclerosis Impact Profile (MSIP) disability scores in the domains muscle and movement, excretion and reproductive functions, mental functions, basic movement activities, activities of daily living, environmental factors, participation in life situations, and the symptoms fatigue, pain, speech, and vision (lower numbers). Upper numbers represent the maximum of the score range. Higher scores indicate a worse condition. Right bars give the actual score, left bars the previous score.

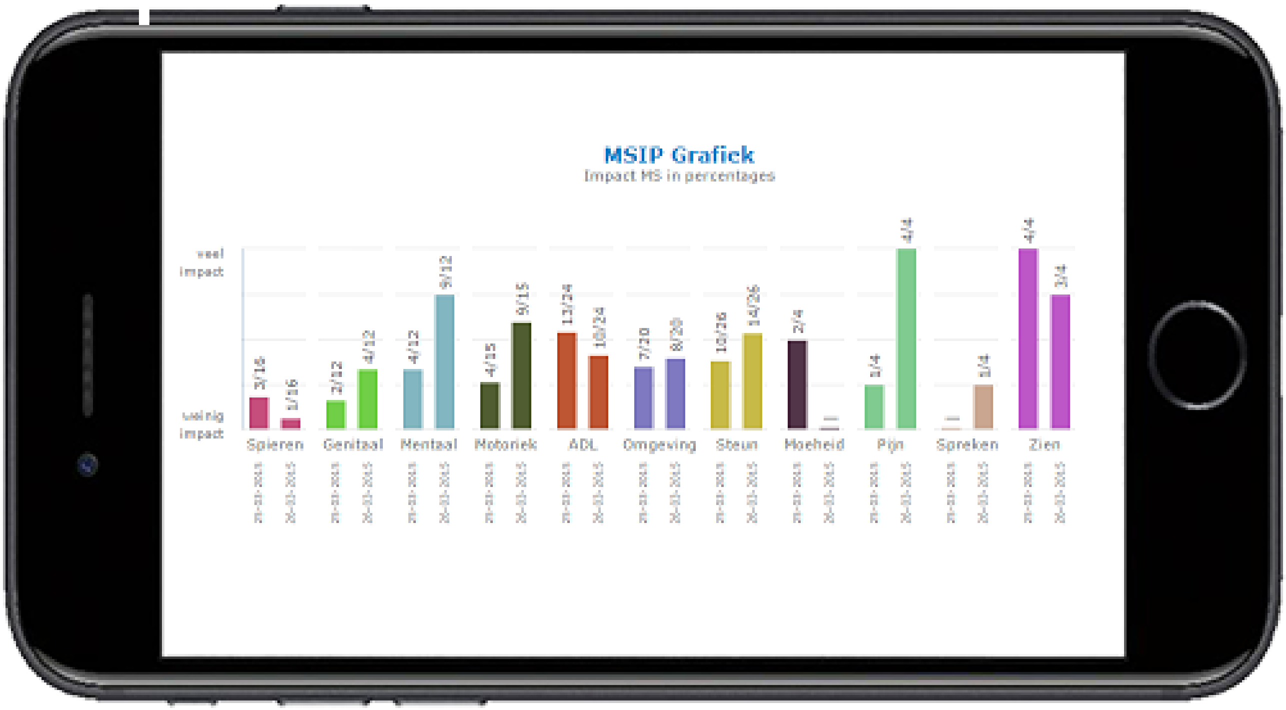
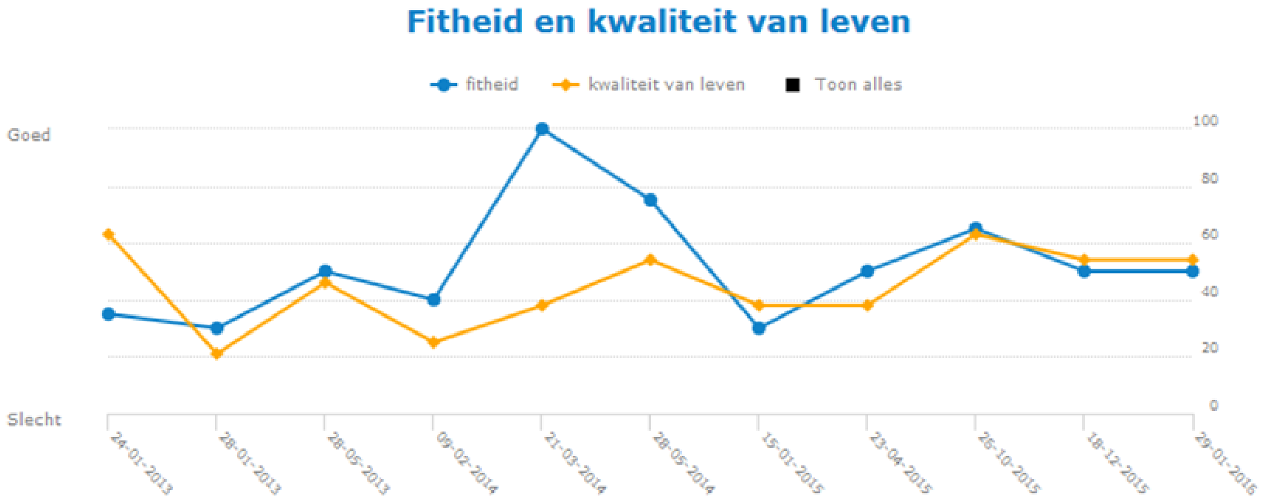


Figure 2. Screenshot of graphic presentation of changes over time in MFIS-5 and LMSQoL scores. The MFIS-5 score (higher is worse) is converted into a fitness (fitheid) score (higher is better) to match the direction of the LMSQoL score. Scores are converted into percentages (0%, minimum score; 100%, maximum score). MFIS-5: Modified Fatigue Impact Scale-5 items; LMSQoL: Leeds Multiple Sclerosis Quality of Life.



Inventories provide overviews and do not generate scores; for example, the Medication and Adherence Inventory gives an update of medication that is taken, the number of missed DMD doses in the last month, and the date and reason of eventual DMD treatment discontinuation. Diaries enable the recording of specific activities or functions and thus give insight into MS-related processes over 24-hour periods [23,24]. The Activity Diary records type and duration of activities and rest periods, whereas the Miction Diary documents the frequencies and quantities of mictions and fluid intakes [23,24]. The combined use of MFIS-5, LMSQoL, and Medication and Adherence Inventory (*Quick Scan*) enables quick self-assessments of fatigue, HRQoL, and adherence to DMD treatment. The HADS, Miction Diary, and Miction Inventory are only available to patients after indication by health care professionals and were therefore not part of the study.

Study Design

This was a prospective, quasi-experimental study. For the MSmonitor group, the baseline assessment was conducted when the participant was registered as a user. In all participants, follow-up assessment was conducted at 4 months. This follow-up period was chosen pragmatically and was dictated by the principal researcher's availability. We considered this period justifiable, as in chronic disorders, a follow-up at 3 to 6 months is generally qualified as *short-term*.

Owing to the study's time frame and the standard 6-month interval between consecutive completions of the MSIP and the MSQoL-54 (Table 1), the usage of these questionnaires was limited to single completions. Notably, to prevent patients from being overburdened, the program makes the MSQoL-54 available 3 months after the MSIP. Hence, 4 components were available for multiple use: the MFIS-5, LMSQoL and Medication and Adherence Inventory (Quick Scan), and the Activity Diary.

Study Setting

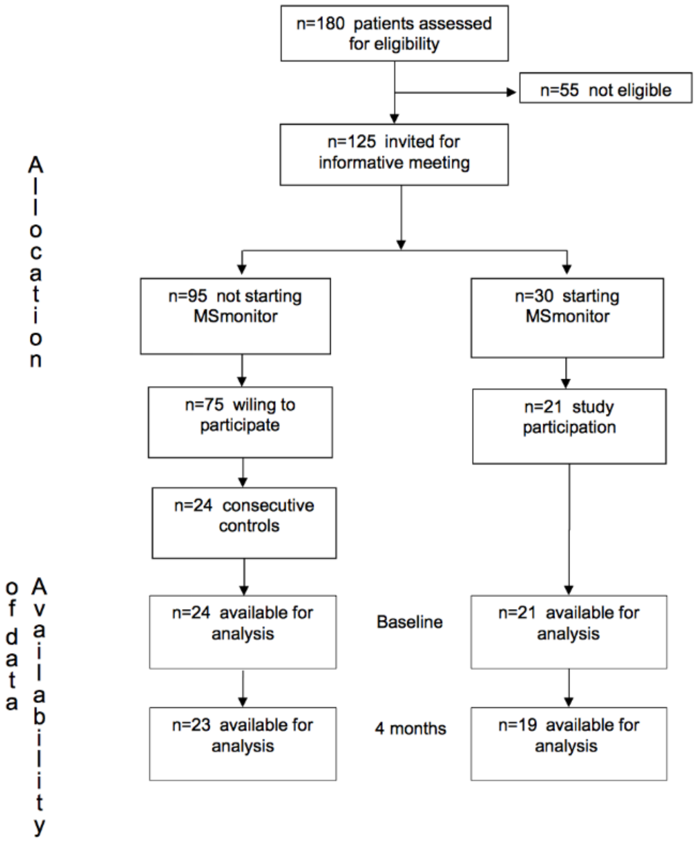
The study was performed in the Neurological Department of the Isala Diaconessenhuis, Meppel, the Netherlands. The Isala Diaconessenhuis is a medium-sized (120 beds) general hospital with 1300 neurological outpatient visits per year, 5000 of which being new referrals.

Recruitment

All patients registered with the diagnosis of MS constituted the study population (n=180; Figure 3).

The exclusion criteria for participation were as follows: diagnosis of clinically isolated syndrome, actual medical doubts about MS diagnosis, serious cognitive impairment, limited knowledge of the Dutch language, and nursing home residents. As a result, 55 patients were not eligible. The remaining 125 eligible patients were invited for a general meeting to be informed about MSmonitor and the study, and patients who had not attended the meeting were informed by phone. A total of 30 patients decided to start with the program and 21 of these were willing to participate in the study. Of the 95 patients who decided not to start with the program, 75 were willing to participate in the study, and out of those, 24 consecutive persons were recruited to form the control group. So, it was actually the patients who decided which study group to join (MSmonitor or control), and this fact explains the quasi-experimental design of the study.

Figure 3. Study Flow Chart.



Ethical Approval and Informed Consent

The study did not qualify for being reviewed according to the Dutch Medical Research Involving Human Subjects Act of 1999. The study was carried out in compliance with the Declaration of Helsinki (Ethical Principles for Medical Research Involving Human Subjects version 2013; 64th World Medical Association General Assembly, Fortaleza, Brazil, October 2013) and the Dutch Medical Research Involving Human Subjects Act of 1999. Patients received no financial incentive or reward to participate. The study was approved by the hospital's coordinator of local evaluation of medical experiments. Patients who agreed to participate signed an informed consent form.

Data Collection

Patient-reported data were obtained by the use of paper-and-pencil questionnaires sent by regular mail, 1 week before the start of study participation and at follow-up. The questionnaires were accompanied by a stamped return envelope addressed to the neurological outpatient department for the attention of the principal researcher. Owing to financial restrictions, it was not feasible to provide e-versions of the questionnaires and to integrate these into the program. Data on MSmonitor utilization were provided by Curavista bv, Geertruidenberg, the Netherlands.

The following aspects of empowerment were assessed using psychometrically validated questionnaires: self-efficacy, participation, autonomy, and self-management (Table 2). We did not use a general or disease-specific empowerment measure, as such measures were not available in Dutch.

Table 2. Questionnaires used to assess aspects of empowerment.

Name	Purpose	Structure	Min[a]-max[b]	Validation
MSSES[c] control	Confidence with managing symptoms and coping with demands of illness	9 items (10-100)	90-900 (higher=more confidence)	Schwartz et al [37] (MS[d])
MSSES function	Confidence with regard to functional abilities	9 items (10-100)	90-900 (higher=more confidence)	Schwartz et al [37] (MS)
IPA[e] limitations	Limitations to participation and autonomy	32 items (0-4)	0-128 (lower=less limitations)	Vazirinejad et al [38] (MS); Karhula et al [39] (MS)
IPA problems	Problems with limitations to participation and autonomy	9 items (0-2)	0-18 (lower=less problems)	Vazirinejad et al [38] (MS); Karhula et al [39] (MS)
PIH[f] coping	Coping	3 items (0-8)	0-24 (lower=better coping)	Petkov et al [40] (CCC)[g]; Lenferink et al [41] (COPD[h])
PIH symptoms	Recognition and management of symptoms	3 items (0-8)	0-24 (lower=better management of symptoms)	Petkov et al [40] (CCC); Lenferink et al [41] (COPD)
PIH adherence	Adherence to treatment	2 items (0-8)	0-16 (lower=better adherence to treatment)	Petkov et al [40] (CCC); Lenferink et al [41] (COPD)
PIH knowledge	Knowledge	4 items (0-8)	0-32 (lower=better knowledge)	Petkov et al [40] (CCC); Lenferink et al [41] (COPD)

[a] Min: minimum score.

[b] Max: maximum score.

[c] MSSES: Multiple Sclerosis Self-Efficacy Scale.

[d] MS: multiple sclerosis.

[e] IPA: Impact on Participation and Autonomy.

[f] PIH: Partners In Health.

[g] CCC: comorbid chronic conditions not including MS.

[h] COPD: chronic obstructive pulmonary disease.

Self-efficacy was assessed by the Multiple Sclerosis Self-Efficacy Scale (MSSES) [37], participation and autonomy were assessed by the Impact on Participation and Autonomy (IPA) questionnaire [38,39,42-44], and self-management behaviors and knowledge were assessed by the revised 12-item Partners In Health (PIH) scale [40,41].

At baseline, the level of education, degree of computer use, and degree of computer skills were assessed via multiple-choice questions (Table 3), as these factors may conceivably influence the speed with which persons become familiar with Web-based programs.

Table 3. Demographics, disease characteristics, level of education, degree of computer use, and degree of computer skills in the MSmonitor group and the control group.

Patient characteristics	MSmonitor (n=21)	Control (n=24)	*P* value
Female, n (%)	17 (81)	17 (71)	.43
Age (years), mean (SD)	45.4 (10.2)	49.3 (11.4)	.23
Disease duration (years), mean (SD)	8.7 (6.4)	12.2 (9.7)	.17
Disease course, n (%)			**.40**
Relapsing remitting	16 (76)	13 (54)	
Secondary progressive	4 (19)	7 (29)	
Primary progressive	1 (4)	3 (12)	
Benign	0 (0)	1 (4)	
Education, n (%)			**.39**
Lower	6 (28)	10 (41)	
Middle	12 (57)	7 (29)	
Higher	3 (14)	7 (29)	
Computer use, n (%)			**.83**
Several times per day	5 (23)	5 (20)	
Daily	9 (42)	10 (41)	
Several times per week	3 (14)	5 (20)	
Once per week	0 (0)	1 (4)	
Rarely or never	4 (19)	3 (12)	
Computer skills, n (%)			**.71**
Rapidly familiar with new programs	7 (33)	7 (29)	
Familiar after some log-ins	8 (38)	11 (46)	
Difficulties with getting familiar	5 (24)	6 (25)	
Impossible to get familiar	1 (5)	0 (0)	

Data Analysis

For all outcomes, the absolute values at baseline and follow-up are presented as mean, standard deviation (SD), minimum, and maximum. As the study's purpose was to investigate whether short-term changes could be observed after the start of MSmonitor usage, we compared in each group the follow-up with the baseline values by using multiple paired *t* tests. The baseline characteristics in the MSmonitor and control groups were tested for differences using *t* tests and χ^2 tests. The analyses were performed at the Department for Health Evidence, Radboud University Medical Centre, Nijmegen, the Netherlands. For all tests, a *P* value of <.05 was considered significant.

Results

Patient Characteristics

A total of 45 patients were included, 21 in the MSmonitor group and 24 in the control group. Demographics, characteristics of disease, level of education, degree of computer use, and degree of computed skills in both groups are presented in Table 3. There were no statistically significant differences with respect to gender, age, duration of disease, course of disease, level of education, degree of computer use, or degree of computer skills. A total of 2 patients in the MSmonitor group and 1 in the control group failed to complete the follow-up questionnaires. Hence, the data analysis set comprised 19 MSmonitor and 23 control patients.

MSmonitor Utilization

At 1 month, all 19 patients had used the Quick Scan. As not all patients started usage immediately after baseline assessment, the second and third Quick Scans were available to 11 and 7 patients, respectively, and these were used by 9 and 7 of them, respectively. Accordingly, the Quick Scan utilization rate was 95% (35/37). The MSIP was used by 14 out of 19 patients, and the MSQoL-54 by 4 out of 5 patients to whom it was available. So, in total, the Quick Scan, MSIP, and MSQoL-54 were available 61 times and were used 53 times, resulting in a combined completion rate of 87% for these 3 components. The Activity Diary was used by 13 (68%) patients; and in these, the mean (minimum, maximum) number of days of usage was 12 (1, 44). In all, the completion frequencies of Quick Scan, MSIP, and MSQoL-54, and the percentage of patients using the Activities Diary, resulted in an overall utilization rate of 83% of the MSmonitor components.

The Medication and Adherence Inventory part of the first Quick Scan showed that 10 of the 19 patients used a DMD and that 2 patients had missed 1 and 2 doses, respectively, in the preceding month. Moreover, all patients completed the first Quick Scan, whereas patients who also used the Activity Diary completed the second Quick Scan more frequently than those who did not use the Activity Diary (8/11 vs 1/8).

Empowerment Outcomes

The mean, SD, minimum, and maximum values of the various outcome scores at baseline and at follow-up are presented in Table 4.

In the MSmonitor group, scores remained unchanged for MSSES control, MSSES function, IPA limitations, IPA problems, PIH coping, PIH recognition and management of symptoms, and PIH adherence to treatment. The mean PIH knowledge score decreased, suggesting an improvement. In the control group, all scores were unchanged.

Table 4. Mean (SD) and minimum-maximum values of Multiple Sclerosis Self-Efficacy Scale, Impact on Participation and Autonomy, and Partners In Health scores at baseline and at follow-up in the MSmonitor and control groups.

Empowerment aspects	MSmonitor (n=19)			Control (n=23)		
	Baseline, mean (SD); minimum-maximum	4 months, mean (SD); minimum-maximum	*P* value	Baseline, mean (SD); minimum-maximum	4 months, mean (SD); minimum-maximum	*P* value
MSSES[a] control	59.8 (19.0), 20.0-87.8	63.4 (17.7), 32.2-94.4	.19	55.3 (19.5), 25.6-96.7	53.7 (20.8), 22.2-91.1	.40
MSSES function	72.4 (22.4), 28.9-100	73.5 (21.8), 30.0-100	.62	67.8 (24.3), 21.3-100	66.5 (23.5), 26.7-100	.17
IPA[b] limitations	2.8 (0.4), 1.7-3.6	2.7 (0.6), 1.6-3.7	.26	2.6 (0.7), 1.1-4.0	2.6 (0.6), 1.5-3.95	.28
IPA problems	0.71 (0.36), 0.00-1.71	0.79 (0.48), 0.00-1.86	.40	0.97 (0.35), 0.22-1.71	0.85 (0.35), 0.00-1.43	.25
PIH[c] coping	19.2 (3.3), 13.0-24.0	18.9 (4.8), 4.0-24.0	.73	18.6 (3.9), 8.0-24.0	18.9 (3.2), 12.0-24	.76
PIH symptoms	21.9 (1.2), 20.0-24.0	21.6 (2.3), 16.0-24.0	.52	20.8 (2.2), 16.0-24.0	20.9 (2.6), 13.0-24	.87
PIH adherence	15.1 (1.4), 11.0-16.0	14.9 (2.2), 8.0-16.0	.80	14.8 (1.4), 12.0-16.0	14.6 (2.3), 7.0-16	.34
PIH knowledge	28.7 (2.0), 25.0-32.0	27.8 (1.7), 24.0-30.0	.02	27.7 (3.4), 21.0-31.0	28.7 (2.2), 24.0-32	.24

[a] MSSES: Multiple Sclerosis Self-Efficacy Scale.

[b] IPA: Impact on Participation and Autonomy.

[c] PIH: Partners In Health.

Discussion

Principal Findings

We conducted a quasi-experimental study in first-time users of MSmonitor to explore the program's early effects on empowerment and found that at 4 months' follow-up, self-efficacy, participation, autonomy, and self-management did not change, whereas knowledge had increased. The increase in the PIH knowledge score was about 0.9 SD baseline, which suggests that the change was clinically meaningful and can therefore be qualified as an improvement. It is, however, not sure to what degree the better knowledge results from the utilization of MSmonitor because at the time of the study, the program had not included an information function with links to websites of patient organizations and health care organizations.

In all, our findings suggest that early improvement of patient empowerment is unlikely to occur after starting MSmonitor. This result is clinically relevant as it may be communicated to first-time users to prevent them from having unrealistic expectations about the program's effects. Similarly, health care professionals should not expect their patients to have a better control of their situation, increased participation in care processes, or improved self-management in the short term [3,8,9]. A lack of utilization is unlikely to explain the unchanged empowerment outcomes, as the program's utilization was high; the overall rate being 83%.

An additional finding was the association between the completion of the Quick Scan and the use of the Activity Diary. Although all patients completed the first Quick Scan, those who also used the Activity Diary evidently completed the second Quick Scan more frequently. This suggests a substantial relationship between repeated self-assessments of fatigue and HRQoL on the one hand and the documentation of activities and resting periods on the other hand and is in agreement with the hypothesis on the role of the program in self-management of fatigue. It may therefore be promising for future research on the effects of MSmonitor to focus on medium- to long-term changes in fatigue and on how patients self-manage their fatigue.

Limitations

The study has several limitations. First, the sample size, and therefore the chance of achieving statistically significant results, was rather low. Nevertheless, it may have been large enough to detect clinically relevant changes, as is suggested by the improved PIH knowledge score. Second, the study group was heterogeneous. Self-management perspectives and goals may differ between relapsing remitting, secondary progressive, and primary progressive patients. Moreover, the failure to detect a change in the PIH adherence to treatment score may relate to the fact that 9 of the 19 patients were not treated with a DMD, whereas only 2 of the DMD-treated patients reported (a low number of) missed doses [45]. Actually, the low incidence of DMD treatment in our patients may be explained by the fact that 1 out of 4 had progressive MS, and that in general 1 out of 4 patients with relapsing-remitting MS is not treated with a DMD [45]. Third, we included 5 patients in the MSmonitor group with reportedly low computer skills and 1 patient with reportedly no such skills, without offering them further training or education. This may have prevented these patients from optimally using the program. Moreover, owing to financial restrictions, the technology was introduced to patients and health care professionals in a single introductory session, without further staff training or education of patients; the absence of an optimal embedding in the daily life of patients and practices of health care providers may have negatively influenced the occurrence of short-term effects. Fourth, patient involvement in the developing process of MSmonitor consisted of receiving patient feedback on a continuous basis via the program's helpdesk and by means of meetings in the hospitals where the program was implemented. However, there were no cocreation sessions or focus group meetings, and this may be considered a limitation. Fifth, the study was not randomized. The fact that it was the participants who decided which group to join may have biased their reporting. As, however, both groups failed to show changes at follow-up (except for knowledge in the MSmonitor group), this limitation seems of minor importance. Finally, some aspects of empowerment were not covered by the questionnaires, such as support and patient-provider interaction [46,47].

Comparison With Prior Work

Few studies have investigated the effect of Web-based self-management and care programs on empowerment in patients with MS, and the results are ambiguous. In a 6-month uncontrolled study (n=31) on the perceived benefits of Web-based MS-related patient-reported outcome collection, nearly 52% of the participants reported improved understanding of their disease [48]. Similarly, in a survey among MSmonitor users (n=55), 46% reported that their insight into symptoms and disabilities had increased since the use of the program [24]. On the other hand, in a 12-month randomized controlled trial (n=206), the expansion of an electronic MS health record with a self-monitoring and self-management system did not result in improved self-efficacy or symptoms [13]. Combined with the findings of this 4-month quasi-experimental study, the available data suggest that patient empowerment is not necessarily affected by the use of MS-related Web-based self-management and care programs.

Notably, self-management is a major issue for patients with MS. A recent study showed that positive expectations about the helpfulness for self-management is an important predictor for the acceptance of MS-related apps [49]. Actually, 26% of the available MS-related apps have been designed for self-management purposes [50]. On the other hand, a recent review showed that the available MS apps fail to sufficiently meet the needs and demands of patients [51]. Although education and personal data management were the frequently included features, remote monitoring and fatigue management were often not present [51], despite the fact that fatigue management functions in mobile health solutions are important to patients with MS [52]. So, it seems that because of its fatigue management, monitoring and—recently added—information functions, MSmonitor compares favorably with the majority of MS apps [52].

The program's overall utilization rate was 83%, whereas in a previous survey among all MSmonitor users, the most frequently used components Medication and Adherence Inventory, Activity Diary, and MSIP were used by 55%, 47%, and 40% of the respondents, respectively [23]. It is known that long-term Web-based self-monitoring in patients with MS is hampered by a declining adherence, both in regular care and in direct-to-patient research settings [48,53]. Thus, in a 6-month study (n=31) with monthly completion of 5 questionnaires, it was found that all questionnaires were completed less frequently in the second 3 months [48]. Interestingly, a recent study suggests that continuous communication with patients may promote the continued use of digital data collection tools [54]. Moreover, adaptation of digital self-monitoring tools to patients' personal situation, giving guidance to increase the value of their data, and integration of digital self-monitoring into treatment plans might also increase the adherence of patients with MS to Web-based programs and apps [55].

Finally, a recent review identified over 100 MS-related apps, but in none was evidence found in the literature on evaluation of the effects [50]. This may be worrisome, as the widespread implementation and utilization of MS-related Web-based programs, including mobile apps, will most likely depend on whether convincing evidence can be obtained regarding their effectiveness and cost-effectiveness [50]. Owing to limited resources, it is unlikely that all available tools will be evaluated in randomized controlled trials [56]. Moreover, the external validity of trial results is not self-evident, given that patient preferences may differ between regions or countries and preferences may change over time and the ongoing development of the tools. Therefore, prospective observational studies in real-world settings and retrospective studies using large databases are increasingly being considered as alternatives for obtaining actionable data [56,57].

Conclusions

In a quasi-experimental study, we investigated short-term changes in empowerment in patients with MS who started using the Web-based program MSmonitor. At 4 months, self-efficacy, participation, autonomy, coping, recognition and management of symptoms, and adherence to treatment did not change. The utilization rate of the program's components was high. Our findings suggest that it may not be justified for first-time users of MSmonitor and their health care providers to expect a short-term improvement in empowerment. Immediate effects might be realized by better informing patients about the option to give the multidisciplinary team access to their data, as this may influence treatment decisions and care at short notice. It may well be that the program becomes effective in the medium to long term because of patients becoming increasingly familiar with the various components and their possibilities. A better adjustment of the program to the expectations and wishes of patients—in terms of content, personalization, and integration into treatment plans—is expected to also enhance empowerment [55].

Acknowledgments

The authors thank Yvonne Kroon, Curavista bv, Geertuidenberg, the Netherlands, for providing the data on MSmonitor utilization. At the time of the study, the MSmonitor program was funded by Teva Netherlands, Curavista bv, and MS4 Research Institute, the Netherlands.

Conflicts of Interest

EN is co-owner of Curavista bv, Geertruidenberg, the Netherlands. PJ has received expense compensation from Curavista bv for serving as chairman of the board of the MSmonitor Foundation, and honoraria from Bayer Netherlands for consultancy activities.

References

1. Oh J, Vidal-Jordana A, Montalban X. Multiple sclerosis: clinical aspects. Curr Opin Neurol 2018 Dec;31(6):752-759. [doi: 10.1097/WCO.0000000000000622] [Medline: 30300239]
2. Dobson R, Giovannoni G. Multiple sclerosis - a review. Eur J Neurol 2019 Jan;26(1):27-40. [doi: 10.1111/ene.13819] [Medline: 30300457]
3. Lejbkowicz I, Caspi O, Miller A. Participatory medicine and patient empowerment towards personalized healthcare in multiple sclerosis. Expert Rev Neurother 2012 Mar;12(3):343-352. [doi: 10.1586/ern.11.161] [Medline: 22364333]
4. Rappaport J. Community psychology. In: Corsini R, editor. Encyclopedia of Psychology. New York: Wiley; 1984:252-254.
5. 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]
6. Agner J, Braun KL. Patient empowerment: a critique of individualism and systematic review of patient perspectives. Patient Educ Couns 2018 Dec;101(12):2054-2064. [doi: 10.1016/j.pec.2018.07.026] [Medline: 30143254]
7. 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]
8. Castro EM, van Regenmortel T, Vanhaecht K, Sermeus W, van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: a concept analysis based on a literature review. Patient Educ Couns 2016 Dec;99(12):1923-1939. [doi: 10.1016/j.pec.2016.07.026] [Medline: 27450481]
9. McAllister M, Dunn G, Payne K, Davies L, Todd C. Patient empowerment: the need to consider it as a measurable patient-reported outcome for chronic conditions. BMC Health Serv Res 2012 Jun 13;12:157 [FREE Full text] [doi: 10.1186/1472-6963-12-157] [Medline: 22694747]
10. Huang T, Sung C, Wang W, Wang B. The effects of the empowerment education program in older adults with total hip replacement surgery. J Adv Nurs 2017 Aug;73(8):1848-1861. [doi: 10.1111/jan.13267] [Medline: 28122161]
11. Chang J, Savage SJ, Waldman DM. Estimating willingness to pay for online health services with discrete-choice experiments. Appl Health Econ Health Policy 2017 Aug;15(4):491-500. [doi: 10.1007/s40258-017-0316-z] [Medline: 28290107]
12. Lejbkowicz I, Paperna T, Stein N, Dishon S, Miller A. Internet usage by patients with multiple sclerosis: implications to participatory medicine and personalized healthcare. Mult Scler Int 2010;2010:640749 [FREE Full text] [doi: 10.1155/2010/640749] [Medline: 22096625]
13. Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee J, et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemed J E Health 2011;17(1):5-13 [FREE Full text] [doi: 10.1089/tmj.2010.0133] [Medline: 21214498]
14. Lavorgna L, Russo A, de Stefano M, Lanzillo R, Esposito S, Moshtari F, et al. Health-related coping and social interaction in people with multiple sclerosis supported by a social network: pilot study with a new methodological approach. Interact J Med Res 2017 Jul 14;6(2):e10 [FREE Full text] [doi: 10.2196/ijmr.7402] [Medline: 28710056]
15. Lavorgna L, Brigo F, Moccia M, Leocani L, Lanzillo R, Clerico M, et al. e-Health and multiple sclerosis: an update. Mult Scler 2018 Nov;24(13):1657-1664. [doi: 10.1177/1352458518799629] [Medline: 30231004]
16. de Jong CC, Ros WJ, van Leeuwen M, Schrijvers G. How professionals share an e-care plan for the elderly in primary care: evaluating the use of an e-communication tool by different combinations of professionals. J Med Internet Res 2016 Nov 24;18(11):e304 [FREE Full text] [doi: 10.2196/jmir.6332] [Medline: 27884811]
17. Rai M, Vigod SN, Hensel JM. Barriers to office-based mental health care and interest in e-communication with providers: a survey study. JMIR Ment Health 2016 Aug 1;3(3):e35 [FREE Full text] [doi: 10.2196/mental.6068] [Medline: 27480108]
18. Yeh P, Hung S, Lee M, Chiou W, Lai C, Tsai W, et al. Implementing web-based ping-pong-type e-communication to enhance staff satisfaction, multidisciplinary cooperation, and clinical effectiveness: A SQUIRE-compliant quality-improving study. Medicine (Baltimore) 2016 Nov;95(44):e5236 [FREE Full text] [doi: 10.1097/MD.0000000000005236] [Medline: 27858876]

19. Haase R, Schultheiss T, Kempcke R, Thomas K, Ziemssen T. Use and acceptance of electronic communication by patients with multiple sclerosis: a multicenter questionnaire study. J Med Internet Res 2012 Oct 15;14(5):e135 [FREE Full text] [doi: 10.2196/jmir.2133] [Medline: 23069209]
20. Driscoll M, Gurka D. Using the electronic medical record to enhance physician-nurse communication regarding patients' discharge status. Nurs Adm Q 2015;39(4):E31-E37. [doi: 10.1097/NAQ.0000000000000127] [Medline: 26340250]
21. O'Malley AS, Reschovsky JD, Saiontz-Martinez C. Interspecialty communication supported by health information technology associated with lower hospitalization rates for ambulatory care-sensitive conditions. J Am Board Fam Med 2015;28(3):404-417 [FREE Full text] [doi: 10.3122/jabfm.2015.03.130325] [Medline: 25957373]
22. Jongen PJ, Sanders E, Zwanikken C, Koeman J, Visser LH, Koopmans P, OPTIVIT study group. Adherence to monthly online self-assessments for short-term monitoring: a 1-year study in relapsing-remitting multiple sclerosis patients after start of disease modifying treatment. Patient Prefer Adherence 2013;7:293-300 [FREE Full text] [doi: 10.2147/PPA.S40173] [Medline: 23589682]
23. Jongen PJ, Sinnige LG, van Geel BM, Verheul F, Verhagen WI, van der Kruijk RA, et al. The interactive web-based program MSmonitor for self-management and multidisciplinary care in multiple sclerosis: concept, content, and pilot results. Patient Prefer Adherence 2015;9:1741-1750 [FREE Full text] [doi: 10.2147/PPA.S93783] [Medline: 26715841]
24. Jongen PJ, Sinnige LG, van Geel BM, Verheul F, Verhagen WI, van der Kruijk RA, et al. The interactive web-based program MSmonitor for self-management and multidisciplinary care in multiple sclerosis: utilization and valuation by patients. Patient Prefer Adherence 2016;10:243-250 [FREE Full text] [doi: 10.2147/PPA.S93786] [Medline: 27042018]
25. Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. Clin Infect Dis 1994 Jan;18(Suppl 1):S79-S83. [doi: 10.1093/clinids/18.supplement_1.s79] [Medline: 8148458]
26. National MS Society. Multiple Sclerosis Quality of Life Inventory: A User's Manual URL: http://walkcoc.nationalmssociety.org/docs/HOM/MSQLI_Manual_and_Forms.pdf [accessed 2020-01-07]
27. Ford HL, Gerry E, Tennant A, Whalley D, Haigh R, Johnson MH. Developing a disease-specific quality of life measure for people with multiple sclerosis. Clin Rehabil 2001 Jun;15(3):247-258. [doi: 10.1191/026921501673658108] [Medline: 11386394]
28. Ensari I, Motl RW, McAuley E. Structural and construct validity of the Leeds Multiple Sclerosis Quality of Life scale. Qual Life Res 2016 Jun;25(6):1605-1611. [doi: 10.1007/s11136-015-1202-5] [Medline: 26660145]
29. 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]
30. Honarmand K, Feinstein A. Validation of the Hospital Anxiety and Depression Scale for use with multiple sclerosis patients. Mult Scler 2009 Dec;15(12):1518-1524. [doi: 10.1177/1352458509347150] [Medline: 19965520]
31. Watson TM, Ford E, Worthington E, Lincoln NB. Validation of mood measures for people with multiple sclerosis. Int J MS Care 2014;16(2):105-109 [FREE Full text] [doi: 10.7224/1537-2073.2013-013] [Medline: 25061435]
32. Wynia K, Middel B, van Dijk JP, de Ruiter H, de Keyser J, Reijneveld SA. The Multiple Sclerosis impact Profile (MSIP). Development and testing psychometric properties of an ICF-based health measure. Disabil Rehabil 2008;30(4):261-274. [doi: 10.1080/09638280701256868] [Medline: 17852244]
33. Wynia K, Middel B, de Ruiter H, van Dijk JP, de Keyser JH, Reijneveld SA. Stability and relative validity of the Multiple Sclerosis Impact Profile (MSIP). Disabil Rehabil 2008;30(14):1027-1038. [doi: 10.1080/09638280701476193] [Medline: 18953748]
34. Vickrey BG, Hays RD, Harooni R, Myers LW, Ellison GW. A health-related quality of life measure for multiple sclerosis. Qual Life Res 1995 Jun;4(3):187-206. [doi: 10.1007/bf02260859] [Medline: 7613530]
35. University of Minnesota Twin Cities. 2018 Jun 25. Research Brief: Today's kids able to delay gratification longer than those in 1960s URL: https://twin-cities.umn.edu/news-events/research-brief-todays-kids-able-delay-gratification-longer-those-1960s [accessed 2020-01-07]
36. Carlson SM, Shoda Y, Ayduk O, Aber L, Schaefer C, Sethi A, et al. Cohort effects in children's delay of gratification. Dev Psychol 2018 Aug;54(8):1395-1407. [doi: 10.1037/dev0000533] [Medline: 29939038]
37. Schwartz CE, Coulthard-Morris L, Zeng Q, Retzlaff P. Measuring self-efficacy in people with multiple sclerosis: a validation study. Arch Phys Med Rehabil 1996 Apr;77(4):394-398. [doi: 10.1016/s0003-9993(96)90091-x] [Medline: 8607766]
38. Vazirinejad R, Lilley JM, Ward CD. The 'Impact on Participation and Autonomy': acceptability of the English version in a multiple sclerosis outpatient setting. Mult Scler 2003 Dec;9(6):612-615. [doi: 10.1191/1352458503ms936oa] [Medline: 14664475]
39. Karhula ME, Salminen A, Hämäläinen PI, Ruutiainen J, Era P, Tolvanen A. Psychometric evaluation of the Finnish version of the impact on participation and autonomy questionnaire in persons with multiple sclerosis. Scand J Occup Ther 2017 Nov;24(6):410-420. [doi: 10.1080/11038128.2016.1272630] [Medline: 28049364]
40. Petkov J, Harvey P, Battersby M. The internal consistency and construct validity of the partners in health scale: validation of a patient rated chronic condition self-management measure. Qual Life Res 2010 Sep;19(7):1079-1085. [doi: 10.1007/s11136-010-9661-1] [Medline: 20437206]

41. Lenferink A, Effing T, Harvey P, Battersby M, Frith P, van Beurden W, et al. Construct validity of the Dutch version of the 12-item Partners in Health Scale: measuring patient self-management behaviour and knowledge in patients with chronic obstructive pulmonary disease. PLoS One 2016;11(8):e0161595 [FREE Full text] [doi: 10.1371/journal.pone.0161595] [Medline: 27564410]
42. Cardol M, de Haan RJ, de Jong BA, van den Bos GA, de Groot IJ. Psychometric properties of the Impact on Participation and Autonomy Questionnaire. Arch Phys Med Rehabil 2001 Feb;82(2):210-216. [doi: 10.1053/apmr.2001.18218] [Medline: 11239312]
43. Cardol M, Beelen A, van den Bos GA, de Jong BA, de Groot IJ, de Haan RJ. Responsiveness of the Impact on Participation and Autonomy questionnaire. Arch Phys Med Rehabil 2002 Nov;83(11):1524-1529. [doi: 10.1053/apmr.2002.35099] [Medline: 12422319]
44. Franchignoni F, Ferriero G, Giordano A, Guglielmi V, Picco D. Rasch psychometric validation of the Impact on Participation and Autonomy questionnaire in people with Parkinson's disease. Eura Medicophys 2007 Dec;43(4):451-461 [FREE Full text] [Medline: 18084167]
45. Minden SL, Kinkel RP, Machado HT, Levin JS, Rosenthal MB, Iezzoni LI. Use and cost of disease-modifying therapies by Sonya Slifka Study participants: has anything really changed since 2000 and 2009? Mult Scler J Exp Transl Clin 2019;5(1):2055217318820888 [FREE Full text] [doi: 10.1177/2055217318820888] [Medline: 30815276]
46. Chiauzzi E, DasMahapatra P, Cochin E, Bunce M, Khoury R, Dave P. Factors in patient empowerment: a survey of an online patient research network. Patient 2016 Dec;9(6):511-523 [FREE Full text] [doi: 10.1007/s40271-016-0171-2] [Medline: 27155887]
47. Kantor D, Bright JR, Burtchell J. Perspectives from the patient and the healthcare professional in multiple sclerosis: social media and participatory medicine. Neurol Ther 2018 Jun;7(1):37-49 [FREE Full text] [doi: 10.1007/s40120-017-0088-2] [Medline: 29222700]
48. Engelhard MM, Patek SD, Sheridan K, Lach JC, Goldman MD. Remotely engaged: lessons from remote monitoring in multiple sclerosis. Int J Med Inform 2017 Apr;100:26-31 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.006] [Medline: 28241935]
49. Apolinário-Hagen J, Menzel M, Hennemann S, Salewski C. Acceptance of mobile health apps for disease management among people with multiple sclerosis: web-based survey study. JMIR Form Res 2018 Dec 12;2(2):e11977 [FREE Full text] [doi: 10.2196/11977] [Medline: 30684408]
50. Salimzadeh Z, Damanabi S, Kalankesh LR, Ferdousi R. Mobile applications for multiple sclerosis: a focus on self-management. Acta Inform Med 2019 Mar;27(1):12-18 [FREE Full text] [doi: 10.5455/aim.2019.27.12-18] [Medline: 31213737]
51. Giunti G, Fernández EG, Zubiete ED, Romero OR. Supply and demand in mHealth apps for persons with multiple sclerosis: systematic search in app stores and scoping literature review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512 [FREE Full text] [doi: 10.2196/10512] [Medline: 29792295]
52. Giunti G, Kool J, Romero OR, Zubiete ED. Exploring the specific needs of persons with multiple sclerosis for mHealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 9;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
53. Jongen PJ, Kremer IE, Hristodorova E, Evers SM, Kool A, van Noort EM, et al. Adherence to web-based self-assessments in long-term direct-to-patient research: two-year study of multiple sclerosis patients. J Med Internet Res 2017 Jul 21;19(7):e249 [FREE Full text] [doi: 10.2196/jmir.6729] [Medline: 28733272]
54. Karnoe A, Kayser L, Skovgaard L. Identification of factors that motivate people with multiple sclerosis to participate in digital data collection in research: sequential mixed methods study. JMIR Hum Factors 2019 Oct 9;6(4):e13295 [FREE Full text] [doi: 10.2196/13295] [Medline: 31599738]
55. Wendrich K, van Oirschot P, Martens MB, Heerings M, Jongen PJ, Krabbenborg L. Toward digital self-monitoring of multiple sclerosis: investigating first experiences, needs, and wishes of people with ms. Int J MS Care 2019;21(6):282-291 [FREE Full text] [doi: 10.7224/1537-2073.2018-083] [Medline: 31889935]
56. Marziniak M, Brichetto G, Feys P, Meyding-Lamadé U, Vernon K, Meuth SG. The use of digital and remote communication technologies as a tool for multiple sclerosis management: narrative review. JMIR Rehabil Assist Technol 2018 Apr 24;5(1):e5 [FREE Full text] [doi: 10.2196/rehab.7805] [Medline: 29691208]
57. Jongen PJ. Observational designs in clinical multiple sclerosis research: particulars, practices and potentialities. Mult Scler Relat Disord 2019 Oct;35:142-149. [doi: 10.1016/j.msard.2019.07.006] [Medline: 31394404]

Abbreviations

DMD: disease-modifying drug
HADS: Hospital Anxiety and Depression Scale
HRQoL: health-related quality of life
IPA: Impact on Participation and Autonomy
LMSQoL: Leeds Multiple Sclerosis Quality of Life

MFIS-5: Modified Fatigue Impact Scale-5 items
MS: multiple sclerosis
MSIP: Multiple Sclerosis Impact Profile
MSQoL-54: Multiple Sclerosis Quality of Life-54
MSSES: Multiple Sclerosis Self-Efficacy Scale
PIH: Partners In Health

Edited by G Eysenbach; submitted 06.04.19; peer-reviewed by A Maier, L Lavorgna, T Risling, G Giunti, A Nguyen, J Wilroy; comments to author 22.05.19; revised version received 12.08.19; accepted 19.12.19; published 09.03.20.

Please cite as:
Jongen PJ, ter Veen G, Lemmens W, Donders R, van Noort E, Zeinstra E
The Interactive Web-Based Program MSmonitor for Self-Management and Multidisciplinary Care in Persons With Multiple Sclerosis: Quasi-Experimental Study of Short-Term Effects on Patient Empowerment
J Med Internet Res 2020;22(3):e14297
URL: https://www.jmir.org/2020/3/e14297
doi:10.2196/14297
PMID:32149713

Original Paper

A Mobile Just-in-Time Adaptive Intervention for Smoking Cessation: Pilot Randomized Controlled Trial

Emily T Hébert[1], DrPH; Chaelin K Ra[1], PhD; Adam C Alexander[1], PhD; Angela Helt[1], MA; Rachel Moisiuc[1], BS; Darla E Kendzor[1], PhD; Damon J Vidrine[2], DrPH; Rachel K Funk-Lawler[3], PhD; Michael S Businelle[1], PhD

[1]Oklahoma Tobacco Research Center, Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

[2]Moffitt Cancer Center, Tampa, FL, United States

[3]Department of Psychiatry and Behavioral Sciences, Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

Corresponding Author:
Emily T Hébert, DrPH
Oklahoma Tobacco Research Center
Stephenson Cancer Center
University of Oklahoma Health Sciences Center
655 Research Parkway, Suite 400
Oklahoma City, OK, 73104
United States
Phone: 1 4052718001 ext 50558
Email: Emily-Hebert@ouhsc.edu

Abstract

Background: Smartphone apps for smoking cessation could offer easily accessible, highly tailored, intensive interventions at a fraction of the cost of traditional counseling. Although there are hundreds of publicly available smoking cessation apps, few have been empirically evaluated using a randomized controlled trial (RCT) design. The Smart-Treatment (Smart-T2) app is a just-in-time adaptive intervention that uses ecological momentary assessments (EMAs) to assess the risk for imminent smoking lapse and tailors treatment messages based on the risk of lapse and reported symptoms.

Objective: This 3-armed pilot RCT aimed to determine the feasibility and preliminary efficacy of an automated smartphone-based smoking cessation intervention (Smart-T2) relative to standard in-person smoking cessation clinic care and the National Cancer Institute’s free smoking cessation app, QuitGuide.

Methods: Adult smokers who attended a clinic-based tobacco cessation program were randomized into groups and followed for 13 weeks (1 week prequitting through 12 weeks postquitting). All study participants received nicotine patches and gum and were asked to complete EMAs five times a day on study-provided smartphones for 5 weeks. Participants in the Smart-T2 group received tailored treatment messages after the completion of each EMA. Both Smart-T2 and QuitGuide apps offer on-demand smoking cessation treatment.

Results: Of 81 participants, 41 (50%) were women and 55 (68%) were white. On average, participants were aged 49.6 years and smoked 22.4 cigarettes per day at baseline. A total of 17% (14/81) of participants were biochemically confirmed 7-day point prevalence abstinent at 12 weeks postquitting (Smart-T2: 6/27, 22%, QuitGuide: 4/27, 15%, and usual care: 4/27, 15%), with no significant differences across groups ($P>.05$). Participants in the Smart-T2 group rated the app positively, with most participants agreeing that they can rely on the app to help them quit smoking, and endorsed the belief that the app would help them stay quit, and these responses were not significantly different from the ratings given by participants in the usual care group.

Conclusions: Dynamic smartphone apps that tailor intervention content in real time may increase user engagement and exposure to treatment-related materials. The results of this pilot RCT suggest that smartphone-based smoking cessation treatments may be capable of providing similar outcomes to traditional, in-person counseling.

Trial Registration: ClinicalTrials.gov NCT02930200; https://clinicaltrials.gov/show/NCT02930200

(J Med Internet Res 2020;22(3):e16907) doi:10.2196/16907

KEYWORDS

smartphones; mobile phones; smoking cessation; just-in-time adaptive intervention; mHealth

Introduction

Background

Although a majority of cigarette smokers are interested in quitting, very few use evidence-based cessation treatments [1]. Best practice guidelines for treating tobacco use and dependence suggest that a combination of counseling and medication is most effective for smoking cessation [2]; however, only 25% of US adult cigarette smokers reported using nicotine patches or gum during their most recent quit attempt, and only 15% sought help from a doctor or other health professional [3]. Smokers have reported a number of barriers to accessing tobacco cessation counseling, including the lack of time, transportation issues, and cost [4,5]. These barriers may be even more burdensome among individuals of a lower socioeconomic status (SES), who have higher rates of tobacco use and are less likely to quit despite similar numbers of quit attempts as those of a higher SES [6,7]. Therefore, improving access to smoking cessation interventions is an important step toward reducing smoking-related health disparities.

Mobile technology has enormous potential to overcome many of the barriers that have hampered the use of other empirically supported smoking cessation treatments among lower SES individuals [8,9]. Smartphone ownership is widespread; 81% of US adults overall and 71% of adults with annual household incomes less than US $30,000 reported owning a smartphone in 2019 [10]. Smartphone apps could offer easily accessible, highly tailored, intensive interventions at a fraction of the cost of traditional smoking cessation counseling [11]. A recent systematic review found that technology-based cessation interventions increased cessation rates compared with standard self-help treatments and produced comparable cessation outcomes among disadvantaged and nondisadvantaged groups [12].

Although there are hundreds of publicly available smoking cessation apps, a few have been empirically evaluated using a randomized controlled trial (RCT) design. In a content analysis of 252 available iPhone and Android apps for smoking cessation, Abroms et al [13] found that very few apps adhered to proven strategies for smoking cessation (eg, suggesting the use of effective medications, connecting to quit lines or clinics) [13]. QuitGuide is a free smartphone app developed by the Tobacco Control Research Branch at the National Cancer Institute (NCI) and based on content from Smokefree.gov [14]. Unlike most available apps, QuitGuide's content adheres to established clinical practice guidelines [2] and includes features such as motivational messages to encourage users to make a quit attempt and detailed information about medications. To date, only one study has examined the efficacy of the QuitGuide app. Bricker et al [15] compared QuitGuide with a smartphone-delivered Acceptance and Commitment Therapy–based app (SmartQuit) in an RCT. The overall self-reported quit rate for participants who were randomized to QuitGuide was 8% compared with 13% for SmartQuit. Thus, there appears to be much room for improvement in phone-based treatments.

In addition to its ability to expand the reach of smoking cessation interventions, mobile technology also allows researchers to examine the dynamic nature of smoking relapse in greater detail than previously possible. Ecological momentary assessment (EMA), in which mobile devices are used to capture moment to moment experiences, allows for the measurement of phenomena in real time within natural settings [16,17]. Using EMA, it is possible to understand the patterns of affect, environment, and social context that individuals experience when undergoing a quit attempt [18]. Although EMA utilizes self-reports, recall bias is greatly reduced by the frequency of measurement, and data are collected in natural contexts rather than laboratory-based settings [19]. A number of factors and cues have been found to be associated with smoking lapse, including urge to smoke [20,21], proximity to others smoking [22], proximity to tobacco retail outlets [23,24], and stress [25-27]. Furthermore, real-time reports of smoking lapse contexts suggest that most lapse episodes occur within minutes of the onset of a craving [28].

Just-in-Time Adaptive Interventions

A recent model for addressing dynamic health behaviors such as smoking lapse is the just-in-time adaptive intervention (JITAI) [29]. JITAIs aim to address moments of vulnerability for unhealthy behaviors (such as high-risk situations) by providing support in real time through mobile technology [30]. JITAIs have been used to target a wide variety of health behaviors, including physical activity [31,32], eating behavior [33], and substance use [34,35]. Although JITAIs for smoking cessation are relatively new, a few studies have shown initial promise. McClure et al [36] found that a mobile intervention that combined self-help content and adaptively tailored advice for managing medication side effects and nicotine withdrawal symptoms was feasible and acceptable among a group of smokers who were ready to quit. Naughton et al [37] demonstrated the feasibility of using geolocation data to trigger support messages to prevent smoking. Using EMA data from smokers undergoing a quit attempt, Businelle et al [26] created a smoking lapse risk estimator that identified 80% of all smoking lapses within 4 hours of the lapse. The algorithm was used in a follow-up study to deliver tailored messages based on a person's momentary risk for smoking lapse, and it was found that urges to smoke and cigarette availability were significantly reduced when tailored urge messages were delivered by the app compared with instances where other types of messages were delivered [38]. Although these JITAIs show great potential for providing widely accessible, innovative treatment for smoking cessation, most JITAIs remain untested. The purpose of this study was to compare, in a pilot RCT, the feasibility and preliminary effectiveness of a smartphone-delivered JITAI for smoking cessation (Smart-Treatment; Smart-T2) with the NCI QuitGuide app and usual care in-person tobacco cessation treatment.

Methods

Participants and Procedure

Individuals were screened for eligibility following a provider referral or self-referral to the Tobacco Treatment Research Program (TTRP), which is located at the University of Oklahoma Health Sciences Center (OUHSC) campus in

Oklahoma City. The TTRP offers free tobacco cessation counseling and pharmacotherapy to the public and facilitates the recruitment, screening, and enrollment of participants into research studies. Referrals are received through the electronic medical record, and via phone, the internet, fax, and word of mouth. Individuals were eligible to participate if they (1) demonstrated an English literacy level greater than the sixth grade, (2) were willing to quit smoking 7 days from their first visit, (3) were ≥18 years of age, (4) had an expired carbon monoxide (CO) level >7 ppm suggestive of current smoking, (5) reported smoking ≥5 cigarettes per day, (6) were willing and able to attend four in-person assessment sessions, and (7) had no contraindications for over-the-counter nicotine replacement therapy (NRT; ie, uncontrolled blood pressure, myocardial infarction within the past 2 weeks, or current pregnancy or plans to become pregnant during the study period). Participants were informed that the study purpose was to compare three smoking cessation treatment approaches and were provided with a detailed outline of study procedures, and written informed consent was obtained.

The study procedure was approved by the Institutional Review Board at the OUHSC. Data collection took place between May 2017 and October 2018. Participants were followed for 13 weeks (1 week prequitting through 12 weeks postquitting) and completed in-person assessments at baseline, on the quit date (1 week after baseline), and at 4- and 12-week postquit visits. All participants were provided with a smartphone (Samsung Galaxy On5) at the baseline visit, were trained to use their assigned app by the study staff, and were asked to carry the phone with them at all times. Study smartphones were used to prompt and deliver EMAs and included (1) a Call Staff function/button that automatically called study staff when/if participants had problems with the phone and (2) a Payment function/button that enabled participants to track their current level of EMA compliance and level of compensation. All data collected through the smartphone app were deidentified and encrypted. Participants were prompted to complete EMAs five times per day (four random assessments and one daily diary) for 5 weeks (1 week prequitting and 4 weeks postquitting). During the EMA period, participants were also asked to self-initiate EMAs when they had an urge to smoke or had already smoked.

Participants were compensated for attending in-person visits and for completing prompted EMAs (ie, daily diary and random EMAs). Specifically, participants received a US $30 gift card for attending and completing each of the first three postquit visits (ie, baseline, quit date, and 4 weeks) and US $50 for completing the 12-week postquit visit. At the 4-week postquit visit, participants received additional compensation based on the percentage of random and daily diary EMAs that they completed. Specifically, those who completed 50% to 74% of all prompted EMAs over the 5-week EMA period received US $50 in gift cards, those who completed 75% to 89% of prompted EMAs received US $100, and those who completed 90% or more of prompted EMAs received US $150. Participants were not compensated for completing self-initiated urge or smoking reports.

Treatment Groups

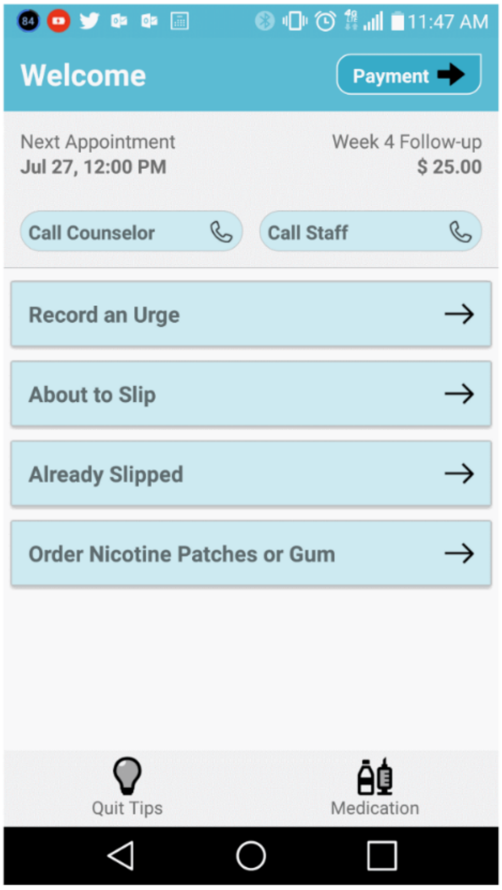
At baseline, participants were randomized into one of the following treatment groups: (1) Smart-T2 phone-based automated smoking cessation treatment, (2) NCI QuitGuide app, or (3) usual tobacco cessation clinic care (TTRP) using a simple computer-generated randomization scheme. All participants were provided with a smartphone preloaded with their assigned smoking cessation app and/or the EMA app for 5 weeks. In addition, all participants received a 2-week supply of over-the-counter NRT (ie, patches and gum) for the initial postquit period.

Smart-Treatment

The Smart-T2 app has been described in detail elsewhere [39]. Briefly, Smart-T2 is a multicomponent adjunctive smoking cessation app featuring (1) an algorithm that evaluates the current risk of smoking lapse based on EMA responses and pushes tailored messages to help participants cope, (2) a "Quit Tips" button offering cessation advice, coping strategies, and quitting benefits, (3) a "Medications" button offering information about smoking cessation medications, (4) a "Phone a Counselor" button that calls the free Oklahoma Tobacco Help Line, (5) daily treatment messages (eg, your quit date is tomorrow), and (6) a button to request additional NRT through the EMA app home screen (Figure 1).

Risk of smoking lapse was estimated in real time using a weighted lapse risk estimation formula developed by Businelle et al [26]. The formula included six variables shown to be associated with lapse, including urge to smoke, stress, recent alcohol consumption, interaction with someone smoking, motivation to quit, and cigarette availability. These lapse risk factors were weighted based on their ability to discriminate moments of high risk for lapse from moments of low risk for lapse (described in detail in a study by Businelle et al [26]). Intervention messages were delivered at the completion of every EMA. During the prequit period, participants received messages that aimed to prepare them for their upcoming quit attempt. During the 4-week postquit period, participants received automated, individually tailored messages based on their current level of risk for imminent smoking lapse and the presence of lapse triggers. When EMA responses indicated low risk for imminent smoking lapse, messages focused on maintaining abstinence motivation and general cessation advice. When EMA responses indicated high risk for imminent smoking lapse or the participant already smoked that day or the day before, or the participant indicated on their first daily assessment that they had a greater than 25% chance of smoking that day, tailored messages focused on ways to cope with current lapse risk symptoms (ie, reported during the current EMA) and were tailored to the highest rate of 4 current lapse triggers (ie, stress, smoking urge, easy access to cigarettes, and low motivation to quit). In addition, when EMA responses indicated a high risk of imminent lapse, participants received a message to chew a piece of nicotine gum to reduce their risk for lapse. When a participant indicated that they lapsed and were no longer interested in quitting smoking, messages focused on treating the lapse as a learning experience and supported a return to abstinence.

Figure 1. The Smart-Treatment app home screen.



Next Appointment
Jul 27, 12:00 PM

Week 4 Follow-up
$ 25.00

QuitGuide

The NCI's QuitGuide app is a free smartphone app that is available through the Smokefree.gov website [14]. The QuitGuide app aims to help smokers understand their smoking patterns and develop the skills needed to quit smoking. Participants can track cravings, mood, smoking triggers, and motivations for quitting. Participants can also access features that provide information about the health consequences of smoking and quitting, smoking cessation medications, ways to handle urges to smoke, developing a multicomponent smoking cessation plan, and coping with lapse. Finally, participants have the option to schedule automated messages to be delivered when they are in a specific location or at a specific time.

Usual Care

Usual tobacco cessation treatment in the TTRP was based on established clinical practice guidelines [2] and included six weekly individual counseling sessions from 1 week before the quit date through 4 weeks postquit date. A total of five unique topics were discussed based on their relevance to the participant at each visit: (1) the impact of tobacco on health/benefits of quitting, (2) stress management strategies, (3) making positive lifestyle changes, (4) developing coping skills, and (5) relapse prevention. The counselor checked in with participants each week about the difficulties and successes they experienced and helped to plan for anticipated challenging situations. Advice and support were provided as needed. In-person counseling was provided at baseline, on the quit date, and at 4-week postquit visits, and all other counseling visits were completed either in-person or via telephone.

Measures

At baseline, participants answered demographic questions including questions on age, sex, race/ethnicity, and smoking history. On the quit date, participants were asked if they smoked "even a puff" since 10:00 PM on the night before their quit date visit. At each in-person visit following the scheduled quit date, participants were asked if they smoked "even a puff" during the past 7 days. Abstinence was verified via expired CO at each visit using a Vitalograph CO monitor. Self-reported abstinence over the specified time period and a CO reading below 6 ppm (10 ppm on the quit date) were required to be considered abstinent. Participants who did not provide biochemical confirmation of abstinence (eg, they did not attend the visit) were considered smoking. On the quit date and at 4-week postquit visits, participants also answered questions to evaluate their satisfaction with their smoking cessation counselor or assigned smoking cessation smartphone apps.

During each EMA, participants answered questions about psychological, social, and environmental factors including stress, urge to smoke, cigarette availability, motivation to quit, recent alcohol consumption, and interaction with someone smoking. Participants were asked to rate on a scale from 1 (strongly disagree) to 5 (strongly agree) if they had an urge to smoke, if they felt stressed, how easily cigarettes were available to them (1 [not at all] to 5 [easily available]), if they were motivated to avoid smoking (1 [strongly agree] to 5 [strongly disagree]), whether they were interacting with anyone who was smoking (yes/no), and whether they drank alcohol in the last hour (yes/no). Participants were also asked to report the likelihood that they would "smoke between now and the end of the day." During daily diary EMAs, participants were asked to report on behaviors and their environment the prior day (eg, how many pieces of nicotine gum they chewed [0 to 8 or more], how many hours that they wore a nicotine patch [I did not wear the patch at all to I wore the patch for at least 22-24 hours], and the number of alcoholic drinks [0 to 8 or more]). Nicotine patch wear time was recoded as a continuous variable by selecting the middle time point of each category used to indicate the period of daily patch wear time (ie, "I did not wear it at all"=0 hours, "Less than 3 hours"=2 hours, "4-6 hours"=5 hours, etc).

Data Analysis

Descriptive statistics were used to summarize participant demographics and engagement with the smartphone app. Comparisons between groups were made using chi-square tests or analyses of variance with a Fisher least significant difference post hoc test, as appropriate. All analyses were conducted in IBM SPSS version 26.

Results

Participants

A total of 98 individuals were assessed for eligibility. Of those, 84 were eligible and consented to participate in the study. Subsequently, 3 individuals dropped out of the study before the baseline visit was completed, and thus all analyses included the remaining 81 participants (Figure 2). Participants were 51% (41/81) women, were mostly white (55/81, 68%), were on average aged 49.6 years, and smoked, on average, 22.4 cigarettes per day at baseline (Table 1). Over the course of the study, 27 participants either withdrew (ie, discontinued participation in the study) or were lost to follow-up (Figure 2). The average age was significantly different across treatment groups (*P*<.007); however, there were no other significant differences in other demographic variables, withdrawal, or loss to follow-up across groups (all *P* values >.05).

Figure 2. Consolidated Standards of Reporting Trials diagram. CO: carbon monoxide.

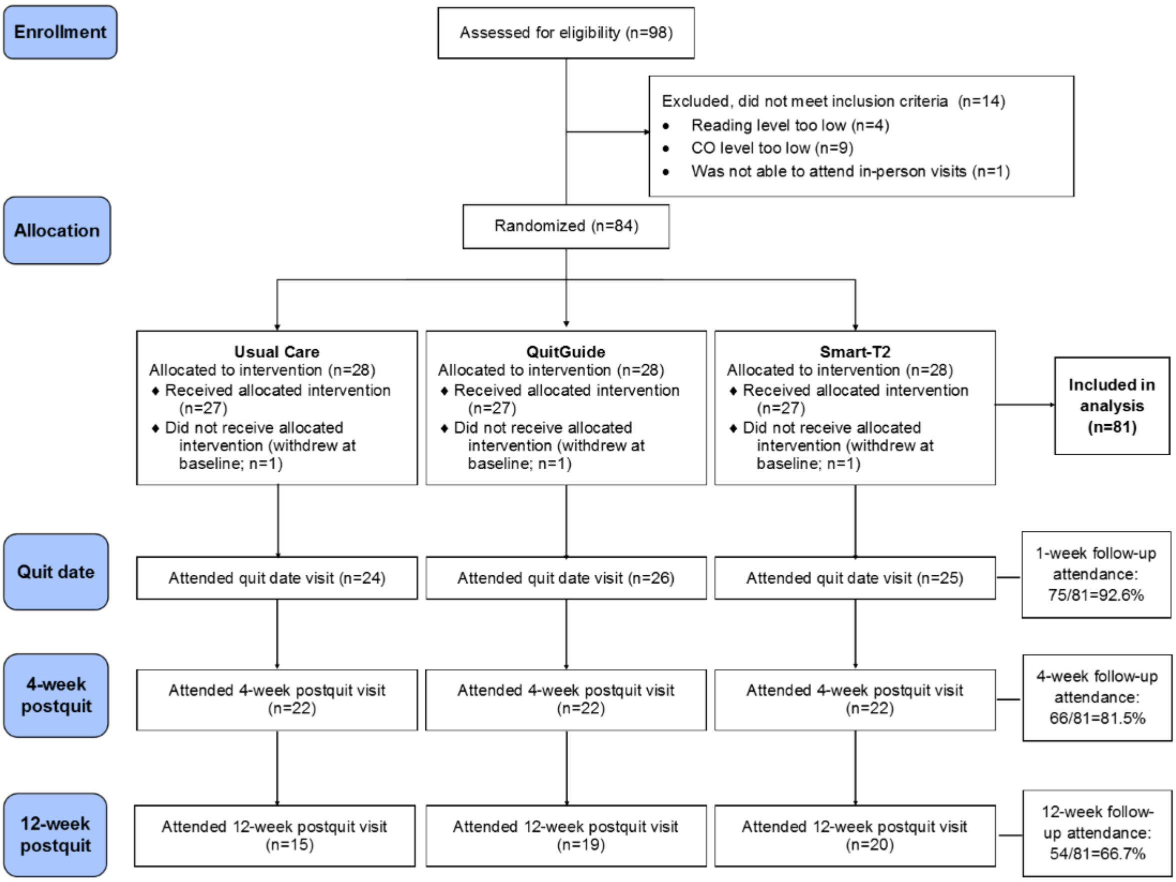


Table 1. Participant demographics.

Characteristic	Total (N=81)	Usual care (n=27)	QuitGuide (n=27)	Smart-Treatment (n=27)
Age (years), mean (SD)	49.6 (11.9)	51.3 (10.1)[a]	44.0 (12.6)[a]	53.6 (11.1)[a]
Gender (female), n (%)	41 (50)	12 (44)	16 (59)	13 (48)
Race, n (%)				
White	55 (67)	18 (66)	19 (70)	18 (66)
Black	14 (17)	6 (22)	6 (22)	2 (7)
Other	12 (14)	3 (11)	2 (7)	7 (25)
Annual household income (<US $30,000), n (%)	40 (49)	15 (55)	14 (51)	11 (40)
Cigarettes smoked per day at baseline, mean (SD)	22.4 (12.6)	21.7 (13.0)	22.9 (14.9)	22.7 (10.0)
Withdrew or lost to follow-up, n (%)	27 (33)	12 (44)	8 (29)	7 (25)

[a] Values 51.3 (10.1) and 53.6 (11.1) are not significantly different from each other but the value 44.0 (12.6) is significantly different from both at the *P*<.05 level.

Ecological Momentary Assessment Completion and Treatment Engagement

Over the study period, participants completed a total of 2384 prompted daily diary and 7688 prompted random EMAs, with an overall compliance rate of 84.0%. Participants self-initiated 3253 EMAs, including reporting cigarette smoking in the prequit period, and lapse and urge to smoke in the postquit period. Compliance with prompted EMAs did not significantly differ across treatment groups. Most phones (72/81, 89%) were returned undamaged. In the usual care group, participants received an average of 3.8 counseling sessions (range 1-6 sessions).

Smart-Treatment

Among the 27 participants in the Smart-T2 group, 14 participants (52%) accessed the on-demand medication tips, and 20 (74%) participants accessed the on-demand quit tips. Once a participant selected a specific category of message, they could click "next" to view multiple messages within that category. The most frequently selected tip types were "Coping with Others Smoking" and "Harms of Smoking" (selected an average of 2.0 times during the intervention period), followed by "Coping with Mood" (selected an average of 1.8 times), and "Medication: Nicotine Gum" (selected an average of 1.5 times). On average, participants viewed the most messages within "Coping with Others Smoking" (mean 60.0 messages, SD 7.1), "General Quitting Advice" (mean 58.3 messages, SD 50.2), and "Harms of Smoking" categories (mean 42.0 messages, SD 26.1). The number of on-demand tips accessed varied over time (Figure 3) with the majority of tips viewed on days 1 and 2 of the study and declining on day 3. There was a sharp increase in the number of tips viewed on day 6 (the day before the scheduled quit day) as well as day 9 (2 days into the quit attempt).

A total of 66% (18/27) of participants in the Smart-T2 group used the Order NRT button to request a refill of NRT and 40% (11/27) of participants used the Call counselor button to reach the Oklahoma Tobacco Cessation Helpline, for an average of 2.2 button pushes during the 5-week EMA period. During the course of the intervention, a total of 3873 messages were delivered to participants in the Smart-T2 group. Each Smart-T2 participant received 145 treatment messages on average during the study. Of the high-risk tailored messages delivered, a majority (869/1638, 53.05%) were related to easy cigarette availability, followed by urge to smoke (630/1638, 38.46%), motivation to quit (85/1638, 5.19%), and stress (54/1638, 3.30%). Figure 4 shows the distribution of tailored messages over the course of the postquit period.

Figure 3. The distribution of on-demand tips accessed over time.

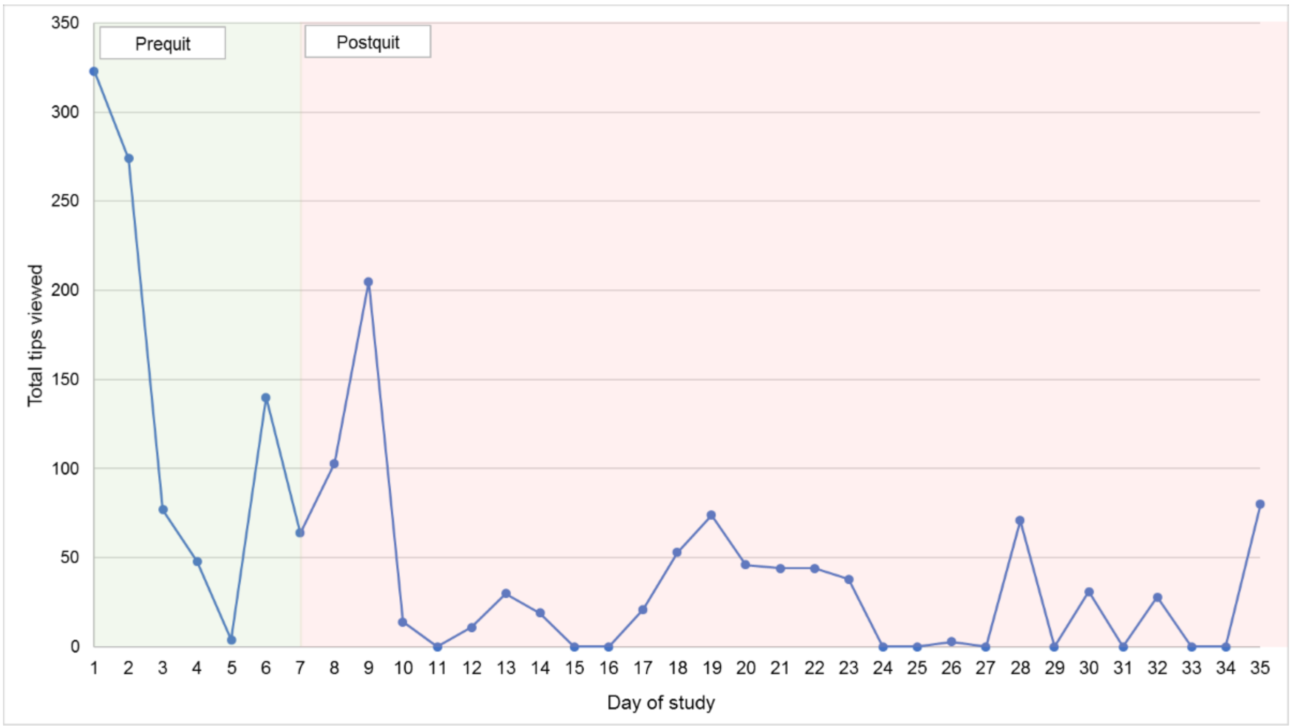
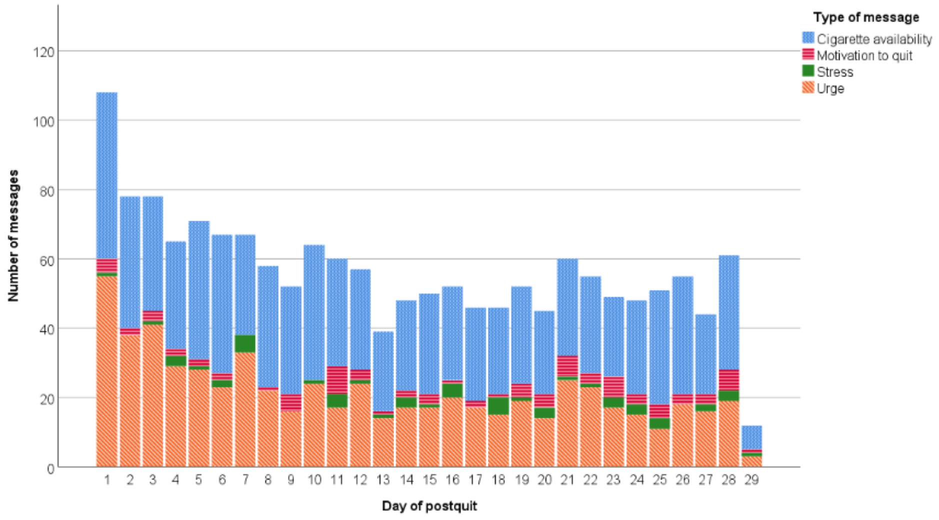


Figure 4. The distribution of high-risk messages over time.



QuitGuide

Among the participants in the QuitGuide treatment group, 78% (21/27) of participants opened the app at an average of 9.9 times (SD 7.4) and interacted with the application for an average of 10.6 days (SD 4.8; range 2-20 days) during the 5-week EMA period. Within each day interacting with QuitGuide, participants completed an average of 5.5 actions (ie, a unique button press that initiates an app feature, eg, opening the journal; SD 5.9; range 1-32 actions). Over the study period, 59% (16/27) of participants used the Manage my Mood feature at an average of 3.0 times, 33% (9/27) of participants accessed the How to Quit feature at an average of 2.1 times, and 41% (11/27) of participants used the journal at an average of 1.7 times. Only 4 participants used the Location Help feature, which allowed them to tag a location in which they would receive a message to prevent lapse or to cope with craving to smoke. In addition, only 2 participants used the Share my Stats feature to share their progress on social media.

Smoking Cessation Outcomes

A total of 26% (21/81) of participants were confirmed abstinent (ie, 7-day point prevalence, intent to treat) at 4 weeks postquitting (Smart-T2: 6/27, 22%, QuitGuide: 7/27, 26%, usual care: 8/27, 30%), and 17% (14/81) participants were confirmed abstinent at 12 weeks postquitting (Smart-T2: 6/27, 22%, QuitGuide: 4/27, 15%, usual care: 4/27, 15%). There were no significant differences in smoking abstinence across treatment groups at any time point.

Nicotine Replacement Therapy Utilization

On each daily diary, participants reported the number of pieces of gum that they chewed and the number of hours that they wore a nicotine patch the previous day. On average, participants reported chewing 5.1 pieces of gum each day (SD 2.5; range 0-8 or more) and wearing a nicotine patch for an average of 19.4 hours each day (SD 5.6; range=0-23). There were no significant differences in the number of pieces of gum chewed or patch wear time across treatment groups. Within the Smart-T2 group, when the risk for lapse was high, participants also received a message stating, “Chewing a piece of nicotine gum right now may reduce your risk for smoking. Will you chew a piece of nicotine gum right now?” The nicotine gum message was delivered 861 times, 31.9 times per participant on average (SD 44.0). Participants responded “yes,” that they would chew a piece a gum in 65.2% of cases.

Treatment Satisfaction

Across all three groups, a majority of participants agreed or strongly agreed with the statements, “I can rely on my (treatment) to provide guidance that will help me to quit smoking and stay quit” (46/65, 70.8%) and “I believe that my [treatment] will help me to quit smoking and stay quit” (45/65, 69.2%. The mean response on a scale from 1 (strongly disagree) to 5 (strongly agree) for both questions was significantly higher in the usual care group compared with the QuitGuide group, but not significantly different from the mean Smart-T2 group response (Table 2). Participants in the usual care group reported a significantly higher mean response agreeing with the statement, “I feel that my [treatment] provides smoking cessation treatment that is personalized to my specific needs,” compared with both the QuitGuide and Smart-T2 groups. Participants in both the usual care and Smart-T2 groups reported significantly higher average responses (ie, strongly agreeing, agreeing) to the statements, “My [treatment] knows how to help me quit smoking” and “I believe I can depend on my [treatment],” and reported significantly lower mean responses (ie, strongly disagreeing, disagreeing) to the question, “Do you find the smartphone application to be annoying?” (Table 2).

Table 2. Treatment satisfaction and app perceptions at week 4 postquit.

Treatment group	Value, mean (SD)	SE	*F* test (*df*)	*P* value
I can rely on my treatment to provide guidance that will help me to quit smoking and stay quit				.04
Usual Care (n=21)	4.33^{a} (0.66)	0.14	3.34 (2,62)	
QuitGuide (n=22)	3.59^{b} (1.14)	0.24	3.34 (2,62)	
Smart-T2^{a} (n=22)	$3.95^{a,b}$ (0.95)	0.20	3.34 (2,62)	
Total (N=65)	3.95 (0.98)	0.12	3.34 (2,62)	
I feel that my treatment provides smoking cessation treatment that is personalized to my specific needs				.01
Usual Care (n=21)	4.33^{a} (0.66)	0.14	5.45 (2,62)	
QuitGuide (n=22)	3.59^{b} (1.05)	0.23	5.45 (2,62)	
Smart-T2 (n=22)	3.55^{b} (0.86)	0.18	5.45 (2,62)	
Total (N=65)	3.82 (0.93)	0.12	5.45 (2,62)	
I believe that my treatment will help me to quit smoking and stay quit				.04
Usual Care (n=21)	4.33^{a} (0.80)	0.17	3.52 (2,62)	
QuitGuide (n=22)	3.64^{b} (1.05)	0.22	3.52 (2,62)	
Smart-T2 (n=22)	4.09^{ab} (0.75)	0.16	3.52 (2,62)	
Total (N=65)	4.02 (0.91)	0.11	3.52 (2,62)	
My treatment knows how to help me to quit smoking				.02
Usual Care (n=21)	4.29^{a} (0.78)	0.17	4.23 (2,62)	
QuitGuide (n=22)	3.59^{b} (1.10)	0.23	4.23 (2,62)	
Smart-T2 (n=22)	4.23^{a} (0.69)	0.15	4.23 (2,62)	
Total (N=65)	4.03 (0.92)	0.11	4.23 (2,62)	
I believe I can depend on my treatment				<.001
Usual Care (n=21)	4.33^{a} (0.80)	0.17	7.48 (2,62)	
QuitGuide (n=22)	3.27^{b} (1.24)	0.27	7.48 (2,62)	
Smart-T2 (n=22)	4.05^{a} (0.65)	0.14	7.48 (2,62)	
Total (N=65)	3.88 (1.02)	0.13	7.48 (2,62)	
Do you find the smartphone application to be annoying?				.01
Usual Care (n=22)	2.45^{a} (1.37)	0.29	5.11 (2,63)	
QuitGuide (n=22)	3.41^{b} (1.40)	0.30	5.11 (2,63)	
Smart-T2 (n=22)	2.23^{a} (1.11)	0.24	5.11 (2,63)	
Total (N=66)	2.70 (1.38)	0.17	5.11 (2,63)	

[a] All scales rated from 1=strongly disagree to 5=strongly agree.

[b] Pairwise comparisons are indicated with a superscript. Values that do not share a letter are significantly different at the *P*<.05 level.

Discussion

Principal Findings

This study demonstrates the feasibility and acceptability of Smart-T2, a smartphone-based, JITAI for smoking cessation. Participants across all three treatment groups (ie, Smart-T2, QuitGuide, and usual care) were largely compliant with the EMA protocol, and a majority of participants in the Smart-T2 group engaged with on-demand treatment content and utilized the app to order additional NRT. Although the study was not powered to detect significant differences in smoking cessation outcomes or NRT use across the three treatment groups, the results of this pilot RCT suggest that smartphone-based smoking cessation treatments may be capable of providing similar outcomes to traditional, in-person counseling. Participants in the Smart-T2 group rated the app positively, with most participants agreeing that they can rely on the app to help them quit smoking and endorsed the belief that the app would help them stay quit, and these responses were not significantly different from the ratings given by participants in the usual care group.

Treatment Engagement

Engagement with digital behavior change interventions has been defined as “the extent (e.g., amount, frequency, duration, depth) of usage and a subjective experience characterized by attention, interest, and affect” [40]. Research has demonstrated that engagement with smoking cessation apps is typically low [41]. For example, Zbikowski et al [41] tracked program utilization for an integrated phone and Web-based tobacco cessation program and found that of 11,143 participants, users logged into an app at an average of 1 to 2 times and completed an average of 2 to 2.5 counseling calls. In the Smart-T2 group, participants consistently engaged with the EMAs for the duration of the study period, and most participants engaged with the on-demand content at least once. In this study, engagement with the QuitGuide app was lower compared with the Bricker et al [15] trial wherein participants self-reported opening the app at an average of 15.2 times compared with an average of 9.9 times in this study. However, it is worth noting that participants assigned to the QuitGuide intervention arm in this study had to manage two separate apps (the QuitGuide app and the EMA app) to access intervention content and complete EMAs, whereas those in the Smart-T2 intervention were reminded about the on-demand content because the home screen was displayed at the completion of each EMA. In addition, participants in Bricker et al [15] self-reported the number of times they opened the app, whereas engagement statistics in this study were objectively recorded in the QuitGuide app and exported.

Within the Smart-T2 group, participants accessed the most messages within the “General Quitting Advice” and “Coping with Others Smoking” categories, on average. The fact that participants were drawn to the general quitting advice category may suggest that many participants were seeking general coping strategies to aid their cessation attempt. However, the high number of tips viewed from the “Coping with Others Smoking” category is consistent with the fact that a majority of automated, tailored messages delivered were related to easy cigarette availability. Over the course of the study, most on-demand tips were viewed within the first 2 days of the prequit period, and on the second and third days of the postquit period. This pattern may suggest that participants desire greater treatment content in the first few days after downloading an app and in the immediate period after the quit date, or that participants were initially curious about treatment content and then lost interest over time. Future interventions could examine if having tips or treatment content that becomes available over time is associated with a more consistent pattern of engagement for the duration of treatment compared with on-demand content that is available all at once.

Treatment Satisfaction

Participant treatment satisfaction in this pilot study may also have implications for future mobile smoking cessation interventions. Participants in the Smart-T2 group and usual care groups found the EMA app significantly less “annoying” compared with those in the QuitGuide group. A possible explanation is that participants in the Smart-T2 group were receiving tailored treatment content at the end of each EMA, whereas participants in the QuitGuide group only received smoking cessation treatment through a separate app. Repeated surveys during the day can be burdensome; however, participants may find the surveys more useful if they know that they are driving the delivery of content that is specifically relevant to their current socioenvironmental context. Across a range of health behaviors, tailored treatments have been shown to be superior to the more commonly used “one-size-fits-all” treatment approach [42-45]. This study further illustrates that dynamically tailored content enhances participant engagement and may increase acceptability of smartphone-based cessation interventions.

It is promising that both app-based interventions (Smart-T2 and QuitGuide) performed at least as well as the traditional, in-person counseling in terms of response rates, loss to follow-up, participant perceptions of the treatment, and engagement. The patterns of engagement and the participant perceptions of the intervention are consistent with findings in the literature. For example, Oliver et al [46] surveyed 224 daily cigarette smokers and asked them to describe the utility of features within smartphone apps for smoking cessation. Features that were rated as most important included gain-framed messages such as “tells me how much my health is improving each day that I don’t smoke” as well as “develops a personalized quit plan for me” and “helps me track my stress and craving levels.” Participants in both the QuitGuide and the Smart-T2 intervention groups interacted most frequently with features that offered general quitting advice, as well as features that helped manage their stress and mood. Both the QuitGuide and the Smart-T2 intervention groups provided tailored treatment with smoking cessation advice with either a personalized quit plan (ie, QuitGuide), or specific coping strategies to deal with the smoking triggers reported by participants in the moment (ie, Smart-T2).

Smoking Cessation Outcomes

Although there has been a proliferation of smartphone-based smoking cessation interventions, very few RCTs have been conducted to test the efficacy of these interventions. A recent review of mobile applications for the treatment of tobacco use and dependence [47] found that only four well-powered studies have tested efficacy or effectiveness of smoking cessation apps, with abstinence rates ranging from 0.9% to 12% at the end of the study [47]. A review of smoking cessation interventions for disadvantaged groups found a similar dearth of high-quality studies that have examined the effectiveness of technology-based smoking cessation among disadvantaged populations [12], yet there is preliminary evidence that they are effective at increasing quit rates at 1-, 3-, 6-, and 18-month follow-ups. The sample size of this pilot trial precludes the ability to make definitive conclusions about the effectiveness of this intervention; however, the current preliminary results of a 12-week biochemically confirmed quit rate of 22.2% in the Smart-T2 group and 14.8% in the QuitGuide group may suggest that combining nicotine therapy with a mobile app could improve quit rates.

To our knowledge, only one other study, Bricker et al [15], has examined the effectiveness of the QuitGuide app, with a self-reported 30-day point prevalence cessation rate of 8% at a 2-month follow-up. In contrast, 14.8% of participants in the QuitGuide group of this study were biochemically confirmed abstinent at 12 weeks postquitting. However, the studies differed in terms of the method of measuring abstinence (ie, self-reporting vs biochemical verification and completers only vs intent to treat) and in terms of treatment protocol (ie, QuitGuide participants in this study were provided with NRT); thus, it is unclear if these results will generalize to a larger, fully powered sample.

Strengths and Limitations

The strength of this study is the RCT design, which allowed for a preliminary comparison of the Smart-T2 app with in-person smoking cessation treatment and the NCI QuitGuide app that adheres to many of the recommendations of the smoking cessation clinical practice guidelines [2]. This study also has several limitations. First, the Smart-T2 app did not save the duration of counseling calls that were initiated through the app. Thus, we were unable to determine if use of the “Call Counselor” button was representative of legitimate calls to the Oklahoma Tobacco Cessation Helpline. Second, this pilot trial had a small sample size, which precludes the ability to make definitive conclusions about the effectiveness of the interventions. A future study will compare the effectiveness of the Smart-T2 app and QuitGuide intervention in a larger trial [48].

Future Directions

The Smart-T2 app and similar interventions may represent a potential way to address the substantial tobacco-related health disparities experienced by low SES smokers who may have limited access to in-person treatment. To our knowledge, no interventions to date besides our recent pilot work [39] have used EMAs to repeatedly assess current smoking lapse risk and automatically deliver tailored treatment content. This preliminary work indicates that smartphones may be used to deliver well-liked, automated, tailored, low burden, and easily accessible interventions to smokers seeking smoking cessation treatments.

Acknowledgments

Programming and technological support was provided through the mobile health shared resource of the Stephenson Cancer Center via an NCI Cancer Center Support Grant (P30CA225520). Data analysis and manuscript preparation were additionally supported through K99DA046564, and through the Oklahoma Tobacco Settlement Endowment Trust grant 092-016-0002. The contents of the manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the sponsoring organizations.

Conflicts of Interest

MB, DK, and DV are inventors of the Insight mHealth Platform, which was used to develop the Smart-T2 app. They receive royalties related to its use.

Multimedia Appendix 1
CONSORT-EHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3120 KB - jmir_v22i3e16907_app1.pdf]

References

1. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults - United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017 Jan 6;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
2. Fiore M, Jaen C, Baker T, Bailey W, Benowitz N, Curry S. Treating Tobacco Use and Dependence: 2008 Update: Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services; 2008.
3. Caraballo RS, Shafer PR, Patel D, Davis KC, McAfee TA. Quit methods used by US adult cigarette smokers, 2014-2016. Prev Chronic Dis 2017 Apr 13;14:E32 [FREE Full text] [doi: 10.5888/pcd14.160600] [Medline: 28409740]
4. Skojac TM, Ackerman MD. Evaluating patient barriers to tobacco cessation treatment. Fed Pract 2014;31(4):22-26 [FREE Full text]
5. Gollust SE, Schroeder SA, Warner KE. Helping smokers quit: understanding the barriers to utilization of smoking cessation services. Milbank Q 2008 Dec;86(4):601-627 [FREE Full text] [doi: 10.1111/j.1468-0009.2008.00536.x] [Medline: 19120982]
6. Kotz D, West R. Explaining the social gradient in smoking cessation: it's not in the trying, but in the succeeding. Tob Control 2009 Feb;18(1):43-46 [FREE Full text] [doi: 10.1136/tc.2008.025981] [Medline: 18936053]
7. Hyland A, Borland R, Li Q, Yong H, McNeill A, Fong GT, et al. Individual-level predictors of cessation behaviours among participants in the International Tobacco Control (ITC) Four Country Survey. Tob Control 2006 Jun;15(Suppl 3):iii83-iii94 [FREE Full text] [doi: 10.1136/tc.2005.013516] [Medline: 16754952]

8. Siahpush M, Yong H, Borland R, Reid JL, Hammond D. Smokers with financial stress are more likely to want to quit but less likely to try or succeed: findings from the International Tobacco Control (ITC) Four Country Survey. Addiction 2009 Aug;104(8):1382-1390 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02599.x] [Medline: 19438837]
9. Copeland AL, Businelle MS, Stewart DW, Patterson SM, Rash CJ, Carney CE. Identifying barriers to entering smoking cessation treatment among socioeconomically disadvantaged smokers. J Smok Cessat 2010;5(2):164-171. [doi: 10.1375/jsc.5.2.164]
10. Pew Research Center. 2019 Jun 12. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-09-17]
11. Naughton F. Delivering 'Just-in-time' smoking cessation support via mobile phones: current knowledge and future directions. Nicotine Tob Res 2017 Mar;19(3):379-383. [doi: 10.1093/ntr/ntw143] [Medline: 27235703]
12. Boland V, Stockings E, Mattick R, McRobbie H, Brown J, Courtney R. The methodological quality and effectiveness of technology-based smoking cessation interventions for disadvantaged groups: a systematic review and meta-analysis. Nicotine Tob Res 2018 Feb 7;20(3):276-285. [doi: 10.1093/ntr/ntw391] [Medline: 28034998]
13. Abroms LC, Westmaas JL, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
14. Smokefree. QuitGuide URL: https://smokefree.gov/tools-tips/apps/quitguide [accessed 2020-02-17]
15. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 1;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
16. Shiffman S, Hufford M, Hickcox M, Paty JA, Gnys M, Kassel JD. Remember that? A comparison of real-time versus retrospective recall of smoking lapses. J Consult Clin Psychol 1997 Apr;65(2):292-300. [doi: 10.1037/0022-006x.65.2.292.a] [Medline: 9086693]
17. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. Br Med J 2002 May 18;324(7347):1193-1194 [FREE Full text] [doi: 10.1136/bmj.324.7347.1193] [Medline: 12016186]
18. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
19. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
20. Shiffman S, Gwaltney CJ, Balabanis MH, Liu KS, Paty JA, Kassel JD, et al. Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment. J Abnorm Psychol 2002 Nov;111(4):531-545. [doi: 10.1037//0021-843x.111.4.531] [Medline: 12428767]
21. Shiffman S, Paty JA, Gnys M, Kassel JA, Hickcox M. First lapses to smoking: within-subjects analysis of real-time reports. J Consult Clin Psychol 1996 Apr;64(2):366-379. [doi: 10.1037//0022-006x.64.2.366] [Medline: 8871421]
22. Shapiro D, Jamner LD, Davydov DM, James P. Situations and moods associated with smoking in everyday life. Psychol Addict Behav 2002 Dec;16(4):342-345. [doi: 10.1037//0893-164x.16.4.342] [Medline: 12503908]
23. Watkins KL, Regan SD, Nguyen N, Businelle MS, Kendzor DE, Lam C, et al. Advancing cessation research by integrating EMA and geospatial methodologies: associations between tobacco retail outlets and real-time smoking urges during a quit attempt. Nicotine Tob Res 2014 May;16(Suppl 2):S93-101 [FREE Full text] [doi: 10.1093/ntr/ntt135] [Medline: 24057995]
24. Reitzel LR, Cromley EK, Li Y, Cao Y, Dela Mater R, Mazas CA, et al. The effect of tobacco outlet density and proximity on smoking cessation. Am J Public Health 2011 Feb;101(2):315-320. [doi: 10.2105/AJPH.2010.191676] [Medline: 21164089]
25. Shiffman S, Balabanis MH, Gwaltney CJ, Paty JA, Gnys M, Kassel JD, et al. Prediction of lapse from associations between smoking and situational antecedents assessed by ecological momentary assessment. Drug Alcohol Depend 2007 Dec 1;91(2-3):159-168 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.05.017] [Medline: 17628353]
26. Businelle MS, Ma P, Kendzor DE, Frank SG, Wetter DW, Vidrine DJ. Using intensive longitudinal data collected via mobile phone to detect imminent lapse in smokers undergoing a scheduled quit attempt. J Med Internet Res 2016 Oct 17;18(10):e275 [FREE Full text] [doi: 10.2196/jmir.6307] [Medline: 27751985]
27. Businelle MS, Ma P, Kendzor DE, Reitzel LR, Chen M, Lam CY, et al. Predicting quit attempts among homeless smokers seeking cessation treatment: an ecological momentary assessment study. Nicotine Tob Res 2014 Oct;16(10):1371-1378 [FREE Full text] [doi: 10.1093/ntr/ntu088] [Medline: 24893602]
28. Ferguson SG, Shiffman S. The relevance and treatment of cue-induced cravings in tobacco dependence. J Subst Abuse Treat 2009 Apr;36(3):235-243. [doi: 10.1016/j.jsat.2008.06.005] [Medline: 18715743]
29. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 2014;13(3):13-17. [doi: 10.1109/mprv.2014.46]
30. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

31. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
32. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: Implications for JITAI optimization. Health Psychol 2015 Dec;34S:1261-1267 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]
33. Rahman T, Czerwinski M, Gilad-Bachrach R, Johns P. Predicting 'About-to-Eat' Moments for Just-in-Time Eating Intervention. In: Proceedings of the 6th International Conference on Digital Health Conference. 2016 Presented at: DH'16; April 11 - 13, 2016; Montréal, Québec, Canada p. 141-150. [doi: 10.1145/2896338.2896359]
34. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
35. Boyer EW, Fletcher R, Fay RJ, Smelson D, Ziedonis D, Picard RW. Preliminary efforts directed toward the detection of craving of illicit substances: the iHeal project. J Med Toxicol 2012 Mar;8(1):5-9 [FREE Full text] [doi: 10.1007/s13181-011-0200-4] [Medline: 22311668]
36. McClure JB, Anderson ML, Bradley K, An LC, Catz SL. Evaluating an adaptive and interactive mHealth smoking cessation and medication adherence program: a randomized pilot feasibility study. JMIR Mhealth Uhealth 2016 Aug 3;4(3):e94 [FREE Full text] [doi: 10.2196/mhealth.6002] [Medline: 27489247]
37. Naughton F, Hopewell S, Lathia N, Schalbroeck R, Brown C, Mascolo C, et al. A context-sensing mobile phone app (Q Sense) for smoking cessation: a mixed-methods study. JMIR Mhealth Uhealth 2016 Sep 16;4(3):e106 [FREE Full text] [doi: 10.2196/mhealth.5787] [Medline: 27637405]
38. Hébert ET, Stevens EM, Frank SG, Kendzor DE, Wetter DW, Zvolensky MJ, et al. An ecological momentary intervention for smoking cessation: The associations of just-in-time, tailored messages with lapse risk factors. Addict Behav 2018 Mar;78:30-35 [FREE Full text] [doi: 10.1016/j.addbeh.2017.10.026] [Medline: 29121530]
39. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. J Med Internet Res 2016 Dec 12;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]
40. 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]
41. Zbikowski SM, Hapgood J, Smucker Barnwell S, McAfee T. Phone and web-based tobacco cessation treatment: real-world utilization patterns and outcomes for 11,000 tobacco users. J Med Internet Res 2008 Nov 14;10(5):e41 [FREE Full text] [doi: 10.2196/jmir.999] [Medline: 19017583]
42. Borland R, Balmford J, Hunt D. The effectiveness of personally tailored computer-generated advice letters for smoking cessation. Addiction 2004 Mar;99(3):369-377. [doi: 10.1111/j.1360-0443.2003.00623.x] [Medline: 14982550]
43. Strecher VJ, Kreuter M, Den Boer DJ, Kobrin S, Hospers HJ, Skinner CS. The effects of computer-tailored smoking cessation messages in family practice settings. J Fam Pract 1994 Sep;39(3):262-270. [Medline: 8077905]
44. Kreuter MW, Bull FC, Clark EM, Oswald DL. Understanding how people process health information: a comparison of tailored and nontailored weight-loss materials. Health Psychol 1999 Sep;18(5):487-494. [doi: 10.1037//0278-6133.18.5.487] [Medline: 10519465]
45. 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]
46. Oliver J, Hallyburton M, Pacek L, Mitchell J, Vilardaga R, Fuemmeler B, et al. What do smokers want in a smartphone-based cessation application? Nicotine Tob Res 2018 Nov 15;20(12):1507-1514 [FREE Full text] [doi: 10.1093/ntr/ntx171] [Medline: 29065202]
47. Vilardaga R, Casellas-Pujol E, McClernon JF, Garrison KA. Mobile applications for the treatment of tobacco use and dependence. Curr Addict Rep 2019 Jun;6(2):86-97 [FREE Full text] [doi: 10.1007/s40429-019-00248-0] [Medline: 32010548]
48. Businelle MS. Grantome. 2019. Smartphone Based Smoking Cessation Intervention for Socioeconomically Disadvantaged Adults URL: http://grantome.com/grant/NIH/R01-CA221819-01A1 [accessed 2020-02-19]

Abbreviations

CO: carbon monoxide
EMA: ecological momentary assessment
JITAI: just-in-time adaptive intervention
NCI: National Cancer Institute
NRT: nicotine replacement therapy
OUHSC: University of Oklahoma Health Sciences Center
RCT: randomized controlled trial

SES: socioeconomic status
Smart-T2: Smart-Treatment
TTRP: Tobacco Treatment Research Program

Edited by G Eysenbach; submitted 05.11.19; peer-reviewed by R Vilardaga, T Wyatt, M Buzzi; comments to author 26.11.19; revised version received 17.01.20; accepted 03.02.20; published 09.03.20.

Please cite as:
Hébert ET, Ra CK, Alexander AC, Helt A, Moisiuc R, Kendzor DE, Vidrine DJ, Funk-Lawler RK, Businelle MS
A Mobile Just-in-Time Adaptive Intervention for Smoking Cessation: Pilot Randomized Controlled Trial
J Med Internet Res 2020;22(3):e16907
URL: https://www.jmir.org/2020/3/e16907
doi:10.2196/16907
PMID:32149716

Original Paper

Efficacy of a Web App for Cognitive Training (MeMo) Regarding Cognitive and Behavioral Performance in People With Neurocognitive Disorders: Randomized Controlled Trial

Philippe Robert[1,2,3], MD, PhD; Valeria Manera[1,2], PhD; Alexandre Derreumaux[2,3], MSc; Marion Ferrandez Y Montesino[2], MSc; Elsa Leone[3], PhD; Roxane Fabre[4], PhD; Jeremy Bourgeois[3], PhD

[1]Cognition Behaviour Technology Lab, Université Côte d'Azur, Nice, France
[2]Association Innovation Alzheimer, Nice, France
[3]Centre Memoire, Centre Hospitalier Universitaire de Nice, Nice, France
[4]Departement de Santé Publique, Centre Hospitalier Universitaire de Nice, Nice, France

Corresponding Author:
Philippe Robert, MD, PhD
Cognition Behaviour Technology Lab
Université Côte d'Azur
Institut Claude Pompidou
10 Rue Moliere
Nice, 06100
France
Phone: 33 0492034770
Email: Philippe.ROBERT@univ-cotedazur.fr

Abstract

Background: Cognitive and behavioral symptoms are the clinical hallmarks of neurocognitive disorders. Cognitive training may be offered to reduce the risks of cognitive decline and dementia and to reduce behavioral symptoms, such as apathy. Information and communication technology approaches, including serious games, can be useful in improving the playful aspect of computerized cognitive training and providing motivating solutions in elderly patients.

Objective: The objective of this study was to assess the effectiveness of employing the MeMo (Memory Motivation) Web app with regard to cognitive and behavioral symptoms in patients with neurocognitive disorders.

Methods: MeMo is a Web app that can be used on any Web browser (computer or tablet). The training activities proposed in MeMo are divided into the following two parts: memory and mental flexibility/attention. The study included 46 individuals (mean age 79.4 years) with a diagnosis of neurocognitive disorders at the Institut Claude Pompidou Memory Center in Nice. This randomized controlled study compared the evolution of cognition and behavior between patients not using MeMo (control group) and patients using MeMo (MeMo group) for 12 weeks (four sessions per week). Patients underwent memory and attention tests, as well as an apathy assessment at baseline, week 12 (end of the training period), and week 24 (12 weeks after the end of the training sessions). In addition, to assess the impact of high and low game uses, the MeMo group was divided into patients who used MeMo according to the instructions (about once every 2 days; active MeMo group) and those who used it less (nonactive MeMo group).

Results: When comparing cognitive and behavioral scores among baseline, week 12, and week 24, mixed model analysis for each cognitive and behavioral score indicated no significant interaction between testing time and group. On comparing the active MeMo group (n=9) and nonactive MeMo group (n=13), there were significant differences in two attention tests (Trial Making Test A [*P*=.045] and correct Digit Symbol Substitution Test items [*P*=.045]) and in the Apathy Inventory (AI) (*P*=.02). Mixed analysis (time: baseline, week 12, and week 24 × number of active days) indicated only one significant interaction for the AI score (*P*=.01), with a significant increase in apathy in the nonactive MeMo group.

Conclusions: This study indicates that the cognitive and behavioral efficacies of MeMo, a Web-based training app, can be observed only with regular use of the app. Improvements were observed in attention and motivation.

Trial Registration: ClinicalTrials.gov NCT04142801; https://clinicaltrials.gov/ct2/show/NCT04142801

(J Med Internet Res 2020;22(3):e17167) doi:10.2196/17167

KEYWORDS

neurocognitive disorders; Alzheimer disease; cognition; motivation; apathy; intervention

Introduction

Memory, attention, and behavioral symptoms are the clinical hallmarks of neurocognitive disorders (NCDs), such as Alzheimer disease and related disorders [1]. While there is no curative treatment for major NCDs, the proactive management of modifiable risk factors can delay or slow down the onset or progression of these diseases [2]. In particular, a number of randomized controlled trials (RCTs) suggested that lifestyle-based multidomain interventions, including management of cardiovascular risk factors and cognitive training, may be feasible and effective in reducing the risk of cognitive decline in older at-risk adults [3]. Cognitive training, which refers to the guided practice of specific standardized tasks designed to enhance particular cognitive functions [4], appears to be a promising option to improve cognitive performance in older adults [5]. For this reason, the World Health Organization guidelines for risk reduction of cognitive decline and dementia indicate that cognitive training may be offered to older adults with normal cognition and with mild cognitive impairment to reduce the risks of cognitive decline and dementia [6]. However, evidence on the efficacy of cognitive training in people with cognitive decline (minor and major NCDs) is still scarce, and there is a lack of RCTs [5,7,8]. Furthermore, most existing studies only focused on the improvement of cognitive function. In NCDs, cognitive symptoms are often associated and are sometimes even preceded by behavioral symptoms, such as apathy [9]. Apathy is a multidimensional syndrome characterized by a significant reduction in goal-directed activity [10,11]. It represents the most common behavioral and psychological symptoms in people with Alzheimer disease and is often observed in Parkinson disease and other dementia-related disorders, such as vascular dementia and frontotemporal dementia [12]. Pharmacological therapies have demonstrated limited efficacy in the management of apathy associated with NCDs and neuropsychiatric conditions [13]. However, nonpharmacological treatments, including cognitive training via serious games and virtual reality, are considered promising approaches for apathy management in people with NCDs [14-16].

The use of information and communication technology (ICT) in the health domain is progressively expanding. Recently, increasing attention is being devoted to the field of NCDs, where ICT is employed to both support and improve the assessment of different functional and cognitive abilities [17-20] and to provide alternative solutions for patient stimulation and rehabilitation. Serious games are mental and physical contests played with a computer in accordance with specific rules, which use entertainment to promote training, education, health, public policy, and strategic communication objectives [21]. One of the targets is to improve the playful aspect of computerized cognitive training in order to improve motivation [22]. ICT can also be useful in providing solutions that can be proposed to patients with NCDs coming for consultations to memory centers. Cognitive training is indeed indicated for some of them according to clinician judgement, but it is often difficult to propose. This is due to several reasons, including logistic constraints (eg, necessity to visit the memory center several times a week to participate in stimulation sessions and clinician availability), the necessity to pay subscription fees, the fact that the effectiveness of existing tools is not scientifically proven, and the fact that most existing free training platforms are available only in English. To overcome these problems, the Alzheimer Innovation Association and the CoBTeK research team located in the Nice Memory Center at the Institut Claude Pompidou have developed MeMo (Memory Motivation), a free multilingual Web app that can be used at home by patients. This Web app has been designed by a multidisciplinary team of neuropsychologists, physicians, engineers, and Web designers. The objective of this study was to assess the effectiveness of employing the MeMo platform with regard to cognitive performance and apathy in patients with mild and major NCDs.

Methods

Web App

MeMo is a Web app that can be used on any Web browser (computer or tablet). On a tablet, the app can be installed on the home screen to look like a native app and can be used offline. Users first create an account, so their progress can be logged in the database and visualized in the form of charts. Clinicians can create professional accounts and follow the game progression of all patients who agree to be added to their professional lists.

The app was developed in HTML5/CSS3 with PHP5/MySQL on backend.gnosis. The platform opened in April 2015. In November 2019, 9654 accounts had been created and 327,838 exercises had been performed. MeMo is accessible at the following URL: https://www.memory-motivation.org/home-2/. It is available in French and English, and an Italian version is in development.

The exercises were designed by neuropsychologists and medical doctors. Every exercise was conceived to train a specific cognitive function in order to allow personalized training according to observed deficits. The training activities proposed in MeMo are divided into two parts. The first part involves memory, which includes the following three activities: “recognition” for visual memory training, “MeMo quiz” for working memory training, and “faces” for associative memory training. The second part involves mental flexibility/attention, which includes the following three activities: “arrows” for processing speed, inhibitory control, and mental flexibility training; “tricky cards” for working memory training; and “jumping squares” for reaction anticipation and inhibitory control training. The exercises were conceived to try to motivate people to play. Based on recommendations published by our research group [22], the exercises were designed to be adapted to the target population (eg, simple rules, clear and simplified graphical user interface, clear feedback on performance, instruction reminders, etc), and they embed game rules to include challenges for the player and allow progression in exercise difficulty according to performance.

After creating an account, users can track the evolution of their performance in the exercises. This also allows therapists to follow the evolution of patient performance over time. Indeed, each exercise has several levels of difficulty. Whenever users obtain the maximum score at one level, they are advanced automatically to the next level.

Study Design and Population

This study aimed to compare the evolution of cognitive performance, as well as the motivational status (presence of apathy) between participants using the Web app (MeMo group) and participants not using it (control group). Participants were recruited at the Nice Memory Center located in the Institut Claude Pompidou. The memory center receives approximatively 1700 patients per year. Consultations are performed by a multidisciplinary team of medical doctors (two psychiatrists, one neurologist, and one geriatrician), five neuropsychologists, and one speech therapist. Inclusion in the MeMo study was proposed by all the staff members. All participants provided informed written consent before starting the study. The study was performed in compliance with the Declaration of Helsinki and was approved by the ethics committee of CCP Sud Ed IV-Montpellier (2016-A00879-42). Individuals were included if they were outpatients consulting with the memory center, were older than 60 years, had a Mini-Mental State Examination (MMSE) [23] score between 16 and 28/30, and met the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition diagnostic criteria for mild or major NCDs. Patients were not included if they were not able to read and write in French, had a hearing or major visual impairment, had a history of premorbid intellectual disability, had already used the MeMo app, and were presently involved in structured memory training activities.

The study included 49 participants. Three of them were included incorrectly, and they have not been included in the population description. Thus, the final sample included 46 participants (20 with mild NCDs and 26 with major NCDs) [24]. In terms of disorder etiology, 32 participants were diagnosed with probable Alzheimer disease and 14 with mixed disorders. Among the 46 participants, 42 had a concomitant pharmacological treatment (memantine or cholinesterase inhibitors). Additionally, 25 had concurrent nonpharmacological interventions, including speech therapy (n=22) and other noncognitive interventions, such as kinesitherapy and musicotherapy (n=8).

Of the 46 patients, 25 were randomized to the MeMo group (11 with mild NCDs and 14 with major NCDs) and 21 were randomized to the control group (9 with mild NCDs and 12 with major NCDs). Patients underwent a neuropsychological and behavioral assessment battery at baseline, week 12 (end of the training period), and week 24 (12 weeks after the end of the training sessions). The assessments included the MMSE and Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE). The IQCODE is widely used as a complementary screening tool. The 26 items of the instrument were proposed in the following two versions: informant version (IQCODEi) and self-completion version (IQCODEs) [25]. Each item is rated on a 5-point scale from 1 (much better) to 5 (much worse), and the ratings are averaged over the 26 items to give a score from 1 to 5, with 3 representing no change in any item. Neuropsychological assessments included the Free and Cued Selective Reminding Test [26] for episodic memory and a set of attention tasks [27], including the Trial Making Test A (TMT A), Stroop test, Digit Symbol Substitution Test (DSST), and Frontal Assessment Battery (FAB) [28]. Behavioral symptoms were assessed using the Neuropsychiatric Inventory (NPI) [29], including the apathy subscale, and the Apathy Inventory (AI; clinician version) for quantitative apathy assessment [30].

Participants randomized to the MeMo group were instructed on how to use the app in the clinic by a researcher involved in the study and were asked to train for 12 weeks. Participants were asked to train regularly, with four sessions of 30 minutes each per week. In every session, participants were asked to perform the following six exercises for 5 minutes each (presented in randomized order): “recognition” for visual memory training; “MeMo quiz” for working memory training; “faces” for associative memory training; “arrows” for processing speed, inhibitory control, and mental flexibility training; “tricky cards” for working memory training; and “jumping squares” for reaction anticipation and inhibitory control training. All exercises started at level 1, and the difficulty was progressively adjusted, so that when participants completed one level, they progressed to the next (during the same session or the following session if time expired). The order of presentation of the exercises was randomized for each session and automatically proposed by the app. During the training, participants could contact the study investigators at any time in case of technical problems or questions about the use of the app. All participants used the French version of the app. During the trial, the medical doctor responsible for the participants had no access to the game results (participants were assigned with anonymous account codes). Neuropsychological assessments at baseline, week 12, and week 24 were performed by a neuropsychologist blinded to the participant group allocation (MeMo and control groups).

Statistical Analysis

Data are presented as mean (SD) for quantitative variables and as frequency and percentage for qualitative variables (sex, education level, diagnosis, and concomitant treatment). Comparisons between the different groups were performed using the Student *t*-test or Wilcoxon-Mann-Whitney test for quantitative variables and the χ^2 or Fisher exact test for qualitative variables. Paired Wilcoxon tests were used to assess the evolution of scores between different times 2 to 2 (baseline, week 12, and week 24) for each diagnosis of cognitive impairment. Mixed models taking into account data from the same patient were used to assess the interaction (time × group). These results were adjusted for age and sex. All scores were used as the dependent variable. In the active MeMo group, an additional analysis was performed according to the instruction that the app had to be used regularly. A *P* value <.05 was considered significant, and 95% CIs were indicated. The analyses were performed using the free software R 3.5.1 (The R Foundation for Statistical Computing, Vienna, Austria).

Results

As presented in Table 1, no significant differences in demographics, diagnosis, and clinical characteristics were found between the control group and MeMo group.

On comparing cognitive and behavioral scores among baseline, week 12, and week 24, there were no significant differences in the control group. In the MeMo group, the TMT A time score significantly increased (66.2 s at baseline to 78.7 s at week 12; *P*=.02), suggesting a slight decrease in participants' attention abilities over the 12 weeks, which is compatible with the fact that participants had neurodegenerative disorders. A slight increase in the total NPI score was also found (8.9 s at baseline to 11.3 s at week 12), but this was almost negligible considering that the NPI score can range from 0 to 144 s. Mixed model analysis for each cognitive and behavioral score indicated no significant interaction between the testing time and group.

In order to better explore the MeMo group, we checked whether participants adhered to the proposed study protocol. Data for 22 participants were retrieved directly from the app. For one participant, the data could not be recovered at week 12 (no update on the tablet), and another participant had an adverse event at week 12. The number of days of use varied from 1 day to 78 days over a maximum period of 80 days. The same variability was observed in the number of games played (mean 527 [SD 661]). Therefore, to better assess the impact of high and low game uses, the population was divided into participants who used MeMo according to the instructions (about once every 2 days; active MeMo group) and those who used it less (nonactive MeMo group). A 40-day threshold was selected because it is close to the number of active days described by the protocol, and moreover, it appropriately splits the participants into two groups (<40 active days for the nonactive MeMo group and 40-78 active days for the active MeMo group). The two groups were not different in terms of age, sex, level of education, presence of concomitant treatments, or results at baseline (NPI, *P*=.45; AI, *P*=.69; and IQCODE, *P*=.39).

Table 2 shows the cognitive and behavioral score changes between baseline and week 12. Significant differences between the active and nonactive MeMo groups were found in two attention tests (TMT A and correct DSST items) and in the AI. Specifically, participants in the nonactive MeMo group showed decreased performance in attention (longer reaction times in the TMT A) and became slightly more apathetic (higher scores in AI) over the 12 weeks as compared with participants in the active MeMo group, who showed stable results over the training duration. Furthermore, participants in the active MeMo group showed improved performance in the DSST over the training, whereas participants in the nonactive MeMo group showed stable results over time.

Mixed analysis (time: baseline, week 12, and week 24 × number of active days) indicated only one significant interaction for the AI score (*P*=.01), with a significant increase in apathy in the nonactive MeMo group at week 12 (adjusted coefficient=1.31, 95% CI 0.56-2.05, *P*=.001) and at week 24 (adjusted coefficient=1.61, 95% CI 0.82-2.41, *P*<.001). The results are illustrated in Figure 1.

Table 3 shows the highest performance levels observed for each MeMo game at week 12. For each game, participants in the active MeMo group achieved a higher level of performance as compared with participants in the nonactive MeMo group (*P*<.02).

Finally, correlations between the evolution of cognitive performance (differences in the scores between baseline and week 12) and the MeMo activities were tested. The only significant correlations were between the score changes in the FAB and the number of activities played (Spearman ρ=0.43, *P*=.047) and number of active minutes (Spearman ρ=0.48, *P*=.02). Additionally, more game play by the MeMo participants (number of active days and number of active minutes) was associated with a higher difference in the FAB scores between posttraining and pretraining performance.

Table 1. Demographics and clinical characteristics of the study participants.

Characteristic	Overall population (n=46)	Control group (n=21)	MeMo[a] group (n=25)	*P* value[b]
Age, mean (SD)	79.4 (6.8)	78.8 (6.6)	79.8 (7.0)	.63
DSM-5[c] diagnosis, n (%)				.94
Major NCDs[d]	26 (56)	12 (57)	14 (56)	
Minor NCDs	20 (43)	9 (42)	11 (44)	
Sex, n (%)				.07
Female	24 (52)	14 (66)	10 (40)	
Male	22 (47)	7 (33)	15 (60)	
Level of education, n (%)				.82
Primary level	10 (21)	4 (19)	6 (24)	
Secondary level	21 (47)	12 (57)	10 (40)	
Superior level	14 (30)	5 (23)	9 (36)	
MMSE[e], mean (SD)	21.4 (3.5)	21.6 (2.7)	21.3 (4.1)	.70
IQCODE[f] patient, mean (SD)	83.8 (5.7)	82.5 (4.8)	84.9 (6.2)	.22
IQCODE caregiver, mean (SD)	93.0 (8.7)	91.3 (8.9)	94.1 (8.6)	.44
FAB[g], mean (SD)	13.0 (2.2)	13.4 (2.0)	12.6 (2.2)	.18
FCSRT[h] immediate recall, mean (SD)	13.4 (4.1)	13.6 (4.6)	13.2 (3.7)	.81
FCSRT delayed recall, mean (SD)	1.3 (2.1)	1.0 (1.1)	1.6 (2.6)	.73
TMT A[i], mean (SD)	65.7 (25.7)	64.7 (24.7)	66.6 (27.0)	.95
STROOP[j], mean (SD)	3.0 (1.0)	2.7 (0.9)	3.3 (1.1)	.049
DSST[k], mean (SD)	32.3 (12.6)	36.3 (13.2)	29.0 (11.3)	.08
NPI[l], mean (SD)	8.2 (8.3)	8.3 (9.4)	8.3 (7.6)	.72
AI[m], mean (SD)	0.7 (1.6)	0.7 (1.3)	0.8 (1.8)	.97

[a] MeMo: Memory Motivation.

[b] Comparisons performed using the Student *t*-test for quantitative variables and the χ^2 or Fisher exact test for qualitative variables.

[c] DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[d] NCDs: neurocognitive disorders.

[e] MMSE: Mini-Mental State Examination.

[f] IQCODE: Informant Questionnaire on Cognitive Decline in the Elderly.

[g] FAB: Frontal Assessment Battery.

[h] FCSRT: Free and Cued Selective Reminding Test.

[i] TMT A: Trail Making Test A; score, time in seconds.

[j] STROOP: Stroop test.

[k] DSST: Digit Symbol Substitution Test.

[l] NPI: Neuropsychiatric Inventory.

[m] AI: Apathy Inventory.

Table 2. Cognitive and behavioral score changes between baseline and week 12 in the active MeMo group and nonactive MeMo group.

Test/scale	Score difference[a]		*P* value[b]
	Nonactive MeMo[c] group (active days <40; n=13)	Active MeMo group (active days ≥40; n=9)	
MMSE[d]	−0.8 (1.8)	−0.7 (2.0)	.84
IQCODE[e] patient	−1.9 (2.8)	−1.2 (4.4)	.46
IQCODE caregiver	0.9 (5.9)	3.8 (2.2)	.39
FAB[f]	−0.2 (2.4)	0.7 (2.4)	.33
FCSRT[g] immediate recall	1.4 (3.0)	1.24.6)	.79
FCSRT delay recall	−0.6 (1.3)	−0.2 (2.1)	.28
TMT A[h]	21.5 (29.9)	0.3 (14.8)	.045
STROOP[i]	−0.1 (1.0)	0.2 (1.5)	.84
DSST[j]	0.5 (7.9)	4.25.1)	.04
NPI[k]	6.8 (9.7)	2.3 (5.1)	.45
AI[l]	1.3 (1.5)	−0.2 (1.1)	.02

[a] Values are presented as mean (SD).

[b] Student *t*-test.

[c] MeMo: Memory Motivation.

[d] MMSE: Mini-Mental State Examination.

[e] IQCODE: Informant Questionnaire on Cognitive Decline in the Elderly.

[f] FAB: Frontal Assessment Battery.

[g] FCSRT: Free and Cued Selective Reminding Test.

[h] TMT A: Trail Making Test A; score, time in seconds.

[i] STROOP: Stroop test.

[j] DSST: Digit Symbol Substitution Test.

[k] NPI: Neuropsychiatric Inventory.

[l] AI: Apathy Inventory.

Figure 1. Apathy Inventory score changes in the active MeMo group and nonactive MeMo group. MeMo: Memory Motivation.

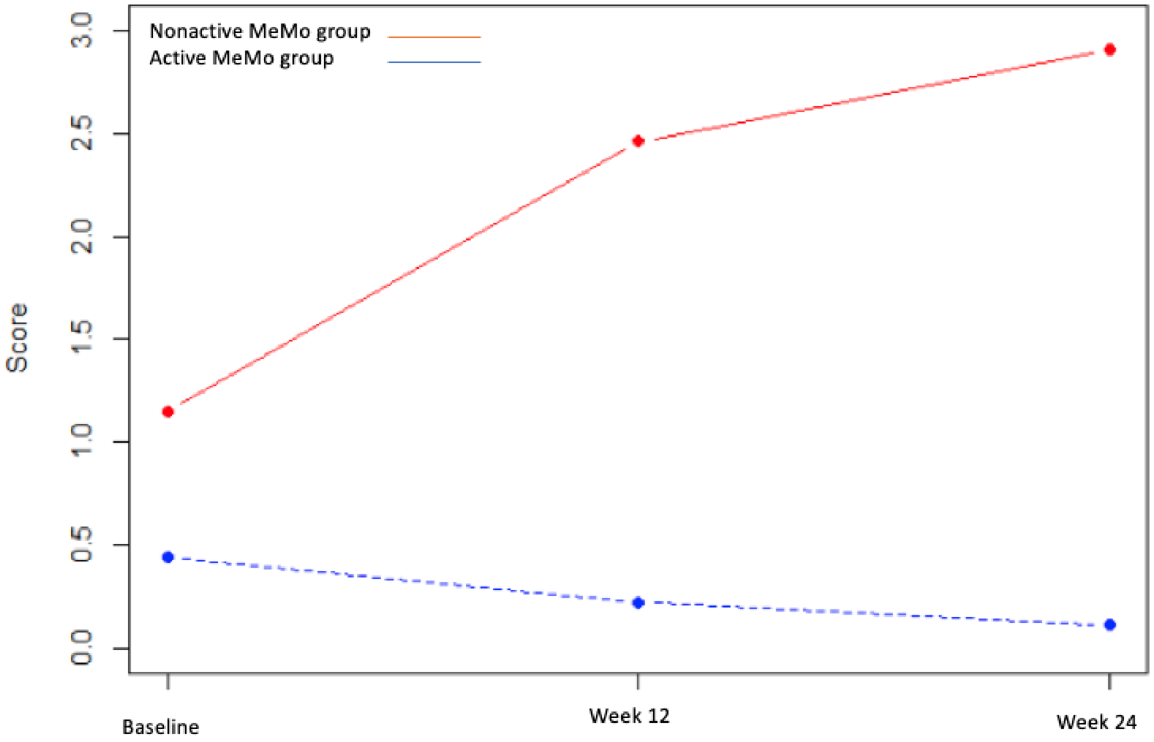


Table 3. Maximum performance score observed for each game in the MeMo app at the end of the study after 12 weeks of use.

Game (maximum possible score)	Maximum performance score[a]		*P* value[b]
	Nonactive MeMo group (active days <40; n=13)	Active MeMo group (active days ≥40; n=9)	
Recognition (10)	3.3 (1.8)	6.2 (2.3)	.007
Quiz (8)	3.3 (2.1)	5.7 (1.7)	.02
Faces (10)	4.2 (2.9)	7.6 (1.1)	.003
Arrows (10)	4.4 (2.6)	8.9 (1.2)	.001
Tricky cards (9)	3.8 (0.7)	5.0 (1.2)	.007
Jumping square (12)	5.1 (2.4)	9.8 (1.5)	<.001

[a] Values are presented as mean (SD).

[b] Wilcoxon-Mann-Whitney test.

Discussion

Cognitive training is regarded as a promising option to slow cognitive decline in the elderly [5] and to improve cognitive function and behavioral symptoms in people with neurodegenerative disorders [7,8]. However, RCTs focused on the efficacy of cognitive training in people with mild and major NCDs having cognitive and behavioral symptoms are still lacking [5]. Furthermore, most existing studies only investigated improvements in cognitive function. Behavioral symptoms, such as apathy (disorder of goal-directed behavior), depression, and agitation, are very common in people with NCDs, and they greatly contribute to increased caregiver burden and reduced patient quality of life [31]. Given that pharmacological treatments have shown limited efficacy in apathy management thus far, there is increasing interest in employing nonpharmacological interventions to reduce apathy [16], including solutions based on new ICTs, such as virtual reality and serious games [14,15]. In this RCT, we tested the efficacy of MeMo, a Web-based app for memory, attention, and mental flexibility training (MeMo group) by comparing it with treatment as usual (control group) in terms of cognitive function and behavioral symptoms, particularly apathy. The results collected from participants with mild and major NCDs showed small but significant differences in attentional and executive function tests and apathy over a 3-month training period. However, these positive effects were only observed for participants who regularly used the app, as indicated in the study protocol (at least 2 days per week; MeMo active subgroup). Specifically, participants in the nonactive MeMo group showed a decline in performance regarding attention and a slight increase in apathy over the 3-month training period, whereas participants in the active MeMo group showed stable results over time. This kind of “dose effect” of cognitive training has been previously observed. In a recent review, Sood et al [32] reported two studies in older subjects suggesting a relationship between the efficacy of the intervention and the play time [33] or the number of gaming sessions over a period of 2 months [34]. Similarly, improvements in attention and executive function have been previously reported following cognitive training. For instance, Sood et al [32] in their recent literature review found that out of 18 studies, eight studies reported a great improvement in attention/working memory. In addition, nine of the reviewed studies assessed depression, a behavioral symptom, as an outcome variable. Four of these studies reported a great reduction in depression among individuals with mild cognitive disorders. Ge et al [7] also reported this noncognitive outcome for depression and anxiety. In this study, we assessed apathy. Apathy is a disorder of motivation defined as a quantitative reduction in goal-directed activity in comparison with the patient’s previous level of functioning [11,35]. Symptoms must persist for at least 4 weeks and affect at least two of the three apathy dimensions (behavior/cognition, emotion, and social interaction). Apathy and depression show some overlap in terms of prevalence and brain circuits, but they can be differentiated [36]. Thus, apathy should be measured as an independent outcome variable. In this study, after 12 weeks, the active MeMo participants were significantly less apathetic as compared with the nonactive MeMo participants (P=.02). This is in line with the findings of preliminary studies performed by our group [15,37] and suggests the interest of employing cognitive training delivered through ICT to increase participants’ motivation to train [22]. As indicated by Booth et al [38], motivation is a core element of the program theory underlying fall prevention interventions in older adults with cognitive impairment. Within the motivation component of this program, two key mechanisms (perceived benefit and support) were shown to influence the extent to which an older adult with cognitive impairment is motivated to undertake an exercise-based intervention. MeMo was designed to increase the intrinsic motivation to keep exercising thanks to several features [22]. First, the game interface was specifically designed for the target population of older adults with cognitive impairment (eg, simplified graphical user interface, simplified instructions with regular reminders, and clear game rules) and was thus easily usable for participants in this study. In addition, the level of difficulty of the exercises was dynamically adapted to the participant’s performance in order to provide an errorless type of training [39] and to keep the participant in a “challenge zone” [20]. Finally, MeMo was tested on a tablet instead of a computer to reduce constraints associated with the use of nonfamiliar technological interfaces, such as a mouse and keyboard, which are often perceived as difficult to use by older adults [20].

Although the exercise design was meant to increase motivation, it should be noted that only 41% (9/22) of the participants used the app as indicated in the guidelines of the research protocol, and the others exercised less. A better training adherence could have been accomplished by scheduling regular in-person phone meetings or discussions with clinicians. This would have probably increased the number of participants that adhered to the protocol guidelines. Indeed, it has been previously suggested that the presence of a clinician is a key factor in determining ICT-based treatment adherence [37]. However, in the RCT, we did not want to add external motivation factors to be able to assess the app efficacy itself. Future studies should test if adding regular meetings with clinicians or support by a caregiver can improve treatment adherence.

Our results are promising; however, the study has several limitations. The major limitation of this study is the small number of participants. Initially, we aimed to include 40 patients in each group considering participant drop out and possible nonexploitability of data. In reality, we managed to include only 46 participants in the time period of the study, although the research was proposed to almost 100 participants who met the inclusion criteria. The main reason for refusal to participate in the trial was concern over the use of new technologies. The mean age of the included participants was 79.4 years. It is reasonable to estimate that in a country like France, the proportion of refusal would have been lower among younger subjects, but this situation will rapidly change. Indeed, people aged over 50 years represent a non-negligible percentage of digital gamers (11% in the United States) according to the 2018 Entertainment Software Association report [40], and this percentage is increasing every year. Playing games is recognized as a source of human enjoyment. Gamified training has been shown to increase motivation, positive mood, and compliance, and therefore, it could substantially drive cognitive benefits [41]. However, the quality of evidence is currently low [6], and the factors boosting motivation to play among older adults are still not completely known. To maximize future brain-gaming tools, it is critical to be able to appreciate what aspects are promising and what aspects are not promising for possible cognitive improvements among older adults with cognitive impairments.

A second limitation of this study is that differences in cognitive performance between active and nonactive MeMo participants were quite small and were limited to the domains of attention and executive functions. Although this is consistent with previous studies [32], it questions the generalizability of the effects of cognitive training to real-life cognitive functions and the general impact on activities of daily living. It should be noted that technology-based games and apps are usually employed in the context of multidomain interventions, and this may increase their benefits in real-life settings [2,3].

In summary, the main result of this study is that the cognitive and behavioral efficacies of a Web-based cognitive training app can be observed only after regular use. In a clinical setting, this points out the importance of reminding about regular use at the moment of prescription and possibly the requirement of regular follow-ups to check prescription adherence.

Acknowledgments

This work was supported by a grant from the Association France Alzheimer. Additional support was received from the Association IA, JL Noisiez Foundation, and Edmond & Lily Safra Foundation. We thank all the patients who participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2822 KB - jmir_v22i3e17167_app1.pdf]

References

1. Cummings J. Cognitive and behavioral heterogeneity in Alzheimer's disease: seeking the neurobiological basis. Neurobiol Aging 2000;21(6):845-861. [doi: 10.1016/s0197-4580(00)00183-4] [Medline: 11124429]
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. The Lancet 2015 Jun;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5]
3. Rosenberg A, Mangialasche F, Ngandu T, Solomon A, Kivipelto M. Multidomain Interventions to Prevent Cognitive Impairment, Alzheimer's Disease, and Dementia: From FINGER to World-Wide FINGERS. J Prev Alzheimers Dis 2020;7(1):29-36. [doi: 10.14283/jpad.2019.41] [Medline: 32010923]
4. Clare L, Woods RT. Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: A review. Neuropsychological Rehabilitation 2004 Sep;14(4):385-401. [doi: 10.1080/09602010443000074]
5. Butler M, McCreedy E, Nelson VA, Desai P, Ratner E, Fink HA, et al. Does Cognitive Training Prevent Cognitive Decline? Ann Intern Med 2017 Dec 19;168(1):63. [doi: 10.7326/m17-1531]
6. World Health Organization. Global Action Plan on the Public Health Response to Dementia, 2017-2025. Geneva: World Health Organization; 2017.
7. Ge S, Zhu Z, Wu B, McConnell ES. Technology-based cognitive training and rehabilitation interventions for individuals with mild cognitive impairment: a systematic review. BMC Geriatr 2018 Sep 15;18(1):213 [FREE Full text] [doi: 10.1186/s12877-018-0893-1] [Medline: 30219036]
8. Hill NT, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized Cognitive Training in Older Adults With Mild Cognitive Impairment or Dementia: A Systematic Review and Meta-Analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
9. van Dalen JW, van Wanrooij LL, Moll van Charante EP, Brayne C, van Gool WA, Richard E. Association of Apathy With Risk of Incident Dementia: A Systematic Review and Meta-analysis. JAMA Psychiatry 2018 Oct 01;75(10):1012-1021 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.1877] [Medline: 30027214]
10. Levy R, Dubois B. Apathy and the functional anatomy of the prefrontal cortex-basal ganglia circuits. Cereb Cortex 2006 Jul;16(7):916-928. [doi: 10.1093/cercor/bhj043] [Medline: 16207933]
11. Robert P, Lanctôt KL, Agüera-Ortiz L, Aalten P, Bremond F, Defrancesco M, et al. Is it time to revise the diagnostic criteria for apathy in brain disorders? The 2018 international consensus group. Eur Psychiatry 2018 Oct;54:71-76 [FREE Full text] [doi: 10.1016/j.eurpsy.2018.07.008] [Medline: 30125783]
12. Husain M, Roiser JP. Neuroscience of apathy and anhedonia: a transdiagnostic approach. Nat Rev Neurosci 2018 Aug 26;19(8):470-484. [doi: 10.1038/s41583-018-0029-9] [Medline: 29946157]
13. Ruthirakuhan M, Herrmann N, Vieira D, Gallagher D, Lanctôt KL. The Roles of Apathy and Depression in Predicting Alzheimer Disease: A Longitudinal Analysis in Older Adults With Mild Cognitive Impairment. Am J Geriatr Psychiatry 2019 Aug;27(8):873-882 [FREE Full text] [doi: 10.1016/j.jagp.2019.02.003] [Medline: 30910421]
14. Benoit M, Guerchouche R, Petit PD, Chapoulie E, Manera V, Chaurasia G, et al. Is it possible to use highly realistic virtual reality in the elderly? A feasibility study with image-based rendering. Neuropsychiatr Dis Treat 2015;11:557-563 [FREE Full text] [doi: 10.2147/NDT.S73179] [Medline: 25834437]
15. Manera V, Petit P, Derreumaux A, Orvieto I, Romagnoli M, Lyttle G, et al. 'Kitchen and cooking,' a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study. Front Aging Neurosci 2015 Mar 17;7:24 [FREE Full text] [doi: 10.3389/fnagi.2015.00024] [Medline: 25852542]
16. Manera V, Abrahams S, Agüera-Ortiz L, Bremond F, David R, Fairchild K, et al. Recommendations for the Nonpharmacological Treatment of Apathy in Brain Disorders. Am J Geriatr Psychiatry 2019 Aug 09. [doi: 10.1016/j.jagp.2019.07.014] [Medline: 31495772]

17. Aalbers T, Baars MA, Olde Rikkert MG, Kessels RP. Puzzling with online games (BAM-COG): reliability, validity, and feasibility of an online self-monitor for cognitive performance in aging adults. J Med Internet Res 2013 Dec 03;15(12):e270 [FREE Full text] [doi: 10.2196/jmir.2860] [Medline: 24300212]
18. König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement (Amst) 2015 Mar;1(1):112-124 [FREE Full text] [doi: 10.1016/j.dadm.2014.11.012] [Medline: 27239498]
19. Manera V, Ben-Sadoun G, Aalbers T, Agopyan H, Askenazy F, Benoit M, et al. Recommendations for the Use of Serious Games in Neurodegenerative Disorders: 2016 Delphi Panel. Front Psychol 2017;8:1243 [FREE Full text] [doi: 10.3389/fpsyg.2017.01243] [Medline: 28790945]
20. Robert PH, König A, Amieva H, Andrieu S, Bremond F, Bullock R, et al. Recommendations for the use of Serious Games in people with Alzheimer's Disease, related disorders and frailty. Front Aging Neurosci 2014 Mar 24;6:54 [FREE Full text] [doi: 10.3389/fnagi.2014.00054] [Medline: 24715864]
21. Zyda M. From visual simulation to virtual reality to games. Computer 2005 Sep;38(9):25-32. [doi: 10.1109/mc.2005.297]
22. Ben-Sadoun G, Manera V, Alvarez J, Sacco G, Robert P. Recommendations for the Design of Serious Games in Neurodegenerative Diseases. Front Aging Neurosci 2018 Feb 2;10:13 [FREE Full text] [doi: 10.3389/fnagi.2018.00013] [Medline: 29456501]
23. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". Journal of Psychiatric Research 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6]
24. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC: American Psychiatric Association; 2013.
25. Jorm AF, Jacomb PA. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. Psychol Med 1989 Nov 09;19(4):1015-1022. [doi: 10.1017/s0033291700005742] [Medline: 2594878]
26. Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. Neurology 1988 Jun 01;38(6):900-903. [doi: 10.1212/wnl.38.6.900] [Medline: 3368071]
27. Lezak M. Neuropsychological Assessment, 3rd Ed. Oxford, United Kingdom: Oxford University Press; 1995.
28. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. Neurology 2000 Dec 12;55(11):1621-1626. [doi: 10.1212/wnl.55.11.1621] [Medline: 11113214]
29. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology 1994 Dec 01;44(12):2308-2314. [doi: 10.1212/wnl.44.12.2308] [Medline: 7991117]
30. Robert PH, Clairet S, Benoit M, Koutaich J, Bertogliati C, Tible O, et al. The apathy inventory: assessment of apathy and awareness in Alzheimer's disease, Parkinson's disease and mild cognitive impairment. Int J Geriatr Psychiatry 2002 Dec;17(12):1099-1105. [doi: 10.1002/gps.755] [Medline: 12461757]
31. Isik AT, Soysal P, Solmi M, Veronese N. Bidirectional relationship between caregiver burden and neuropsychiatric symptoms in patients with Alzheimer's disease: A narrative review. Int J Geriatr Psychiatry 2019 Sep 10;34(9):1326-1334. [doi: 10.1002/gps.4965] [Medline: 30198597]
32. Sood P, Kletzel S, Krishnan S, Devos H, Negm A, Hoffecker L, et al. Nonimmersive Brain Gaming for Older Adults With Cognitive Impairment: A Scoping Review. Gerontologist 2019 Nov 16;59(6):e764-e781. [doi: 10.1093/geront/gny164] [Medline: 30605502]
33. Basak C, Boot WR, Voss MW, Kramer AF. Can training in a real-time strategy video game attenuate cognitive decline in older adults? Psychol Aging 2008 Dec;23(4):765-777 [FREE Full text] [doi: 10.1037/a0013494] [Medline: 19140648]
34. Miller KJ, Dye RV, Kim J, Jennings JL, O'Toole E, Wong J, et al. Effect of a computerized brain exercise program on cognitive performance in older adults. Am J Geriatr Psychiatry 2013 Jul;21(7):655-663. [doi: 10.1016/j.jagp.2013.01.077] [Medline: 23602310]
35. Manera V, Fabre R, Stella F, Loureiro JC, Agüera-Ortiz L, López-Álvarez J, et al. A survey on the prevalence of apathy in elderly people referred to specialized memory centers. Int J Geriatr Psychiatry 2019 Oct;34(10):1369-1377. [doi: 10.1002/gps.5125] [Medline: 30993719]
36. Benoit M, Robert PH. Imaging correlates of apathy and depression in Parkinson's disease. J Neurol Sci 2011 Nov 15;310(1-2):58-60. [doi: 10.1016/j.jns.2011.07.006] [Medline: 21831400]
37. Ben-Sadoun G, Sacco G, Manera V, Bourgeois J, König A, Foulon P, et al. Physical and Cognitive Stimulation Using an Exergame in Subjects with Normal Aging, Mild and Moderate Cognitive Impairment. JAD 2016 Aug 08;53(4):1299-1314. [doi: 10.3233/jad-160268]
38. Booth V, Harwood R, Hancox JE, Hood-Moore V, Masud T, Logan P. Motivation as a mechanism underpinning exercise-based falls prevention programmes for older adults with cognitive impairment: a realist review. BMJ Open 2019 Jun 19;9(6):e024982 [FREE Full text] [doi: 10.1136/bmjopen-2018-024982] [Medline: 31221867]
39. Lee GY, Yip CC, Yu EC, Man DW. Evaluation of a computer-assisted errorless learning-based memory training program for patients with early Alzheimer's disease in Hong Kong: a pilot study. Clin Interv Aging 2013;8:623-633 [FREE Full text] [doi: 10.2147/CIA.S45726] [Medline: 23766638]

40. Entertainment Software Association. Entertainment Software Association. 2019. 2019 Essential Facts About the Computer and Video Game Industry URL: https://www.theesa.com/esa-research/2019-essential-facts-about-the-computer-and-video-game-industry/ [accessed 2020-02-17]
41. Mishra J, Anguera JA, Gazzaley A. Video Games for Neuro-Cognitive Optimization. Neuron 2016 Apr 20;90(2):214-218 [FREE Full text] [doi: 10.1016/j.neuron.2016.04.010] [Medline: 27100194]

Abbreviations

AI: Apathy Inventory
DSST: Digit Symbol Substitution Test
FAB: Frontal Assessment Battery
ICT: information and communication technology
IQCODE: Informant Questionnaire on Cognitive Decline in the Elderly
MeMo: Memory Motivation
MMSE: Mini-Mental State Examination
NCD: neurocognitive disorder
NPI: Neuropsychiatric Inventory
RCT: randomized controlled trial
TMT A: Trial Making Test A

Edited by G Eysenbach; submitted 23.11.19; peer-reviewed by D Fulford, A Piau, S Ge; comments to author 16.12.19; revised version received 16.01.20; accepted 04.02.20; published 11.03.20.

Please cite as:
Robert P, Manera V, Derreumaux A, Ferrandez Y Montesino M, Leone E, Fabre R, Bourgeois J
Efficacy of a Web App for Cognitive Training (MeMo) Regarding Cognitive and Behavioral Performance in People With Neurocognitive Disorders: Randomized Controlled Trial
J Med Internet Res 2020;22(3):e17167
URL: https://www.jmir.org/2020/3/e17167
doi:10.2196/17167
PMID:32159519

Review

The Feasibility and Effectiveness of Web-Based Advance Care Planning Programs: Scoping Review

Doris van der Smissen[1], MSc; Anouk Overbeek[1,2], PhD; Sandra van Dulmen[3,4,5], PhD; Lisette van Gemert-Pijnen[6], PhD; Agnes van der Heide[1], MD, PhD; Judith AC Rietjens[1], PhD; Ida J Korfage[1], PhD

[1]Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands
[2]Department of Criminal Law, Erasmus University Rotterdam, Rotterdam, Netherlands
[3]Nivel (Netherlands institute for health services research), Utrecht, Netherlands
[4]Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, Netherlands
[5]Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway
[6]Centre for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

Corresponding Author:
Doris van der Smissen, MSc
Department of Public Health
Erasmus MC, University Medical Center Rotterdam
P O Box 2040
Rotterdam, 3000 CA
Netherlands
Phone: 31 107038470
Email: d.vandersmissen@erasmusmc.nl

Abstract

Background: Advance care planning (ACP) is a process with the overall aim to enhance care in concordance with patients' preferences. Key elements of ACP are to enable persons to define goals and preferences for future medical treatment and care, to discuss these with family and health care professionals, and to document and review these if appropriate. ACP is usually conducted through personal conversations between a health care professional, a patient, and—if appropriate—family members. Although Web-based ACP programs have the potential to support patients in ACP, their effectiveness is unknown.

Objective: This study aimed to assess the feasibility and effectiveness of Web-based, interactive, and person-centered ACP programs.

Methods: We systematically searched for quantitative and qualitative studies evaluating Web-based, interactive, and person-centered ACP programs in seven databases including EMBASE, Web of Science, Cochrane Central and Google Scholar. Data on the characteristics of the ACP programs' content (using a predefined list of 10 key elements of ACP), feasibility, and effectiveness were extracted using a predesigned form.

Results: Of 3434 titles and abstracts, 27 studies met the inclusion criteria, evaluating 11 Web-based ACP programs—10 were developed in the United States and one in Ireland. Studied populations ranged from healthy adults to patients with serious conditions. Programs typically contained the exploration of goals and values (8 programs), exploration of preferences for treatment and care (11 programs), guidance for communication about these preferences with health care professionals or relatives (10 programs), and the possibility to generate a document in which preferences can be recorded (8 programs). Reportedly, participants were satisfied with the ACP programs (11/11 studies), considering them as easy to use (8/8 studies) and not burdensome (7/8 studies). Designs of 13 studies allowed evaluating the effectiveness of five programs. They showed that ACP programs significantly increased ACP knowledge (8/8 studies), improved communication between patients and their relatives or health care professionals (6/6 studies), increased ACP documentation (6/6 studies), and improved concordance between care as preferred by the patients and the decisions of clinicians and health care representatives (2/3 studies).

Conclusions: Web-based, interactive, and person-centered ACP programs were mainly developed and evaluated in the United States. They contained the key elements of ACP, such as discussing and documenting goals and preferences for future care. As participants considered programs as easy to use and not burdensome, they appeared to be feasible. Among the 13 studies that measured the effectiveness of programs, improvement in ACP knowledge, communication, and documentation was reported.

The concordance between preferred and received care is yet understudied. Studies with high-quality study designs in different health care settings are warranted to further establish the feasibility and effectiveness of Web-based ACP programs.

(J Med Internet Res 2020;22(3):e15578) doi:10.2196/15578

KEYWORDS

advance care planning; Web-based intervention; decision aids; patient education; eHealth; health communication; patient-centered care; chronic disease

Introduction

Background

Contemporary conceptualization of advance care planning (ACP) defines ACP as a process that enables persons to define goals and preferences for future medical treatment and care and to discuss these with family and health care professionals [1]. Furthermore, persons may record and review these preferences if appropriate [1]; therefore, preferences can be accessed when these persons are not able to speak for themselves. The overall aim of ACP is to improve concordance between preferred and received care. ACP may be useful in any stage of life but becomes more targeted when a person's health condition worsens or when age increases [1]. ACP is usually conducted through a structured, personal conversation between a health care professional, a patient, and—if appropriate—family members. Since the 1990s, evidence has amassed, showing that ACP interventions have potentially beneficial outcomes for patients and health care systems. These include increased completion of advance directives (ADs): documents in which preferences for future medical treatment and care can be recorded [2,3]. Furthermore, these beneficial outcomes include better alignment of care to expressed preferences, better quality of communication in clinical consultations, improved quality of life, reduction of unwanted hospital admissions, and increased use of palliative care [2,3].

Health care professionals and patients generally underline the importance of ACP [4]. Given that the number of people with chronic conditions is increasing [5] and that ACP can be relevant in the early stages of disease, ACP will become relevant for a growing number of people. Nevertheless, the implementation of ACP in practice faces several challenges [4,6,7]. The (facilitated) ACP process takes time, and supporting patients in this process is, therefore, costly [4,6]. This limits the upscaling of and accessibility to the ACP process. Engagement in ACP is further limited by the delay in its initiation because of the barriers experienced by health care professionals and patients: health care professionals report concerns about taking away patients' hope and uncertainty about timing of ACP, whereas patients expect health care professionals to initiate ACP [4,6]. Further barriers to engagement in ACP are physicians' lack of training in having ACP conversations and lack of continuity of care [4,6]. Furthermore, people with chronic diseases may not have the time and energy for face-to-face conversations, for example, because of treatment burden, even if these conversations would help them [8-10]. Still, there are patients and healthy individuals who experience a clear need to engage in ACP. For instance, 398 of 502 (79.3%) Belgian citizens aged 64 years and older indicated to be willing to take the initiative to start the ACP process, for example, by completion of an AD [11].

One way of overcoming the barriers to wider implementation of ACP may be Web-based ACP programs [12-14]. They can be accessed on the Web at any preferred time, have the potential to reach a larger audience, are relatively easy to implement, and are scalable. Moreover, a Web-based format of ACP may be an addition to the ACP process as facilitated by professionals, as it can be delivered stepwise and tailored and can include interactive elements and videos. Web-based ACP programs should not replace discussions with health care professionals or with ACP facilitators, but they may support patients to prepare these discussions and to consider their values, beliefs, and care preferences in their own time and environment. Ample research in other domains has shown that Web-based health programs can be effective in improving health outcomes such as physical activity [15], patient empowerment [15,16], and depression [17]. They have the potential to be cost-effective [18]. Patients perceive Web-based health programs usually to be useful and helpful [19]. Therefore, Web-based ACP programs may have the potential to support patients in ACP.

Several reviews described person-centered tools, including decision aids targeted at adult patients and their relatives as well as healthy individuals. These tools are related to ACP, shared decision making, and end-of-life care. Some studies in these reviews also included Web-based or computerized programs [14,20-24]. None of these reviews focused specifically on empirically evaluated, Web-based, and available programs for ACP and their feasibility and effectiveness. In addition, none of the reviews focused on interactive programs, which guide users through the information and in which users are enabled to interact with the information. Only Butler et al [14] focused specifically on ACP decision aids. Most of the reviews (except for the study by Butler et al [14]) focused on specific populations instead of providing an overview of available ACP programs for the general population as well as for patients.

Scope of This Review

The overall aim of this scoping review was to assess the feasibility and effectiveness of Web-based, interactive, and person-centered ACP programs. This review focuses on the following research questions: (1) What are the functionalities of Web-based ACP programs?, (2) What is the content of Web-based ACP programs?, (3) How feasible are Web-based ACP programs?, and (4) How effective are Web-based ACP programs?

Methods

Methodological Framework

Scoping studies “aim to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available and can be undertaken as stand-alone projects in their own right, especially where an area is complex or has not been reviewed comprehensively before” [25]. Scoping reviews can be used to explore the literature within a research area of interest by addressing broad research questions. This exploration can be done regardless of the methodological quality of the studies or risk of bias [26,27]. We used a systematic approach, namely, the methodological framework for scoping reviews by Arksey and O’Malley [26]. The five stages of the framework for scoping reviews are (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results [26].

Search Strategy

The search strategy was developed in collaboration with a medical librarian. We systematically searched for empirical studies written in the English language that evaluated Web-based, interactive, and person-centered ACP programs. We searched in EMBASE on July 24, 2017, and in Medical Literature Analysis and Retrieval System Online Epub (MEDLINE Epub [Ovid]), Web of Science, Cochrane Central, PsycINFO (Ovid), Cumulative Index to Nursing and Allied Health Literature (CINAHL [EBSCO]), and Google Scholar on July 28, 2017, and updated this search on April 16, 2019. Multimedia Appendix 1 presents the search strategy.

Study Selection

Duplicates were removed. Two reviewers (DS and AO) independently screened titles, abstracts, and full text of articles to identify relevant studies, assisted by the program Covidence (operated by Veritas Health Innovation Ltd) [28]. Articles were included when they fulfilled the inclusion criteria, as presented in Textbox 1. In addition, we handsearched the references of included articles and other possibly relevant articles. When DS and AO could not reach consensus about inclusion or exclusion, other authors were consulted (IK and JR). Disagreements were readily resolved.

Textbox 1. Inclusion criteria for the full-text papers.

1. The study has an original empirical quantitative or qualitative research design. Reviews and conference abstracts were excluded.

2. The study evaluates a program that:

- supports the completion of one or more elements of advance care planning (ACP), defined as enabling persons to define, discuss, record, and review goals and preferences for future medical treatment and care [1];
- is accessible and available on the internet;
- is interactive, defined as guiding users through the ACP process in which they are enabled to interact with information instead of only reading text; and
- is person centered, defined as being targeted at adult patients, relatives, and/or healthy individuals in general rather than solely at clinicians or medical students.

3. Language of the publication should be English.

Data Extraction and Outcomes of Interest

Data extraction was performed by DS and AO using a predesigned form. Data were extracted from the ACP programs and from the studies evaluating the ACP programs.

Advance Care Planning Programs

Functionalities of the ACP programs were extracted based on the Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth (CONSORT-EHEALTH) checklist [29], which is developed to ensure that electronic health (eHealth) interventions in randomized controlled trials (RCTs) are reported in sufficient detail for replication. We extracted the programs’ target group and accessibility, for example, whether it was possible to access the program without registration. Furthermore, we extracted whether the programs were free of charge; were tailored to the users’ information needs; provided feedback on responses; showed progress information; had the possibility of giving input, for example, to answer questions; contained hyperlinks to (external) Web pages; contained a text-to-speech option; contained videos; could be used without assistance; addressed the privacy policy; and addressed log data analysis (tracking behavior of users in a Web-based program).

The content of the ACP programs was extracted based on the European Association for Palliative Care (EAPC) consensus definition of ACP [1]. In this review, the EAPC ACP task force defined 12 key elements of ACP, which we summarized into 10 elements, such as providing information about ACP, addressing the readiness/timing for ACP, addressing exploration of values and goals, and addressing recording of ACP and ACP communication [1].

Advance Care Planning Studies

The following study characteristics were extracted: first author and year, country, participants and setting, study design, intervention and outcome measures, and results of the studies on feasibility and effectiveness of ACP programs.

The feasibility of the ACP programs was extracted based on the framework of Bowen et al [30]. We extracted the acceptability of the burden of the program, ease of use, understandability of the text in the program, and the acceptability of the program. To briefly address the implementation of the programs, we extracted whether further developments or research of the programs were described. Furthermore, we extracted outcomes as recommended by the CONSORT-EHEALTH checklist [29], namely, participation rates among the contacted participants, completion rates of those who provided consent completing the entire program, whether the use of log data of users was described, and whether user feedback was obtained.

To report on the effectiveness of ACP programs, we used the outcome measures of ACP that were recommended by the EAPC ACP task force, such as ACP knowledge; self-efficacy; identification of goals, values, and preferences; helpfulness in ACP (for making decisions); health care use; and decision concordance between the patients' preferences and health care professionals' decisions [1].

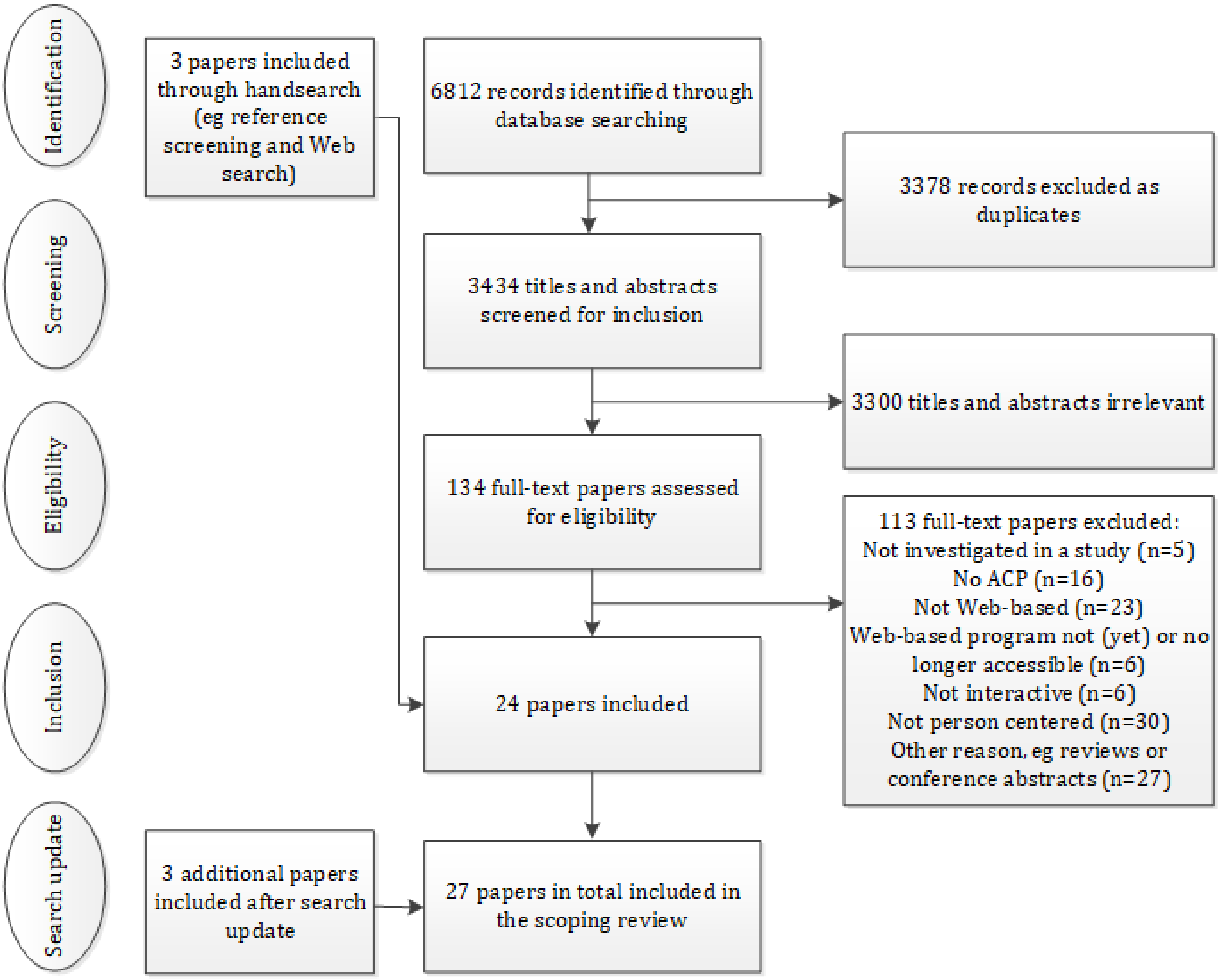
Results

Inclusion of Papers

The search resulted in 6812 records (see Figure 1). After removing duplicates, 3434 titles and abstracts remained. On the basis of the inclusion criteria, 3300 titles and abstracts were found to be irrelevant and were excluded. Next, 134 studies were screened full text, of which 113 studies were excluded (see Figure 1 for details on exclusion). Twenty-one studies were identified as relevant. Handsearch of systematic reviews and potentially relevant other references resulted in the inclusion of three further studies. Overall, in 2017, 24 studies were included for data extraction. On the basis of their initial independent scoring, DS and AO had an agreement for 110 of the 134 full texts (82.1%) about inclusion or exclusion. The interrater reliability is considered moderate (kappa=0.52). Disagreements about inclusion or exclusion were readily resolved, and it was seldom necessary to consult other authors.

In 2019, 983 new references were identified, of which 36 were screened for full-text review. Three articles were included, which analyzed two programs that were already described in this review. This resulted in a total number of 27 included articles.

Figure 1. Flowchart of the inclusion of papers.



Functionalities of the Web-Based Advance Care Planning Programs

Table 1 presents an overview of the functionalities of the Web-based ACP programs, and Multimedia Appendix 2 presents the links to the Web-based programs. The 27 included studies evaluated 11 Web-based programs—8 programs were targeted at patients or healthy individuals, two programs were targeted at patients and relatives, and one program was targeted at patients and health care professionals. Related to the accessibility of the Web-based programs, we found that six programs were accessible without registration, and 10 programs were free of charge. Ten programs could be used without assistance of a health care professional. Related to the interactivity, we found that all programs provided the possibility for the users to give input, for example, by answering questions. Eight programs included an indicator of the progress of users in completing the program. Seven programs could tailor to users’ information needs by providing additional information if preferred, and six programs contained hyperlinks to (external) Web pages. Three of the programs provided (specific) feedback on responses of users, for example, by giving a personalized response to answered questions in the program. Related to the layout, eight programs contained videos, and two had a text-to-speech option to play text in audio. Finally, eight programs described their privacy policy in the program and reported that they analyzed log data of users. The program PREPARE For Your Care (PREPARE) addresses 11 of 12 functionalities, and the program Making Your Wishes Known addresses 10 functionalities. The programs Death over Dinner, MyDirectives, and Think Ahead address nine functionalities each, and all other programs address five to eight functionalities.

Table 1. Functionalities of the Web-based advance care planning programs.

Programs	Target group	Accessible (without registration)	Free of charge	Can be used without assistance	Possibility to give input (eg, answer questions)	Shows progress information	Tailored to users' information needs	Contains hyperlinks to (external) Web pages	Provides feedback on responses	Contains videos	Text-to-speech option	Privacy policy addressed	Log data analysis addressed
ACP Decisions [31]	P[a] and H[b]	x[c]	✓[d]	x	✓	x	x	x	x	✓	x	✓	✓
Death over Dinner [32]	P and R[e]	✓	✓	✓	✓	✓	✓	✓	✓	✓	x	x	x
Five Wishes [33]	P	x	x	✓	✓	✓	x	✓	x	✓	x	✓	✓
Making Your Wishes Known [34-45]	P	x	✓	✓	✓	✓	✓	x	✓	✓	✓	✓	✓
MyDirectives [38,46]	P	x	✓	✓	✓	✓	✓	✓	x	✓	x	✓	✓
MyICUGuide [47]	P and R	✓	✓	✓	✓	✓	x	x	x	x	x	✓	✓
NVLivingWill [48]	P	x	✓	✓	✓	✓	✓	✓	x	x	x	x	x
Plan Your Lifespan [49]	P	✓	✓	✓	✓	✓	✓	✓	x	✓	x	x	x
PREPARE For Your Care [38,50-55]	P	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	✓	✓
The Letter Project Advance Directive [56]	P	✓	✓	✓	✓	x	x	x	x	x	x	✓	✓
Think Ahead [57]	P	✓	✓	✓	✓	x	✓	✓	x	✓	x	✓	✓
Total[f]	-	6	10	10	11	8	7	6	3	8	2	8	8

[a] P: patients.

[b] H: health care professionals.

[c] x: not addressed in the program.

[d] ✓: addressed in the program.

[e] R: relatives.

[f] Total number of programs that addressed the functionalities.

Content of the Web-Based Advance Care Planning Programs

Table 2 presents an overview of the content of the Web-based ACP programs, and Multimedia Appendix 2 presents the links to the Web-based programs. Target groups, for example, patients, were involved in the development of seven programs [31,34-45,47,49-57]. Four programs had a theory base [34-45,47,50,57], for example, Making Your Wishes Known was based on the Multi-Attribute Utility Theory, and PREPARE was based on behavior change theories. Related to the key elements for ACP, we found that almost all programs provided information about ACP (10 programs) and included attention for readiness for ACP or for adequate timing of ACP (10 programs). Furthermore, the exploration of goals and values for future treatment and care was addressed by eight programs. In all programs, attention was paid to treatment and care options and treatment and care preferences. Furthermore, all programs addressed the potential appointment of a health care representative (ie, someone who can make decisions on behalf of the patient when he or she is unable to do so) and paid attention to the recording of ACP: eight programs included the possibility to generate a document in which patients can record their goals, values, and preferences. In nine programs, users were encouraged to share this document with their relatives or health care professionals. Ten programs addressed how to communicate preferences with health care professionals or with relatives.

Table 2. Inclusion of the recommended key elements for advance care planning in the Web-based advance care planning programs.

Programs	Provides information about ACP[a]	Addresses readiness/timing for ACP	Addresses exploration of values/goals	Addresses treatment and care options	Addresses treatment and care preferences	Addresses appointment of a health care representative	Addresses recording of ACP	Generates document	Encourages to share the document	Addresses ACP communication
ACP Decisions [31]	✓[b]	✓	✓	✓	✓	✓	✓	x[c]	x	✓
Death over Dinner [32]	✓	x	x	✓	✓	✓	✓	x	✓	✓
Five Wishes [33]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Making Your Wishes Known [34-45]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MyDirectives [38,46]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MyICUGuide [47]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NVLivingWill [48]	✓	✓	x	✓	✓	✓	✓	✓	✓	✓
Plan Your Lifespan [49]	✓	✓	x	✓	✓	✓	✓	✓	x	✓
PREPARE For Your Care [38,50-55]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
The Letter Project Advance Directive [56]	x	✓	✓	✓	✓	✓	✓	✓	✓	x
Think Ahead [57]	✓	✓	✓	✓	✓	✓	✓	x	✓	✓
Total[d]	10	10	8	11	11	11	11	8	9	10

[a] ACP: advance care planning.

[b] ✓: addressed in the program.

[c] x: not addressed in the program.

[d] Total number of programs that addressed the elements.

Evaluation of the Web-Based Advance Care Planning Programs

Most programs were evaluated in one study [31-33,47-49,56,57], two studies evaluated MyDirectives [38,46], seven studies evaluated PREPARE [38,50-55], and 12 studies evaluated Making Your Wishes Known in [34-45]. Multimedia Appendix 3 presents an overview of the characteristics of the studies. All programs were developed in the United States, except for Think Ahead, which was developed in Ireland [57]. All studies were published in the period from 2007 to 2018, of which nine studies were published in 2017. In total, 25 of the studies have a quantitative design [31,33-49,51-57], and two study designs are qualitative [32,50]. Nine studies allowed comparison of outcomes before/after an intervention [31,35,40,42,43,45,53-55], and eight studies allowed comparison between intervention and control groups [31,39,45,49,52,54-56]. Studied populations ranged from healthy adults to patients with serious conditions. The sample sizes of the quantitative studies ranged from 17 to 3119, and participation rates ranged from 14% to 100%. The use of validated measures, if applicable, is indicated in Multimedia Appendix 3.

Feasibility of the Web-Based Advance Care Planning Programs

Multimedia Appendix 4 presents an overview of the feasibility of the Web-based ACP programs of the 25 quantitative studies. The participation rate among contacted participants was over 60% in six studies and ranged from 14% to 58% in 12 studies. Seven studies did not report on participation rates. The completion rate considering the entire program ranged from 31% to 72% in five studies and ranged from 83% to 100% in the other 17 studies. Three studies did not report on completion rates. One paper used log data analysis [58] to assess feasibility, seven studies obtained (qualitative) user feedback, and 12 studies described further developments of the program or planned future research on the program, which may indicate further implementation/continued use of the programs.

Thirteen of the 25 quantitative studies evaluated one or more of the four predefined elements of feasibility [30] in 6 of the 11 programs: acceptability of the burden of the program (8 studies), ease of use (8 studies), understandability of the text (4 studies), and acceptability of the program (2 studies). With the exception of one study with mixed results [57], outcomes indicated that users found the burden acceptable, the program easy to use, and the text understandable. The program was found acceptable in one study [47].

In one qualitative study, participants reported ease of use and understandability of the text as well as barriers because of confusing layout and emotive language [50]. However, the authors concluded that the program was acceptable, applicable, and understandable.

Outcomes of the Studies on Web-Based Advance Care Planning Programs

Multimedia Appendix 3 presents an overview of the outcomes of the quantitative and qualitative studies. The 25 quantitative studies reported on evaluations of ACP as recommended by the EAPC ACP task force [1], such as the identification of goals, values, and preferences (18 studies) or documentation of preferences in a PDF output document or an AD (18 studies). Often, these aspects were part of the ACP program. In approximately half of the studies, ACP communication (13 studies), satisfaction with the program (11 studies), and ACP helpfulness (11 studies) were evaluated. Less than half of the studies evaluated ACP knowledge (11 studies) and quality of ACP/accuracy in reflecting wishes (7 studies). Few studies evaluated ACP readiness (6 studies), self-efficacy (6 studies), ACP revision over time (4 studies), (decision) concordance between the patients' preferences and the health care professionals' decisions (2 studies) or the health care representatives' decisions (1 study), and health care use (1 study).

The research designs of 13 of the 25 quantitative studies allowed for the determination of the effectiveness of Web-based ACP programs using an RCT design and/or before and after designs. Eight studies applied an RCT design [31,39,45,49,52,54-56], and nine studies compared follow-up results of the intervention group with baseline (before and after design) [31,34,40,42,43,45,53-55] (see Multimedia Appendix 3). These studies evaluated the effectiveness of five Web-based ACP programs. Outcomes of these 13 quantitative studies indicate significantly increased ACP knowledge (8/8 studies); ACP communication (6/6 studies); ACP documentation (6/6 studies); identification of goals, values, and preferences (4/4 studies); self-efficacy (4/5 studies); and ACP readiness (4/5 studies). These outcomes are visualized in Table 3 and Figure 2 [31,34,39,40,42,43,45,49,52-56]. The remaining 12 quantitative studies were cross-sectional.

The extent to which programs were evaluated differed. For example, the programs ACP Decisions, MyICUGuide, and NVLivingWill were each evaluated considering one of the predefined outcome measures [1] in one study, whereas Making Your Wishes Known was evaluated in 12 studies in different settings considering 10 outcome measures, and PREPARE was evaluated in seven studies in different settings considering eight outcome measures.

The two qualitative studies indicated that participants were satisfied with the programs, which helped them to communicate about ACP [32,50]. In one of these qualitative studies, which evaluated the program PREPARE, participants gained more knowledge about ACP, although the section about values and beliefs was considered less relevant [50].

Figure 2. Results of quantitative studies assessing the effectiveness of Web-based advance care planning programs comparing the intervention group (Web-based advance care planning program) with baseline or control groups (N=13). ACP: advance care planning.

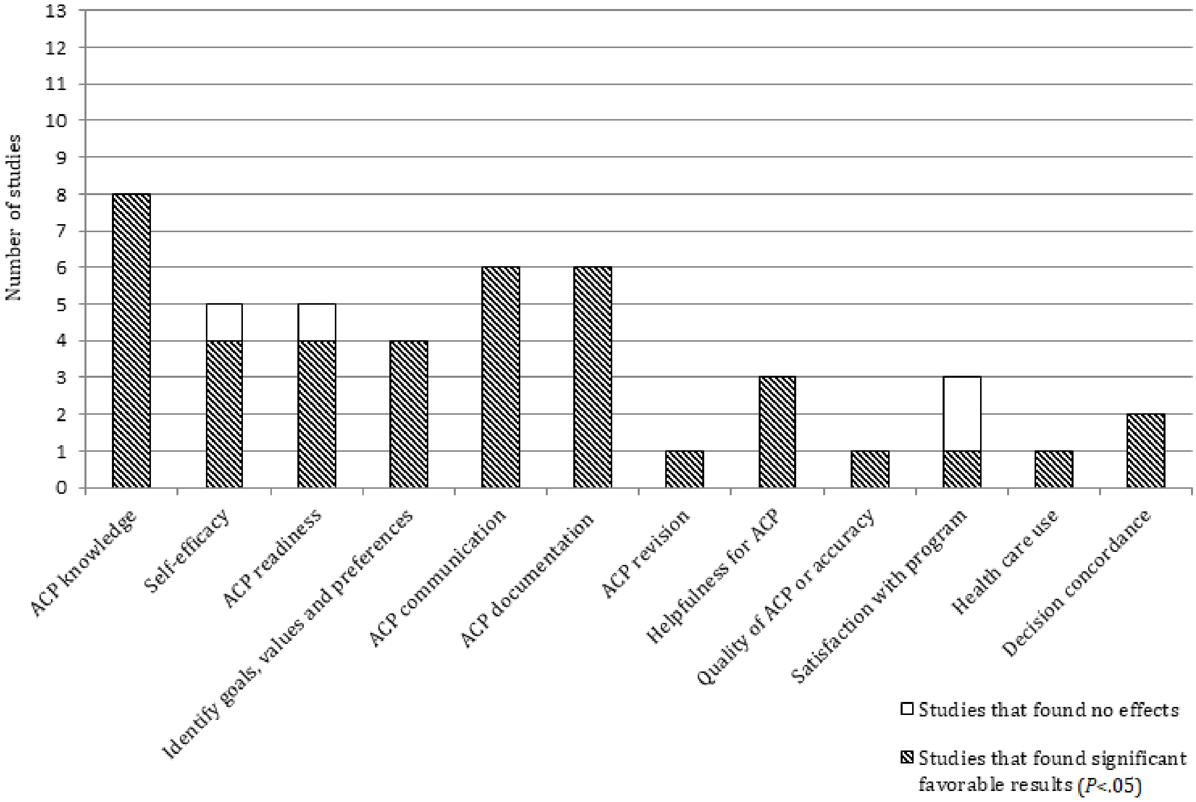


Table 3. Results of quantitative studies assessing the effectiveness of Web-based advance care planning programs comparing the intervention group (Web-based advance care planning program) with baseline and/or control groups.

Studies	Program	ACP[a] knowledge	Self-efficacy	ACP readiness	Identification of goals, values and preferences	ACP communication	ACP documentation	ACP revision	ACP helpfulness (making decisions)	Quality of ACP or accuracy	Satisfaction	Health care use	Decision concordance
Volandes et al [31]	ACP Decisions	—[b]	—	—	—	—	+[c]	—	—	—	—	↓[d]	—
Green and Levi [34]	Making Your Wishes Known	—	—	—	—	—	—	—	—	+	—	—	—
Markham et al [40]	Making Your Wishes Known	+	—	—	—	—	—	—	—	—	—	—	—
Green et al [39]	Making Your Wishes Known	+	—	—	+	+	—	—	+	—	+	—	—
Van Scoy et al [42]	Making Your Wishes Known	+	—	—	—	—	—	—	—	—	—	—	—
Levi et al [43]	Making Your Wishes Known	+	—	—	—	—	—	—	—	—	—	—	+
Green et al [45]	Making Your Wishes Known	—	NS[e]	—	—	—	—	—	—	—	—	—	+
Lindquist et al [49]	Plan Your Lifespan	—	—	NS	—	+	+	+	—	—	—	—	—
Sudore et al [53]	PREPARE For Your Care	+	+	+	+	+	+	—	—	—	—	—	—
Sudore et al [52]	PREPARE For Your Care	+	+	+	+	+	+	—	—	—	NS	—	—
Lum et al [54]	PREPARE For Your Care	+	+	+	—	+	+	—	—	—	—	—	—
Sudore et al [55]	PREPARE For Your Care	+	+	+	—	+	+	—	+	—	NS	—	—
Periyakoil et al [56]	The Letter Project Advance Directive	—	—	—	+	—	—	—	+	—	—	—	—
Total[f]	13	8	5	5	4	6	6	1	3	1	3	1	2

[a] ACP: advance care planning.

[b] —: Not examined by statistically comparing groups.

[c] +: significant increase.

[d] ↓: significant decrease.

[e] NS: effectiveness examined, but no significant differences found.

[f] Total number of studies in which the outcome measures were examined by comparing the intervention group (Web-based advance care planning program) with baseline and/or control groups.

Discussion

Principal Findings

This scoping review provides an overview of Web-based, interactive, and person-centered ACP programs, including their functionalities, content, feasibility, and effectiveness.

This review identified 11—mainly developed in the United States—programs, many of which contain videos, provide tailored information, and can be used without assistance. Most of the programs contain the key elements of ACP [1], such as information about ACP, goals and preferences for future treatment and care, and included the possibility to generate a document in which patients can record their goals, values, and preferences. The extent to which programs contain functionalities such as *text-to-speech* differs. The program PREPARE, for instance, has 11 of such functionalities, whereas ACP Decisions contains five functionalities. Furthermore, the extent to which programs were evaluated differed between studies. For example, the programs ACP Decisions, MyICUGuide, and NVLivingWill were each evaluated in one study, whereas the program Making Your Wishes Known was evaluated in 12 studies in different settings.

Reportedly, programs were easy to use and not burdensome to participants. However, the feasibility of the programs was evaluated in only 13 of 27 studies for six programs, the evaluation was often limited to one or two outcome measures, and the response and completion rates were relatively low for some studies. In general, reportedly, participants were satisfied with the ACP programs. Some outcome measures, such as quality or accuracy of the program in representing wishes, health care use, its concordance with patients’ preferences, and the revision of preferences and documents over time, were less often evaluated. Overall, the studies with RCT or before and after designs comparing the intervention group with baseline or a control group showed that Web-based ACP programs are a promising approach to support patients in ACP by showing significant improvement in ACP knowledge, ACP communication, and ACP documentation.

There seems to be no link between the outcomes and the content of the programs because almost all programs address the key elements of ACP. Although many studies found results in a favorable direction, only the minority of the studies, namely, 13 of 27, use strong research designs in which groups were statistically compared.

Comparison With Prior Research

This review focuses on providing an overview of Web-based, interactive, and person-centered ACP programs that are currently available. Although previous reviews did not have this specific focus, the reviews describe similar content and outcomes for Web-based programs related to ACP as in the this review, such as identification of preferences and treatment options, completion of ADs, the appointment of a health care representative, and they report satisfaction with the programs and increase of knowledge after using the program [14,20,21]. This review found some Web-based programs, which were also identified by the prior reviews, such as Making Your Wishes Known and PREPARE. As Butler et al [14] described, it seems that many Web-based programs in ACP, end-of-life care, and palliative care are available in the gray literature as well. For example, the interactive ACP program *My Decisions* from the United Kingdom [59] and the palliative care communication program *Tell Us* from the United States [60] are available on the internet but seemed not to be investigated in a study (when our search strategy was conducted). Therefore, the evaluation of Web-based ACP programs seems to be a challenge/opportunity for future research.

Strengths and Limitations

This review has several strengths. We used a systematic approach, namely, the methodological framework for scoping reviews by Arksey and O’Malley [26]. The EAPC definition and recommendations for ACP allowed for a structured evaluation of the content and the effectiveness of the programs [1]. The framework of Bowen et al [30] and the CONSORT-EHEALTH checklist [29] allowed for a structured evaluation of the feasibility of the programs. The search was systematically conducted and performed with broad search terms in seven databases. Two reviewers independently screened the titles, abstracts, and full text of articles to select relevant studies.

Some limitations should also be mentioned. Importantly, it should be taken into account that the content and layout of Web-based programs are continually changing. Although the WebCite tool allowed us to archive the websites’ homepages (Multimedia Appendix 2), it is possible that (parts of the) programs have changed in the period between our review of the websites and the publication of this study. Furthermore, we only included Web-based programs which were evaluated in a study.

Recommendations for Future Research

First, as most of the Web-based ACP programs are developed in the United States, we recommend the development of evidence-based, Web-based, interactive, and person-centered ACP programs in countries outside the United States. Ideally, to allow for proper scaling up of ACP, these programs should be tailored to local cultural and legal circumstances. To enhance the quality of Web-based ACP programs, we recommend that these Web-based ACP programs contain all key elements of ACP. Second, several important outcomes of ACP were often not reported. More clarity on which outcome measure to report, and when, would be useful. In addition, support in how to assess key outcome measures, such as concordance between preferred and received care, is needed because this important outcome measure is difficult to measure. Namely, it is not always clearly stated in medical files whether provided treatments had a curative or a palliative intent [61]. When this is not mentioned, it is difficult to determine whether treatments aligned with preferences [61]. Furthermore, it is difficult to establish a baseline measure of patients' goals, and when goals are not documented, the concordance with these goals cannot be evaluated [62]. In addition, patients' preferences may change during the study period, which complicates the use of this measure in practice [62]. In addition, it could be that Web-based ACP affects care in the long run, which further complicates its measurement. We recommend further research into this topic, for instance, by developing a core outcome set for ACP. Third, we recommend evaluating the feasibility of ACP programs. A clear guideline of evaluating feasibility in eHealth tools/Web-based programs is not yet available. Therefore, preferably that evaluation should be based on the framework of Bowen et al [30] and the CONSORT-EHEALTH checklist [29], which indicate important outcome measures. Fourth, we recommend the use of proper research designs, such as RCTs or before and after research designs, allowing for further determination of the effectiveness and feasibility of Web-based ACP programs. Future studies may evaluate how stakeholders other than patients perceive the role of Web-based ACP programs in the health care process, for example, general practitioners. We strongly recommend comparing the feasibility and effectiveness of the programs to ACP by health care professionals or to ACP supported by facilitators because the effectiveness of the programs in health care practice is still unknown. Finally, related to safety and technology, the safety of the generated documents by the Web-based ACP programs is still unknown, and it is also unknown whether Web-based ACP programs can be used among underserved groups who have possibly less access to these technologies, for example, patients with low eHealth or health literacy skills.

Conclusions

This scoping review shows that Web-based, interactive, and person-centered ACP programs are mainly developed and evaluated in the United States. The Web-based programs contained the key elements of ACP, such as discussing and documenting goals and preferences for future care. In general, studies report that Web-based ACP programs tend to be feasible. Only 13 studies measured the programs' effectiveness, and they showed significant improvement in ACP knowledge, communication, and documentation. The key outcome of ACP—concordance between preferred and received treatment and care—is yet understudied. Studies with high-quality study designs in diverse cultural contexts on feasibility and effectiveness are warranted to further establish the effectiveness of important outcomes. Furthermore, it is unknown how programs are used in practice, including attitudes of health care professionals toward Web-based ACP programs.

Overall, we conclude that Web-based, interactive, and person-centered ACP programs are promising to support patients in ACP. Web-based ACP programs may improve accessibility to ACP, allowing people to start with ACP in their own time and environment. Web-based ACP programs may, therefore, help to overcome the time and emotional barriers in the initiation of ACP and to scale up ACP.

Acknowledgments

This study was funded by the Netherlands Organization for Health Research and Development. The authors thank Gerdien de Jonge, Wichor Bramer, Sabrina Gunput, and Elise Krabbendam (biomedical information specialists, Medical Library, Erasmus MC) for their support in literature searching.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy scoping review.
[DOCX File , 19 KB - jmir_v22i3e15578_app1.docx]

Multimedia Appendix 2
Links to the Web-based advance care planning programs.
[DOCX File , 16 KB - jmir_v22i3e15578_app2.docx]

Multimedia Appendix 3
Data extraction of the 27 included studies about Web-based advance care planning programs, alphabetical order of the program names and the year the article was published.
[DOCX File , 34 KB - jmir_v22i3e15578_app3.docx]

Multimedia Appendix 4
Feasibility of the Web-based advance care planning programs of the quantitative studies.

[DOCX File , 20 KB - jmir_v22i3e15578_app4.docx]

References

1. Rietjens JA, Sudore RL, Connolly M, van Delden JJ, Drickamer MA, Droger M, European Association for Palliative Care. Definition and recommendations for advance care planning: an international consensus supported by the European Association for Palliative Care. Lancet Oncol 2017 Sep;18(9):e543-e551. [doi: 10.1016/S1470-2045(17)30582-X] [Medline: 28884703]
2. Brinkman-Stoppelenburg A, Rietjens JA, van der Heide A. The effects of advance care planning on end-of-life care: a systematic review. Palliat Med 2014 Sep;28(8):1000-1025. [doi: 10.1177/0269216314526272] [Medline: 24651708]
3. Houben CH, Spruit MA, Groenen MT, Wouters EF, Janssen DJ. Efficacy of advance care planning: a systematic review and meta-analysis. J Am Med Dir Assoc 2014 Jul;15(7):477-489. [doi: 10.1016/j.jamda.2014.01.008] [Medline: 24598477]
4. Jimenez G, Tan WS, Virk AK, Low CK, Car J, Ho AH. Overview of systematic reviews of advance care planning: summary of evidence and global lessons. J Pain Symptom Manage 2018 Sep;56(3):436-59.e25 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.05.016] [Medline: 29807158]
5. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
6. Jabbarian LJ, Zwakman M, van der Heide A, Kars MC, Janssen DJ, van Delden JJ, et al. Advance care planning for patients with chronic respiratory diseases: a systematic review of preferences and practices. Thorax 2018 Mar;73(3):222-230. [doi: 10.1136/thoraxjnl-2016-209806] [Medline: 29109233]
7. Jimenez G, Tan WS, Virk AK, Low CK, Car J, Ho AH. State of advance care planning research: a descriptive overview of systematic reviews. Palliat Support Care 2018 Jul 30:1-11. [doi: 10.1017/S1478951518000500] [Medline: 30058506]
8. May CR, Eton DT, Boehmer K, Gallacher K, Hunt K, MacDonald S, et al. Rethinking the patient: using Burden of Treatment Theory to understand the changing dynamics of illness. BMC Health Serv Res 2014 Jun 26;14:281 [FREE Full text] [doi: 10.1186/1472-6963-14-281] [Medline: 24969758]
9. Tran VT, Montori VM, Eton DT, Baruch D, Falissard B, Ravaud P. Development and description of measurement properties of an instrument to assess treatment burden among patients with multiple chronic conditions. BMC Med 2012 Jul 4;10:68 [FREE Full text] [doi: 10.1186/1741-7015-10-68] [Medline: 22762722]
10. Eton DT, Yost KJ, Lai J, Ridgeway JL, Egginton JS, Rosedahl JK, et al. Development and validation of the Patient Experience with Treatment and Self-management (PETS): a patient-reported measure of treatment burden. Qual Life Res 2017 Feb;26(2):489-503 [FREE Full text] [doi: 10.1007/s11136-016-1397-0] [Medline: 27566732]
11. Scholten G, Bourguignon S, Delanote A, Vermeulen B, van Boxem G, Schoenmakers B. Advance directive: does the GP know and address what the patient wants? Advance directive in primary care. BMC Med Ethics 2018 Jun 11;19(1):58 [FREE Full text] [doi: 10.1186/s12910-018-0305-2] [Medline: 29890967]
12. Houben C. Semantic Scholar. Maastricht: Maastricht University; 2018. Advance Care Planning in Life-Limiting Illnesses URL: https://pdfs.semanticscholar.org/6782/021adbb7883c042c22eaa942f2f1aaf90325.pdf?_ga=2.6522868.732736689.1578897898-1679671381.1567599385 [accessed 2020-01-13]
13. Chiarchiaro J, Arnold RM, White DB. Reengineering advance care planning to create scalable, patient- and family-centered interventions. J Am Med Assoc 2015 Mar 17;313(11):1103-1104. [doi: 10.1001/jama.2015.0569] [Medline: 25781435]
14. Butler M, Ratner E, McCreedy E, Shippee N, Kane RL. Decision aids for advance care planning: an overview of the state of the science. Ann Intern Med 2014 Sep 16;161(6):408-418. [doi: 10.7326/M14-0644] [Medline: 25069709]
15. Lustria ML, Noar SM, Cortese J, van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
16. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010 Jun 24;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
17. Van't Hof E, Cuijpers P, Stein DJ. Self-help and internet-guided interventions in depression and anxiety disorders: a systematic review of meta-analyses. CNS Spectr 2009 Feb;14(2 Suppl 3):34-40. [doi: 10.1017/s1092852900027279] [Medline: 19238128]
18. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
19. Apolinário-Hagen J, Harrer M, Kählke F, Fritsche L, Salewski C, Ebert DD. Public attitudes toward guided internet-based therapies: web-based survey study. JMIR Ment Health 2018 May 15;5(2):e10735 [FREE Full text] [doi: 10.2196/10735] [Medline: 29764797]
20. Austin CA, Mohottige D, Sudore RL, Smith AK, Hanson LC. Tools to promote shared decision making in serious illness: a systematic review. JAMA Intern Med 2015 Jul;175(7):1213-1221 [FREE Full text] [doi: 10.1001/jamainternmed.2015.1679] [Medline: 25985438]

21. Cardona-Morrell M, Benfatti-Olivato G, Jansen J, Turner RM, Fajardo-Pulido D, Hillman K. A systematic review of effectiveness of decision aids to assist older patients at the end of life. Patient Educ Couns 2017 Mar;100(3):425-435. [doi: 10.1016/j.pec.2016.10.007] [Medline: 27765378]
22. Ostherr K, Killoran P, Shegog R, Bruera E. Death in the digital age: a systematic review of information and communication technologies in end-of-life care. J Palliat Med 2016 Apr;19(4):408-420 [FREE Full text] [doi: 10.1089/jpm.2015.0341] [Medline: 26713368]
23. Saunders CH, Patel K, Kang H, Elwyn G, Kirkland K, Durand M. Serious choices: a systematic environmental scan of decision aids and their use for seriously ill people near death. J Hosp Med 2019 May;14(5):294-302. [doi: 10.12788/jhm.3110] [Medline: 30794146]
24. Baik D, Cho H, Creber RM. Examining interventions designed to support shared decision making and subsequent patient outcomes in palliative care: a systematic review of the literature. Am J Hosp Palliat Care 2019 Jan;36(1):76-88 [FREE Full text] [doi: 10.1177/1049909118783688] [Medline: 29925244]
25. Mays N, Roberts E, Popay J. Synthesising research evidence. In: Fulop N, Allen P, Clarke A, Black N, Anderson S, editors. Studying the Organisation and Delivery of Health Services: Research Methods. London: Routledge; 2001:188-220.
26. 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]
27. 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]
28. Covidence Systematic Review Software. Melbourne, Australia: Veritas Health Innovation; 2018. URL: https://www.covidence.org/home [accessed 2019-05-21]
29. 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]
30. 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]
31. Volandes AE, Paasche-Orlow MK, Davis AD, Eubanks R, El-Jawahri A, Seitz R. Use of video decision aids to promote advance care planning in Hilo, Hawai'I. J Gen Intern Med 2016 Sep;31(9):1035-1040 [FREE Full text] [doi: 10.1007/s11606-016-3730-2] [Medline: 27194151]
32. South AL, Elton J. Contradictions and promise for end-of-life communication among family and friends: death over dinner conversations. Behav Sci (Basel) 2017 Apr 20;7(2):pii: E24 [FREE Full text] [doi: 10.3390/bs7020024] [Medline: 28425929]
33. Chovan WL. Implementing a pilot program of advanced directives in nursing homes. Psychol Rep 2007 Feb;100(1):195-198. [doi: 10.2466/pr0.100.1.195-198] [Medline: 17451025]
34. Green MJ, Levi BH. Development of an interactive computer program for advance care planning. Health Expect 2009 Mar;12(1):60-69 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00517.x] [Medline: 18823445]
35. Hossler C, Levi BH, Simmons Z, Green MJ. Advance care planning for patients with ALS: feasibility of an interactive computer program. Amyotroph Lateral Scler 2011 May;12(3):172-177. [doi: 10.3109/17482968.2010.509865] [Medline: 20812887]
36. Levi BH, Heverley SR, Green MJ. Accuracy of a decision aid for advance care planning: simulated end-of-life decision making. J Clin Ethics 2011;22(3):223-238 [FREE Full text] [Medline: 22167985]
37. Schubart JR, Levi BH, Camacho F, Whitehead M, Farace E, Green MJ. Reliability of an interactive computer program for advance care planning. J Palliat Med 2012 Jun;15(6):637-642 [FREE Full text] [doi: 10.1089/jpm.2011.0489] [Medline: 22512830]
38. Holland DE, Vanderboom CE, Dose AM, Ingram CJ, Delgado A, Austin CM, et al. Nurse-led patient-centered advance care planning in primary care. J Hosp Palliat Nurs 2017;19(4):368-375. [doi: 10.1097/NJH.0000000000000358]
39. Green MJ, Schubart JR, Whitehead M, Farace E, Lehman E, Levi BH. Advance care planning does not adversely affect hope or anxiety among patients with advanced cancer. J Pain Symptom Manage 2015 Jun;49(6):1088-1096. [doi: 10.1016/j.jpainsymman.2014.11.293] [Medline: 25542552]
40. Markham SA, Levi BH, Green MJ, Schubart JR. Use of a computer program for advance care planning with African American participants. J Natl Med Assoc 2015 Feb;107(1):26-32. [doi: 10.1016/S0027-9684(15)30006-7] [Medline: 27282525]
41. Schubart JR, Green MJ, van Scoy LJ, Lehman E, Farace E, Gusani NJ, et al. Advanced cancer and end-of-life preferences: curative intent surgery versus noncurative intent treatment. J Palliat Med 2015 Dec;18(12):1015-1018 [FREE Full text] [doi: 10.1089/jpm.2015.0021] [Medline: 26262942]
42. van Scoy LJ, Green M, Dimmock A, Bascom R, Boehmer J, Hensel J, et al. High satisfaction and low decisional conflict with advance care planning among chronically ill patients with advanced chronic obstructive pulmonary disease or heart failure using an online decision aid: A pilot study. Chronic Illn 2016 Sep;12(3):227-235. [doi: 10.1177/1742395316633511] [Medline: 27055468]

43. Levi BH, Simmons Z, Hanna C, Brothers A, Lehman E, Farace E, et al. Advance care planning for patients with amyotrophic lateral sclerosis. Amyotroph Lateral Scler Frontotemporal Degener 2017 Aug;18(5-6):388-396 [FREE Full text] [doi: 10.1080/21678421.2017.1285317] [Medline: 28631959]
44. Schubart JR, Camacho F, Green MJ, Rush KA, Levi BH. Assessing the internal consistency and temporal stability of advance directives generated by an interactive, online computer program. BMJ Support Palliat Care 2017 Mar;7(1):67-72 [FREE Full text] [doi: 10.1136/bmjspcare-2014-000814] [Medline: 25743440]
45. Green MJ, van Scoy LJ, Foy AJ, Stewart RR, Sampath R, Schubart JR, et al. A randomized controlled trial of strategies to improve family members' preparedness for surrogate decision-making. Am J Hosp Palliat Care 2018 Jun;35(6):866-874 [FREE Full text] [doi: 10.1177/1049909117744554] [Medline: 29186982]
46. Fine RL, Yang Z, Spivey C, Boardman B, Courtney M. Early experience with digital advance care planning and directives, a novel consumer-driven program. Proc (Bayl Univ Med Cent) 2016 Jul;29(3):263-267 [FREE Full text] [doi: 10.1080/08998280.2016.11929430] [Medline: 27365867]
47. van Scoy LJ, Chiarolanzio PJ, Kim C, Heyland DK. Development and initial evaluation of an online decision support tool for families of patients with critical illness: A multicenter pilot study. J Crit Care 2017 Jun;39:18-24. [doi: 10.1016/j.jcrc.2016.12.022] [Medline: 28131020]
48. Klugman CM, Usatine RP. An evaluation of 2 online advance directive programs. Am J Hosp Palliat Care 2013 Nov;30(7):657-663. [doi: 10.1177/1049909112463116] [Medline: 23087248]
49. Lindquist LA, Ramirez-Zohfeld V, Sunkara PD, Forcucci C, Campbell DS, Mitzen P, et al. PlanYourLifeSpan.org - an intervention to help seniors make choices for their fourth quarter of life: results from the randomized clinical trial. Patient Educ Couns 2017 Nov;100(11):1996-2004 [FREE Full text] [doi: 10.1016/j.pec.2017.06.028] [Medline: 28689855]
50. Cresswell MA, Robinson CA, Fyles G, Bottorff JL, Sudore R. Evaluation of an advance care planning web-based resource: applicability for cancer treatment patients. Support Care Cancer 2018 Mar;26(3):853-860. [doi: 10.1007/s00520-017-3901-7] [Medline: 28929291]
51. Ouchi K, Knabben V, Rivera-Reyes L, Ganta N, Gelfman LP, Sudore R, et al. Preparing older adults with serious illness to formulate their goals for medical care in the emergency department. J Palliat Med 2017 Apr 1;20(4):404-408 [FREE Full text] [doi: 10.1089/jpm.2016.0109] [Medline: 27797638]
52. Sudore RL, Boscardin J, Feuz MA, McMahan RD, Katen MT, Barnes DE. Effect of the PREPARE website vs an easy-to-read advance directive on advance care planning documentation and engagement among veterans: a randomized clinical trial. JAMA Intern Med 2017 Aug 1;177(8):1102-1109 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1607] [Medline: 28520838]
53. Sudore RL, Knight SJ, McMahan RD, Feuz M, Farrell D, Miao Y, et al. A novel website to prepare diverse older adults for decision making and advance care planning: a pilot study. J Pain Symptom Manage 2014 Apr;47(4):674-686 [FREE Full text] [doi: 10.1016/j.jpainsymman.2013.05.023] [Medline: 23972574]
54. Lum HD, Barnes DE, Katen MT, Shi Y, Boscardin J, Sudore RL. Improving a full range of advance care planning behavior change and action domains: the prepare randomized trial. J Pain Symptom Manage 2018 Oct;56(4):575-81.e7 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.06.007] [Medline: 29940209]
55. Sudore RL, Schillinger D, Katen MT, Shi Y, Boscardin WJ, Osua S, et al. Engaging diverse English- and Spanish-speaking older adults in advance care planning: the PREPARE randomized clinical trial. JAMA Intern Med 2018 Dec 1;178(12):1616-1625 [FREE Full text] [doi: 10.1001/jamainternmed.2018.4657] [Medline: 30383086]
56. Periyakoil VS, Neri E, Kraemer H. A randomized controlled trial comparing the letter project advance directive to traditional advance directive. J Palliat Med 2017 Sep;20(9):954-965. [doi: 10.1089/jpm.2017.0066] [Medline: 28471713]
57. O'Shea B, Martin D, Brennan B, Bailey O, McElwee O, O'Leary F, et al. Are we ready to 'think ahead'? Acceptability study using an innovative end of life planning tool. Ir Med J 2014 May;107(5):138-140. [Medline: 24908856]
58. Suneetha K, Krishnamoorthi R. Identifying user behavior by analyzing web server access log file. Int J Comput Sci Netw Secur 2009;9(4):327-332 [FREE Full text]
59. My Decisions. England and Wales: Compassion in Dying URL: https://mydecisions.org.uk/ [accessed 2019-04-12]
60. Dy SM, Roy J, Ott GE, McHale M, Kennedy C, Kutner JS, et al. Tell Us: a web-based tool for improving communication among patients, families, and providers in hospice and palliative care through systematic data specification, collection, and use. J Pain Symptom Manage 2011 Oct;42(4):526-534 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.12.006] [Medline: 21458214]
61. Overbeek A, Korfage IJ, Jabbarian LJ, Billekens P, Hammes BJ, Polinder S, et al. Advance care planning in frail older adults: a cluster randomized controlled trial. J Am Geriatr Soc 2018 Jul;66(6):1089-1095. [doi: 10.1111/jgs.15333] [Medline: 29608789]
62. Halpern SD. Goal-concordant care - searching for the holy grail. N Engl J Med 2019 Oct 24;381(17):1603-1606. [doi: 10.1056/NEJMp1908153] [Medline: 31644843]

Abbreviations

ACP: advance care planning

AD: advance directive
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth
EAPC: European Association for Palliative Care
eHealth: electronic health
RCT: randomized controlled trial
PREPARE: PREPARE For Your Care

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by C Pasarelu, G Jimenez; comments to author 10.09.19; revised version received 04.11.19; accepted 16.12.19; published 17.03.20.

Please cite as:
van der Smissen D, Overbeek A, van Dulmen S, van Gemert-Pijnen L, van der Heide A, Rietjens JAC, Korfage IJ
The Feasibility and Effectiveness of Web-Based Advance Care Planning Programs: Scoping Review
J Med Internet Res 2020;22(3):e15578
URL: http://www.jmir.org/2020/3/e15578/
doi:10.2196/15578
PMID:32181750

Original Paper

The Effectiveness of a Guided Internet-Based Tool for the Treatment of Depression and Anxiety in Pregnancy (MamaKits Online): Randomized Controlled Trial

Hanna M Heller[1,2*], MD; Adriaan W Hoogendoorn[1,2,3*], PhD; Adriaan Honig[1,2,4*], MD, PhD; Birit F P Broekman[1,2,4*], MD, PhD; Annemieke van Straten[2,5*], PhD

[1]Department of Psychiatry, Amsterdam Universitair Medische Centra, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[2]Amsterdam Public Health Research Institute, Amsterdam, Netherlands
[3]GGZ inGeest, Amsterdam, Netherlands
[4]Department of Psychiatry, OLVG Hospital, Amsterdam, Netherlands
[5]Department of Clinical Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[*]all authors contributed equally

Corresponding Author:
Hanna M Heller, MD
Department of Psychiatry
Amsterdam Universitair Medische Centra
Vrije Universiteit Amsterdam
de Boelelaan 1117
Amsterdam, 1081 HV
Netherlands
Phone: 31 204440196
Email: hm.heller@amsterdamumc.nl

Abstract

Background: Pregnant women with symptoms of depression or anxiety often do not receive adequate treatment. In view of the high incidence of these symptoms in pregnancy and their impact on pregnancy outcomes, getting treatment is of the utmost importance. A guided internet self-help intervention may help to provide more women with appropriate treatment.

Objective: This study aimed to examine the effectiveness of a guided internet intervention (MamaKits online) for pregnant women with moderate to severe symptoms of anxiety or depression. Assessments took place before randomization (T0), post intervention (T1), at 36 weeks of pregnancy (T2), and 6 weeks postpartum (T3). We also explored effects on perinatal child outcomes 6 weeks postpartum.

Methods: This randomized controlled trial included pregnant women (<30 weeks) with depressive symptoms above threshold (ie, Center for Epidemiological Studies Depression scale [CES-D] >16) or anxiety above threshold (ie, Hospital Anxiety and Depression Scale-Anxiety subscale [HADS-A] >8) or both of them. Participants were recruited via general media and flyers in prenatal care waiting rooms or via obstetricians and midwives. After initial assessment, women were randomized to (1) MamaKits online in addition to treatment as usual or (2) treatment as usual (control condition). MamaKits online is a 5-week guided internet intervention based on problem solving treatment. Guidance was was provided by trained students pursuing a Master's in Psychology. Outcomes were based on a Web-based self-report. Women in the control condition were allowed to receive the intervention after the last assessment (6 weeks postpartum).

Results: Of the 159 included women, 79 were randomized to MamaKits online, 47% (79/37) of whom completed the intervention. Both groups showed a substantial decrease in affective symptoms on the CES-D, HADS-A, and Edinburgh Postnatal Depression Scale over time. In the intervention group, affective symptoms decreased more than that in the control group, but between-group effect sizes were small to medium (Cohen *d* at T3=0.45, 0.21, and 0.23 for the 3 questionnaires, respectively) and statistically not significant. Negative perinatal child outcomes did not differ between the 2 groups (χ^2_1=0.1; *P*=.78). Completer analysis revealed no differences in outcome between the treatment completers and the control group. The trial was terminated early for reasons of futility based on the results of an interim analysis, which we performed because of inclusion problems.

Conclusions: Our study did show a significant reduction in affective symptoms in both groups, but the differences in reduction of affective symptoms between the intervention and control groups were not significant. There were also no differences in perinatal child outcomes. Future research should examine for which women these interventions might be effective or if changes in the internet intervention might make the intervention more effective.

Trial Registration: Netherlands Trial Register NL4162; https://tinyurl.com/sdckjek

(J Med Internet Res 2020;22(3):e15172) doi:10.2196/15172

KEYWORDS

pregnancy; depression; anxiety; internet; pregnancy outcome; treatment

Introduction

Background

Depression and anxiety are common problems in women in the perinatal period. Major depressive disorder and anxiety disorders affect 7% to 15% of women during pregnancy [1-4]. The prevalence of symptoms of depression and anxiety is even higher, as they occur in 18% to 20% of the pregnant women [1,5,6]. Depression and anxiety are both associated with poor pregnancy outcomes [7], postpartum depression [4,8,9], and negative influences on child development [10-14]. Hence, effective treatment of these disorders is of the utmost importance.

Psychotherapeutic interventions such as cognitive behavioral therapy and interpersonal therapy have proven effective in treating perinatal depression and anxiety [15-17]. However, the implementation of effective treatment interventions is often hampered by factors relating to the pregnancy, for example, nonrecognition of the symptoms because of overlapping symptomatology with pregnancy itself [18], or by feelings of stigmatization, lack of time, problems with transportation, or difficulties arranging childcare [11,19,20]. Some of these barriers may be overcome by providing guided internet-based self-help interventions. Indeed, Web-based interventions are easier to access, have no waiting lists, allow anonymity, and can be carried out whenever and wherever the patient wants [19,21,22]. Moreover, because the therapeutic input is smaller than that in regular face-to-face treatments, internet-based interventions are likely to be less costly and more scalable. This is especially advantageous for disorders that are characterized by a combination of high prevalence and a low treatment-seeking rate, which is the case for pregnant women with depressive and anxiety symptoms. Although internet-based interventions proved effective in the general population [23,24] as well as postpartum [25], recent studies of internet-based interventions during pregnancy [26-28] showed varying success. Outcomes may have been influenced by differences in the methodology, content, and duration of these internet-based therapies.

Previously, we developed an internet-based problem solving treatment (PST) consisting of five modules and support provided by a trained coach, which proved effective for depressed and anxious people in general [29,30]. PST is a generic treatment that is used for different kinds of psychiatric problems, such as depression [31] and anxiety [32]. The core assumption of PST is that affective symptoms are generated when people become overwhelmed by practical problems they face in their daily lives. In PST, participants make a list of all their worries and problems and learn structured ways to resolve those problems. This approach makes them feel less overwhelmed, which in turn alleviates their mood.

Although the effectiveness of face-to-face PST has been firmly established [31], there is no evidence yet whether online guided PST might be effective in reducing symptoms of depression and anxiety in pregnant women.

Objectives

For this study, we adapted the Web-based guided PST to provide an effective, easily accessible intervention for above-threshold affective symptoms in pregnant women. We hypothesized that the intervention would be effective (1) in reducing depressive and anxiety symptoms post intervention during pregnancy, at the end of pregnancy, and at 6 weeks postpartum and (2) in improving perinatal child outcomes, such as preterm birth, growth restriction, and breastfeeding initiation.

Methods

Study Design

We performed a randomized controlled trial with an intervention condition (internet-based PST) and a control group (care-as-usual). For ethical reasons, the participants in the control condition were also offered access to the intervention, but only after the last follow-up (6 weeks postpartum). Both groups were allowed to use concurrent treatment (care-as-usual) as well. The use of additional care was monitored through self-report.

The study protocol, information brochure, and informed consent form were approved by the Medical Ethics Committee of the VU University Medical Center (registration number 2013.275) and registered with the Dutch Trial Registry (NL4162). The tenets of the Declaration of Helsinki were observed. An extensive description of the study protocol can be found elsewhere [33].

Participants

All participants were self-referred. They were recruited through articles and advertisement in national newspapers and magazines and through social media, pregnancy websites, and websites of patient’s associations. Information flyers and posters were also distributed in maternity clinics and in clinics for primary care nationwide. Pregnant women with symptoms of depression and/or anxiety were advised to visit our study website, where they could find more information about the study and were given the opportunity to register online. After registration, they received an informed consent letter by post. After returning the signed informed consent form, they were invited to complete the first Web-based questionnaire.

Inclusion and Exclusion Criteria

Women aged 18 years and older were eligible if they were pregnant for less than 30 weeks, showed symptoms of depression or anxiety or both, and had sufficient access to the internet. Symptoms of depression were measured with the initial Web-based questionnaire using the Center for Epidemiological Studies Depression scale (CES-D) [34], and symptoms of anxiety were assessed using the Hospital Anxiety and Depression Scale-Anxiety subscale (HADS-A) [35]. Women were eligible to participate if their score on the CES-D was at least 16 or the score on the HADS-A was 8 or more. Women with severe depressive or anxiety symptoms (CES-D ≥25 or HADS-A ≥12) were also allowed to participate. However, we advised them to contact their general practitioner as well to check if another treatment or additional treatment was needed. We did not exclude them because internet-based PST has also proven to be effective for severe depressive and anxiety symptoms [29,36,37]. However, women were excluded if they reported intentions to harm themselves or to attempt suicide (assessed by one question of the Web Screening Questionnaire) [38].

During the trial, participants in the intervention group were allowed to receive additional care-as-usual, such as psychiatric treatment including psychotherapy or psychopharmacological drugs. Any additional treatments were monitored through participants' self-reports at every assessment.

Randomization

Women who were included in the study were randomized in a 1:1 ratio to the intervention condition versus the control condition. An independent researcher created a computer-generated randomization scheme based on blocks of 10 and provided the next randomization outcome to one of the coaches. This procedure ensured allocation concealment. The research assistant informed all the participants on the randomization outcome by email. The participants of the intervention group also received the name of the website and details on where and how they could log-in to start the intervention.

Intervention

An existing evidence-based internet version of PST [29] was used. We adapted this version for pregnant women by adding one session of psychoeducation on pregnancy and affective symptoms and adjusting all the existing case examples to pregnancy-related case examples. The adapted Web-based intervention for pregnant women was named MamaKits online. The course consists of 5 modules, and participants are advised to try to carry out one module each week. Each module consists of information, examples of other pregnant women with depressive or anxiety symptoms carrying out the intervention, and homework assignments.

The intervention consists of 3 steps: (1) participants describe what really matters to them, (2) participants write down all their current worries and problems, and (3) participants make a plan for the future, in which they describe how they will try to accomplish those things that matter most to them. After that, they categorize the problems into three types: unimportant problems (problems unrelated to the things that matter to them), problems that can potentially be solved, and problems that cannot be solved (eg, the loss of a loved one). The core of the intervention consists of a structured approach to solve the potentially solvable problems. This approach consists of 6 steps: (1) write down a clear definition of the problem, (2) generate multiple solutions to the problem, (3) select the best solution, (4) work out a systematic plan for this solution, (5) carry out the solution, and (6) evaluate whether the solution has resolved the problem.

After each module, trained coaches (students pursuing Master's in Psychology) provided feedback on the assignments via secured email. All coaches were trained for 4 hours in PST and providing feedback via secured email. They were trained by an experienced psychotherapist, who also provided the coaches with regular supervision. On average, the coaches gave 20 min of feedback per patient per module. The feedback was directed to helping the patient work through the intervention; the coaches answered questions if something was not clear and provided feedback on homework assignments. If a participant was delayed in submitting the homework, the coach sent a reminder by email, with a maximum of three emails and one phone call after that.

Measures

Assessments

Assessments took place at baseline (T0), 10 weeks after baseline (T1), 4 weeks before the expected date of delivery (T2), and 6 weeks postpartum (T3). Participants who started the intervention after 24 weeks' gestation were not assessed at T2, as the period between T1 and T2 would have been too short to expect any effects. All assessments were based on self-report and took place online. At baseline, we additionally collected demographic data and data on current mental treatment, parity, pregnancy duration, and previous and current pregnancy complications. At T1, we collected additional data about treatment satisfaction, and at T3, we additionally collected data on perinatal child outcomes.

Primary Outcomes

Primary outcomes were reduction in symptoms of depression and anxiety and perinatal child outcomes. Depression was measured with the Dutch version of the CES-D [34]. This scale has 20 self-rated items, each of which is scored from 0 to 3. The total score range is 0 (no depressive symptoms) to 60 (high number of depressive symptoms). The validity of the CES-D has been tested in different populations, including pregnant women [39,40] and also online [41]. Scores of 16 and higher represent a clinically significant level of depressive symptoms with a sensitivity of 0.82 to 1.00 and a specificity of 0.69 to 0.88 [37,38].

Anxiety was measured with the Dutch version of the HADS-A [35]. The HADS-A is a 7-item anxiety subscale of the HADS with item responses on a 0 to 3 scale. Total score range is 0 to 21. Higher scores indicate more anxiety. The questionnaire has been found to be reliable in the internet version [42]. The HADS-A has an optimal cut-off ≥8 with a sensitivity of 0.89 and a specificity of 0.75 [43].

Perinatal child outcomes were assessed through self-report and analyzed by calculating the differences between the percentages of women in the intervention and control condition who delivered preterm (gestational age <37 weeks), whose babies had a low birth weight for gestational age (weight ≤tenth percentile, according to the guidelines by the Dutch Association of Gynecologists and Obstetricians, based on data of the Dutch National Birth Register), who delivered with an emergency cesarean section or vacuum extraction, or who did not continue breastfeeding until 6 weeks postpartum.

Secondary Outcomes

Secondary outcomes were reduction in symptoms of depression as measured with the Edinburgh Postnatal Depression Scale (EPDS), additional psychological health care use, and treatment satisfaction. The EPDS [44] is a 10-item depression scale developed for women primarily in the postpartum period, but also in pregnancy. Depending on the trimester, the cut-off score varied worldwide from 6.5 to 14.5, and in the Netherlands, it varied from 10 to 11 [45,46]. Item response varies from 0 to 3, and the total score range is 0 to 30 [45]. Information about additional mental health care was obtained using the Trimbos/institute for Medical Technology Assessment, Erasmus University Rotterdam, questionnaire for Costs associated with Psychiatric Illness [47].

We also used the Client Satisfaction Questionnaire (CSQ-8). The CSQ-8, a questionnaire with 8 items measured on a 4-point scale, has good psychometric properties in the Dutch population [48]. We added several questions about the intervention, the website, and the feedback of the coach. These questions could be answered through visual analog scales (VASs).

Sample Size Considerations

The between-group effect size (Cohen *d*) at post test (T1) was assumed to be at least 0.40, as was demonstrated in previous studies using the same internet-based PST [29,30]. Using an alpha of .05 (2-tailed), a statistical power (1-beta) of 0.80, and an attrition rate of 30% (as seen in other internet-based therapies in depressed patients) [30], we calculated that we needed to enroll 143 respondents in each arm.

After reviewing the literature, we assumed that symptoms of major depressive disorder and any anxiety disorder affect 7% to 15% of women during pregnancy [1-4] and that about 17% of the pregnant women have mild affective symptoms in pregnancy [6]. On the basis of a yearly birth rate of 171,341 in the Netherlands [49], at least about 29,127 women would be eligible for screening. With an expected response rate of 1%, 291 women would be included. Therefore, inclusion was expected to be completed within 1 year.

Statistical Analysis

All data were analyzed according to intention-to-treat analysis (comprising all the participants who were randomized) as well as per-protocol analysis (focusing on the participants who completed the intervention, ie, a subset of the intention-to-treat sample).

Mean total scores (standard deviations) of the 3 questionnaires (CES-D, HADS-A, and EPDS) were computed for the intervention and control arms separately at different time points (T0, T1, T2, and T3). The internet-based PST intervention effect was tested with linear mixed model (LMM) analyses, while correcting for baseline differences in the depressive and anxiety symptoms. LMM analysis can handle missing data owing to dropout under the assumption that the data are missing-at-random. Adverse perinatal child outcomes were defined as having experienced any negative outcome and were also evaluated by means of chi-square tests. Statistical analyses were carried out with SPSS (version 24; IBM, Armonk, New York) and Stata (version 15; StataCorp, College Station, Texas) software.

Results

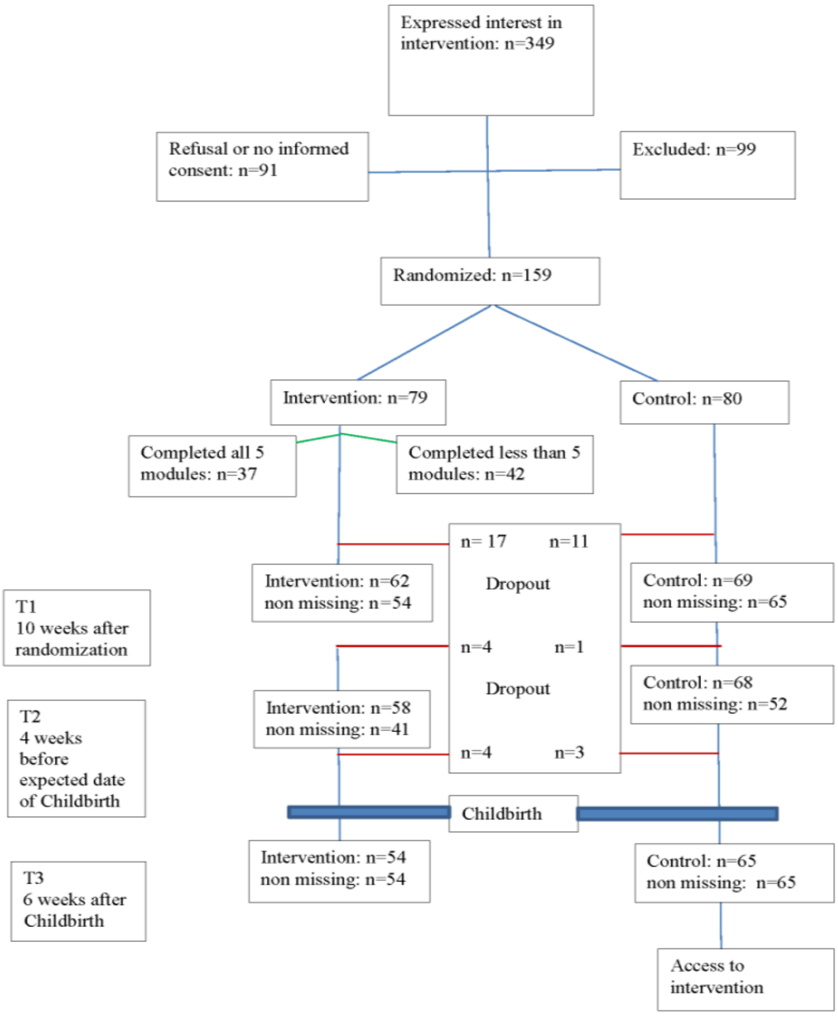
Inclusion, Study Flow, Study Termination, and Dropout

The inclusion period was extended from 1 to 3 years owing to a low inclusion rate (March 2014 until January 2017). After 3 years, we performed an interim analysis, which had not been planned in the study protocol, to decide if inclusion of additional participants (and applying for additional funding) would be worthwhile or not. We developed an interim analysis protocol, which was approved by the ethical board. We evaluated the intervention effect on the first primary outcome CES-D at posttest (T1) when 153 participants had been randomized. According to the interim analysis protocol, the trial would be stopped for efficacy if the estimated intervention effect (in terms of standardized mean difference) exceeded 0.54 (in other words, extra patients would not be needed because the power was enough to establish the effect with significance). Inclusion would also be stopped, for futility, if the intervention effect was below 0.29 (in other words, continuing with our previously planned number of patients would not be useful because even if this number was reached, the power would be insufficient to demonstrate the effect with significance). As the interim analysis provided an estimated effect size of 0.035, the inclusion of participants was terminated prematurely. Although the inclusion of new participants stopped, all measurements continued as scheduled for the participants already included.

At the time of closure, a total of 349 women had expressed interest in the intervention. Of those women, 99 were excluded because they did not fulfill the inclusion criteria (eg, due to being pregnant beyond 30 weeks or due to not exceeding the required threshold for depression or anxiety scores). Of the remaining 250 women, 91 did not want to participate because of several reasons (eg, they already felt better or started another type of therapy). Of the originally planned 286 women, 159 were included in the study, as another 6 women were in the process of inclusion during the interim analysis. Of these 159 women, 79 were randomly allocated to the intervention group and 80 to the control group (Figure 1). Study dropout was 14% (11/80) in the control arm versus 22% (17/79) in the experimental arm at T1 (*P*=.20), 15% (12/80) versus 27% (21/79) at T2 (*P*=.07), and 19% (15/80) versus 32% (25/79) at T3 (*P*=.06). Overall, 60% (48/80) of the control group versus 43% (34/79) of the intervention group responded in all waves (T0, T1, T2, and T3) and 21% (17/80) of the control group versus 25% (20/79) of the intervention group missed either T1 or T2 (or both).

Of the 79 participants who were randomized to the intervention group, 37 (47%) completed all 5 modules of the intervention, 39 (49%) women completed at least four modules, 50 (63%) women completed at least three modules, 67 (72%) women completed at least two modules, 70 (89%) women completed at least one module, and 9 (11%) women did not even complete the first module. Reasons for nonadherence included being too busy (n=7), feeling better (n=4), need for other treatment (too sick; n=5), not being motivated (n=8), difficulties in confessing to the computer (n=1), intervention not meeting the expectations (n=3), and other reason/no reason given (n=14). There were no statistically significant differences between the baseline scores of treatment completers (having done all five modules) and noncompleters (having done less than five modules). The number of women using additional therapy was similar in both groups (*P*=.68, *P*=.82, and *P*=.73 at T0, T1, and T3, respectively).

Figure 1. Flowchart of participants throughout the trial. Intervention: still in intervention group, but not everyone participated in this assessment; Control: still in control group, but not everyone participated in this assessment; Nonmissing: did participate in this assessment; Dropout: not in study anymore.



Description of Participants

In total, 159 women were randomized. Differences in baseline demographics between the internet-based PST group and the control group were small (Table 1). Most women were of native Dutch origin (134/159, 84.2%), highly educated (120/159, 75.4%), and employed (111/159, 69.8%). Differences between the intervention and the control group with respect to baseline severity scores of depression and anxiety (primary and secondary outcomes) were also small and nonsignificant.

Table 1. Sociodemographic and clinical characteristics at baseline for the intervention group and the control group (primary and secondary outcomes).

Demographic factors	Intervention, n=79	Control, n=80
Age (years), mean (SD)	32.08 (4.61)	31.94 (4.83)
Background, n (%)		
Dutch	72 (91)	62 (78)
Other	7 (9)	18 (23)
Education[a], n (%)		
Low	4 (5)	0 (0.0)
Middle	14 (18)	21 (26)
High	61 (77)	59 (74)
Marital status, n (%)		
In a relationship	76 (96)	76 (95)
Living together	71 (90)	73 (91)
Employed, n (%)	57 (72)	54 (68)
Pregnancy, n (%)		
Duration by study entrance		
<12 weeks	5 (6)	11 (14)
>12 and <26 weeks	48 (61)	44 (55.0)
> 26 weeks	26 (33)	25 (31)
Nulliparous	42 (53)	36 (45)
Complications in previous pregnancy[b]	29 (60)	39 (72)
Complications in this pregnancy	9 (11)	7 (9)
Previous mental health[c], n (%)		
Depressive disorder	24 (30)	29 (36)
Anxiety disorder	20 (25)	25 (31)
Other mental problems	9 (11)	2 (3)
No diagnosis	31 (39)	30 (38)
Current treatment, n (%)		
Psychological treatment	31 (39)	34 (43)
Psychotropic medication	12 (15)	14 (18)
Affective symptoms, mean (SD)		
Primary outcomes		
Center for Epidemiological Studies Depression	28.84 (7.54)	27.94 (9.04)
Hospital Anxiety and Depression Scale-Anxiety	11.44 (3.50)	11.89 (3.38)
Secondary outcome		
Edinburgh Postnatal Depression Scale	14.27 (4.91)	13.96 (4.94)

[a] Dutch Standard Classification of Education: 2006–Edition 2016/’17, CBS, Statistics Netherlands.

[b] First pregnancies excluded.

[c] Note that women can be both in the category “depressive disorder” and in the category “anxiety disorder.”

Effects on Mood Within the Intention-to-Treat Sample

In the intervention group, large within-group effect sizes in primary and secondary outcomes were found between T0 and T1, T2 and T3 (Table 2). However, within-group effect sizes in the control group were also large (Figure 2). Differences in effects, as measured in the between-group effect sizes, were small and statistically insignificant (Table 3). The only exception to this finding was the medium effect size of the CES-D outcome on T3, but this was not significant either (*d*=0.45; *P*=.06).

Table 2. Mean scores (standard deviations) for affective symptoms (primary and secondary outcomes) considering the intervention group and the control group at baseline, 10 weeks after randomization, 4 weeks before expected birth date, and 6 weeks after child birth.

Condition	Baseline		10 weeks after randomization		4 weeks before expected birth date		6 weeks after child birth	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
Primary outcomes								
Center for Epidemiological Studies Depression								
Intervention	79	28.8 (7.5)	54	19.5 (10.2)	41	19.7 (11.1)	54	13.8 (10.3)
Waitlist control	80	27.9 (9.0)	65	18.6 (9.4)	52	18.6 (10.0)	65	16.8 (11.9)
Hospital Anxiety and Depression Scale-Anxiety								
Intervention	79	11.4 (3.5)	54	8.4 (4.2)	41	7.9 (4.4)	54	7.1 (4.4)
Waitlist control	80	11.9 (3.4)	65	8.6 (3.7)	52	7.9 (4.1)	65	7.9 (4.5)
Secondary outcomes								
Edinburgh Postnatal Depression Scale								
Intervention	79	14.3 (4.9)	54	9.5 (5.6)	41	9.0 (5.5)	54	8.0 (5.2)
Waitlist control	80	14.0 (4.9)	65	8.9 (5.5)	52	8.2 (5.2)	65	8.7 (5.9)

Figure 2. Predicted Center for Epidemiological Studies Depression and Hospital Anxiety and Depression Scale severity scores (primary outcomes) estimated using linear mixed models correcting for baseline differences. Measurements taken at T0: baseline; T1: 10 weeks after baseline; T2: 4 weeks before the expected date of delivery; T3: 6 weeks postpartum.

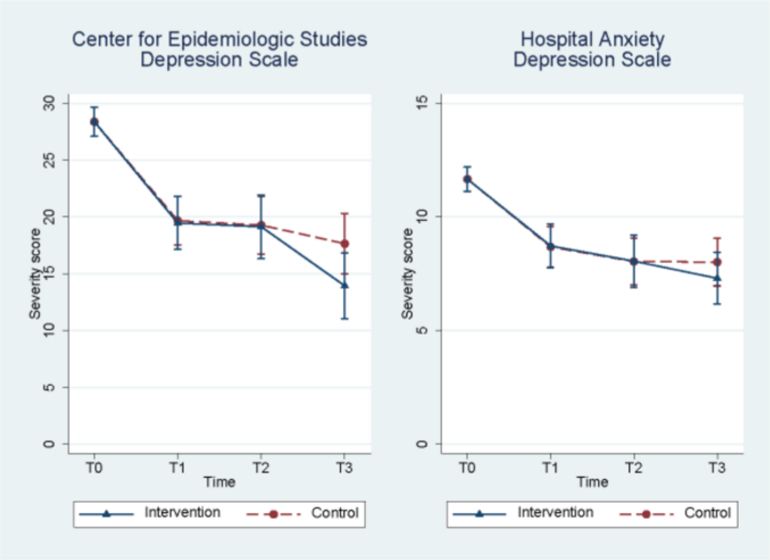


Table 3. Estimated effects (unstandardized), test results, and effect sizes of the differences in primary and secondary outcomes (Center for Epidemiological Studies Depression, Hospital Anxiety and Depression Scale-Anxiety, and Edinburgh Postnatal Depression Scale) within groups and between the intervention group and the control group, using linear mixed model analysis at baseline, 10 weeks after randomization, 4 weeks before expected birth date, and 6 weeks after child birth after correction for scores at baseline.

Condition[a]	Estimated effect[b]	Test results		Effect size[c] (Cohen *d*)		
		Test statistic, *z*	*P* value	Within control condition	Within intervention condition	Between groups
Primary outcomes						
Center for Epidemiological Studies Depression						
Intervention×T1[d]	−0.21	−0.14	.89	−1.05	−1.07	−0.03
Intervention×T2[e]	−0.14	−0.08	.94	−1.10	−1.11	−0.02
Intervention×T3[f]	−3.71	−1.87	.06	−1.29	−1.74	−0.45
Hospital Anxiety and Depression Scale-Anxiety						
Intervention×T1	0.04	0.07	.95	−0.87	−0.86	0.01
Intervention×T2	0.02	0.02	.98	−1.06	−1.05	0.01
Intervention×T3	−0.71	−0.92	.36	−1.06	−1.27	−0.21
Secondary outcomes						
Edinburgh Postnatal Depression Scale						
Intervention×T1	0.00	0.01	.10	−0.91	−0.91	0.00
Intervention×T2	0.10	0.11	.91	−1.09	−1.07	0.02
Intervention×T3	−1.12	−1.13	.26	−1.02	−1.25	−0.23

[a] The test on all three parameters tests the null hypothesis that all three intervention-by-timepoint interaction terms are zero, meaning that the course of the outcome variable within the intervention group is identical to the course of the outcome variable within the waitlist control group.

[b] Estimated effects (unstandardized) are the parameter estimates of the intervention-by-timepoint interaction terms and reflect the additional increase (or decrease) within the intervention group compared with the increase (or decrease) in the waitlist control group.

[c] Effect sizes (Cohen *d*) are standardized effects, obtained by dividing the unstandardized estimated effects by the standard deviation of the primary outcomes.

[d] T1: 10 weeks after randomization.

[e] T2: 4 weeks before expected birth date.

[f] T3: 6 weeks after child birth.

Effects on Perinatal Child Outcomes Within the Intention-to-Treat Sample

The analyses of perinatal child outcomes revealed that 50.4% (60/119) of the women experienced one or more negative perinatal child outcomes or early cessation of breastfeeding. There was no statistically significant difference in these perinatal outcomes between the intervention group and the control group (Table 4).

Table 4. Perinatal child outcomes in the intervention group compared with the control group.

Perinatal child outcomes	Intervention, n (%)	Control, n (%)	*P* value
Preterm birth	4 (7)	1 (2)	.12
Small for gestational age	1 (2)	4 (6)	.24
Emergency cesarean section	7 (13)	9 (14)	.89
Vacuum extraction	7 (13)	9 (14)	.89
No breastfeeding initiation	9 (17)	5 (8)	.13
Stopped breastfeeding early	9 (17)	11 (17)	.31
At least one negative perinatal child outcome (including no or early cessation of breastfeeding)	28 (52)	32 (49)	.78

Per-Protocol Analysis of Treatment Completers (Mood and Perinatal Outcomes)

Of all 79 intervention patients, 37 (47%) completed the whole intervention. We examined the effects for those patients compared with the controls and found no significant differences in any of the outcome measures. LMM analyses revealed predominantly small nonsignificant differences between group effect sizes, with the exception of the CES-D on T3, which had a significant, medium to high effect size (CES-D T1: *d*=−0.25 and *P*=.21, CES-D T3: *d*=−0.53 and *P*=.04, HADS T1: *d*=−0.04 and *P*=.85, HADS T3: *d*=−0.41 and *P*=.09, EPDS T1: *d*=−0.09 and *P*=.66, and EPDS T3: *d*=−0.27 and *P*=.25).

Client Satisfaction

The CSQ-8 was completed at T1 by 53 intervention participants. The majority of the participants 87% (46/53) were satisfied with the help they received and 74% (39/53) would recommend the intervention to others. The total intervention was rated 7.1 (SD 1.6) on a 10-point VAS. The website was rated as fairly good to excellent by 83% (44/53) of the participants, and the feedback of coaches was also rated as fairly good to excellent by 83% (44/53) of the participants.

Additional Psychological Health Care

Both groups used additional psychological health care interventions, and in all cases, these interventions consisted of outpatient care. There were no statistically significant differences between the groups in the use of additional psychological health care. This was 42% (25/54) in the intervention group and 46% (27/65) in the control group (*P*=.60).

Discussion

Principal Findings

To the best of our knowledge, this randomized controlled trial is the first to investigate the effects of offering Web-based guided PST to pregnant women with symptoms of depression and anxiety, with the purpose of reducing barriers for effective therapy. In both the intervention group and the control group, symptoms decreased significantly over time, till 6 weeks postpartum. Although this difference was more pronounced in the intervention group than in the control group, the between-group differences were small and not statistically significant. The only statistically significant difference was shown in the per-protocol analysis at T3 on depression. We consider this result a *lucky finding*, and therefore, we do not think that this result is clinically meaningful. Except for this outlier, the outcomes of the questionnaires did not differ much. The differences in outcomes on the CES-D were larger than those on the EPDS. This might be explained by the fact that the EPDS also contains anxiety items (question 4 and 5), assuming that the intervention had a smaller effect on anxiety than on depression. There were also no differences between the groups in perinatal child outcomes. Attrition was high, with 47% (37/79) women completing the whole intervention and 63% (50/79) women completing more than three modules. In both groups, many women (52/119, 43.6%) used additional psychological treatment.

Our results are not in line with those of other studies on the effects of face-to-face PST in pregnant women, or with those of studies on the effects of internet-based PST in general, or with those of other Web-based therapies for pregnant women. Studies on face-to-face PST delivered perinatally did show medium to high effect [16] on depression, and studies on internet PST among people recruited in the general population showed moderate effects on both depression and anxiety [29,30]. Of 2 other studies on Web-based cognitive behavioral treatment (CBT) for depression during pregnancy [26,50], one showed favorable effects on the follow-up of anxiety but not on depression [50], whereas the other showed a large effect on depression [26] and only a small nonsignificant effect on anxiety.

There might be several reasons why our findings are not in line with those of previous studies. One possible explanation could be the intervention itself. Although the pregnant women who were included were generally satisfied with the intervention, a considerable proportion of these women dropped out. This proportion was larger than that in the intervention of Loughnan [50]. Although we do not know the reasons for the high dropout rate, one reason might have been that the women had sufficiently recovered and did not need more therapy. However, the dropout rate might also indicate that the treatment was not optimal or not sufficiently adapted to the population. The participants might have preferred additional modules (eg, with psychoeducation about changing relationships and role transition) and more supplementary resources, or they might have preferred another type of treatment (eg, CBT), one more like the treatments offered in the above-mentioned studies [26,50]. Another possibility is that the women in our study might have preferred face-to-face therapy, which is the default treatment in the Netherlands. The other 2 trials on Web-based CBT for depression during pregnancy were performed in Australia and Sweden, where people might be more familiar with electronic health because of their inability to commute to health care facilities if they live in remote areas [26,50]. Nevertheless, the fact that almost half of all participants in this study did complete the whole intervention indicates that an internet treatment might be a useful addition to the existing mental health services in the Netherlands.

The second possible reason why our findings differ from those of previous studies is the difference in measuring techniques. In the study that found a significant treatment effect for depression, symptoms were measured with the Montgomery Åsberg Depression Rating Scale Self-report version [26]. This instrument might be more sensitive to picking up relevant changes, but as far as we know, it has not been validated in pregnancy, and the changes could also be related to the improvement of symptoms of pregnancy itself. Besides, both studies on Web-based CBT for depression during pregnancy also used the EPDS as secondary outcome, resulting in small nonsignificant treatment effects.

The third possible reason for the lack of effect in this study is the remarkable improvement in the control group. This suggests that the improvement in both groups might rather be explained by spontaneous recovery. In general, people seek treatment when they are feeling at their worst. It is not unusual that symptoms improve spontaneously afterward [51]. This improvement might also be explained by the use of additional psychological services. A considerable part of the intervention group as well as the control group used other psychological treatments, and the majority of them started treatment before the intervention and continued after starting the intervention.

The fourth possible reason is that the patients in our study were relatively healthy. They had less severe depressive symptoms than those in the studies on Web-based interventions that showed greater effects [26,27]. Studies with patients with more severe symptoms often demonstrate higher effects than studies with patients with less severe symptoms [37].

The fifth possible reason is that the intervention might have been offered at the wrong moment during pregnancy. Most participants (92/159, 57.8%) were in the second trimester of their pregnancy, and several systematic reviews concluded that interventions carried out toward the end of pregnancy or in the postpartum period might be more effective [16,52]. However, in view of the negative consequences of anxiety and depression in pregnancy, an early intervention is of the utmost importance. Although we did not offer our intervention later in pregnancy or in the postpartum period, as recommended, we did meet the other 2 mentioned requirements of a successful treatment, which are an individual approach and an approach targeted at an at-risk population [52].

Strengths

Our study has several strengths. First, we created and tested the first internet version of evidence-based PST in a perinatal setting. Second, we had a relatively long follow-up of 20 weeks. Third, we used an array of different outcome measures, including perinatal child outcomes. Fourth, we allowed women of both groups to use concurrent treatment, including treatment as usual, which makes the results of our study compatible with clinical practice.

Limitations

Despite all our efforts to increase the number of women included in the study (by seeking publicity and prolonging the study period by 2 years), the required number of participants was not obtained. Second, adherence to the intervention was limited. Third, perinatal child outcomes were self-reported, which makes them less objective. The fourth limitation is that there was a sampling bias of mostly native Dutch, employed, and highly educated women, which makes the results less representative of the general population. The fifth limitation is that because of trial reasons, and to keep the population more homogeneous, women in the last 10 weeks of pregnancy were excluded from starting the intervention because they might not be able to finish the treatment before delivery. This is a limitation because by setting this limit, we excluded a group of women who could have benefited from the intervention, and we also possibly reduced the inclusion rate. Furthermore, as we mentioned earlier, interventions carried out toward the end of pregnancy or in the postpartum period might have been more effective [16,52]. We, therefore, might have been able to demonstrate larger effects if the intervention had been offered during this period.

The sixth limitation is that due to the small inclusion sample, the prevalence of negative perinatal child outcomes is probably less reliable.

Clinical Implications

The aim of our study was to improve the care for pregnant women with symptoms of depression or anxiety or both by offering a Web-based intervention with the intention to overcome perceived barriers to treatment. Although inclusion was low, attrition was high, and outcome differences between the intervention group and the control group were mostly nonsignificant, we still recommend investigating how adherence and the effectiveness might be improved by adjusting the Web-based intervention, as satisfaction with the offered modules was high and the intervention is easily applicable at low cost.

Conclusions

To the best of our knowledge, this is the first study to examine a Web-based PST intervention in pregnant women. Although this study did not show a significant reduction in depression and anxiety in comparison with a control condition, Web-based interventions remain a practical, cost-effective, complementary, or alternative therapy modality for face-to-face treatment. Future research is needed to see if the intervention might be more successful if it is offered later in pregnancy or if it is better adapted to the pregnant population or both.

Acknowledgments

The authors disclosed receipt of the following financial support: the randomized controlled trial was supported by *the Stichting tot SteunVCVGZ.*

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist. (V 1.6.1).
[PDF File (Adobe PDF File), 361 KB - jmir_v22i3e15172_app1.pdf]

References

1. Dennis CL, Falah-Hassani K, Shiri R. Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis. Br J Psychiatry 2017 May;210(5):315-323. [doi: 10.1192/bjp.bp.116.187179] [Medline: 28302701]
2. Falah-Hassani K, Shiri R, Dennis C. The prevalence of antenatal and postnatal co-morbid anxiety and depression: a meta-analysis. Psychol Med 2017 Sep;47(12):2041-2053. [doi: 10.1017/S0033291717000617] [Medline: 28414017]
3. Fisher J, de Mello MC, Patel V, Rahman A, Tran T, Holton S, et al. Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review. Bull World Health Organ 2012 Feb 1;90(2):139G-149G [FREE Full text] [doi: 10.2471/BLT.11.091850] [Medline: 22423165]
4. Gelaye B, Rondon MB, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. Lancet Psychiatry 2016 Oct;3(10):973-982 [FREE Full text] [doi: 10.1016/S2215-0366(16)30284-X] [Medline: 27650773]
5. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol 2005 Nov;106(5 Pt 1):1071-1083. [doi: 10.1097/01.AOG.0000183597.31630.db] [Medline: 16260528]
6. Josefsson A, Berg G, Nordin C, Sydsjö G. Prevalence of depressive symptoms in late pregnancy and postpartum. Acta Obstet Gynecol Scand 2001 Mar;80(3):251-255. [doi: 10.1034/j.1600-0412.2001.080003251.x] [Medline: 11207491]
7. Jarde A, Morais M, Kingston D, Giallo R, MacQueen GM, Giglia L, et al. Neonatal outcomes in women with untreated antenatal depression compared with women without depression: a systematic review and meta-analysis. JAMA Psychiatry 2016 Aug 1;73(8):826-837. [doi: 10.1001/jamapsychiatry.2016.0934] [Medline: 27276520]
8. Postpartum Depression: Action Towards Causes Treatment (PACT) Consortium. Heterogeneity of postpartum depression: a latent class analysis. Lancet Psychiatry 2015 Jan;2(1):59-67 [FREE Full text] [doi: 10.1016/S2215-0366(14)00055-8] [Medline: 26359613]
9. Robertson E, Grace S, Wallington T, Stewart DE. Antenatal risk factors for postpartum depression: a synthesis of recent literature. Gen Hosp Psychiatry 2004;26(4):289-295. [doi: 10.1016/j.genhosppsych.2004.02.006] [Medline: 15234824]
10. Barker ED, Jaffee SR, Uher R, Maughan B. The contribution of prenatal and postnatal maternal anxiety and depression to child maladjustment. Depress Anxiety 2011 Aug;28(8):696-702. [doi: 10.1002/da.20856] [Medline: 21769997]
11. Goodman JH. Women's attitudes, preferences, and perceived barriers to treatment for perinatal depression. Birth 2009 Mar;36(1):60-69. [doi: 10.1111/j.1523-536X.2008.00296.x] [Medline: 19278385]
12. Grigoriadis S, Graves L, Peer M, Mamisashvili L, Tomlinson G, Vigod SN, et al. Maternal anxiety during pregnancy and the association with adverse perinatal outcomes: systematic review and meta-analysis. J Clin Psychiatry 2018 Sep 4;79(5):pii: 17r12011 [FREE Full text] [doi: 10.4088/JCP.17r12011] [Medline: 30192449]
13. Grigoriadis S, VonderPorten EH, Mamisashvili L, Tomlinson G, Dennis C, Koren G, et al. The impact of maternal depression during pregnancy on perinatal outcomes: a systematic review and meta-analysis. J Clin Psychiatry 2013 Apr;74(4):e321-e341. [doi: 10.4088/JCP.12r07968] [Medline: 23656857]
14. Verbeek T, Bockting CL, van Pampus MG, Ormel J, Meijer JL, Hartman CA, et al. Postpartum depression predicts offspring mental health problems in adolescence independently of parental lifetime psychopathology. J Affect Disord 2012 Feb;136(3):948-954 [FREE Full text] [doi: 10.1016/j.jad.2011.08.035] [Medline: 21930302]
15. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, et al. Interventions to prevent perinatal depression: US Preventive Services Task Force recommendation statement. J Am Med Assoc 2019 Feb 12;321(6):580-587. [doi: 10.1001/jama.2019.0007] [Medline: 30747971]
16. Sockol LE. A systematic review of the efficacy of cognitive behavioral therapy for treating and preventing perinatal depression. J Affect Disord 2015 May 15;177:7-21. [doi: 10.1016/j.jad.2015.01.052] [Medline: 25743368]
17. van Ravesteyn LM, den Berg MP, Hoogendijk WJ, Kamperman AM. Interventions to treat mental disorders during pregnancy: a systematic review and multiple treatment meta-analysis. PLoS One 2017;12(3):e0173397 [FREE Full text] [doi: 10.1371/journal.pone.0173397] [Medline: 28358808]
18. Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Prevalence of depression during pregnancy: systematic review. Obstet Gynecol 2004 Apr;103(4):698-709. [doi: 10.1097/01.AOG.0000116689.75396.5f] [Medline: 15051562]
19. Kim DR, Hantsoo L, Thase ME, Sammel M, Epperson CN. Computer-assisted cognitive behavioral therapy for pregnant women with major depressive disorder. J Womens Health (Larchmt) 2014 Oct;23(10):842-848 [FREE Full text] [doi: 10.1089/jwh.2014.4867] [Medline: 25268672]
20. O'Mahen HA, Flynn HA. Preferences and perceived barriers to treatment for depression during the perinatal period. J Womens Health (Larchmt) 2008 Oct;17(8):1301-1309. [doi: 10.1089/jwh.2007.0631] [Medline: 18816202]
21. Maloni JA, Przeworski A, Damato EG. Web recruitment and internet use and preferences reported by women with postpartum depression after pregnancy complications. Arch Psychiatr Nurs 2013 Apr;27(2):90-95. [doi: 10.1016/j.apnu.2012.12.001] [Medline: 23540519]
22. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018 Jun 5;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
23. 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]

24. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
25. Lee EW, Denison FC, Hor K, Reynolds RM. Web-based interventions for prevention and treatment of perinatal mood disorders: a systematic review. BMC Pregnancy Childbirth 2016 Feb 29;16:38 [FREE Full text] [doi: 10.1186/s12884-016-0831-1] [Medline: 26928898]
26. Forsell E, Bendix M, Holländare F, von Schultz BS, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
27. Haga SM, Drozd F, Lisøy C, Wentzel-Larsen T, Slinning K. Mamma Mia - a randomized controlled trial of an internet-based intervention for perinatal depression. Psychol Med 2019 Aug;49(11):1850-1858 [FREE Full text] [doi: 10.1017/S0033291718002544] [Medline: 30191779]
28. Krusche A, Dymond M, Murphy SE, Crane C. Mindfulness for pregnancy: a randomised controlled study of online mindfulness during pregnancy. Midwifery 2018 Oct;65:51-57. [doi: 10.1016/j.midw.2018.07.005] [Medline: 30099285]
29. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008 Mar 25;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
30. Warmerdam L, van Straten A, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: randomized controlled trial. J Med Internet Res 2008 Nov 20;10(4):e44 [FREE Full text] [doi: 10.2196/jmir.1094] [Medline: 19033149]
31. Cuijpers P, de Wit L, Kleiboer A, Karyotaki E, Ebert DD. Problem-solving therapy for adult depression: an updated meta-analysis. Eur Psychiatry 2018 Feb;48:27-37. [doi: 10.1016/j.eurpsy.2017.11.006] [Medline: 29331596]
32. Kleiboer A, Donker T, Seekles W, van Straten A, Riper H, Cuijpers P. A randomized controlled trial on the role of support in Internet-based problem solving therapy for depression and anxiety. Behav Res Ther 2015 Sep;72:63-71. [doi: 10.1016/j.brat.2015.06.013] [Medline: 26188373]
33. Heller HM, van Straten A, de Groot CJM, Honig A. The (cost) effectiveness of an online intervention for pregnant women with affective symptoms: protocol of a randomised controlled trial. BMC Pregnancy Childbirth 2014 Aug 14;14:273 [FREE Full text] [doi: 10.1186/1471-2393-14-273] [Medline: 25123230]
34. Bouma J, Ranchor VA, Sanderman R, van Sonderen E. Measuring Symptoms of Depression With the CES-D: A Manual. Groningen: University of Groningen; 1995.
35. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychol Med 1997 Mar;27(2):363-370. [doi: 10.1017/s0033291796004382] [Medline: 9089829]
36. Boeschoten RE, Dekker J, Uitdehaag BM, Polman CH, Collette EH, Cuijpers P, et al. Internet-based self-help treatment for depression in multiple sclerosis: study protocol of a randomized controlled trial. BMC Psychiatry 2012 Sep 11;12:137 [FREE Full text] [doi: 10.1186/1471-244X-12-137] [Medline: 22967202]
37. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. Br Med J 2013 Feb 26;346:f540 [FREE Full text] [doi: 10.1136/bmj.f540] [Medline: 23444423]
38. Gega L, Kenwright M, Mataix-Cols D, Cameron R, Marks IM. Screening people with anxiety/depression for suitability for guided self-help. Cogn Behav Ther 2005;34(1):16-21. [doi: 10.1080/16506070410015031] [Medline: 15844684]
39. Breedlove G, Fryzelka D. Depression screening during pregnancy. J Midwifery Womens Health 2011;56(1):18-25. [doi: 10.1111/j.1542-2011.2010.00002.x] [Medline: 21323846]
40. Tandon SD, Cluxton-Keller F, Leis J, Le H, Perry DF. A comparison of three screening tools to identify perinatal depression among low-income African American women. J Affect Disord 2012 Jan;136(1-2):155-162 [FREE Full text] [doi: 10.1016/j.jad.2011.07.014] [Medline: 21864914]
41. van Ballegooijen W, Riper H, Cuijpers P, van Oppen P, Smit JH. Validation of online psychometric instruments for common mental health disorders: a systematic review. BMC Psychiatry 2016 Feb 25;16:45 [FREE Full text] [doi: 10.1186/s12888-016-0735-7] [Medline: 26915661]
42. Whitehead L. Methodological issues in internet-mediated research: a randomized comparison of internet versus mailed questionnaires. J Med Internet Res 2011 Dec 4;13(4):e109 [FREE Full text] [doi: 10.2196/jmir.1593] [Medline: 22155721]
43. Olssøn I, Mykletun A, Dahl AA. The Hospital Anxiety and Depression Rating Scale: a cross-sectional study of psychometrics and case finding abilities in general practice. BMC Psychiatry 2005 Dec 14;5:46 [FREE Full text] [doi: 10.1186/1471-244X-5-46] [Medline: 16351733]
44. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
45. Bergink V, Kooistra L, den Berg MP, Wijnen H, Bunevicius R, van Baar A, et al. Validation of the Edinburgh Depression Scale during pregnancy. J Psychosom Res 2011 Apr;70(4):385-389. [doi: 10.1016/j.jpsychores.2010.07.008] [Medline: 21414460]

46. Kozinszky Z, Dudas RB. Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period. J Affect Disord 2015 May 1;176:95-105. [doi: 10.1016/j.jad.2015.01.044] [Medline: 25704562]
47. Bouwmans C, Hakkaart-van Royen L. Research Gate. Rotterdam: Erasmus School of Health Policy & Management; 2002. Manual Trimbos/iMTA Questionnaire for Costs Associated with Psychiatric Illness (in Dutch) URL: https://www.researchgate.net/publication/254758194_Manual_TrimbosiMTA_Questionnaire_for_Costs_Associated_with_Psychiatric_Illness_TIC-P_in_Dutch [accessed 2020-01-21]
48. de Brey H. A cross-national validation of the client satisfaction questionnaire: the Dutch experience. Eval Program Plann 1983;6(3-4):395-400. [doi: 10.1016/0149-7189(83)90018-6] [Medline: 10267266]
49. StatLine. The Hague, The Netherlands: CBS StatLine; 2019. Birth; key figures URL: https://opendata.cbs.nl/statline/#/CBS/en/dataset/37422eng/table?ts=1579432841643 [accessed 2020-01-19]
50. Loughnan SA, Sie A, Hobbs MJ, Joubert AE, Smith J, Haskelberg H, et al. A randomized controlled trial of 'MUMentum Pregnancy': Internet-delivered cognitive behavioral therapy program for antenatal anxiety and depression. J Affect Disord 2019 Jan 15;243:381-390. [doi: 10.1016/j.jad.2018.09.057] [Medline: 30266030]
51. Spijker J, de Graaf R, Bijl RV, Beekman AT, Ormel J, Nolen WA. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry 2002 Sep;181:208-213. [doi: 10.1192/bjp.181.3.208] [Medline: 12204924]
52. Dennis C, Dowswell T. Psychosocial and psychological interventions for preventing postpartum depression. Cochrane Database Syst Rev 2013 Feb 28(2):CD001134. [doi: 10.1002/14651858.CD001134.pub3] [Medline: 23450532]

Abbreviations

CBT: cognitive behavioral treatment
CES-D: Center for Epidemiological Studies Depression scale
CSQ-8: Client Satisfaction Questionnaire
EPDS: Edinburgh Postnatal Depression Scale
HADS-A: Hospital Anxiety and Depression Scale-Anxiety
PST: problem solving treatment
LMM: linear mixed model
VAS: visual analog scale

Edited by G Eysenbach; submitted 25.06.19; peer-reviewed by H O'Mahen, T Muto, M Brouwer; comments to author 03.08.19; revised version received 26.10.19; accepted 16.12.19; published 23.03.20.

Please cite as:
Heller HM, Hoogendoorn AW, Honig A, Broekman BFP, van Straten A
The Effectiveness of a Guided Internet-Based Tool for the Treatment of Depression and Anxiety in Pregnancy (MamaKits Online): Randomized Controlled Trial
J Med Internet Res 2020;22(3):e15172
URL: http://www.jmir.org/2020/3/e15172/
doi:10.2196/15172
PMID:32202505

Original Paper

The Association of Therapeutic Alliance With Long-Term Outcome in a Guided Internet Intervention for Depression: Secondary Analysis From a Randomized Control Trial

Juan Martín Gómez Penedo[1], PhD; Anna Margarete Babl[1], PhD; Martin grosse Holtforth[2], PhD; Fritz Hohagen[3], MD; Tobias Krieger[1], PhD; Wolfgang Lutz[4], PhD; Björn Meyer[5], PhD; Steffen Moritz[6], PhD; Jan Philipp Klein[3], MD; Thomas Berger[1], PhD

[1]Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland
[2]Department of Psychosomatic Medicine, Inselspital, University of Bern, Bern, Switzerland
[3]Department of Psychiatry and Psychotherapy, Lübeck University, Lübeck, Germany
[4]Department of Clinical Psychology and Psychotherapy, University of Trier, Trier, Germany
[5]Department of Psychology, City, University of London, London, United Kingdom
[6]Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:
Juan Martín Gómez Penedo, PhD
Department of Clinical Psychology and Psychotherapy
University of Bern
Fabrikstrasse 8
Bern, 3012
Switzerland
Phone: 41 31 631 5424
Email: jmgomezpenedo@gmail.com

Abstract

Background: Therapeutic alliance has been well established as a robust predictor of face-to-face psychotherapy outcomes. Although initial evidence positioned alliance as a relevant predictor of internet intervention success, some conceptual and methodological concerns were raised regarding the methods and instruments used to measure the alliance in internet interventions and its association with outcomes.

Objective: The aim of this study was to explore the alliance-outcome association in a guided internet intervention using a measure of alliance especially developed for and adapted to guided internet interventions, showing evidence of good psychometric properties.

Methods: A sample of 223 adult participants with moderate depression received an internet intervention (ie, Deprexis) and email support. They completed the Working Alliance Inventory for Guided Internet Intervention (WAI-I) and a measure of treatment satisfaction at treatment termination and measures of depression severity and well-being at termination and 3- and 9-month follow-ups. For data analysis, we used two-level hierarchical linear modeling that included two subscales of the WAI-I (ie, *tasks and goals* agreement with the program and *bond* with the supporting therapist) as predictors of the estimated values of the outcome variables at the end of follow-up and their rate of change during the follow-up period. The same models were also used controlling for the effect of patient satisfaction with treatment.

Results: We found significant effects of the *tasks and goals* subscale of the WAI-I on the estimated values of residual depressive symptoms (γ_{02}=−1.74, standard error [SE]=0.40, 95% CI −2.52 to −0.96, t_{206}=−4.37, P<.001) and patient well-being (γ_{02}=3.10, SE=1.14, 95% CI 0.87-5.33, t_{198}=2.72, P=.007) at the end of follow-up. A greater score in this subscale was related to lower levels of residual depressive symptoms and a higher level of well-being. However, there were no significant effects of the *tasks and goals* subscale on the rate of change in these variables during follow-up (depressive symptoms, P=.48; patient well-being, P=.26). The effects of the *bond* subscale were also nonsignificant when predicting the estimated values of depressive symptoms and well-being at the end of follow-up and the rate of change during that period (depressive symptoms, P=.08; patient well-being, P=.68).

Conclusions: The results of this study point out the importance of attuning internet interventions to patients' expectations and preferences in order to enhance their agreement with the tasks and goals of the treatment. Thus, the results support the notion that responsiveness to a patient's individual needs is crucial also in internet interventions. Nevertheless, these findings need to be replicated to establish if they can be generalized to different diagnostic groups, internet interventions, and supporting formats.

(J Med Internet Res 2020;22(3):e15824) doi:10.2196/15824

KEYWORDS

internet interventions; guidance; alliance; Working Alliance Inventory for Guided Internet Intervention; only interventions; tasks and goals; bond

Introduction

Several meta-analyses positioned therapeutic alliance as a robust predictor of outcomes in face-to-face psychotherapy [1-3]. However, alliance effects do not seem to be limited to the field of traditional psychotherapy. Alliance also predicted outcomes in other health-related interventions, such as pharmacotherapy [4,5]. The increasing development of internet interventions and the evidence for their efficacy and effectiveness in treating diverse mental disorders [6-8] raised the question of what role therapeutic alliance might play in such treatments, especially in those providing guidance from trained supporters (guided internet interventions) [9]. In the last years, several studies addressed this question scientifically. Although some authors found that the alliance could be less important in internet interventions than in traditional face-to-face therapies [10], a recent meta-analysis reported similar effect sizes (r=0.275) for the alliance-outcome relationship in online interventions as in traditional face-to-face therapies [1]. Among the 18 studies included in that meta-analysis [1], 15 analyzed the alliance specifically in guided internet interventions. To measure the alliance, most of the studies used the same instruments usually administered in face-to-face psychotherapy (ie, the Working Alliance Inventory [WAI]) [11] with very slight modifications (eg, talking about a treatment instead of therapy) and focused on the effects of the relationship between the patient and the supporting therapist [10,12-21]. This approach of measuring alliance follows its classical conceptualization in psychotherapy research as tripartite, consisting of (1) the patient-therapist emotional bond, (2) patient agreement with the tasks of therapy, and (3) patient agreement with the therapeutic goals that the patient and therapist seek in treatment [22,23]. However, it was not considered that in these trials, treatment tasks and goals were not set in collaboration with the supporting therapist but were proposed by the online program, which might be a limiting factor. Some studies did this the other way around; they used adapted instruments to measure the alliance between the patient and the online intervention only [18,24-26].

Regardless of whether the measuring instruments focused on the supporting therapist or the program, all the abovementioned studies have one thing in common. They lack an exploration of the psychometric properties (ie, validity and reliability) of the measuring instruments used in the specific context of guided online interventions. Only Kiluk et al [24] presented an exploratory analysis of the adapted version of the WAI (ie, WAI-Tech) with some evidence of internal consistency (Cronbach alpha) and external validity (ie, significant correlations with the original WAI and with patient treatment satisfaction). However, the small sample size of that study (n=34) limited the interpretability of the findings and prevented the provision of further evidence for the psychometric properties of the scale (eg, construct validity). Thus, beside conceptual issues identified in most previous alliance-outcome studies of guided internet interventions, concerns might be raised regarding the validity and reliability of the instruments used to measure the alliance construct.

In this context, Berger et al [27], as well as Scherer et al [28], presented a compromise between the two previous approaches of analyzing alliance-outcome associations in guided internet interventions (focus on the supporting therapist or the online program). The authors took the original version of the WAI and adapted it to guided internet interventions, exploring the bond with the supporting therapist but the tasks and goals with the online program. With this version of the instrument, they captured the most relevant aspects of the alliance, considering both the importance of the relationship with the supporting therapist and patient attunement with the online program. Recently, Gómez Penedo et al [29] systematized the efforts by Berger et al [27] and Scherer et al [28], presenting the Working Alliance Inventory for Guided Internet Interventions (WAI-I) and exploring the psychometric properties of the scale. The findings provide evidence for adequate internal consistency, external validity, and construct validity (based on a confirmatory factor analysis) of the WAI-I.

Given the cumulative evidence showing that residual depressive symptoms are some of the main predictors of relapse [30-32], in this study, we will analyze how the alliance during treatment is associated with long-term outcomes (ie, 9-month follow-up) after a guided internet intervention for patients with moderate depressive symptoms. We will focus on both analyzing the effects of the alliance on the changes produced during the follow-up period (ie, deterioration or further improvement) and evaluating the residual depressive symptoms at the 9-month follow-up. Furthermore, we will analyze the same effects on the well-being of patients. Beside responding to a general call for further studies clarifying the role of alliance in online interventions [1,9], especially with adapted and psychometrically sound instruments [29], the aim of this study was to explore the alliance-outcome association using the WAI-I, a measure of alliance especially developed for and adapted to guided internet interventions.

Methods

Participants

For this study, we drew on a dataset from the EVIDENT study [33,34], a large multicenter randomized controlled trial analyzing the effects of the internet intervention Deprexis [35]. In that trial, 509 patients were assigned to the online intervention. Of these 509 participants, 317 presented with moderate depressive symptoms (ie, score between 10 and 14 in the Patient Health Questionnaire [PHQ]) [36] and therefore additionally received weekly email support from trained clinicians. In our study, we analyzed a sample of 223 patients who received both Deprexis and email support, and completed the WAI-I at treatment termination. Patients in this study were aged between 18 and 65 years (mean 44.48 [SD 10.68] years) and were German speaking. Furthermore, most participants were female (157/223, 70.4%), had high school education (107/223, 48.0%), and were in a romantic relationship at the beginning of the treatment (151/223, 67.7%). The exclusion criteria were a lifetime diagnosis of a bipolar disorder or schizophrenia and acute suicidality established by a telephone diagnostic interview.

Internet Intervention

Deprexis is an internet intervention that demonstrated effectiveness when treating depression, showing medium effect sizes at posttreatment [37]. Deprexis has 10 modules (in addition to one introductory and one summary module) developed consistent with cognitive-behavioral treatment manuals. Within the modules, the program provides simulated dialogues, explaining different techniques and key concepts and delivering examples, and illustrations for easy understanding. Participants were requested to complete different exercises within the program and provide feedback in order to attune the interventions to each participant. Although the modules were presented sequentially, patients could repeat them as often as they wanted. In this sample, the participants spent a mean of 520 (SD 314) minutes in the program and completed a mean of 9.74 (SD 4.51) modules.

All the participants in this sample received standardized email support that consisted of weekly feedback regarding their activity on Deprexis during the last week. The main goal of this support was to enhance participants' motivation and engagement with the internet intervention. The support was implemented via a secured email system included in the internet intervention and was delivered by master's students in clinical psychology and psychotherapy, psychotherapists in training, and licensed psychotherapists who received an intensive 4-hour training in the program and feedback strategies, using example cases. The instructions provided to the supporters were in line with those used in a similar previous trial [38]. An expert on internet interventions supervised their tasks by periodically revising the messages from the supporters and providing feedback to them. The study participants were able to contact the supporters directly or respond to their messages.

The sample of this study received a mean of 12.11 (SD 3.23) messages from the supporting clinicians and read a mean of 9.65 (SD 4.88) of those messages. Additionally, the patients sent a mean of 1.99 (SD 2.84) messages to the supporting clinicians (54.7% of the sample sent at least one message). Further details on the internet intervention Deprexis are presented in articles by Meyer et al [35] and Klein et al [33,34].

Measures

Working Alliance Inventory for Guided Internet Interventions

The WAI-I is an instrument derived from the Working Alliance Inventory-Short revised [39], and it was specifically adapted to guided internet interventions [27,29]. The WAI-I is a 12-item self-reported measure rated on a 5-point Likert scale ranging from 1 (never) to 5 (always). The instrument has two dimensions. One dimension explores the emotional bond between the patient and the supporting therapist, with items like "I feel that the psychologist who supports me in the online program appreciates me." The second dimension analyzes patient agreement with the tasks and goals of the internet intervention, with items like "I believe the way the online program is working with my problem is correct" and "The goals of the online program are important goals for me." The instrument showed evidence of adequate internal consistency, external validity, and construct validity for a two-factor solution (based on a confirmatory factor analysis) [29]. In this sample, the Cronbach alpha of the *bond* subscale was .89, whereas that of the *tasks and goals* subscale was .93.

Patient Satisfaction Questionnaire (Zurich Satisfaction Questionnaire-8)

To measure patient satisfaction with treatment, we used the Zurich Satisfaction Questionnaire-8 (ZUF-8) questionnaire [40] adapted for internet interventions [33]. This instrument is a self-reported measure of eight items rated on a 4-point Likert scale ranging from 1 (low satisfaction) to 4 (high satisfaction). The ZUF-8 adapted to the German culture showed good psychometric properties with evidence of internal consistency, concurrent validity, and construct validity [40]. In the sample of this study, the ZUF-8 showed a Cronbach alpha of .92.

Patient Health Questionnaire-9

The PHQ-9 is a widely used outcome measure for the treatment of depression [41]. It has nine self-reported items representing the nine Diagnostic and Statistical Manual of Mental Disorders-version IV criteria of depression that are completed on a 4-point Likert scale from 0 (not at all) to 3 (nearly every day). Previous studies showed that it is a reliable and valid instrument to measure depression severity [41]. In this study, the PHQ-9 presented good internal consistency during follow-up, with a Cronbach alpha of .82.

Short-Form Health Survey-12

The Short-Form Health Survey-12 (SF-12) is an instrument for measuring health-related quality of life [42]. For this study, we used the mental health subscale (SF-P) of this measure, which consisted of six items, with higher scores representing a higher quality of life in terms of mental health. The items from SF-12 have a Likert scale that varies from 3 to 6 response categories, depending on the item. The individual item responses are then transformed into a 0 to 100 scale and then aggregated into different dimensions or subscales, with higher scores representing greater health well-being. Thereafter, the aggregated scores are standardized according to a normative population by computing *t*-scores that have a mean of 50 and an SD of 10 [34]. The SF-12 showed evidence of reliability and validity [42]. In this sample, the SF-P presented good internal consistency, with a Cronbach alpha of .84.

Procedure

Patients completed the alliance measure (WAI-I) only once at posttreatment. We decided to measure alliance only at the end of therapy, as some studies showed that patients might have difficulties to complete it early in treatment, because of the limited interaction with the supporter during the intervention [16]. At treatment termination, patients also completed the ZUF-8 as a general measure of patient satisfaction. Furthermore, they completed the PHQ-9 and SF-12 at posttreatment and at 3- and 9-month follow-ups. The Ethics Committee of the German Psychological Association approved the procedure of the study (Deutsche Gesellschaft für Psychologie, reference number SM 04_2012). All patients completed an electronic informed consent form before baseline assessments.

Analytic Strategy

For the analyses in this study, we used hierarchical linear models (HLMs) to deal with the dependency of the observations owing to the nestedness of the data [43]. Considering that repeated measures during follow-up were nested within patients, we ran two-level HLMs, accounting for within-patient and between-patient variabilities. These models accommodate missing data, allowing to retain in the analyses all patients with at least one measurement point, which mimics an intent-to-treat approach.

We first ran two-level fully unconditional models with PHQ-9 and SF-12 values during follow-up as the dependent variables. In the next step, we ran an unconditional time-as-only predictor model, with time as a level 1 predictor centered at the 9-month follow-up and representing the evolution during the follow-up period (posttreatment=−1; end of follow-up=0). Thereafter, we ran a conditional model that included WAI-I scores in the *bond* and *tasks* and *goals* subscales as separate level 2 predictors of the intercept (ie, estimated score of the outcome variables at the 9-month follow-up) and the linear slope of time (ie, evolution of the outcome variables during follow-up). Finally, we ran the exact same models but included either (1) patient satisfaction with treatment or (2) participant-supporter interaction indicators (ie, number of messages sent by the participant, number of messages sent by the supporter, and number of messages read by the participant) as a level 2 predictor to control for its effect.

Results

Sample Details

To characterize the sample, we calculated the mean and SD of each of the targeted variables at posttreatment. We have presented these descriptive statistics in Table 1.

Furthermore, in Multimedia Appendix 1, we present the correlations among these variables at posttreatment. Beside these correlations among the targeted variables of the study, we found significant correlations between the use of the program in minutes and both the *bond* subscale (*r*=0.18, *P*=.01) and the *tasks and goals* subscale (*r*=0.16, *P*=.02) of the WAI-I. In addition, the number of modules performed by the participants was significantly related to the *bond* subscale (*r*=0.22, *P*=.001) and the *tasks and goals* subscale (*r*=0.15, *P*=.02).

Table 1. Sample details at posttreatment for the variables in the study.

Measures	Value, mean (SD)
WAI-I[a]	
T&G[b] subscale	3.17 (0.91)
Bond subscale	3.56 (1.15)
ZUF-8[c]	
Total scale	3.13 (0.59)
PHQ-9[d]	
Total scale	7.44 (4.32)
SF-12[e]	
SF-P[f]	38.22 (11.75)

[a] WAI-I: Working Alliance Inventory for Internet Interventions.

[b] T&G: task and goal subscale.

[c] ZUF-8: Zurich Satisfaction Questionnaire-8.

[d] PHQ-9: Patient Health Questionnaire-9.

[e] SF-12: Short-Form Health Survey-12.

[f] SF-P: mental health well-being subscale of the Short-Form Health Survey-12.

Fully Unconditional Model

We present the results of all conducted models in Table 2. The fully unconditional model estimated a mean level of 7.14 units in the PHQ-9 during follow-up (γ_{00}=7.14, standard error [SE]=0.25, 95% CI 6.65-7.63, t_{219}=64.05, *P*<.001). Furthermore, the model for SF-P estimated a mean level of 39.23 units during follow-up (γ_{00}=39.23, SE=0.68, 95% CI 37.90-40.56, t_{217}=57.65, *P*<.001).

Table 2. Results of the unconditional, time-as-only predictor, and conditional main effect hierarchical linear models.

Fixed effects	Estimated score at the end of follow-up (β_{0j})		Change during follow-up (β_{1j})	
	γ	SE[a]	γ	SE
PHQ[b] as outcome				
Fully unconditional				
Intercept	7.14[c]	0.25	—[d]	—
Time-as-only predictor				
Intercept	6.85[c]	0.29	−0.49[e]	0.27
Main effects of alliance				
Intercept	6.84[c]	0.28	−0.49[e]	0.27
WAI-I[f] T&G[g]	−1.74[c]	0.40	0.27	0.39
WAI-I bond	0.57[e]	0.32	0.54[e]	0.31
SF-P[h] as outcome				
Fully unconditional				
Intercept	39.23[c]	0.68	—	—
Time-as-only predictor				
Intercept	40.04[c]	0.80	1.33[e]	0.75
Main effects of alliance				
Intercept	40.19[c]	0.79	1.39[e]	0.76
WAI-I T&G	3.10[i]	1.14	−1.26	1.10
WAI-I bond	−0.74	0.91	−0.36	0.89

[a] SE: standard error.

[b] PHQ: Patient Health Questionnaire.

[c] *P*<.001.

[d] Not applicable.

[e] *P*<.10.

[f] WAI-I: Working Alliance Inventory for Internet Interventions.

[g] T&G: task and goal subscale.

[h] SF-P: mental health well-being subscale of the Short-Form Health Survey-12.

[i] *P*<.01.

Unconditional Time-as-Only Predictor Model

When predicting PHQ-9 scores, the inclusion of the time variable as a level 1 predictor significantly improved the fit of the fully unconditional model (χ^2_3=9.70, *P*=.02). The time-as-only predictor model for PHQ-9 estimated a residual depression symptom score of 6.85 at the 9-month follow-up (γ_{00}=6.85, SE=0.29, 95% CI 6.28-7.42, t_{209}=54.84, *P*<.001). This model also showed that the change in the PHQ-9 score during follow-up approached significance (γ_{10}=−0.49, SE=0.27, 95% CI −1.02 to 0.04, t_{203}=−1.80, *P*=.08). The computation of the CIs for the random effects showed significant random effects for both the estimated residual depressive symptoms at the 9-month follow-up (SD 3.33, 95% CI 2.91-3.98) and the change during that period (SD 1.73, 95% CI 0.74-2.80). The results revealed that the findings of the participants significantly varied around the average estimated residual depressive symptoms at the end of the 9-month follow-up and the average rate of change during follow-up, suggesting the inclusion of level 2 predictors to explain this variance.

For the models predicting SF-P, the inclusion of the time variable as a level 1 predictor did not significantly increase the fit of the fully unconditional model (χ^2_3=4.17, *P*=.24). This time-as-only predictor model showed an estimated value of 40.04 for the SF-P at the 9-month follow-up (γ_{00}=40.04, SE=0.80, 95% CI 38.47-41.61, t_{200}=50.19, *P*<.001). Furthermore, the change in the SF-P during follow-up approached significance (γ_{10}=1.33, SE=0.75, 95% CI −0.14 to 2.80, t_{190}=1.77, *P*=.08). The calculation of CIs showed significant random effects for both the estimated well-being at the end of follow-up (SD 8.88, 95% CI 7.35-10.55) and the rate of change during follow-up (SD 3.20, 95% CI 0.09-6.32). Thus, the findings of the participants significantly varied around the average estimated well-being at the end of follow-up and the average change during follow-up, suggesting the inclusion of level 2 predictors to explain this variance.

Conditional Models: Alliance Main Effects

The conditional model with the PHQ-9 as an outcome variable and the alliance subscales as level 2 predictors significantly improved the fit of the time-as-only predictor model (χ^2_4=26.71, *P*<.001). This model showed a significant effect of the *tasks and goals* subscale on the estimated PHQ-9 value at the end of follow-up (γ_{02}=−1.74, SE=0.40, 95% CI −2.52 to −0.96, t_{206}=−4.37, *P*<.001). A 1-unit greater score in the *tasks and goals* subscale was associated with a 1.74 lower score in the PHQ-9 at the end of follow-up. However, there was no significant effect of the *tasks and goals* subscale on PHQ-9 change during follow-up (γ_{12}=0.27, SE=0.39, 95% CI −0.49 to 1.03, t_{197}=0.71, *P*=.48). Additionally, the *bond* subscale did not have a significant effect on the estimated score of the PHQ-9 at the end of follow-up (γ_{01}=0.57, SE=0.32, 95% CI −0.06 to 1.20, t_{204}=1.79, *P*=.07) or the change in the PHQ-9 during follow-up (γ_{11}=0.54, SE=0.31, 95% CI −0.07 to 1.15, t_{197}=1.75, *P*=.08).

Furthermore, the conditional model with the SF-P as the outcome variable and the alliance subscales as level 2 predictors significantly improved the model fit as compared with the time-as-only predictor model (χ^2_4=20.59, *P*<.001). This conditional model also showed a significant effect of the *tasks and goals* subscale on the estimated SF-P score at the end of follow-up (γ_{02}=3.10, SE=1.14, 95% CI 0.87-5.33, t_{198}=2.72, *P*=.007). A 1-unit greater *tasks and goals* score at posttreatment was associated with a 3.10-unit higher score in the SF-P at the end of the 9-month follow-up. There was no significant effect of the *tasks and goals* subscale on the development of the SF-P during follow-up (γ_{12}=−1.26, SE=1.10, 95% CI −3.42 to 0.90, t_{188}=−1.14, *P*=.26). Moreover, there was no significant effect of the *bond* subscale on the estimated score of the SF-P at the end of follow-up (γ_{01}=−0.74, SE=0.91, 95% CI −2.52 to 1.04, t_{191}=−0.82, *P*=.41) or the evolution of the SF-P during follow-up (γ_{11}=−0.36, SE=0.89, 95% CI −2.10 to 1.38, t_{191}=0.41, *P*=.68).

Conditional Models: Alliance Main Effects Controlling for Patient Satisfaction and Patient-Supporter Interaction

The results of the conditional models estimating the alliance effects controlling for either patient satisfaction or patient-supporter interaction are presented in Multimedia Appendix 2. When running the same conditional models to predict PHQ-9 scores presented above on controlling for patient satisfaction with treatment, there was a significant improvement in the conditional model fit (χ^2_2=8.61, *P*=.01). This model comparison test suggested the importance of controlling for patient satisfaction when estimating the alliance effects on the PHQ-9. The results of this model showed that there was still a significant effect of the *tasks and goals* subscale on the estimated PHQ-9 value at the end of follow-up (γ_{02}=−1.78, SE=0.58, 95% CI −2.92 to −0.64, t_{203}=−3.09, *P*=.002). The other effects of the alliance were nonsignificant as in the previous models. Furthermore, the effect of patient satisfaction was not significant when predicting the estimated PHQ-9 scores at the 9-month follow-up (γ_{03}=0.09, SE=0.85, 95% CI −1.58 to 1.76, t_{200}=0.11, *P*=.91) but was significant when predicting the change produced in the PHQ-9 during follow-up (γ_{13}=2.00, SE=0.81, 95% CI 0.41-3.59, t_{193}=2.48, *P*=.58). When controlling for the alliance subscale effects, a 1-unit greater score in the ZUF-8 at posttreatment (ie, patient satisfaction with treatment) was associated with a 2-unit increase in the PHQ-9 score during the follow-up period.

However, when running a conditional model exploring alliance effects on the PHQ controlling for patient-supporter interaction indicators (ie, number of messages sent by the participant, number of messages sent by the supporter, and number of messages read by the participant), there was no significant improvement in the model fit (χ^2_6=0.56, *P*=.99). This test suggested not to include participant-supporter interactions as covariates in conditional alliance models. Nevertheless, it is worth highlighting that the model controlling for participant-supporter interaction indicators still showed significant effects of the *tasks and goals* subscale on the estimated PHQ-9 value at the end of follow-up (γ_{02}=−1.97, SE=0.44, 95% CI −2.82 to −1.12, t_{141}=−4.49, *P*<.001).

Furthermore, inclusion of patient satisfaction in the conditional models predicting SF-P did not significantly improve the model fit from the conditional models that included only alliance subscales (χ^2_2=0.39, *P*=.82). This model comparison test again suggested not to include patient satisfaction when estimating alliance effects on the SF-P during follow-up, keeping as the final models the conditional models introduced in the section presented above (ie, conditional models with alliance-only main effects). As can be seen in Multimedia Appendix 1, the results of the model predicting the SF-P and controlling for patient satisfaction suggested no significant effects of either alliance subscales or patient satisfaction.

Additionally, the models controlling for participant-supporter interaction did not improve the conditional model fit when predicting the SF-P (χ^2_6=6.43, *P*=.38). Nonetheless, the results of the models controlling for participant-supporter interaction still presented a significant effect of the *tasks and goals* subscale on the estimated SF-*P* value at the end of follow-up (γ_{02}=3.09, SE=1.30, 95% CI 0.58-5.61, t_{139}=2.38, *P*=.02).

Discussion

Responding to a general call for further studies clarifying the role of alliance in online interventions [1,9], particularly with psychometrically sound instruments [29], the aim of this study was to analyze the alliance-outcome association in a guided internet intervention for participants with moderate depression, using an adapted version of the WAI to measure the alliance in this type of approach. The results of the model showed significant effects of tasks and goals on the estimated scores of the PHQ-9 and SF-P at the end of the 9-month follow-up. These results suggest that when participants report a greater agreement with the therapeutic activities and goals proposed by the internet intervention at posttreatment, their residual depressive symptoms are lower and psychological well-being is higher at the end of follow-up. These findings are in line with the results of several studies showing the overall importance of an alliance for internet interventions [1] and the specific relevance of the alliance with an online program for treatment outcomes [18,25]. However, different from previous studies that analyzed overall alliance with internet interventions, according to suggestions from several authors [9,27,28], in this study, alliance was measured disaggregating the bond with trained supporters and the agreement with internet interventions regarding tasks and goals. Furthermore, in this study, the effects of the alliance were associated with residual depressive symptoms and patient well-being 9 months after treatment termination. However, the *tasks and goals* subscale did not exhibit a significant effect on the rate of change during the follow-up period.

Additionally, the results of this study did not show significant effects of the *bond* subscale on PHQ-9 or SF-P-estimated values at the end of follow-up or the rate of change during the follow-up period. These findings are in line with theories suggesting that in internet interventions, the agreement of tasks and goals with the intervention might be more relevant than the bond with trained supporters [9,27].

In conclusion, the results of this study point out the importance of attuning internet interventions with patients' expectations and preferences in order to enhance their agreement with the tasks and goals of the treatment. Being responsive to patients' needs has been presented as a fundamental process of change in face-to-face psychotherapy [44-46].

The results of this study further support this notion in internet interventions. Some online programs, such as Deprexis, use patient feedback to select specific content presented to patients [35]. However, this treatment personalization is only within modules to treat depression. Considering the high rates of comorbidities in patients with depression [47], one extension that could enhance the intervention is incorporation of modules that address other relevant symptomatologies in these patients (eg, anxiety) [27]. In the last years, there were several developments in this direction, tailoring internet interventions to the symptom profile of the patient and offering individually prescribed treatment modules accordingly [27,48-50]. The inclusion of these personalization strategies to account for possible and very likely comorbidities beyond depressive symptoms might enhance participant agreement with the tasks and goals of the program, because the intervention would further include meaningful activities to address relevant problems for the patients. Indeed, in a study comparing tailored and standardized disorder-specific internet interventions, participants who received the tailored condition rated the agreement of the tasks and goals with the program substantially and significantly higher ($P<.001$) as compared with those who received the standardized condition [27]. Future studies might need to further explore evidence-based responsiveness and treatment goals according to patient markers and analyze their associations with internet intervention processes, such as agreement on the tasks and goals of the intervention, and with acute as well as long-term outcomes.

Several limitations characterize this study. For instance, with only one assessment point for the alliance, differential effects of the general level of the alliance during treatment (ie, between-patient effects or trait-like components of alliance) and effects of the modifications of the alliance during the intervention (ie, within-patient effects or state-like components of alliance) could not be established [51,52]. Furthermore, as shown by Crits Christoph et al [53], a single measure of alliance might be less reliable than an aggregation of several measurements. The reason for assessing alliance only once at posttreatment was that contact with the supporting therapist was minimal and an adequate dose of interaction between the patient and the supporting therapist was necessary for reliable evaluation. However, this argument applies mainly to the *bond* subscale but not necessarily the *tasks and goals* agreement subscale with internet interventions (the component of the alliance that showed significant effects on long-term outcomes). Future studies would benefit from the analysis of the agreement of the tasks and goals with the intervention early in treatment and the use of repeated measures that would allow for a more fine-grained and sound analysis of the association between alliance and outcome in guided internet interventions. Additionally, although we measured alliance only at posttreatment to maximize contact between the participants and supporters before the assessment, their contact during the intervention might be too limited to fully capture the potential effects of the bond with the supporter. Future studies would need to further explore the effects of the therapeutic relationship or bond with the supporter in internet interventions where there is greater participant-supporter contact. Furthermore, the analyses of the study were conducted using a subsample of the EVIDENT randomized controlled trial (ie, patients with moderate depression who received Deprexis with email support). Further studies will need to explore whether the association between task and goal agreement and long-term outcomes can be generalized to other populations with different diagnoses (eg, anxiety disorders), different levels of severity (eg, mild or severe depression), other internet interventions, and other forms of therapist support (eg, phone support).

Despite these limitations, our study presents evidence supporting the importance of the agreement of the tasks and goals with the program in guided internet interventions for long-term outcomes. The results could inform the field to improve guided internet interventions by developing responsive strategies to adapt the interventions to patients' specific expectations and needs and thus enhance treatment outcomes.

Conflicts of Interest

JPK received funding for clinical trials (German Federal Ministry of Health, Servier), payments for presentations on internet interventions (Servier), and payments for workshops and books (Beltz, Elsevier, Hogrefe, and Springer) on psychotherapy for chronic depression and psychiatric emergencies. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1
Correlations between the targeted variables of the study at posttreatment.
[DOCX File , 15 KB - jmir_v22i3e15824_app1.docx]

Multimedia Appendix 2
Results of the hierarchical linear models of the conditional main effects controlling for satisfaction or participant-supporter interaction.
[DOCX File , 20 KB - jmir_v22i3e15824_app2.docx]

References

1. 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]
2. Horvath AO, Del Re AC, Flückiger C, Symonds D. Alliance in individual psychotherapy. Psychotherapy (Chic) 2011 Mar;48(1):9-16. [doi: 10.1037/a0022186] [Medline: 21401269]
3. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J Consult Clin Psychol 2000 Jun;68(3):438-450. [Medline: 10883561]
4. Krupnick JL, Sotsky SM, Simmens S, Moyer J, Elkin I, Watkins J, et al. The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol 1996 Jun;64(3):532-539. [doi: 10.1037//0022-006x.64.3.532] [Medline: 8698947]
5. Zuroff DC, Blatt SJ. The therapeutic relationship in the brief treatment of depression: contributions to clinical improvement and enhanced adaptive capacities. J Consult Clin Psychol 2006 Feb;74(1):130-140. [doi: 10.1037/0022-006X.74.1.130] [Medline: 16551150]
6. Andersson G, Carlbring P, Titov N, Lindefors N. Internet Interventions for Adults with Anxiety and Mood Disorders: A Narrative Umbrella Review of Recent Meta-Analyses. Can J Psychiatry 2019 Jul;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
7. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
8. 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]
9. Berger T. The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
10. Andersson G, Paxling B, Wiwe M, Vernmark K, Felix CB, Lundborg L, et al. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. Behav Res Ther 2012 Sep;50(9):544-550. [doi: 10.1016/j.brat.2012.05.003] [Medline: 22728647]
11. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. Journal of Counseling Psychology 1989 Apr;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]
12. Anderson RE, Spence SH, Donovan CL, March S, Prosser S, Kenardy J. Working alliance in online cognitive behavior therapy for anxiety disorders in youth: comparison with clinic delivery and its role in predicting outcome. J Med Internet Res 2012 Jun 28;14(3):e88 [FREE Full text] [doi: 10.2196/jmir.1848] [Medline: 22789657]
13. Andersson E, Ljótsson B, Hedman E, Enander J, Kaldo V, Andersson G, et al. Predictors and moderators of Internet-based cognitive behavior therapy for obsessive–compulsive disorder: Results from a randomized trial. Journal of Obsessive-Compulsive and Related Disorders 2015 Jan;4:1-7 [FREE Full text] [doi: 10.1016/j.jocrd.2014.10.003]
14. Bergman Nordgren L, Carlbring P, Linna E, Andersson G. Role of the working alliance on treatment outcome in tailored internet-based cognitive behavioural therapy for anxiety disorders: randomized controlled pilot trial. JMIR Res Protoc 2013 Jan 18;2(1):e4 [FREE Full text] [doi: 10.2196/resprot.2292] [Medline: 23612437]
15. Hedman E, Andersson E, Lekander M, Ljótsson B. Predictors in Internet-delivered cognitive behavior therapy and behavioral stress management for severe health anxiety. Behav Res Ther 2015 Jan;64:49-55. [doi: 10.1016/j.brat.2014.11.009] [Medline: 25540862]

16. Jasper K, Weise C, Conrad I, Andersson G, Hiller W, Kleinstäuber M. The working alliance in a randomized controlled trial comparing Internet-based self-help and face-to-face cognitive behavior therapy for chronic tinnitus. Internet Interventions 2014 Apr;1(2):49-57. [doi: 10.1016/j.invent.2014.04.002]
17. Knaevelsrud C, Maercker A. Does the quality of the working alliance predict treatment outcome in online psychotherapy for traumatized patients? J Med Internet Res 2006 Dec 19;8(4):e31 [FREE Full text] [doi: 10.2196/jmir.8.4.e31] [Medline: 17213049]
18. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial. Internet Interventions 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
19. Preschl B, Maercker A, Wagner B. The working alliance in a randomized controlled trial comparing online with face-to-face cognitive-behavioral therapy for depression. BMC Psychiatry 2011 Dec 06;11:189 [FREE Full text] [doi: 10.1186/1471-244X-11-189] [Medline: 22145768]
20. Richards D, Timulak L, Hevey D. A comparison of two online cognitive-behavioural interventions for symptoms of depression in a student population: The role of therapist responsiveness. Counselling and Psychotherapy Research 2013 Sep;13(3):184-193. [doi: 10.1080/14733145.2012.733715]
21. Wagner B, Brand J, Schulz W, Knaevelsrud C. Online working alliance predicts treatment outcome for posttraumatic stress symptoms in Arab war-traumatized patients. Depress Anxiety 2012 Jul;29(7):646-651. [doi: 10.1002/da.21962] [Medline: 22678971]
22. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research & Practice 1979;16(3):252-260. [doi: 10.1037/h0085885]
23. Horvath A, Bedi R. The Alliance. In: Norcross JC, editor. Psychotherapy Relationships That Work. New York: Oxford University Press; 2002:37-70.
24. Kiluk BD, Serafini K, Frankforter T, Nich C, Carroll KM. Only connect: The working alliance in computer-based cognitive behavioral therapy. Behav Res Ther 2014 Dec;63:139-146 [FREE Full text] [doi: 10.1016/j.brat.2014.10.003] [Medline: 25461789]
25. Heim E, Rötger A, Lorenz N, Maercker A. Working alliance with an avatar: How far can we go with internet interventions? Internet Interv 2018 Mar;11:41-46 [FREE Full text] [doi: 10.1016/j.invent.2018.01.005] [Medline: 30135758]
26. Ormrod JA, Kennedy L, Scott J, Cavanagh K. Computerised cognitive behavioural therapy in an adult mental health service: a pilot study of outcomes and alliance. Cogn Behav Ther 2010;39(3):188-192. [doi: 10.1080/16506071003675614] [Medline: 20485996]
27. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. Psychotherapy (Chic) 2014 Jun;51(2):207-219. [doi: 10.1037/a0032527] [Medline: 24041199]
28. Scherer S, Alder J, Gaab J, Berger T, Ihde K, Urech C. Patient satisfaction and psychological well-being after internet-based cognitive behavioral stress management (IB-CBSM) for women with preterm labor: A randomized controlled trial. J Psychosom Res 2016 Jan;80:37-43. [doi: 10.1016/j.jpsychores.2015.10.011] [Medline: 26721546]
29. Gómez Penedo JM, Berger T, Grosse Holtforth M, Krieger T, Schröder J, Hohagen F, et al. The Working Alliance Inventory for guided Internet interventions (WAI-I). J Clin Psychol 2019 Jun 25. [doi: 10.1002/jclp.22823] [Medline: 31240727]
30. Bockting CL, Spinhoven P, Koeter MW, Wouters LF, Schene AH, Depression Evaluation Longitudinal Therapy Assessment Study Group. Prediction of recurrence in recurrent depression and the influence of consecutive episodes on vulnerability for depression: a 2-year prospective study. J Clin Psychiatry 2006 May;67(5):747-755. [Medline: 16841624]
31. Verhoeven FE, Wardenaar KJ, Ruhé HG, Conradi HJ, de Jonge P. Seeing the signs: Using the course of residual depressive symptomatology to predict patterns of relapse and recurrence of major depressive disorder. Depress Anxiety 2018 Feb;35(2):148-159. [doi: 10.1002/da.22695] [Medline: 29228458]
32. Wojnarowski C, Firth N, Finegan M, Delgadillo J. Predictors of depression relapse and recurrence after cognitive behavioural therapy: a systematic review and meta-analysis. Behav Cogn Psychother 2019 Sep;47(5):514-529. [doi: 10.1017/S1352465819000080] [Medline: 30894231]
33. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. The EVIDENT-trial: protocol and rationale of a multicenter randomized controlled trial testing the effectiveness of an online-based psychological intervention. BMC Psychiatry 2013 Sep 28;13:239 [FREE Full text] [doi: 10.1186/1471-244X-13-239] [Medline: 24074299]
34. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. Effects of a Psychological Internet Intervention in the Treatment of Mild to Moderate Depressive Symptoms: Results of the EVIDENT Study, a Randomized Controlled Trial. Psychother Psychosom 2016;85(4):218-228. [doi: 10.1159/000445355] [Medline: 27230863]
35. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009 May 11;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]
36. Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]

37. Twomey C, O'Reilly G, Meyer B. Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: A meta-analysis. Psychiatry Res 2017 Oct;256:371-377. [doi: 10.1016/j.psychres.2017.06.081] [Medline: 28686935]
38. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
39. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
40. Schmidt J, Lamprecht F, Wittmann WW. [Satisfaction with inpatient management. Development of a questionnaire and initial validity studies]. Psychother Psychosom Med Psychol 1989 Jul;39(7):248-255. [Medline: 2762479]
41. 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]
42. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
43. Raudenbush S, Bryk AS. Hierarchical Linear Models: Applications And Data Analysis Methods (Advanced Quantitative Techniques in The Social Sciences). Thousand Oaks: Sage Publications, Inc; 2001.
44. Constantino M, Boswell J, Bernecker S, Castonguay L. Context-responsive psychotherapy integration as a framework for a unified clinical science: Conceptual and empirical considerations. J Unified Psychother Clin Sci 2013;2:1-20.
45. Gomez Penedo JM, Constantino MJ, Coyne AE, Westra HA, Antony MM. Markers for context-responsiveness: Client baseline interpersonal problems moderate the efficacy of two psychotherapies for generalized anxiety disorder. J Consult Clin Psychol 2017 Oct;85(10):1000-1011. [doi: 10.1037/ccp0000233] [Medline: 28703605]
46. Stiles W, Honos-Webb L, Surko M. Responsiveness in psychotherapy. Clin Psychol Sci Pract 1998;5:439-458. [doi: 10.1111/j.1468-2850.1998.tb00166.x]
47. Kessler RC, Zhao S, Blazer DG, Swartz M. Prevalence, correlates, and course of minor depression and major depression in the National Comorbidity Survey. J Affect Disord 1997 Aug;45(1-2):19-30. [doi: 10.1016/s0165-0327(97)00056-6] [Medline: 9268772]
48. Andersson G, Estling F, Jakobsson E, Cuijpers P, Carlbring P. Can the patient decide which modules to endorse? An open trial of tailored internet treatment of anxiety disorders. Cogn Behav Ther 2011;40(1):57-64. [doi: 10.1080/16506073.2010.529457] [Medline: 21337215]
49. Carlbring P, Maurin L, Törngren C, Linna E, Eriksson T, Sparthan E, et al. Individually-tailored, Internet-based treatment for anxiety disorders: A randomized controlled trial. Behav Res Ther 2011 Jan;49(1):18-24. [doi: 10.1016/j.brat.2010.10.002] [Medline: 21047620]
50. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. PLoS One 2012;7(5):e36905 [FREE Full text] [doi: 10.1371/journal.pone.0036905] [Medline: 22615841]
51. Falkenström F, Finkel S, Sandell R, Rubel JA, Holmqvist R. Dynamic models of individual change in psychotherapy process research. J Consult Clin Psychol 2017 Jun;85(6):537-549. [doi: 10.1037/ccp0000203] [Medline: 28394170]
52. Zilcha-Mano S. Is the alliance really therapeutic? Revisiting this question in light of recent methodological advances. Am Psychol 2017;72(4):311-325. [doi: 10.1037/a0040435] [Medline: 28481579]
53. Crits-Christoph P, Gibbons MB, Hamilton J, Ring-Kurtz S, Gallop R. The dependability of alliance assessments: the alliance-outcome correlation is larger than you might think. J Consult Clin Psychol 2011 Jun;79(3):267-278 [FREE Full text] [doi: 10.1037/a0023668] [Medline: 21639607]

Abbreviations

PHQ-9: Patient Health Questionnaire-9
SE: standard error
SF-12: Short-Form Health Survey-12
SF-P: mental health subscale of the Short-Form Health Survey-12
WAI-I: Working Alliance Inventory for Guided Internet Interventions
ZUF-8: Zurich Satisfaction Questionnaire-8

Edited by G Eysenbach; submitted 09.08.19; peer-reviewed by K Tomoko, L Bücker, A Graham; comments to author 17.09.19; revised version received 15.10.19; accepted 24.01.20; published 24.03.20.

Please cite as:
Gómez Penedo JM, Babl AM, grosse Holtforth M, Hohagen F, Krieger T, Lutz W, Meyer B, Moritz S, Klein JP, Berger T
The Association of Therapeutic Alliance With Long-Term Outcome in a Guided Internet Intervention for Depression: Secondary Analysis From a Randomized Control Trial
J Med Internet Res 2020;22(3):e15824
URL: https://www.jmir.org/2020/3/e15824
doi:10.2196/15824
PMID:32207689

Original Paper

Brief Web-Based Intervention for Depression: Randomized Controlled Trial on Behavioral Activation

Lena Jelinek[1], PhD; Sönke Arlt[1,2], MD; Steffen Moritz[1], PhD; Johanna Schröder[1,3], PhD; Stefan Westermann[1], PhD; Barbara Cludius[1], PhD

[1]Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Department of Psychiatry and Psychotherapy, Evangelical Hospital Alsterdorf, Hamburg, Germany

[3]Institute for Sex Research, Sexual Medicine and Forensic Psychiatry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:
Lena Jelinek, PhD
Department of Psychiatry and Psychotherapy
University Medical Center Hamburg-Eppendorf
Martinistrasse 52
Hamburg, 20246
Germany
Phone: 49 407410 ext 58087
Email: ljelinek@uke.de

Abstract

Background: Web-based interventions have been shown to be effective for the treatment of depression. However, interventions are often complex and include a variety of elements, making it difficult to identify the most effective component(s).

Objective: The aim of this pilot study was to shed light on mechanisms in the online treatment of depression by comparing a single-module, fully automated intervention for depression (internet-based behavioral activation [iBA]) to a nonoverlapping active control intervention and a nonactive control group.

Methods: We assessed 104 people with at least mild depressive symptoms (Patient Health Questionnaire-9, >4) via the internet at baseline (t_0) and 2 weeks (t_1) and 4 weeks (t_2) later. After the t_0 assessment, participants were randomly allocated to one of three groups: (1) iBA (n=37), (2) active control using a brief internet-based mindfulness intervention (iMBI, n=32), or (3) care as usual (CAU, n=35). The primary outcome was improvement in depressive symptoms, as measured using the Patient Health Questionnaire-9. Secondary parameters included changes in activity, dysfunctional attitudes, and quality of life

Results: While groups did not differ regarding the change in depression from t_0 to t_1 (η_p^2=.007, P=.746) or t_0 to t_2 (η_p^2=.008, P=.735), iBA was associated with a larger decrease in dysfunctional attitudes from t_0 to t_2 in comparison to CAU (η_p^2=.053, P=.04) and a larger increase in activity from t_0 to t_1 than the pooled control groups (η_p^2=.060, P=.02). A change in depression from t_0 to t_2 was mediated by a change in activity from t_0 to t_1. At t_1, 22% (6/27) of the participants in the iBA group and 12% (3/25) of the participants in the iMBI group indicated that they did not use the intervention.

Conclusions: Although we did not find support for the short-term efficacy of the single-module iBA regarding depression, long-term effects are still conceivable, potentially initiated by changes in secondary outcomes. Future studies should use a longer intervention and follow-up interval.

Trial Registration: DKRS (#DRKS00011562)

(J Med Internet Res 2020;22(3):e15312) doi:10.2196/15312

KEYWORDS

affective disorders; depressive symptoms; brief psychotherapy; internet; world wide web

Introduction

Major depressive disorder (MDD) is a disabling disorder that affects millions of people worldwide. Besides the traditional interventions of pharmacotherapy and face-to-face psychotherapy, psychological online interventions (POI) have been established to treat patients with MDD over the last 20 years. Evidence for the efficacy of POI in the treatment of depression is accumulating [1-4], with a recent meta-analysis reporting a pooled effect size of $g=-0.90$ for interventions using an online or mobile setting, in comparison to the use of a waitlist, for adults with a confirmed diagnosis of unipolar depression [5]. The effects are similar to face-to-face treatment [6]. For subthreshold depression, however, a meta-analysis by Zhou et al [7] only showed a small effect, with a standardized mean difference of –0.28.

POIs for depression vary substantially with regard to content; level of guidance/support (guided, unguided, or self-guided [8,9]); duration, with most interventions lasting between 4 and 12 weeks [5]; the degree to which patients can tailor or self-select specific intervention components; and the degree to which interventions focus on a single stand-alone technique or incorporate multiple therapeutic techniques (eg, cognitive behavioral-based interventions for depression usually include psychoeducation, behavior monitoring, cognitive restructuring, and behavioral activation). These differences make it difficult to investigate the mechanisms of change and identify which techniques are the most important for change, as demonstrated for depression [10]. This is often due to the methodological characteristics in the general design of psychotherapeutic trials that can complicate analyses by focusing on the overall effect [11,12], obstructing the larger aim of expanding existing knowledge about change processes and further refining interventions.

Brief interventions reduced to their essentials are a potential alternative to investigate mechanisms of change. A first attempt to evaluate a focused single-module brief intervention for depression was made by Lüdtke et al [13]. However, cognitive restructuring and behavior activation were combined in one module, preventing a thorough investigation of the mechanisms of change.

Generally, cognitive restructuring and behavior activation are two standard interventions in cognitive behavioral therapy (CBT) for depression. Most commonly, treatment starts with interventions aimed at increasing the level of (positive) activities in patients by, for example, monitoring daily activities and mood, scheduling activities, and coping with problems. This "behavioral activation" (BA) has been adapted from behavior therapy for depression, which is also referred to as Lewinsohn's model [14], and incorporated in CBT [15] to increase patients' access to positive reinforcement through positive activities. For a more thorough history of BA, we would like to refer to Quigley and Dobson [16]. In 1996, Jacobson et al [17] showed that BA improved depression and dysfunctional beliefs similar to a complete CBT program, fueling the refinement of theory and practice in the use of BA as a stand-alone treatment. Refined BA models were subsequently introduced by Martell et al [18] and Lejuez et al [19], among others.

As a stand-alone intervention for depression, BA has been found similarly effective to CBT in face-to-face treatment. It is regarded as particularly cost-effective [20]. The Canadian Network for Mood and Anxiety Treatments even recommends BA as a first-line treatment for depression [21]. Still, as with other interventions, the mechanisms of actions are unclear for BA [22].

Regarding internet-delivered BA (iBA), a recent meta-analysis by Huguet et al [23] supports its effectiveness, although conclusions were compromised by the low quality of many of the studies. In this meta-analysis, iBA interventions lasted between 6 and 17 weeks. One of the studies [24] compared BA to physical activity and a waitlist control, and depression decreased more in the treatment groups than in the waitlist control group. There was no relationship between the number of modules used and the decrease in symptoms. Thus, it remains unclear how many modules or sessions of iBA are needed to achieve a response.

Another intervention that is often used in combination with CBT in depression is mindfulness (ie, mindfulness-based cognitive therapy [MBCT]). In face-to-face treatment, some evidence suggests the effectiveness of MBCT for depression [25]. When mindfulness-based interventions, not limited to MBCT, are administered online, meta-analyses have shown small to moderate effects with a Hedge g between 0.29 [26] and 0.61 [27], representing a general decrease in depressive symptoms. However, the effects were not significant in the meta-analysis by Sevilla-Llewellyn-Jones et al [27] when only participants with depression were considered ($g=-0.69$, 95% CI −1.694 to −0.313, $P=.19$).

To the best of our knowledge, there is only one internet-delivered randomized controlled trial [28] that compared BA and a mindfulness-based intervention as an active control group. Smartphone-administered BA and mindfulness interventions were compared over a treatment period of 8 weeks. Overall, both interventions showed a similar effect on depression. However, whereas BA was more effective for participants with a higher severity of depression at baseline (ie, Patient Health Questionnaire-9 [PHQ-9] total score ≥10 and a diagnosis of moderate major depression), mindfulness was more effective for those with lower symptom severity.

The aim of this study was to use a maximum-focused approach to assess the short-term effects of brief Web-based interventions and the potential mechanisms of change. We investigated a brief iBA module compared with an active control in the form of internet-delivered mindfulness (iMBI) and a nonactive control in the form of a waitlist with full access to care as usual (CAU). With regard to overlap in psychoeducation content and specific interventions such as the techniques and skills taught to the participants, we regarded the overlap between iBA and iMBI as minimal. We assessed participants at baseline (t_0) and then 2 weeks (t_1) and 4 weeks (t_2) later. The primary outcome was the improvement in depressive symptoms, as assessed with the PHQ-9, from t_0 to t_2. We chose the 4-week interval, instead of the 2-week interval, for the primary outcome to optimize the comparability of our results to other studies, which often use longer intervals of 8-9 weeks [5], and to investigate lasting change in depression. Secondary parameters included change from t_0 to t_1 in depressive symptoms, as measured with the PHQ-9; change in activity, as measured with the Behavioral Activation for Depression Scale (BADS); mindfulness, as measured with the Kentucky Inventory of Mindfulness Skills (KIMS); dysfunctional cognitive biases, as measured with the Dysfunctional Attitude Scale (DAS); and subjective quality of life, as measured with the World Health Organization Quality of Life (WHOQOL). We wanted to (1) pilot test the efficacy of the brief iBA module and (2) explore the mechanisms of change. Based on Ly et al [28], we expected a reduction in both the primary and secondary outcomes in favor of the experimental intervention (iBA), when compared with the active (iMBI) and nonactive (CAU) control groups. Moreover, we conducted mediation analyses to investigate the mechanisms of change. We expected that an improvement in depressive symptoms at t_2 would be mediated by change in activation (BADS) between t_0 and t_1. To further explore the mechanisms of change, we also included change in mindfulness using the KIMS, which mediates changes in MBCT [29], and dysfunctional cognitive biases using the DAS, which has been generally been associated with change in depression [30], as mediators in the exploratory analyses.

Methods

Design

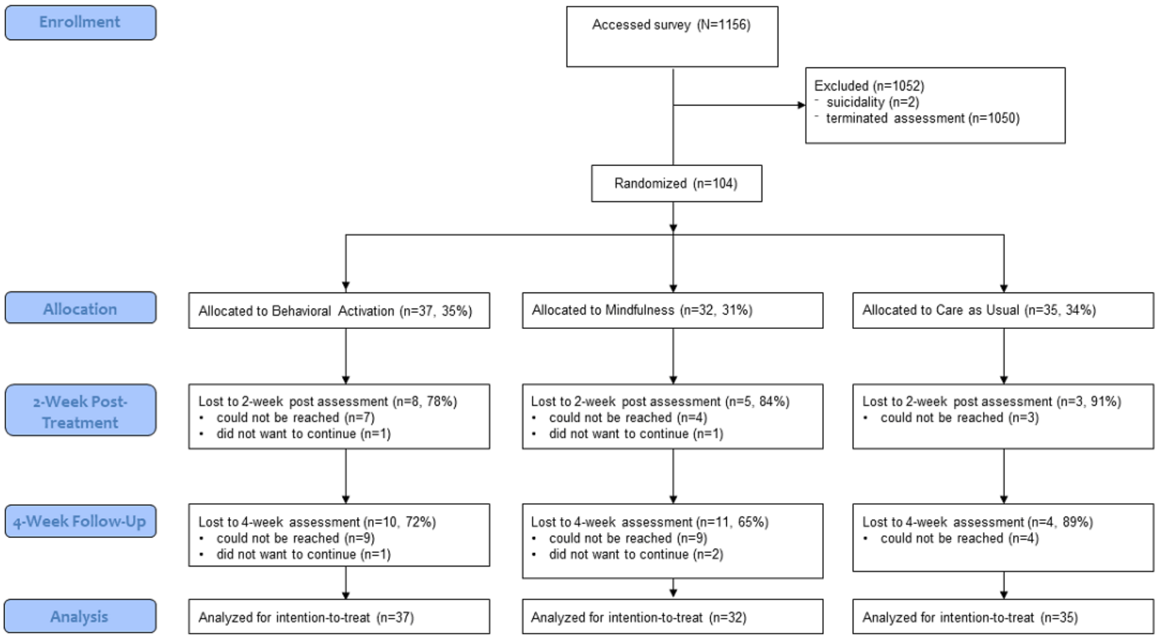
We performed a parallel randomized controlled trial. Participants were assessed using the online survey program EFS Survey developed by Questback (Oslo, Norway). Following the assessment at t_0, participants were randomly allocated to iBA, which included CAU; iMBI, which also included CAU; or CAU only. CAU included full access to treatment as usual. Participants were reassessed after 2 weeks (t_1) and 4 weeks (t_2). All participants provided electronic informed consent. The trial was approved by the Ethics Committee of the German Psychological Association (LJ032018_amd_102016) and was preregistered with the German Clinical Trials Register (DRKS; #DRKS00011562). Some changes were made after the trial was registered. The aim of the original study was to recruit treatment-seeking outpatients through medical staff at the hospital. As this led to very low participation, we decided to recruit through online advertisements instead. Therefore, depressive symptoms according to the PHQ-9 instead of a depressive disorder as diagnosed by a therapist/practitioner was used as an inclusion criterion.

Participants and Procedures

Recruitment took place between April 18, 2018, and May 27, 2018, via a Google AdWords campaign. The link to the study was also sent to participants from previous studies who had provided written consent to be contacted again via email. Inclusion criteria were depressive symptoms according to self-report (PHQ-9 total score >4 indicating mild depression), age between 18 and 65 years, internet access, internet literacy, sufficient German language skills, and informed consent, including the willingness to participate in 3 online assessments and a 2-week intervention. Exclusion criteria were lifetime psychotic or manic symptoms and suicidality as indicated by a score >2 on the Beck Depression Inventory II suicide item. In case of exclusion due to suicidal tendencies, help was provided in the form of emergency phone numbers and Web addresses of health services. Group allocation was performed by a person who did not possess any other information on the respective participant. Participants were randomized at t_0 to one of the 3 groups, with an allocation ratio of 1:1:1 based on a fixed randomization plan conducted by a statistician. This has been referred to as centralized assignment [31]. After randomization, participants were informed about their allocation via email. This email gave participants assigned to the iBA and iMBI groups a link, code, and password to access the appropriate Web-based treatment. They could use the treatment at the times, pace, and frequency they chose to meet their needs. All participants in the intention-to-treat (ITT) group were sent a link for re-assessment at 2 and 4 weeks later. To increase the retention rate, participants were reminded every 2 days to complete the survey (up to 3 times in total). After completion of the 4-week assessment, all participants received the links, codes, and passwords to access both treatments.

In total, 1156 people accessed the survey (see Figure 1). The majority did not finish the baseline survey (n=1050), and the assessment was automatically terminated for 2 participants because they met exclusion criteria (n=2 suicidality; see exclusion criteria) leaving 104 participants for the ITT group.

Figure 1. Flowchart of participation in the study.



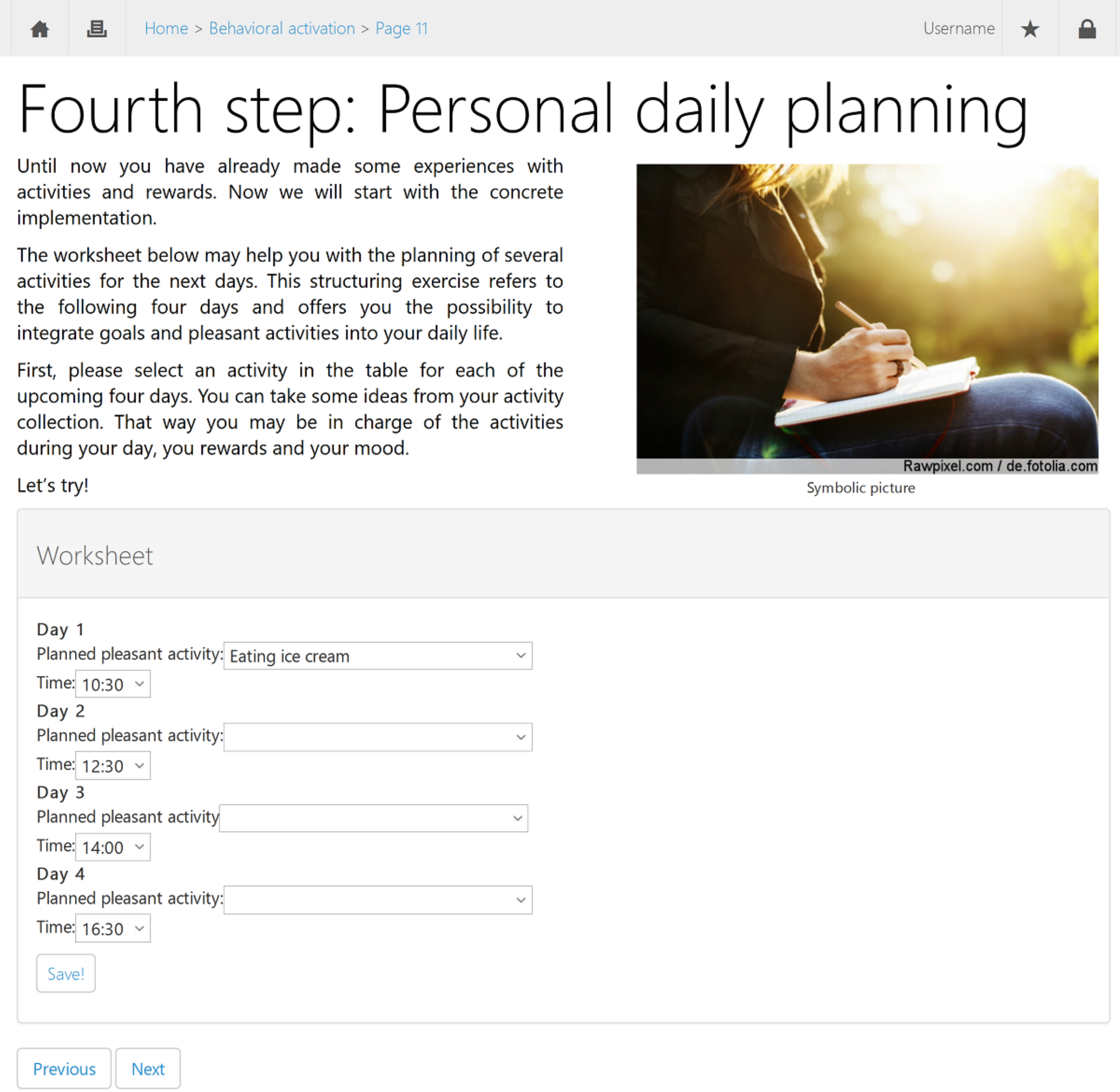
Interventions

Both Web-based interventions were unguided. All information on how to use the interventions was provided within each program; we did not provide any additional information, including in the reminders. The interventions consisted of psychoeducational information as well as worksheets. Worksheets could be saved on the computer or printed for daily use. For iMBI, that also included audio files. Participants were also allowed to log into the module repeatedly to access clean worksheets or to reread the (psychoeducation) information provided.

Experimental Intervention: Internet-Based Behavioral Activation

iBA is a Web-based intervention that focuses on the development of behavioral activities. We used a version based on the German version of the manual by Martell et al [32]. The intervention starts with psychoeducation on the interplay between mood and behavior in depression. Then, participants learn how to monitor their mood and daily activities, sensibly plan activities in their daily schedule, and anticipate problems that may occur when they try to perform the planned activities. Most of the information is accompanied by worksheets that can be filled out and saved on a computer or printed for daily use. Depending on the individual’s reading rate and personal processing time, it takes about 60 minutes to become familiar with the information in the module. However, participants are advised to take their time and use the program daily, including completing the work sheets and performing the planned activities. For a screenshot of the intervention in a Web browser, see Figure 2.

Figure 2. Screenshot of the experimental, internet-based behavioral activation intervention (translated).



Fourth step: Personal daily planning

Until now you have already made some experiences with activities and rewards. Now we will start with the concrete implementation.

The worksheet below may help you with the planning of several activities for the next days. This structuring exercise refers to the following four days and offers you the possibility to integrate goals and pleasant activities into your daily life.

First, please select an activity in the table for each of the upcoming four days. You can take some ideas from your activity collection. That way you may be in charge of the activities during your day, you rewards and your mood.

Let’s try!

Worksheet

Day 1
Planned pleasant activity: Eating ice cream
Time: 10:30
Day 2
Planned pleasant activity:
Time: 12:30
Day 3
Planned pleasant activity
Time: 14:00
Day 4
Planned pleasant activity:
Time: 16:30

Active Control Group: Internet-Based Mindfulness

iMBI is a Web-based intervention introducing and teaching mindfulness practice. We used a version based on the German version of the mindfulness manual by Segal et al [33]. It starts with an introduction into the concept of mindfulness (eg, the “automatic pilot”) and an explanation of how mindfulness can help with depression. Then, participants independently carry out mindfulness exercises, starting with a short exercise on mindful listening. They use worksheets to detect situations in which they are not mindful (on automatic pilot), how to mindfully deal with disturbing thoughts and feelings (“thoughts are not facts”), and how to enhance self-care through mindfulness. In the module, a variety of audio files are provided to optimize mindfulness practice: breathing exercise (5:50 min), body scan (16:15 min), and inner smile (4:40 min). Depending on the reading rate and personal processing time, it takes about 60 minutes to read the content of the module and become familiar with the worksheets and exercises. Similar to the iBA intervention, participants are advised to use the exercises and worksheets in their daily life and to use the program repeatedly. In this study, to keep the practice time similar to that of the iBA group, participants were asked to practice only 1 of the 3 mindfulness exercises repeatedly over the following week.

Care as Usual Only

CAU included full access to treatment as usual provided by the German health care system. As an incentive, participants received both Web-based interventions (iBA and iMBI) for their personal use after finishing the 4-week assessment.

Measures

The primary outcome was improvement in depressive symptoms, as assessed with the PHQ-9, from t_0 to t_2. Secondary outcomes included changes from t_0 to t_1 in depressive symptoms (PHQ-9), activity (BADS), mindfulness (KIMS), dysfunctional attitudes (DAS), and subjective quality of life (WHOQOL). Negative effects were assessed by calculating the reliable change index for clinically significant deterioration in the PHQ-9 from t_0 to t_1 [34].

Primary Outcome

The severity of depression was measured using the German version of the PHQ-9 [35]. The PHQ-9 represents the depression module of the PHQ-D and uses 9 items rated on a 4-point Likert scale. Good validity and reliability have been reported [36,37].

Secondary Outcomes

The German version of the BADS [38] was used to assess levels of behavioral activation. It includes 25 items summarized as a total score. Internal consistency and test-retest reliability are considered acceptable [39]. Dispositional mindfulness was assessed with the German version of the 20-item version of the KIMS [40]. The psychometric properties of the KIMS, including sensitivity of change, are considered good [41]. The German 18-item version of the DAS form B (DAS-18B) was used to assess dysfunctional beliefs. Items are rated on a 7-point Likert scale. Reliability and validity of the scale are good [42]. Quality of life was assessed with the global item of the German version of the WHOQOL-BREF [43].

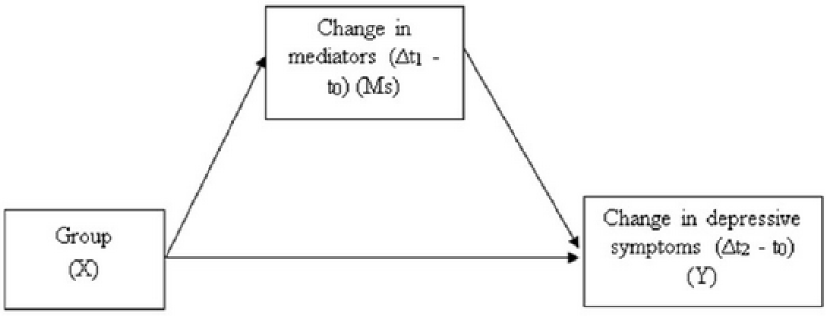
Data Analysis

SPSS 25.0 (IBM Corp, Armonk, NY) software was used for all analyses. We conducted complete-cases (CC) as well as ITT analyses. CC analyses were based on data from participants who were assessed at all 3 assessment points. For ITT analyses, which included all randomized participants, missing data were imputed by multiple imputations based on the assumption that data were missing at random, conditional on information on treatment, sex, age, and all relevant outcomes across the 3 assessment time points. We created 100 imputed datasets.

To investigate efficacy, we conducted analyses of covariance (ANCOVA) with treatment (iBA, iMBI, CAU) as the independent factor, the baseline level of the respective outcome as the covariate, and the level of the outcome at t_1 and t_2 as the dependent variable. To follow up on group differences, control groups were pooled to compare the experimental iBA group with the control iMBI and CAU groups. For post-hoc tests at the individual group level, we used uncorrected *t* tests. Effect sizes for ANCOVAs are reported according to Kinnear and Gray [44], with $\eta_p^2 \approx .01$, $\eta_p^2 \approx .06$, and $\eta_p^2 \approx .14$ corresponding to small, medium, and large effects, respectively. We also calculated Cohen *d* as an effect size for change in outcomes over time within each of the groups, with $|d| \approx .2$, $|d| \approx .5$, and $|d| \approx .8$ corresponding to small, medium, and large effects, respectively.

To investigate the mechanisms of change involved in the interventions, we conducted mediation analyses [45]. We coded iBA as 1 and pooled the control groups (iMBI and CAU) and coded them as 0, with the treatment effect referring to the effects of iBA above and beyond the pooled control groups. We computed standardized, residualized change scores for change in PHQ-9 from t_0 to t_2 and for change in mediators (BADS, KIMS, DAS) from t_0 to t_1 (see Figure 3). The mediation analyses were performed using the SPSS macro PROCESS developed by Hayes (version v3.1) [46] and met all the criteria defined by Kraemer et al [47] for the use of mediators within a randomized controlled trial. We bootstrapped the results 5000 times to correct for potential biases of nonnormality in the sample. For mediation analyses, we used a different strategy to impute missing data, that is, the expectation-maximization algorithm. This was necessary because multiple imputation data sets cannot be used in PROCESS, which requires a single data set. The mediation hypothesis is confirmed when the effect range (lower limit of the 95% CI [LLCI] to upper limit of the 95% CI [ULCI]) does not include zero.

Figure 3. Differences between baseline (t_0) and the 2-week post-treatment (t_1) and the 4-week follow-up (t_2) assessments were calculated as standardized residualized change scores. Group was defined as experimental (internet-based behavioral activation) vs the pooled control groups (internet-based mindfulness intervention and care as usual). Mediators were measured using the Behavioral Activation for Depression Scale, Kentucky Inventory of Mindfulness Skills, and Dysfunctional Attitude Scale (German 18-item version of form B). Depressive symptoms were measured using the Patient Health Questionnaire-9.



Results

Demographic Characteristics

Table 1 shows the demographic characteristics of the 3 groups. The participants were mostly (80/104, 76.9%) female, were in their mid-forties on average, and had almost 12 years of formal education. Previous psychological treatment experiences, including internet interventions and use of self-help books, were high but similar between groups (*P*>.40). Randomization was successful, with no significant differences in the demographic characteristics between groups.

Table 1. Baseline demographics.

Variable	iBA[a] (n=37)	iMBI[b] (n=32)	CAU[c] (n=35)	Total sample (N=104)	Statistic	*P* value
Age (years), mean (SD)	45.84 (10.29)	44.91 (10.98)	47.80 (9.54)	46.21 (10.24)	*F*(2,101)=0.702	.50
Sex, n (%)						
Female	32 (86)	25 (78)	23 (66)	80 (76.9)	χ^2_2=4.409	.11
Male	5 (14)	7 (22)	12 (34)	24 (23.1)		
Years of formal education, mean (SD)	11.97 (1.46)	11.84 (1.37)	11.62 (1.50)	11.81 (1.44)	*F*(2,101)=0.517	.60
Number of psychotherapy sessions, mean (SD)	92.27 (93.07)	92.00 (122.12)	74.83 (66.96)	86.32 (95.14)	*F*(2,101)=0.380	.69
Number of hospitalizations, mean (SD)	1.03 (1.26)	1.38 (2.14)	1.40 (2.20)	1.26 (1.89)	*F*(2,101)=0.434	.65
Number of self-help books, mean (SD)	6.35 (6.83)	10.72 (26.51)	7.03 (6.77)	7.92 (15.70)	*F*(2,101)=0.745	.48
Number of online interventions, mean (SD)	1.19 (1.51)	0.84 (1.11)	1.00 (1.14)	1.02 (1.27)	*F*(2,101)=0.531	.53

[a] iBA: internet-based behavioral activation.

[b] iMBI: internet-based mindfulness intervention.

[c] CAU: care as usual.

Adherence and Completion

At t_1, 6 participants in the iBA group (6/27, 22%) and 3 participants in the iMBI group (3/25, 12%) indicated that they had not used the respective online module at all; due to technical difficulties, 2 participants in the iBA group and 2 participants in the iMBI group were not asked whether they used the respective intervention. In the ITT group (N=104), 88 participants (85%) completed the t_1 assessment, and 79 (76%) completed the t_2 assessment.

Outcome Analyses

The assessment values for all 3 assessment points for the CC group are displayed in Table 2. Regarding the primary outcome (PHQ-9), the proposed superiority of iBA was confirmed in neither the CC sample nor the ITT sample (Tables 3 and 4). In the iBA group, however, depression decreased with a medium effect from t_0 to t_2 (d=−0.53).

Table 2. Primary and secondary outcome measures in the complete-cases group.

	Baseline (t_0), mean (SD)			2-weeks post-treatment (t_1), mean (SD)			4-week follow-up (t_2), mean (SD)		
	iBA[a] (n=37)	iMBI[b] (n=32)	CAU[c] (n=35)	iBA (n=29)	iMBI (n=27)	CAU (n=32)	iBA (n=27)	iMBI (n=21)	CAU (n=31)
PHQ-9[d]	10.46 (4.32)	12.50 (5.04)	11.00 (4.27)	9.73 (5.46)	11.37 (6.15)	9.90 (3.14)	8.70 (5.72)	11.57 (5.27)	9.26 (4.14)
BADS[e]	71.97 (18.91)	72.50 (24.62)	73.97 (20.32)	85.76 (22.90)	77.19 (23.95)	78.59 (21.63)	86.85 (27.22)	81.19 (23.84)	84.29 (26.70)
KIMS[f]	117.68 (15.71)	112.78 (17.11)	114.97 (19.11)	119.97 (15.98)	115.67 (17.19)	113.03 (18.22)	119.89 (16.40)	115.24 (19.10)	112.97 (19.18)
DAS-18B[g]	66.68 (19.65)	69.19 (20.47)	69.29 (18.41)	62.14 (14.97)	64.44 (19.72)	67.68 (19.99)	56.44 (20.49)	64.10 (20.62)	67.94 (19.12)
WHOQOL[h]	3.08 (0.68)	2.91 (0.86)	2.94 (0.73)	3.28 (0.88)	3.04 (0.71)	3.12 (0.66)	3.11 (1.01)	3.05 (0.74)	2.94 (0.77)

[a] iBA: internet-based behavioral activation.

[b] iMBI: internet-based mindfulness intervention.

[c] CAU: care as usual.

[d] PHQ-9: Patient Health Questionnaire-9.

[e] BADS: Behavioral Activation for Depression Scale.

[f] KIMS: Kentucky Inventory of Mindfulness Skills.

[g] DAS-18B: German 18-item version of the Dysfunctional Attitude Scale form B.

[h] WHOQOL: World Health Organization Quality of Life Scale (global item).

Table 3. Outcome analyses, including the effect sizes for and between-group differences in the outcome measures between the baseline (t_0) and 2-week post-treatment (t_1) assessments in the complete-cases group.

	Effect size (*d*)			Between-group differences[a]			
	iBA[b]	iMBI[c]	CAU[d]	Statistic	Effect size (η_p^2)	90% CI	*P* value
PHQ-9[e]	–0.23	–0.31	–0.35	*F*(2,84)=0.29	.007	0.000- 0.041	.75 (81[f])
BADS[g]	0.96	0.25	0.30	*F*(2,84)=2.73	.061	0.000-0.145	.07[h] (.09[f])
KIMS[i]	0.27	0.31	–0.19	*F*(2,84)=3.00	.067	0.000-0.152	.06[j] (.03[f])
DAS-18B[k]	–0.37	–0.37	–0.14	*F*(2,84)=1.11	.026	0.000-0.088	.34 (.14[f])
WHOQOL[l]	0.31	0.16	0.26	*F*(2,84)=0.56	.013	0.000-0.061	.57 (.56[f])

[a] Assessed using analyses of covariance (ANCOVA) with baseline scores as covariates and uncorrected *t* tests for post-hoc tests.

[b] iBA: internet-based behavioral activation.

[c] iMBI: internet-based mindfulness intervention.

[d] CAU: care as usual.

[e] PHQ-9: Patient Health Questionnaire-9.

[f] Intention-to-treat analyses based on multiple imputation.

[g] BADS: Behavioral Activation for Depression Scale.

[h] BA>iMBI.

[i] KIMS: Kentucky Inventory of Mindfulness Skills.

[j] iBA>CAU.

[k] DAS-18B: German 18-item version of the Dysfunctional Attitude Scale form B.

[l] WHOQOL: World Health Organization Quality of Life Scale (global item).

Table 4. Outcome analyses, including the effect sizes for and between-group differences in the outcome measures between the baseline (t_0) and 4-week follow-up (t_2) assessments in the complete-cases group.

	Effect size (*d*)			Between-group differences[a]			
	iBA[b]	iMBI[c]	CAU[d]	Statistic	Effect size (η_p^2)	90% CI	*P* value
PHQ-9[e]	–0.53	–0.24	–0.53	*F*(2,75)=0.31	.008	0.000-0.048	.74 (.99[f])
BADS[g]	0.98	0.44	0.64	*F*(2,75)=0.27	.007	0.000-0.044	.76 (.52[f])
KIMS[h]	0.24	0.25	–0.18	*F*(2,75)=1.08	.028	0.000-0.096	.34 (.11[f])
DAS-18B[i]	−0.87	–0.41	–0.12	*F*(2,75)=3.62	.088	0.004-0.186	.03[j] (.008[f])
WHOQOL[k]	0.05	0.18	0	*F*(2,75)=1.07	.028	0.000-0.095	.35 (.59[f])

[a] Assessed using analyses of covariance (ANCOVA) with baseline scores as covariates and uncorrected *t* tests for post-hoc tests.

[b] iBA: internet-based behavioral activation.

[c] iMBI: internet-based mindfulness intervention.

[d] CAU: care as usual.

[e] PHQ-9: Patient Health Questionnaire-9.

[f] Intention-to-treat analyses based on multiple imputation.

[g] BADS: Behavioral Activation for Depression Scale.

[h] KIMS: Kentucky Inventory of Mindfulness Skills.

[i] DAS-18B: German 18-item version of the Dysfunctional Attitude Scale form B.

[j] iBA>CAU.

[k] WHOQOL: World Health Organization Quality of Life Scale (global item).

For secondary outcomes, a significant medium effect was found for improvement in dysfunctional attitudes (DAS-18B) from t_0 to t_2 (CC and ITT analyses). Improvement in the DAS-18B was larger in the iBA group than in the pooled control groups (η_p^2=.053, *P*=.04), which was explained by the larger decrease in dysfunctional attitudes (DAS-18B) in the iBA than in the CAU group in the single group comparisons. On trend level, the groups in the CC sample also differed in the improvements in level of activation (BADS) and mindfulness skills (KIMS) from t_0 to t_1; the differences were significant in the ITT analyses. When the control groups were pooled, activity increased more in the iBA group than the pooled control groups (η_p^2=.060, *P*=.02), but the differences in KIMS were no longer significant (*P*=.11). Single comparisons showed that the increase in KIMS was larger in the iBA than in the CAU group and that the level of activation (BADS) increased in the iBA group more than in the iMBI group. The groups did not differ in the change in quality of life (WHOQOL) over time.

Regarding changes in the secondary outcomes in the iBA group, the effects were large for increase in activity (BADS, d between 0.96 and 0.98), and the decrease in dysfunctional attitudes reached a medium to large effect (*d* between 0.37 and 0.87). See Tables 3 and 4 for more details.

The deterioration rates were similar between the groups (χ^2_2=1.47, *P*=.48), with a mean deterioration rate of 6.8%.

Mediation

Intercorrelations between all outcome parameters at all assessment points are shown in Table 5.

Table 5. Zero-order correlations for the primary and secondary outcomes at baseline (t_0), 2-week post-treatment (t_1), and 4-week follow-up (t_2).

	PHQ-9[a], t_0	DAS-18B[b], t_0	KIMS[c], t_0	BADS[d], t_0	PHQ-9, t_1	DAS-18B, t_1	KIMS, t_1	BADS, t_1	PHQ-9, t_2	DAS-18B, t_2	KIMS, t_2	BADS, t_2
PHQ-9, t_0												
r	—[e]	.15	–.23	–.53	.73	.17	–.25	–.46	.70	.15	–.30	–.35
P value[f]	—	—	—	—	—	—	—	—	—	—	—	—
DAS-18B, t_0												
r	—	—	–.57	–.33	.11	.80	–.54	–.22	.09	.82	–.50	–.28
P value[f]	.134	—	—	—	—	—	—	—	—	—	—	—
KIMS, t_0												
r	—	—	—	.54	–.27	–.52	.85	.43	–.25	–.49	.83	.41
P value[f]	.017	.001	—	—	—	—	—	—	—	—	—	—
BADS, t_0												
r	—	—	—	—	–.52	–.38	.46	.71	–.52	–.28	.50	.68
P value[f]	.001	.001	.001	—	—	—	—	—	—	—	—	—
PHQ-9, t_1												
r	—	—	—	—	—	.23	–.34	–.71	.83	.15	–.35	–.51
P value[f]	.001	.284	.006	.001	—	—	—	—	—	—	—	—
DAS-18B, t_1												
r	—	—	—	—	—	—	–.58	–.34	.16	.87	–.51	–.33
P value[f]	.085	.001	.001	.001	.019	—	—	—	—	—	—	—
KIMS, t_1												
r	—	—	—	—	—	—	—	.50	–.26	–.51	.89	.43
P value[f]	.011	.001	.001	.001	.001	.001	—	—	—	—	—	—
BADS, t_1												
r	—	—	—	—	—	—	—	—	–.61	–.23	.59	.76
P value[f]	.001	.027	.001	.001	.001	.001	.001	—	—	—	—	—
PHQ−9, t_2												
r	—	—	—	—	—	—	—	—	—	.17	–.37	–.64
P value[f]	.001	.352	.011	.001	.001	.115	.009	.001	—	—	—	—
DAS-18B, t_2												
r	—	—	—	—	—	—	—	—	—	—	–.53	–.33
P value[f]	.141	.001	.001	.004	.120	.001	.001	.022	.095	—	—	—
KIMS, t_2												
r	—	—	—	—	—	—	—	—	—	—	—	.55
P value[f]	.002	.001	.001	.001	.001	.001	.001	.001	.001	.001	—	—
BADS, t_2												
r	—	—	—	—	—	—	—	—	—	—	—	—
P value[f]	.001	.005	.001	.001	.001	.001	.001	.001	.001	.001	.001	—

[a] PHQ-9: Patient Health Questionnaire-9.

[b] DAS-18B: German 18-item version of the Dysfunctional Attitude Scale form B.

[c] BADS: Behavioral Activation for Depression Scale.

[d] KIMS: Kentucky Inventory of Mindfulness Skills.

[e] Not applicable.

[f] two-tailed.

From the 3 potential mediators (changes in BADS, KIMS, and DAS-18B) from t_0 to t_1, only the change in BADS acted as a mediator for the change in depression (PHQ-9) from t_0 to t_2. Change in BADS was the only significant mediator; the changes in KIMS and DAS-18B did not mediate the change in depression (see Table 6). For the change in depression as measured by BADS, the indirect effect was −.20 (SE 0.09, LLCI=-0.40, ULCI=−0.03), with a remaining nonsignificant direct effect of the treatment of 0.10 (SE 0.20, *P*=.60, LLCI=−0.29, ULCI=0.50).

Table 6. Mediation analysis: Effects of treatment (internet-based behavioral activation, coded as 1, vs the pooled control groups, coded as 0) on the mediators and the effects of the mediators (including treatment) on depression, as measured by the Patient Health Questionnaire-9 (PHQ-9).

	Mediators									Outcome		
	DAS-18B[a]			KIMS[b]			BADS[c]			PHQ-9		
	beta	SE	*P* value	beta	SE	*P* value	beta	SE	*P* value	beta	SE	*P* value
Treatment	–.23	0.20	.255	.32	0.20	.116	.49	0.20	.015	.10	.20	.601
DAS-18B	—[d]	—	—	—	—	—	—	—	—	.07	.10	.472
KIMS	—	—	—	—	—	—	—	—	—	.14	.10	.174
BADS	—	—	—	—	—	—	—	—	—	–.41	.10	<.001
R^2	.01	—	—	.02	—	—	.06	—	—	.15	—	—

[a] DAS-18B: German 18-item version of the Dysfunctional Attitude Scale form B.

[b] KIMS: Kentucky Inventory of Mindfulness Skills.

[c] BADS: Behavioral Activation for Depression Scale.

[d] Not applicable.

When we recalculated the mediation analyses to only compare the iBA and iMBI groups with regard to potential mediators, only the indirect effect of the BADS was significant again (DAS-18B: beta=−.02, SE 0.05, LLCI=−0.14, ULCI=0.08; KIMS: beta=.01, SE 0.04, LLCI=−0.07, ULCI=0.11; BADS: beta = −.18, SE 0.10, LLCI=−0.41, ULCI=−0.002; Total: beta=–17, SE 0.09, LLCI=–.36, ULCI=–.01).

Discussion

The aim of this study was to pilot test the efficacy of iBA, a brief Web-based module using BA, in comparison to active (iMBI) and nonactive (CAU) control interventions and to explore the mechanisms of change.

Primary Outcome

Contrary to our expectation, the reduction in the primary outcome of depressive symptoms (PHQ-9) was similar between all the groups. Moreover, the response rate and effect size were smaller than in previous studies that showed BA is efficacious in both face-to-face [20] and online [23] interventions. In this study, the iBA group experienced a symptom decrease of 17% and a medium effect of *d*=−0.53 from t_0 to t_2, compared with a 7% decrease and *d* of −0.24 in the iMBI group and a 16% decrease and *d* of −0.53 in the CAU group. Thus, reasons for the present null findings need to be discussed.

First, to assess the mechanisms of change, we introduced BA as a single-module, internet-based intervention (iBA). The recent meta-analysis by Huguet et al [23] showed efficacy for iBA; however, 3 of the 9 studies combined iBA with other interventions. Therefore, it is unclear whether the treatment effect was attributable to iBA or another active component. Moreover, not one but several BA sessions were used, covering periods between 6 and 17 weeks. Another important difference between our study and those in the meta-analysis of Huguet et al is we used an unguided, fully automated intervention, as opposed to guided interventions. To the best of our knowledge, Lüdtke et al [13] conducted the only study that used a single-module, unguided internet-based intervention for depression, but it combined behavior activation with cognitive restructuring and also found no effect on depressive symptoms. Although previous studies have shown an effect of unguided interventions on depression [9], the results of our study and that by Lüdtke et al may indicate that iBA cannot be learned without therapist contact or that a single-module approach is insufficient to cause change in proximal measures. Generally, there is some evidence for better effects in guided than in unguided POI. Thus, a guided therapy may potentially increase the effects of iBA [49]; however, similar effects have been found between guided and unguided POI [50]. Taken together, the current evidence is not sufficient to support decisions regarding whether more redundancy, the repetition of the same content in more sessions, or a more complex, longer intervention with additional content is needed to affect depression.

Second, with assessments at 2-week intervals, the intervention or assessment interval was relatively short in comparison to the intervals of 6 to 17 weeks used in other study designs [23]. To allow more accurate conclusions about the treatment mechanisms, we chose this maximum-focused, one-module approach with a short interval with the aim of measuring rapid effects at the beginning of the treatment, which have also been reported in other clinical trials on depression, particularly with the use of antidepressants [51–53].

Although adherence is generally considered challenging in POIs, with completion rates as low as 80% and drop-out rates as high as 37% [5]. Although this study had better adherence, with completion rates of 85% at t_1 and 76% at t_2 and drop-out rates of 22% in the iBA group and 12% in the iMBI group, the 2-week intervention period was potentially too short for participants to use the intervention properly and transfer what they learned into everyday life. Indeed, 9 participants indicated that they did not access the modules at all (6/27, 22% in the iBA group and 3/25, 12% in the iMBI group). Unfortunately, we were not able to track user engagement and had to rely on self-report. Finally, we do not know if or how participants in the iBA group put the positive activities into practice. It is possible that a significant increase in physical activity is necessary for iBA to be effective in depression, as indicated by findings of an anti-inflammatory effect of BA with exercise on depression [51]. Hypothetically, a longer intervention interval in combination with automated support, such as email reminders, for the intervention as well as exercises (ie, physical activity) are important to explore whether the present, single-module iBA intervention or the intervention period and usage needs to be modified. Moreover, it would be helpful to assess the effects of engagement by monitoring the activities performed, to what degree physical activity was included, and how often each participant logged into the website and engaged with the content.

Secondary Outcomes

Regarding the secondary outcomes, we found tentative support for the efficacy of the iBA. There was some support for the short-term superiority (ie, t_0 to t_1) of the iBA compared with the control groups based on the differences in behavior activation, as measured by the BADS, at trend level. This was further supported by the large effect sizes within the iBA group for change in activation (d=0.96-0.98). The post-hoc test showed that activation was greater with iBA than with iMBI. These differences, however, disappeared over the 4-week assessment interval. Surprisingly, the groups also differed at trend level with regard to change in mindfulness skills from t_0 to t_1, with a slightly higher improvement in the iBA group than in the iMBI group. However, the effect sizes were small and similar in the iBA and iMBI groups (d between 0.24 and 0.31). The changes in mindfulness, as assessed by the KIMS, in the iBA group may be considered surprising. However, Ly et al [28] also reported similar effects of BA and mindfulness in a measure of emotional acceptance/experiential avoidance.

Dysfunctional attitudes, as measured by the DAS-18B, decreased to a larger degree from t_0 to t_2 in the iBA group than in the control groups, particularly in comparison with the CAU group. This is in line with theoretical assumptions underlying CBT and results by Jacobsen et al [17] that show that change in activation leads to change in dysfunctional attitudes in depression. Because change in dysfunctional attitudes may also lead to change in depression in the long term, we consider that delayed effects on depressive symptoms in the iBA group are still plausible. However, due to the short study interval, we do not have any further follow-up data with which to verify this assumption.

Finally, we did not find group differences regarding an improvement in quality of life. This is unsurprising, considering the lack of difference in depression symptoms, which may be due to the short assessment interval.

Mechanisms of Change

Although no treatment effect was found on the primary outcome, it is not only considered valid but highly important to study treatment processes such as mechanisms of change [45]. As planned, we entered the 3 potential treatment mechanisms in the mediation analysis: behavior activation, mindfulness, and dysfunctional attitudes. The only significant mediator for the effect of the intervention on PHQ-9 was behavior activation (BADS), further underlining the importance of behavior activation in the treatment of depression.

At first glance, this disagrees with the findings by Forand et al [52], who also used an online sample and highlighted the role of attitude modification. However, the intervention interval was much longer in their study (8 weeks vs 2 weeks), and it is still possible that effects will emerge at a later point in time in our sample. Moreover, the researchers used the Competencies Cognitive Therapy Scale to measure dysfunctional attitudes, while we used the DAS-18B, and the scales may not correlate. Generally, our results correspond with theoretical assumptions and evidence suggesting a link between an increase in the BADS and decrease in depressive symptoms [22]. Still, we did not find the expected treatment effect of the experimental intervention (iBA). Findings of the mediation analyses may thus not allow conclusions about the mechanisms of iBA, but they may be relevant to a general understanding of change in depression. We measured several other potential mechanisms (DAS-18B and KIMS) and concurrently considered them in the mediation model. Importantly, and as expected, short-term changes in the DAS-18B and KIMS did not mediate the treatment effects in this study. However, other mechanisms of change (eg, change in rumination) may play an important role in the treatment of depression, but these were not investigated in this study in order to minimize the burden on participants.

Additional mechanisms of change should be assessed in future studies. The BA model extends to possible mechanisms such as reinforcement, mood, and avoidance behavior [22]. This suggests that the relationship between behavior activation and depression is mediated by an initial increase in positive reinforcement, as found with undergraduate students [53]. These aspects of the BA theory should be considered in future studies.

General Limitations

Some general limitations need to be considered. Most importantly, we did not verify the diagnostic criteria for major depression disorder in the sample. However, as the PHQ-9 has excellent psychometric properties and the cut-off of >4 is established in the literature, a change in the results is unlikely if a structured interview were to be included. Also, depressive symptoms were rather mild in the current sample, with means between 10.46 and 12.50 at baseline, and fluctuating symptoms and spontaneous remission are frequent in depression, with the clinical consequence that guidelines on MDD suggest "watchful waiting" as the first-line management for mild depression [54,55]. Also, the two modules (iBA and iMBI) differed in content; iMBI included an audio file (mindfulness exercise). This may have produced additional and unnecessary variance between interventions and should be controlled for in future dismantling studies. Moreover, the current sample was quite experienced in psychological interventions for depression, and activating or mindfulness treatment elements may have already been experienced by the participants in previous therapies and thus might not have the same effects as in treatment-naïve patients with MDD (see Table 1). The increasing number of available online interventions for psychological disorders may affect potential treatment samples for POIs. Expectations of the efficacy of a POI may generally be low if the participant had not benefited from an intervention before or had relapsed again after an intervention, ultimately leading to the self-fulfilling prophecy that POIs do not help. This is also reflected by evidence showing that attitudes toward online interventions and expectations moderate treatment outcome [56,57]. In addition, more than 75% of the participants were female, potentially compromising the generalizability of the findings. However, as depression occurs twice as often in women than in men [58], it seems particularly necessary to evaluate interventions in women. Finally, due to recruitment difficulties with the initial study protocol involving the use of the program in an outpatient clinic to bridge waiting times, changes in the trial protocol with regard to recruitment sources were unavoidable, and this also needs to be considered.

Conclusion

To conclude, we did not find any evidence for the short-term efficacy of the iBA intervention on depression in treatment-experienced participants with mild to moderate levels of depressive symptoms. However, in comparison to the pooled control group, iBA was more effective with regard to the secondary outcomes, specifically an increase in activity over 2 weeks and a decrease in dysfunctional attitudes over 4 weeks. Furthermore, we found some support for short-term behavior activation being an important mediator for change in depression. Additional studies on brief interventions are needed to assess the mechanisms of change in POIs. It would also be helpful to test whether using longer intervention intervals, including automated support such as reminders to use the interventions or human support through a guided intervention enhances the treatment effects of iBA.

Acknowledgments

We would like to thank Katrin Duttlinger and Christopher Lau for their support. The study was not externally funded.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHEALTH checklist V1.6.1
[PDF File (Adobe PDF File), 2566 KB - jmir_v22i3e15312_app1.pdf]

References

1. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
2. Dèttore D, Pozza A, Andersson G. Efficacy of technology-delivered cognitive behavioural therapy for OCD versus control conditions, and in comparison with therapist-administered CBT: Meta-analysis of randomized controlled trials. Cogn Behav Ther 2015;44(3):190-211. [doi: 10.1080/16506073.2015.1005660] [Medline: 25705787]
3. 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]
4. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
5. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. J Affect Disord 2017;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
6. 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;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]

7. Zhou T, Li X, Pei Y, Gao J, Kong J. Internet-based cognitive behavioural therapy for subthreshold depression: A systematic review and meta-analysis. BMC Psychiatry 2016;16(1):356 [FREE Full text] [doi: 10.1186/s12888-016-1061-9] [Medline: 27769266]
8. Schröder J, Berger T, Westermann S, Klein JP, Moritz S. Internet interventions for depression: New developments. Dialogues Clin Neurosci 2016;18(2):203-212 [FREE Full text] [Medline: 27489460]
9. Morgan C, Mason E, Newby JM, Mahoney AEJ, Hobbs MJ, McAloon J, et al. The effectiveness of unguided internet cognitive behavioural therapy for mixed anxiety and depression. Internet Interv 2017;10:47-53 [FREE Full text] [doi: 10.1016/j.invent.2017.10.003] [Medline: 30135752]
10. Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Hollon SD. Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. Psychother Res 2019;29(1):15-29. [doi: 10.1080/10503307.2017.1395922] [Medline: 29115185]
11. Laurenceau J, Hayes AM, Feldman GC. Some methodological and statistical issues in the study of change processes in psychotherapy. Clin Psychol Rev 2007;27(6):682-695 [FREE Full text] [doi: 10.1016/j.cpr.2007.01.007] [Medline: 17328996]
12. A-Tjak JGL, Davis ML, Morina N, Powers MB, Smits JAJ, Emmelkamp PMG. A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. Psychother Psychosom 2015;84(1):30-36. [doi: 10.1159/000365764] [Medline: 25547522]
13. Lüdtke T, Westermann S, Pult LK, Schneider BC, Pfuhl G, Moritz S. Evaluation of a brief unguided psychological online intervention for depression: A controlled trial including exploratory moderator analyses. Internet Interv 2018;13:73-81 [FREE Full text] [doi: 10.1016/j.invent.2018.06.004] [Medline: 30206522]
14. Lewinsohn PM. A behavioural approach to depression. In: Freedman RJ, Katz MM, editors. The psychology of depression: Contemporary theory and research. Oxford: John Wiley & Sons; 1974:157-174.
15. Beck AT, Rush AJ, Shaw BF, Emery GM. Cognitive therapy of depression. New York: The Guilford Press; 1979.
16. Quigley L, Dobson KS. Behavioral activation treatments for depression. In: Hofmann & SG, Asmundson GJG, editors. The science of cognitive behavioral therapy. San Diego, CA, US: Elsevier Academic Press; 2017:291-318.
17. Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, Gollan JK, et al. A component analysis of cognitive-behavioral treatment for depression. J Consult Clin Psychol 1996;64(2):295-304. [Medline: 8871414]
18. Martell CR, Herman-Dunn R, Dimidjian S. Behavioral activation for depression. New York: The Guilford Press; 2010.
19. Lejuez CW, Hopko DR, Hopko SD. A brief behavioral activation treatment for depression. Treatment manual. Behav Modif 2001;25(2):255-286. [doi: 10.1177/0145445501252005] [Medline: 11317637]
20. Richards DA, Ekers D, McMillan D, Taylor RS, Byford S, Warren FC, et al. Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (COBRA): A randomised, controlled, non-inferiority trial. Lancet 2016;388(10047):871-880 [FREE Full text] [doi: 10.1016/S0140-6736(16)31140-0] [Medline: 27461440]
21. Parikh SV, Quilty LC, Ravitz P, Rosenbluth M, Pavlova B, Grigoriadis S, CANMAT Depression Work Group. Clinical guidelines for the management of adults with major depressive disorder: Section 2. Psychological treatments. Can J Psychiatry 2016;61(9):524-539 [FREE Full text] [doi: 10.1177/0706743716659418] [Medline: 27486150]
22. Manos RC, Kanter JW, Busch AM. A critical review of assessment strategies to measure the behavioral activation model of depression. Clin Psychol Rev 2010;30(5):547-561. [doi: 10.1016/j.cpr.2010.03.008] [Medline: 20444531]
23. Huguet A, Miller A, Kisely S, Rao S, Saadat N, McGrath PJ. A systematic review and meta-analysis on the efficacy of Internet-delivered behavioral activation. J Affect Disord 2018;235:27-38. [doi: 10.1016/j.jad.2018.02.073] [Medline: 29649708]
24. Nyström MBT, Stenling A, Sjöström E, Neely G, Lindner P, Hassmén P, et al. Behavioral activation versus physical activity via the internet: A randomized controlled trial. J Affect Disord 2017;215:85-93 [FREE Full text] [doi: 10.1016/j.jad.2017.03.018] [Medline: 28319696]
25. Goldberg SB, Tucker RP, Greene PA, Davidson RJ, Kearney DJ, Simpson TL. Mindfulness-based cognitive therapy for the treatment of current depressive symptoms: A meta-analysis. Cogn Behav Ther 2019:1-18. [doi: 10.1080/16506073.2018.1556330] [Medline: 30732534]
26. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
27. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-Based mindfulness interventions for mental health treatment: Systematic review and meta-analysis. JMIR Ment Health 2018;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]
28. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: A randomised controlled trial. BMJ Open 2014;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
29. van der Velden AM, Kuyken W, Wattar U, Crane C, Pallesen KJ, Dahlgaard J, et al. A systematic review of mechanisms of change in mindfulness-based cognitive therapy in the treatment of recurrent major depressive disorder. Clin Psychol Rev 2015;37:26-39 [FREE Full text] [doi: 10.1016/j.cpr.2015.02.001] [Medline: 25748559]

30. Faissner M, Kriston L, Moritz S, Jelinek L. Course and stability of cognitive and metacognitive beliefs in depression. Depress Anxiety 2018;35(12):1239-1246. [doi: 10.1002/da.22834] [Medline: 30277625]
31. Moritz S, Paulus AM, Hottenrott B, Weierstall R, Gallinat J, Kühn S. Imaginal retraining reduces alcohol craving in problem drinkers: A randomized controlled trial. J Behav Ther Exp Psychiatry 2019;64:158-166. [doi: 10.1016/j.jbtep.2019.04.001] [Medline: 31071483]
32. Martell CR, Dimidjian S, Hermann-Dunn R. Verhaltensaktivierung bei Depression: Eine Methode zur Behandlung von Depression. [Behavioral activation in depression: An intervention to treat depression]. Stuttgart: Kohlhammer; 2015.
33. Segal ZV, Teasdale JD, Williams M. Das MBCT-Arbeitsbuch: Ein 8-Wochen-Programm zur Selbstbefreiung von Depressionen und emotionalem Stress. [The MBCT workbook: An 8-week program to free yourself from depression and emotional distress]. Freiburg: Arbor; 2015.
34. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [Medline: 2002127]
35. IBM Corp. IBM. 2017. IBM SPSS Statistics URL: https://www.ibm.com/products/spss-statistics
36. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: Sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect Disord 2004;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8] [Medline: 15183601]
37. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
38. Fuhr K, Hautzinger M, Krisch K, Berking M, Ebert DD. Validation of the Behavioral Activation for Depression Scale (BADS)-Psychometric properties of the long and short form. Compr Psychiatry 2016;66:209-218. [doi: 10.1016/j.comppsych.2016.02.004] [Medline: 26995255]
39. Teismann T, Ertle A, Furka N, Willutzki U, Hoyer J. The German version of the Behavioral Activation for Depression Scale (BADS): A psychometric and clinical investigation. Clin Psychol Psychother 2016;23(3):217-225. [doi: 10.1002/cpp.1948] [Medline: 25772711]
40. Höfling V, Ströhle G, Michalak J, Heidenreich T. A short version of the Kentucky Inventory of Mindfulness Skills. J Clin Psychol 2011;67(6):639-645. [doi: 10.1002/jclp.20778] [Medline: 21294117]
41. Baum C, Kuyken W, Bohus M, Heidenreich T, Michalak J, Steil R. The psychometric properties of the Kentucky Inventory of Mindfulness Skills in clinical populations. Assessment 2010;17(2):220-229. [doi: 10.1177/1073191109356525] [Medline: 20040728]
42. Rojas R, Geissner E, Hautzinger M. DAS-18 Form A und Form B. [DAS-18 form A and form B]. Diagnostica 2015;61(4):173-183. [doi: 10.1026/0012-1924/a000123]
43. Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004;13(2):299-310. [Medline: 15085902]
44. Kinnear P, Gray C. SPSS 16 made simple. New York: Psychology Press; 2009.
45. Lemmens LHJM, Müller VNLS, Arntz A, Huibers MJH. Mechanisms of change in psychotherapy for depression: An empirical update and evaluation of research aimed at identifying psychological mediators. Clin Psychol Rev 2016;50:95-107. [doi: 10.1016/j.cpr.2016.09.004] [Medline: 27770716]
46. Hayes A. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York: Guilford Publications; 2013.
47. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002;59(10):877-883. [Medline: 12365874]
48. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [Medline: 11556941]
49. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
50. Pasarelu CR, Andersson G, Bergman NL, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: A systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
51. Euteneuer F, Dannehl K, Del Rey A, Engler H, Schedlowski M, Rief W. Immunological effects of behavioral activation with exercise in major depression: An exploratory randomized controlled trial. Transl Psychiatry 2017;7(5):e1132 [FREE Full text] [doi: 10.1038/tp.2017.76] [Medline: 28509904]
52. Forand NR, Barnett JG, Strunk DR, Hindiyeh MU, Feinberg JE, Keefe JR. Efficacy of guided iCBT for depression and mediation of change by cognitive skill acquisition. Behav Ther 2018;49(2):295-307 [FREE Full text] [doi: 10.1016/j.beth.2017.04.004] [Medline: 29530267]
53. Takagaki K, Okamoto Y, Jinnin R, Mori A, Nishiyama Y, Yamamura T, et al. Mechanisms of behavioral activation for late adolescents: Positive reinforcement mediate the relationship between activation and depressive symptoms from pre-treatment to post-treatment. J Affect Disord 2016;204:70-73. [doi: 10.1016/j.jad.2016.06.046] [Medline: 27341422]

54. Whiteford HA, Harris MG, McKeon G, Baxter A, Pennell C, Barendregt JJ, et al. Estimating remission from untreated major depression: A systematic review and meta-analysis. Psychol Med 2013;43(8):1569-1585. [doi: 10.1017/S0033291712001717] [Medline: 22883473]
55. Härter M, Klesse C, Bermejo I, Schneider F, Berger M. Unipolar depression: Diagnostic and therapeutic recommendations from the current S3/National Clinical Practice Guideline. Dtsch Arztebl Int 2010;107(40):700-708 [FREE Full text] [doi: 10.3238/arztebl.2010.0700] [Medline: 21031129]
56. Schröder J, Berger T, Meyer B, Lutz W, Späth C, Michel P, et al. Impact and change of attitudes toward Internet interventions within a randomized controlled trial on individuals with depression symptoms. Depress Anxiety 2018;35(5):421-430. [doi: 10.1002/da.22727] [Medline: 29489038]
57. Cludius B, Schröder J, Moritz S. Expectancy effects in self-help depression treatment: First evidence that the rationale given for an online study impacts the outcome. Behav Cogn Psychother 2018;46(2):195-208. [doi: 10.1017/S1352465817000571] [Medline: 29081332]
58. Kuehner C. Why is depression more common among women than among men? Lancet Psychiatry 2017;4(2):146-158. [doi: 10.1016/S2215-0366(16)30263-2] [Medline: 27856392]

Abbreviations

BA: behavior activation.
BADS: Behavioral Activation for Depression Scale.
CAU: care as usual.
CBT: cognitive behavioral therapy.
CC: complete cases.
DAS: Dysfunctional Attitude Scale.
iBA: internet-based behavioral activation.
iMBI: internet-based mindfulness intervention.
ITT: intention-to-treat.
KIMS: Kentucky Inventory of Mindfulness Skills.
MDD: Major depressive disorder.
MBCT: mindfulness-based cognitive therapy.
POI: psychological online intervention.
WHOQOL: World Health Organization Quality of Life

Edited by G Eysenbach; submitted 01.07.19; peer-reviewed by F Lanfranchi, M Economides; comments to author 10.09.19; revised version received 23.10.19; accepted 16.12.19; published 26.03.20.

Please cite as:
Jelinek L, Arlt S, Moritz S, Schröder J, Westermann S, Cludius B
Brief Web-Based Intervention for Depression: Randomized Controlled Trial on Behavioral Activation
J Med Internet Res 2020;22(3):e15312
URL: http://www.jmir.org/2020/3/e15312/
doi:10.2196/15312
PMID:32213470

Original Paper

Improving Self-Care in Patients With Coexisting Type 2 Diabetes and Hypertension by Technological Surrogate Nursing: Randomized Controlled Trial

Calvin Kalun Or[1], PhD; Kaifeng Liu[1], PhD; Mike K P So[2], PhD; Bernard Cheung[3], MB BChir, PhD; Loretta Y C Yam[4,5], MBBS; Agnes Tiwari[6], RN, PhD; Yuen Fun Emmy Lau[4], MBChB; Tracy Lau[4], RN; Pui Sze Grace Hui[7], MBBS; Hop Chun Cheng[7], RN; Joseph Tan[8], PhD; Michael Tow Cheung[1,5], PhD

[1]Department of Industrial and Manufacturing Systems Engineering, University of Hong Kong, Hong Kong, Hong Kong

[2]Department of Information Systems, Business Statistics, and Operations Management, Hong Kong University of Science and Technology, Hong Kong, Hong Kong

[3]Department of Medicine, University of Hong Kong, Hong Kong, Hong Kong

[4]Ambulatory Diabetes Centre, Pamela Youde Nethersole Eastern Hospital, Hong Kong, Hong Kong

[5]Jockey Club School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, Hong Kong

[6]School of Nursing, Hong Kong Sanatorium and Hospital, Hong Kong, Hong Kong

[7]Diabetes Mellitus Centre, Tung Wah Eastern Hospital, Hong Kong, Hong Kong

[8]DeGroote School of Business, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Calvin Kalun Or, PhD
Department of Industrial and Manufacturing Systems Engineering
University of Hong Kong
Room 8-7, Haking Wong Building
Hong Kong
Hong Kong
Phone: 852 3917 2587
Email: klor@hku.hk

Abstract

Background: Technological surrogate nursing (TSN) derives from the idea that nurse-caregiver substitutes can be created by technology to support chronic disease self-care.

Objective: This paper begins by arguing that TSN is a useful and viable approach to chronic disease self-care. The analysis then focuses on the empirical research question of testing and demonstrating the effectiveness and safety of prototype TSN supplied to patients with the typical complex chronic disease of coexisting type 2 diabetes and hypertension. At the policy level, it is shown that the data allow for a calibration of TSN technology augmentation, which can be readily applied to health care management.

Methods: A 24-week, parallel-group, randomized controlled trial (RCT) was designed and implemented among diabetic and hypertensive outpatients in two Hong Kong public hospitals. Participants were randomly assigned to an intervention group, supplied with a tablet-based TSN app prototype, or to a conventional self-managing control group. Primary indices—hemoglobin A_{1c}, systolic blood pressure, and diastolic blood pressure—and secondary indices were measured at baseline and at 8, 12, 16, and 24 weeks after initiation, after which the data were applied to test TSN effectiveness and safety.

Results: A total of 299 participating patients were randomized to the intervention group (n=151) or the control group (n=148). Statistically significant outcomes that directly indicated TSN effectiveness in terms of hemoglobin $_{1c}$ were found in both groups but not with regard to systolic and diastolic blood pressure. These findings also offered indirect empirical support for TSN safety. Statistically significant comparative changes in these primary indices were not observed between the groups but were suggestive of an operational calibration of TSN technology augmentation. Statistically significant changes in secondary indices were obtained in one or both groups, but not between the groups.

Conclusions: The RCT's strong behavioral basis, as well as the importance of safety and effectiveness when complex chronic illness is proximately self-managed by layperson patients, prompted the formulation of the empirical joint hypothesis that TSN would improve patient self-care while satisfying the condition of patient self-safety. Statistical and decision analysis applied to the experimental outcomes offered support for this hypothesis. Policy relevance of the research is demonstrated by the derivation of a data-grounded operational calibration of TSN technology augmentation with ready application to health care management.

Trial Registration: ClinicalTrials.gov NCT02799953; https://clinicaltrials.gov/ct2/show/NCT02799953

(J Med Internet Res 2020;22(3):e16769) doi:10.2196/16769

KEYWORDS

technological surrogate nursing; eHealth; complex chronic disease; diabetes; hypertension; self-care; patient safety

Introduction

The management of chronic disease requires dedicated and joint efforts from health care professionals and patients. Current literature has emphasized the importance of empowering the latter in the direction of self-care [1-4]. Individuals with long-term conditions, such as diabetes and hypertension, are required to maintain proper levels of blood glucose (BG) and blood pressure (BP), follow medication instructions, and lead healthy lifestyles. However, continuing challenges over compliance, motivation, and organization have been noted in the literature [5-9]. Technology is increasingly being called upon to assist in such endeavors. In particular, electronic health (eHealth) protocols are suggested to facilitate patient-doctor-nurse communication, mutually informed decision making, access to health care resources, and the organization, interpretation, and dissemination of health data [10-17]. The ready availability of technology on the supply side prompts a question on the demand side: Would such interventions be accepted and used by chronically ill patients to the extent of significantly improving self-care? Unless this question is resolved endogenously and the success factors explicated, chronic disease management could turn out to be a costly and unrewarding task.

Earlier studies of technological support in chronic disease self-management have focused on single health conditions and have reported mixed results [18-21]. Following de Boer et al [22], we recognize that chronic disease is generally complex, with the coexistence of type 2 diabetes and hypertension being the *common case*. Additionally, it is recognized that individuals who are ill in this way would be best served by care that is given at just the right time, is available all the time, is up to date, and given with a response that is friendly and immediate. As resource limitations render this humanly impracticable, it can be asked whether substitutes produced by technology would encourage and empower chronically ill patients to self-provide the desired care. In answer, we submit that technology would be able to create substitute nurse-caregivers, under the umbrella designation of technological surrogate nursing (TSN). TSN would effectively and safely support self-care on the part of chronically ill patients at the levels of attention, immediacy, and timeliness determined by medical, engineering, and economic considerations. Since TSN can be supplied to meet a wide range of specifications, it can be deemed to be typical among technology-based interventions dedicated to self-care.

In this paper, research supporting our thesis is approached on an empirical level. As noted above, where self-care is required for chronic illnesses, the *common case* is professionally regarded to be that of coexisting type 2 diabetes and hypertension. The first efforts under our empirical research question are, therefore, directed toward testing the effectiveness of TSN when applied to self-care under the conditions of representative chronic disease and representative technology. Safety is submitted to be an equally important requirement in the self-management of complex chronic diseases, where health care responsibility is shared between health care professionals and patients, but proximate action is undertaken by the patients with limited and asymmetric layperson knowledge and medical expertise. Taking the above considerations into account, a randomized controlled trial (RCT) was designed and applied to parallel groups of patients with both diabetes and hypertension to enable the empirical analysis of the effectiveness along with the safety of a prototype eHealth TSN. The two fundamental requirements, together with the experiment's strong behavioral basis, prompted the formulation of the joint hypothesis that TSN would improve patient self-care while satisfying the condition of patient self-safety. In the following sections, our experiment and its outcomes are presented and interpreted and empirical support offered for this joint hypothesis. In addition, policy relevance is demonstrated by the derivation from the data of an operational calibration of TSN technology augmentation with ready application to health care management, such as cost-benefit analysis.

The exposition is organized along the lines of the research question of this paper, in particular with regard to experimental methodology and outcomes interpretation and analysis.

Methods

Trial Design and Participants

We designed and implemented a 24-week, parallel-group RCT to test TSN effectiveness and safety among patients with coexisting type 2 diabetes and hypertension. Participants in the experiment were recruited from two diabetes outpatient clinics of two public hospitals in Hong Kong. The individuals were aged 18 years and over; had received a physician-confirmed diagnosis of type 2 diabetes and hypertension at least one month prior; were prescribed oral medication in consequence; were, by declaration, able to self-manage chronic conditions; and were able to read Chinese and speak Cantonese. Excluded from the study were individuals with visual, cognitive, or physical impairments or with unstable or life-threatening illnesses that precluded self-management.

The study was approved by the Hong Kong East Cluster Research Ethics Committee (reference: HKEC-2015-058). All participants provided written informed consent.

Recruitment and Follow-Up

Outpatients in the diabetes clinics were invited to an information session, where the trial was introduced and eligibility determined. Selected individuals were then visited at home, where the research protocols were explained, eligibility confirmed, and written informed consent obtained. Baseline BP and BG were measured under seated conditions after 5 minutes of sitting rest. BP measurements were made twice, 1 minute apart, by automated 2-in-1 BG and BP monitors; averages of the two readings were computed as examination values. BG levels in terms of hemoglobin A_{1c} (HbA_{1c}) were measured by applying point-of-care HbA_{1c} analyzers. These data were applied to subsequent stratification and analysis. A questionnaire was administered to collect demographic and other health-related information. Participants were randomized to an intervention group (IG) or a control group (CG) as described in the following subsection. Follow-up home visits were made at 8, 12, 16, and 24 weeks postrandomization. HbA_{1c} was measured at 12 and 24 weeks, while BP and other health-related data were recorded at 8, 16, and 24 weeks.

Randomization and Masking

Participants were stratified into four groups according to baseline HbA_{1c} and systolic BP (SBP): (1) HbA_{1c} ≤8.0% and SBP ≤159 mmHg, (2) HbA_{1c} ≤8.0% and SBP ≥160 mmHg, (3) HbA_{1c} >8.0% and SBP ≤159 mmHg, and (4) HbA_{1c} >8.0% and SBP ≥160 mmHg. Parameter values were based on previous research [23,24]. Individuals in each subgroup were randomized by means of randomly permuted blocks of four and six with sequentially numbered, sealed opaque envelopes, all of which were implemented centrally by a research team member over the telephone. Participants and researchers were all blinded to the randomization; however, given the nature of the intervention, blinding was not possible postrandomization.

Prototype Technological Surrogate Nursing and Other Experimental Equipment

Participants randomized to the IG were supplied with prototype eHealth TSN developed by the research team with reference to clinician inputs, needs, and expectations in chronic disease self-care identified in usability evaluations [25,26] and other literature [27,28], as well as funding constraints. Running on a tablet computer, this protocol offered, within limits of its design and with due attention to safety, timely and interactive access to procedures and resources dedicated to the self-management of type 2 diabetes and hypertension. In particular, measurement and recording of BG and BP were enabled through Bluetooth-connected BG and BP monitors. These data, presented in structured tables and charts, were readily retrievable and reviewable. Authorized outside individuals (eg, caregivers and family members) were allowed access to the TSN platform through a secure Web portal. A module was available to provide text- and video-based learning resources pertaining to the causes and prevention of type 2 diabetes and hypertension, self-care, exercise, diet, health plans, and stress management. An audio function was programmable to emit action reminders at predetermined times. CG participants were supplied with BG and BP monitors of the same type, with logbooks as specified under conventional self-management protocols. Individuals in both groups were trained and encouraged to apply the tools in question to self-monitor and self-record BG and BP, and to adhere to other self-care activities recommended by primary care providers.

Outcome Measures

TSN effectiveness was measured, firstly, by indices representing the fundamental clinical manifestations under complex diabetes and hypertension: BG, as indicated by HbA_{1c}, and BP, as indicated by SBP and diastolic BP (DBP) [23,24]. TSN effectiveness, at a level once-removed from the clinical level, was measured in terms of the following secondary indices:

1. Medication adherence, defined by the frequency of failure to follow medication during the past 2 months and evaluated by five items adapted from the Morisky Medication Adherence Scale [29]: 1 (never), 2 (rarely), 3 (sometimes), 4 (always), and 5 (all the time). Item scores were reversed during data analysis to allow higher scores to correspond to greater adherence.
2. General adherence to treatment, defined by whether the individual followed prescribed treatments during the past 2 months and evaluated by the five-item Medical Outcomes Study General Adherence Scale [30]: 1 (none of the time), 2 (a little of the time), 3 (some of the time), 4 (a good bit of the time), 5 (most of the time), and 6 (all of the time).
3. Adherence to disease-specific activities, defined by the frequency of performance of disease-specific activities during the past 2 months and evaluated using the 15-item Medical Outcomes Study Disease-Specific Adherence Scale [30]: 1 (none of the time), 2 (a little of the time), 3 (some of the time), 4 (a good bit of the time), 5 (most of the time), and 6 (all of the time).
4. Diabetes knowledge and hypertension knowledge assessed in terms of responses to 11 and 25 true or false questions, respectively [31].
5. Self-efficacy for coping with chronic disease, defined as confidence in performing self-management and assessed using five items adapted from a validated scale [32], with scores ranging from 1 (not at all confident) to 10 (totally confident).

Frequencies of self-monitoring of BG and BP were recorded as the number of times per week. As shown in the Discussion section, a method is suggested under which the primary and secondary indices would indirectly indicate the presence or absence of patient self-safety.

Statistical Analysis

According to a priori power calculations based on an assumed 10% loss to follow-up, a sample of 147 individuals per group would provide over 80% power (two-sided alpha=.05) to detect a between-group mean difference of 5 mmHg for SBP (SD 14.5), 4 mmHg for DBP (SD 10), and 0.5% for HbA_{1c} (SD 1.25). Independent *t* tests and chi-square tests were applied to compare differences in baseline characteristics between the IG and the CG members. We applied both intention-to-treat (without imputation) analysis and per-protocol analysis, under linear mixed-modelling of mean changes in indices within and between groups. Effects were classified according to time (ie, follow-ups versus baseline), treatment group (ie, IG versus CG), and group-by-time interaction. Self-monitoring frequency was compared between the two groups using the Mann-Whitney U test. Sensitivity analysis was introduced after adjusting for baseline variables.

Results

Study Characteristics

Between March and October 2016, 151 patients were randomized to the IG and 148 patients were randomized to the CG (see Figure 1 for the trial flowchart). As shown in Table 1, there were no statistically significant differences between the groups with regard to baseline characteristics, except for experience in the use of computer-based self-monitoring systems (*P*=.02).

Figure 1. Trial flowchart. CG: control group; IG: intervention group.

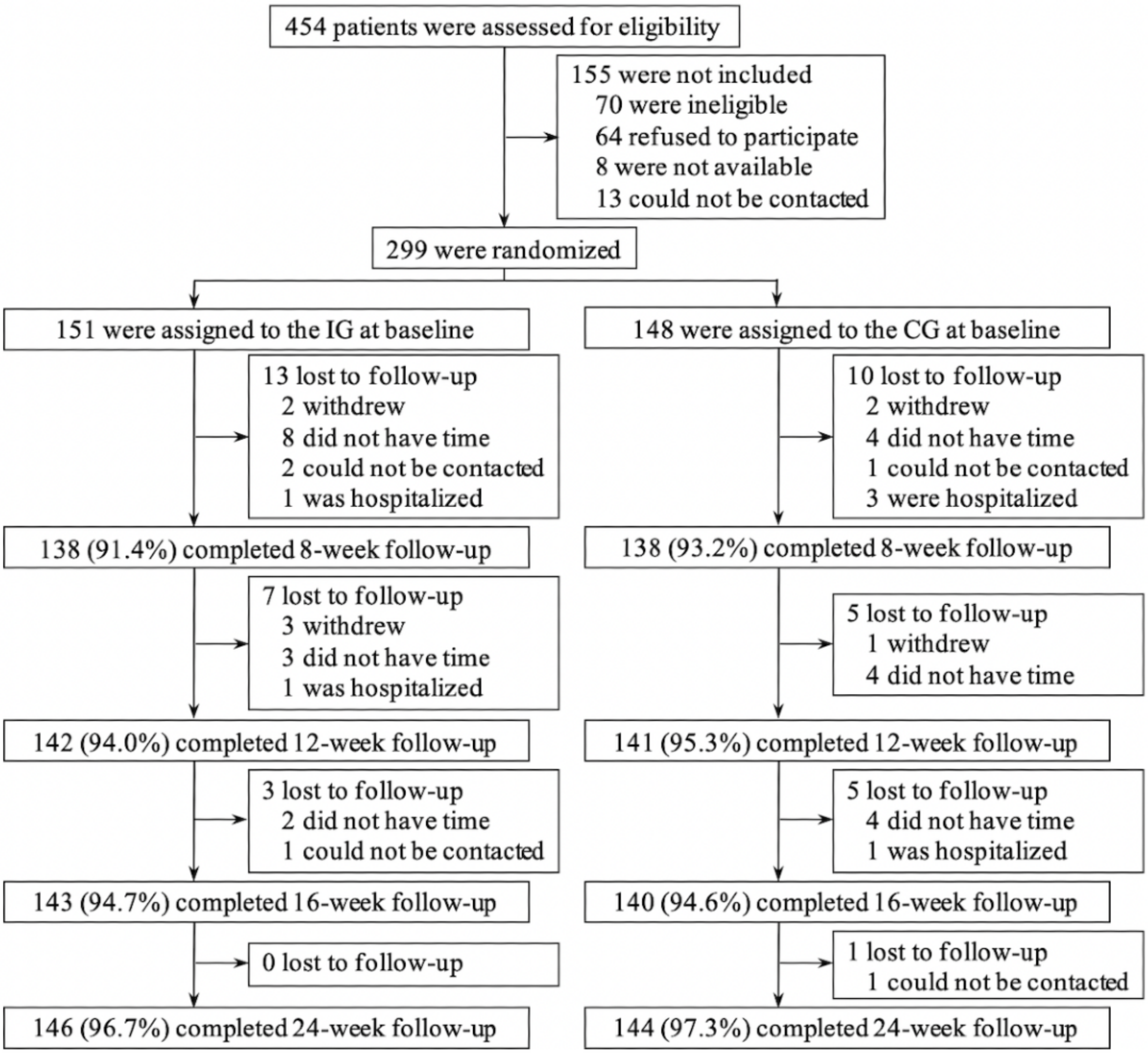


Table 1. Baseline characteristics by study group.

Characteristic	Intervention group (n=151)	Control group (n=148)	*P* value
Age (years), mean (SD)	63.9 (10.2)	63.7 (9.6)	.90
Sex, n (%)			**.09**
Male	104 (68.9)	88 (59.5)	
Female	47 (31.1)	60 (40.5)	
Education, n (%)			**.09**
No schooling completed	0 (0)	4 (2.7)	
Some primary school	10 (6.6)	14 (9.4)	
Completed primary school	25 (16.6)	21 (14.2)	
Some secondary school	23 (15.2)	23 (15.5)	
Completed secondary school	66 (43.7)	47 (31.8)	
Diploma, advanced diploma, associate degree, or the equivalent	10 (6.6)	21 (14.2)	
Bachelor’s degree	14 (9.3)	12 (8.1)	
Master’s degree	3 (2.0)	5 (3.4)	
Doctoral degree	0 (0)	1 (0.7)	
Habitation status, n (%)			**.61**
Living alone	21 (13.9)	15 (10.1)	
Living with family	129 (85.4)	132 (89.2)	
Other	1 (0.7)	1 (0.7)	
Duration of diabetes (years), mean (SD)	15.6 (10.0)	16.6 (11.3)	.42
Duration of hypertension (years), mean (SD)	12.9 (8.7)	12.9 (9.9)	>.99
Experience using computers, tablets, or mobile phones, n (%)			**.85**
Yes	117 (77.5)	116 (78.4)	
No	34 (22.5)	32 (21.6)	
Time spent using a computer, tablet, or mobile phone (hours per day), mean (SD)	2.4 (3.1)	3.2 (4.1)	.06
Experience using computer-based self-management systems, n (%)			**.02**
Yes	3 (2.0)	12 (8.1)	
No	148 (98.0)	136 (91.9)	
Experience with self-management support programs and training, n (%)			**.23**
Yes	100 (66.2)	88 (59.5)	
No	51 (33.8)	60 (40.5)	
Hemoglobin $_{1c}$ (%), mean (SD)	8.02 (1.54)	7.99 (1.23)	.88
Systolic blood pressure (mmHg), mean (SD)	137.6 (17.7)	137.4 (15.4)	.92
Diastolic blood pressure (mmHg), mean (SD)	76.2 (9.8)	74.7 (10.3)	.20

Intention-to-Treat Analysis

Primary Outcomes

Outcomes with direct reference to TSN effectiveness are displayed in Table 2. Mean changes in HbA_{1c} were found to be statistically significant at all assessment times in both the IG and the CG (see Figure 2). Comparing HbA_{1c} mean changes between the two groups, statistically significant outcomes were not observed. With regard to SBP and DBP, mean changes in each group were mixed in statistical significance (see Figure 3). Differences in SBP and DBP mean changes between the IG and the CG were not statistically significant. With regard to TSN patient safety, indirect implications of the above findings will be pursued below in the Discussion section.

Table 2. Results of intention-to-treat analysis.

Outcomes	Intervention group (n=151), mean (95% CI)	Control group (n=148), mean (95% CI)	Between-group difference with regard to change in outcome from baseline (95% CI)	*P* value
Primary outcomes				
Hemoglobin A_{1c} (%)				
Baseline	8.02 (7.80-8.23)	7.99 (7.77-8.21)	N/A[a]	
12 weeks	7.72 (7.50-7.94)[b]	7.65 (7.43-7.87)[b]	0.05 (−0.23 to 0.33)	.74
24 weeks	7.57 (7.35-7.79)[b]	7.64 (7.42-7.86)[b]	−0.09 (−0.37 to 0.19)	.52
Systolic blood pressure (mmHg)				
Baseline	137.6 (134.8-140.3)	137.4 (134.6-140.2)	N/A	
8 weeks	135.2 (132.3-138.0)	132.7 (129.9-135.6)[b]	2.25 (−1.87 to 6.37)	.28
16 weeks	135.9 (133.0-138.7)	134.8 (132.0-137.7)	0.84 (−3.26 to 4.93)	.69
24 weeks	138.1 (135.3-140.9)	134.6 (131.7-137.4)	3.38 (−0.68 to 7.44)	.10
Diastolic blood pressure (mmHg)				
Baseline	76.2 (74.6-77.7)	74.7 (73.1-76.2)	N/A	
8 weeks	74.6 (73.0-76.2)[b]	73.4 (71.8-75.0)	−0.31 (−2.26 to 1.64)	.76
16 weeks	74.8 (73.3-76.4)	74.1 (72.5-75.7)	−0.80 (−2.74 to 1.14)	.42
24 weeks	76.1 (74.5-77.6)	74.2 (72.6-75.8)	0.33 (−1.59 to 2.25)	.73
Secondary outcomes				
Medication adherence (score)				
Baseline	4.52 (4.45-4.59)	4.53 (4.46-4.60)	N/A	
8 weeks	4.64 (4.57-4.71)[b]	4.57 (4.50-4.64)	0.08 (−0.01 to 0.16)	.07
16 weeks	4.58 (4.51-4.65)[b]	4.53 (4.46-4.60)	0.06 (−0.02 to 0.14)	.14
24 weeks	4.58 (4.51-4.65)[b]	4.56 (4.49-4.63)	0.03 (−0.05 to 0.12)	.40
General adherence to treatment (score)				
Baseline	4.17 (4.03-4.32)	4.00 (3.86-4.15)	N/A	
8 weeks	4.15 (4.00-4.30)	3.95 (3.80-4.10)	0.03 (−0.16 to 0.22)	.78
16 weeks	4.17 (4.02-4.32)	4.03 (3.88-4.18)	−0.03 (−0.22 to 0.16)	.75
24 weeks	4.27 (4.12-4.42)	3.98 (3.83-4.13)	0.12 (−0.07 to 0.31)	.21
Adherence to disease-specific activities (score)				
Baseline	3.51 (3.42-3.61)	3.51 (3.41-3.61)	N/A	
8 weeks	3.55 (3.45-3.65)	3.65 (3.55-3.75)[b]	−0.10 (−0.22 to 0.03)	.12
16 weeks	3.73 (3.63-3.83)[b]	3.66 (3.56-3.76)[b]	0.06 (−0.06 to 0.18)	.35
24 weeks	3.72 (3.62-3.82)[b]	3.63 (3.53-3.73)[b]	0.08 (−0.04 to 0.20)	.19
Diabetes knowledge (%)				
Baseline	78.5 (76.7-80.3)	79.1 (77.2-80.9)	N/A	
8 weeks	81.5 (79.7-83.4)[b]	82.6 (80.7-84.5)[b]	−0.53 (−3.21 to 2.15)	.70
16 weeks	84.2 (82.4-86.1)[b]	84.1 (82.2-86.0)[b]	0.65 (−2.01 to 3.31)	.63
24 weeks	84.4 (82.6-86.3)[b]	85.4 (83.5-87.2)[b]	−0.40 (−3.04 to 2.24)	.77
Hypertension knowledge (%)				
Baseline	72.4 (70.7-74.2)	70.9 (69.1-72.6)	N/A	

Outcomes	Intervention group (n=151), mean (95% CI)	Control group (n=148), mean (95% CI)	Between-group difference with regard to change in outcome from baseline (95% CI)	*P* value
8 weeks	73.2 (71.4-75.0)	73.6 (71.7-75.4)[b]	−1.95 (−4.38 to 0.47)	.11
16 weeks	75.6 (73.8-77.4)[b]	74.9 (73.1-76.7)[b]	−0.82 (−3.22 to 1.59)	.50
24 weeks	76.7 (74.9-78.5)[b]	76.2 (74.4-78.0)[b]	−1.01 (−3.40 to 1.38)	.41
Self-efficacy for coping with chronic disease (score)				
Baseline	7.31 (7.09-7.53)	6.98 (6.75-7.21)	N/A	
8 weeks	7.38 (7.15-7.61)	7.17 (6.94-7.40)	−0.12 (−0.39 to 0.15)	.38
16 weeks	7.45 (7.22-7.68)	7.24 (7.02-7.47)[b]	−0.12 (−0.39 to 0.14)	.36
24 weeks	7.49 (7.26-7.72)	7.24 (7.02-7.47)[b]	−0.08 (−0.35 to 0.18)	.55

[a] N/A: not applicable.

[b] Significant difference from baseline.

Figure 2. Mean hemoglobin A_{1c} (HbA_{1c}) levels during the 24-week study period.

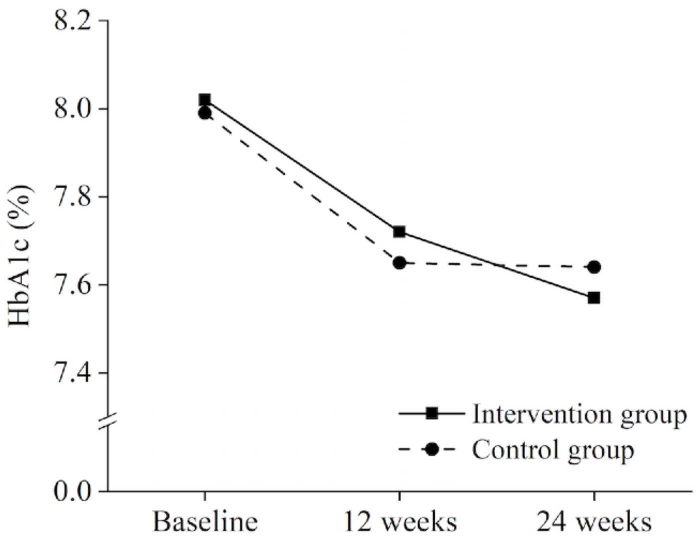
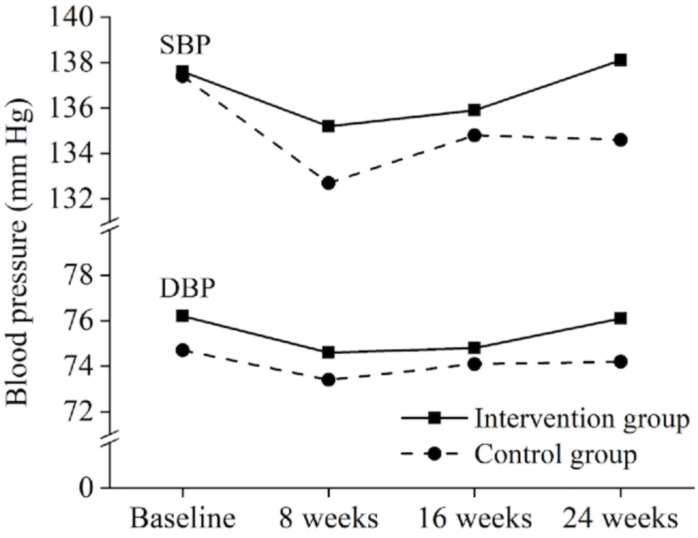


Figure 3. Mean systolic blood pressure (SBP) and diastolic blood pressure (DBP) during the 24-week study period.



Secondary Outcomes

As shown in Table 2, statistically significant direct improvements in TSN effectiveness as measured by adherence to disease-specific activities, diabetes knowledge, and hypertension knowledge were observed in both the IG and the CG at 16 and 24 weeks. Mixed results were obtained for the shorter times. Improvements in medication adherence were significant in the IG at all follow-up points, and improvements in self-efficacy for coping with chronic diseases were significant for the CG at 16 and 24 weeks. Outcomes for general adherence to treatment were all statistically nonsignificant. Significant differences in the secondary indices were not observed between the groups at any follow-up points. Indirect implications of the above findings for TSN patient safety will be pursued in the Discussion section.

Per-Protocol and Sensitivity Analyses

Per-protocol analysis showed that in the IG, the statistically nonsignificant intention-to-treat outcomes for DBP (at 16 weeks) and for self-efficacy in coping with chronic disease (at 24 weeks) became significant, with the opposite change applying to medication adherence at 16 weeks and 24 weeks (see Multimedia Appendix 1). Sensitivity analysis showed that the results were unchanged after adjusting for baseline variables, with the exception that in the CG, the statistically nonsignificant change in self-efficacy in coping with chronic disease at 8 weeks became significant (see Multimedia Appendix 2).

Self-Monitoring Frequency

Over the 24-week study period, participants in both groups displayed similar mean frequencies (times per week) with regard to self-monitoring of BG (IG mean 4.67, SD 4.45, median 3.65; CG mean 4.47, SD 3.67, median 3.62; *P*=.94) and with regard to self-monitoring of BP (IG mean 5.26, SD 6.03, median 3.46; CG mean 4.07, SD 3.54, median 3.17; *P*=.16).

Discussion

Principal Findings

Statistically significant outcomes that directly indicate TSN effectiveness in terms of HbA_{1c} were found for both the IG and the CG, but not with regard to SBP and DBP. Significant comparative changes in the three primary indices were not observed between the groups. Significant changes in secondary indices were obtained in one or both groups, but not between the groups.

Interpretation of Individual Outcomes

In line with its empirical orientation, this study is focused on measurement, outcome interpretation, and hypothesis testing. On the analytical level, it is sufficient for such purposes that the data were derived from the impact of TSN-supported or conventional self-management on diabetic and hypertensive patient behavior, which in turn was structured under the experimental design and manifested empirically by way of the primary and secondary indices. With regard to the interpretation of outcomes, the RCT’s strong behavioral basis prompts a reference to the axiom of rationality in economics and decision science. Under this axiom, individual behavior and changes therein are required to not consistently inflict self-harm [33]. In health care, the idea of *no harm* is embedded in a *sine qua non*, which requires that any treatment or situations must be safe for the individuals concerned. A similar *sine qua non* in the aviation industry is that of passenger safety.

The requirement of *no harm* is especially important in the self-management of complex chronic illnesses, such as diabetes and hypertension, where health care professionals and patients share the health care responsibility but proximate action is undertaken by patients with asymmetric and limited layperson knowledge of the interdependent medical manifestations and effects. In view of this action and knowledge asymmetry, *no harm* can be operationally interpreted to represent the minimum level of safety perceived to be acceptable by the rational patient with regard to their own health care−related activities. As this would be the case with every individual involved in every self-managed health care situation, *no harm* can be formalized linguistically to mean the lower bound of self-safety. Mathematically, *no harm* can, therefore, be understood to be the denoting description of the minimal element in a set of self-managed health situations ordered by the perception-based asymmetric binary relation *≤*, read from right to left as "(perceived to be) more than or equally safe."

To facilitate the interpretation of outcomes in terms of the health care *sine qua non*, a short mathematical discussion is presented [34]. The concepts introduced above can be operationalized in terms of a numerical mapping to represent the total or partial ordering *≤*. A number of technical problems then arise, as follows: the existence of many levels of self-safety leads to the questions of defining its maximum commensurability between factors entering such a definition; the possibility that an upper limit cannot be meaningfully delineated so that, in defiance of practical experience, the set of health situations must be left open ended (ie, specified to be of countably infinite cardinality); differentiation between *maximum safety* and neighboring *very safe* situations; and the order-intransitivity that emerges when this distinction is not clearly established. Until problems of this nature are resolved, the domain of self-managed health situations would be incompletely defined and partially ordered. Numerical representation of the self-safety relation *≤* would require continuous real mapping over such a domain, which additionally must be monotonic and unique under linear transformation in any completely defined and ordered subdomain. The specification of a probability function and a random variable to represent patient self-safety under such circumstances would involve additional applications of measure theory.

Under the study's empirical approach, these analytical difficulties can be mitigated by careful experimental design and data analysis. Reference to the formal description presented above suggests that the *sine qua non* of safety can be satisfied in the strong form, which involves the condition of at least *no harm*, or in the weak form, which involves the condition of *no harm*. To operationalize and test the empirical existence of the *sine qua non* in the case of chronic disease self-management, where the outcomes have an important dependence on layperson knowledge and action, it would be sufficient to demonstrate that its weak form requiring the singleton instance of *no harm*, or patient-perceived minimum safety, is supported by the data. The measurement problem would then be considerably less demanding than if the strong form of the *sine qua non* is involved; witness the case of health care quality assurance, where accommodation of the strong form would lead to incompletely resolved problems of cardinal measurement and numerical comparability of incremental safety over many situations [35].

An additional dimension to the measurement problem arises in the case of complex chronic diseases, such as diabetes and hypertension (ie, that illness is generally manifested in more than one form). Numerical representation of patient self-safety would, therefore, require reference to different health manifestations under a given complex chronic disease (eg, BG and BP in the case of diabetes and hypertension). As shown in the Methods section, BG and BP manifestations can properly be measured in terms of HbA_{1c}, SBP, and DBP indices. Instead of adding complexity to the numerical representation of self-safety, availability of these index numbers suggests a method to reduce the difficulties by carefully interpreting the experimental design and experimental data.

To ease the exposition, we first present three supporting observations. First, because an index number is a mapping function from a set of data to a set of real numbers, the suggested method exploits the mathematical property that under a function, each element in its range—an index value in this case—is the image of a distinct element in the domain—a data point in this case. Second, since index functions are operational and, hence, mathematically constructible, it is possible to focus attention on one computed point in its range—an index value—and one point in its domain—a data point—so that a 1:1 relationship would follow *ipso facto*. Third, it was pointed out earlier that difficult analytical problems must be resolved before a random variable can be constructed to characterize self-safety under sampling. Though statistical tests applicable to such conditions are not yet available, a second-best approach can properly be sought to extract from the data the wherewithal to indicate whether the health care *sine qua non* is present in weak form.

A second-best method to exploit index number statistical results to indicate minimum patient self-safety can be conveniently explained by example. As reported in Table 2, results involving the HbA_{1c} index were statistically significant in the IG at all assessment dates. TSN can then be concluded to have impacted positively on patient self-care. Positive impact on patient well-being follows immediately, especially as it has been shown that lowering HbA_{1c} by 1% can translate to as much as a 40% reduction in the risk of microvascular complications and a 20% reduction in the risk of diabetes-related deaths [36]. Individuals reaping health benefits of this nature can properly be deemed to be at least unharmed; therefore, with regard to TSN, minimum patient self-safety would be assured upon indirect interpretation of the HbA_{1c} index results. Similar reasoning would apply to the other statistically significant index number outcomes in the IG (see Table 2) (ie, DBP for 8 weeks; medication adherence for 8, 16, and 24 weeks; adherence to disease-specific activities for 16 and 24 weeks; diabetes knowledge for 8, 16, and 24 weeks; and hypertension knowledge for 16 and 24 weeks). Referring to the IG outcome for SBP at 24 weeks, it is suggested that since the incremental average in question was measured to be numerically positive and, hence, clinically negative, nonsignificance would instead become the statistical property of concern. Given that this was indeed found to be the case so that statistically significant evidence of self-harm was not forthcoming, it is proper to indirectly infer that TSN was accompanied by minimum patient self-safety in this instance.

The remaining index number outcomes in the IG, though statistically nonsignificant (see Table 2), are amenable to *empirically necessary reasoning* under the second-best method. To explain by example, we first refer to the CG data to draw an inference regarding patient self-safety on the following methodological grounds. Given that self-care was practiced among the CG members following established health care procedures, it can be assumed *ipso facto* that, independent of effectiveness, the resulting outcomes would be at least not harmful and, hence, at least minimally safe. Average index data from the group can, therefore, be properly assumed to have been produced under the auspices of patient self-management characterized by at least *no harm*; hence, they can be sifted for necessary indications of minimum self-safety even in the absence of formal statistical analysis. Notice that the *consistent* condition of the underlying rationality axiom referred to above is accommodated by averaging in the index measurements.

Recall the idea, which was broached above, that index numbers measuring health manifestations such as BP can be exploited to reduce difficulties when measuring patient safety. Consider then how the SBP and DBP primary indices would perform under the suggested data gleaning—as shown earlier, HbA_{1c} has been pre-empted into the group of indices appropriate for indirect interpretation. From Table 2, it is seen that in the CG, DBP at 8, 16, and 24 weeks showed hypertension improvements before statistical testing, so that the incremental index averages (ie, 1.26, 0.53, and 0.44, respectively) expressed clinically positive effects at these dates. By virtue of the fact that only CG data entered into these computations, at least *no harm*—and *a fortiorino harm*—can be properly assumed to have accompanied the self-care that produced the effects in question. Combining these observations and the mathematical properties of functions noted earlier, it is argued that numerical representation of minimum patient self-safety can be attributed to the DBP index when applied to CG average incremental data. This is true in the sense that such a computation would provide the basis for a clinically meaningful index value to be assigned in order to functionally image a data point known to have been produced from a *no harm* situation. By implication, under the 1:1 linguistic relationship, this index value would also numerically represent the denoting minimum self-safety itself. It is thus possible, under this kind of *empirically necessary reasoning*, to interpret the DBP reading of 1.26 mmHg to be the numerical representation at 8 weeks of the minimum self-safety denoting a *no harm* situation embedded in the CG data at the same time. Similar reasoning suggests that the other DBP index values of 0.53 and 0.44 in the CG can be interpreted to numerically represent minimum self-safety at 16 weeks and 24 weeks, respectively. Repeating the argument, the SBP incremental index values of 4.65, 2.54, and 2.81 in Table 2 can be understood to functionally image *no harm* data points for the CG and, hence, to numerically represent the denoting minimum self-safety at 8, 16, and 24 weeks, respectively.

Empirically necessary indications of minimum patient self-safety obtained on the basis of CG data can be carried over to the IG by reference to the between-group comparative findings in Table 2; they can also be carried over by exploiting the fact that the individuals involved were originally part of a single cohort satisfying the same basic health criteria (see the Methods section). Given this shared origin and the random nature of patient assignment to the IG and the CG, behavior in the two groups can properly be assumed to derive from similar perceptions regarding safety when self-managing diabetes and hypertension. Returning to the result that differences in SBP and DBP outcomes between the groups were statistically nonsignificant and focusing on the clinically positive cases, we can now interpret the index averages in question to be empirically equivalent, pairwise. It is further noted that between the two groups, individual indices can be matched 1:1 with assessment date as the parameter. In Table 2, it is therefore seen that clinically positive incremental index averages for DBP at 8, 16, and 24 weeks in the CG can be matched with empirically equivalent, clinically positive incremental index averages for DBP at 8, 16, and 24 weeks in the IG. Understanding this exercise in terms of empirically necessary reasoning, indications of minimum patient self-safety obtained in the CG are thus carried over to the IG. Repeating the argument would show that clinically positive incremental index averages for SBP at 8 and 16 weeks in the CG can be matched with empirically equivalent clinically positive incremental index averages for SBP at 8 and 16 weeks in the IG, so a carryover can again be made. The 24-week IG reading for SBP was found to be clinically negative, so it cannot be matched with the clinically positive 24-week SBP reading in CG. Instead, as already shown, it was interpreted indirectly.

Therefore, it is observed that a second-best method, based on index numbers and applying indirect statistical interpretation or empirically necessary reasoning, can be found to extract support for the presence of minimum self-safety from the experimental data, even when formal statistical tests are not available. To the extent that the secondary index outcomes satisfy the method's conditions, it would again apply. Similar to the previous argument, findings of statistically nonsignificant differences in secondary outcomes between the IG and the CG (see Table 2) would allow clinically positive incremental index averages to be identified and matched using assessment date as the parameter. It is seen from Table 2 that minimum self-safety would then be indicated in the cases of medication adherence, adherence to disease-specific activities, diabetes knowledge, hypertension knowledge, and self-efficacy for coping with chronic disease. Application to the remaining secondary outcome of general adherence to treatment is methodologically excluded.

General adherence to treatment covers a wider scope than the other secondary indices, which are all focused on diabetes and hypertension. The activities in question—dieting, exercise, and weight control—are important to self-management under other chronic conditions that may afflict the participants under study (eg, smoking, obesity, and chronic obstructive pulmonary disease) and that may be differently supported by technology or not at all. Asynchronous progress in these directions would impact the patient's diabetic and hypertensive status in a distributed lag, which cannot be meaningfully compared with the single-dated effects measured under the other secondary indices. Therefore, it is proposed that until the research is properly extended, general adherence to treatment should be excluded from outcomes interpretation involving the other secondary indices, on methodological grounds of differences in scope, behavioral basis, and *ceteris paribus* specifications.

Further interpretation of the BP-based outcomes is suggested if a maintained hypothesis implicit in the experimental design is recalled, that additional knowledge is a necessary condition for consistent changes in behavior. Referring to the statistically nonsignificant SBP and DBP outcomes and the statistically significant knowledge-related outcomes in Table 2, an anomaly would seem to emerge, which is that individuals learned more about hypertension and yet significant impact on BP control did not follow. A resolution is suggested by noting that because of diminishing returns on hypertension treatment [37], small changes in BP such as the ones observed are likely in the case of long-term patients. We also note the view that it is not difficult for the layperson to understand the rudimentary aspects of hypertension [38]. It is then additionally likely that the long-term patients entering each group already possessed working knowledge of hypertension. It follows that with little addition to hypertension knowledge, the behavior of long-term patients would be little affected, to the extent of producing statistically insignificant changes in already-low SBP and DBP measurements.

Effectiveness Along With Patient Safety Under Technological Surrogate Nursing

It was noted in the previous section that empirical analysis of TSN effectiveness and TSN self-safety should properly proceed in parallel. Under the health care *sine qua non*, the first question asked in practice is whether *no harm* would hold in an absolute sense inside the case with minimum reference to outside situations. This suggests that the real-world evaluation of TSN effectiveness should also be approached from an absolute perspective, under which the RCT outcomes are interpreted in the capacity of empirical analysis subject to strong other-things-being-equal conditions. This procedure is in line with common scientific practice, under which research is generally evaluated along two headings depending on *ceteris paribus* conditions: absolutely or relatively. Safety is similarly important as effectiveness in chronic disease self-management due to layperson-patient involvement. In addition, the experimental data were measured in terms of independent indices under clinically delineated conditions; hence, they are, mensuration-wise, consistent with the absoluteness required by the health care *sine qua non*. Consequently, it is methodologically permissible for TSN self-safety to formally appear together with TSN effectiveness to be empirically tested under a joint hypothesis. We, therefore, propose the following: other things being equal, TSN would improve self-care among chronically ill diabetic and hypertensive patients while satisfying the *sine qua non* of patient self-safety.

It has been shown that 13 statistically significant primary and secondary outcomes supported the empirical existence of direct positive incremental impact of TSN on patient self-care and indirect positive incremental impact on patient self-safety. In addition, 10 primary and secondary outcomes offered empirically necessary indications of accompanying minimum patient self-safety. It was shown that minimum patient self-safety yielded sufficient empirical support for the health care *sine qua non*. It was also shown that the first (ie, absolute outcomes evaluation) heading noted above covered the primary indices HbA_{1c} (2 cases) and DBP (1 case), as well as the secondary indices medication adherence (3 cases), adherence to disease-specific activities (2 cases), diabetes knowledge (3 cases), and hypertension knowledge (2 cases). In addition, it was shown that the outcomes under the second (ie, relative outcomes evaluation) heading covered the primary indices SBP (3 cases) and DBP (2 cases), as well as the secondary indices adherence to disease-specific activities (1 case), hypertension knowledge (1 case), and self-efficacy for coping with chronic disease (3 cases). This listing is seen to be exhaustive, given that the number of possible outcomes forthcoming in the IG sum to the same total of 23 (see Table 2), with one secondary index being excluded on methodological grounds. Therefore, it is submitted that the IG data offer statistically significant and empirically necessary support to the joint hypothesis presented above.

Comparative Findings and Calibration of Technological Surrogate Nursing Technology Augmentation

Under the second scientific approach of relative evaluation, the key statistical finding is that, as compared between the IG and the CG, there was no difference in effectiveness of self-care (see Table 2). Self-safety was demonstrated to be empirically equivalent between the groups and so would play a neutral role in the present argument. As a matter of research methodology, a null result of this nature can properly claim a place in the literature [39]. As discussed in the Methods section, since dedicated technological support was absent in the RCT's CG, it is further suggested that the result's empirical implications are not exhausted but can be usefully pursued under the *quaesitum* of adding technology to self-care. If the RCT is interpreted as a *thought experiment*, we can understand the comparative data to reflect the *as-if* effects of adding technology to diabetes and hypertension self-management, which, up to then, did not make use of it. Thought experiments and as-if reasoning, which were famously exploited by Einstein, would be familiar to physicists [40]. Health care is generally asymmetrical, as witnessed by the near-impossibility of recreating the patient's initial state by reversing treatment. In this case, however, individuals receiving prototype TSN improved in self-care and, yet, individuals without it were not worse off. The apparent anomaly is resolved under the thought experiment by suggesting that the technology in question is so elementary that adding it to standard (ie, control) practice did not produce empirically significant effects. The following question then emerges: Would further addition of technology to the TSN prototype (ie, technology augmentation) increase effectiveness?

Analysis under this question can proceed operationally with support from the comparative data. First, it is suggested that the prototype technology represents the lowest meaningful level in TSN design. The IG data were, therefore, obtained under such a technological condition. Second, we recall the seminal Michelson-Morley experiment and its consequences for calibration in physics [40]. An experimentally grounded base level is, therefore, proposed by analogy, with reference to which technology augmentation in TSN can be operationally calibrated in two dimensions: its degree in any given case being measured by engineering index comparison with the base technology, and its effectiveness being measured against base-outcomes indices.

For example, a flexible internet-based, tethered or mobile, TSN platform with artificial intelligence (AI) capabilities can be determined to be m-times more advanced than the base (ie, prototype) version; this is in terms of an index of information technology and computational requirements and its effectiveness measured incrementally under an index covering the base outcomes. Applying the calibration over a properly chosen range of cases would yield a numerical representation of TSN technology augmentation in terms of a ranking of technology and a ranking of effectiveness. Since expert knowledge can be relied on to supply engineering indices for incremental comparison, the first ranking is cardinal and monotonic with origin determined by base technology. Given that the second ranking is constructed from incremental magnitudes, it is naturally cardinal with origin determined by base effectiveness. However, since patient acceptance is involved, this ranking may or may not be monotonic; it is possible for individuals to feel overloaded with technology, in which case incrementally negative effects leading to local nonmonotonicity would emerge.

The suggested calibration is operational, simple (ie, only two constructs), and empirically grounded. It can readily be introduced into health care management, such as TSN cost-benefit analysis. Following the principle of bounded rationality [33], decision complexity would be reduced by eliminating nonmeaningful alternatives through the extraction of a monotonic subordering from the effectiveness ranking; decision complexity would be further reduced by delineation of benefits with reference to this subordering and of costs with reference to the parallel-technology ranking. The solution to the cost-benefit analysis would then determine optimal TSN and optimal choice of technology; since the latter may or may not be determined to be at the most advanced level, the question of whether introducing more technology into health care would be rational in an already technology-loaded society can be resolved under clearly specified conditions.

Suggestions for Paradigm Development

The data suggest a potentially important implication for the treatment of complex chronic disease and, hence, for TSN paradigm development: the emergence of differential effects. It is noted that in each experimental group, the same individuals receiving the same medication and the same self-care were involved when measuring HbA_{1c}, SBP, and DBP. Changes in these indices can, therefore, be validly compared within the group. Note, the influence of baseline values is excluded under incremental measurement. If the patterns revealed by pairwise comparisons in each group (see Table 2)—falling HbA_{1c} against little change in SBP and falling HbA_{1c} against little change in DBP—are interpreted in terms of health manifestations under complex chronic diabetes and hypertension, it is suggested that patients relatively more afflicted with diabetes would be more responsive to self-management than patients suffering relatively more from hypertension. Though the observation is *chartist*, it suggests the empirical existence of differential effects. If formal statistical evidence is forthcoming, empirical grounds follow to justify paradigm development to extend TSN to selectable multitasking under an umbrella of complex chronic disease (eg, obesity, heart failure, chronic obstructive pulmonary diseases, and chronic skeletal-joint problems, in addition to diabetes and hypertension), so as to encourage technology-induced behavioral change along the lines of greater patient response.

The increasing and increasingly successful applications of AI to health care point to a potentially fruitful direction for TSN paradigm development [41-44]. Referring to the classic cybernetic observation that the individual represents a spatially and temporally local pocket of decreasing entropy (ie, increasing organization), and that disease impacts negatively on this status [45], the *human-machine team* [46] of diabetic and hypertensive patients and prototype TSN presented in this paper can be interpreted to be an eHealth system constructed to slow down the latter process. Formally, this system can be imagined to contain a second-order, cybernetic, human subsystem in communication and interaction with a first-order, cybernetic, machine subsystem. The eHealth systems dedicated to different patients would be embedded in a larger system constructed to manage complex chronic diabetes and hypertension.

A suggestion for paradigm development follows immediately, under which the prototype TSN would be upgraded to a second-order, cybernetic subsystem by the introduction of more advanced monitoring and feedback mechanisms; more importantly, the TSN would be upgraded by the introduction of the AI necessary to converse, as well as interact, with its chronically ill human partner, to remember and analyze the results, to learn, and to persuade in an intelligent and friendly manner. The incremental information and knowledge acquired thereby can then be applied to enhance TSN user experience and safety, increase compliance, and induce behavioral change toward more ordered (ie, lower entropy) and healthier living on the part of the patient [47,48]. In particular, deep-learning AI can first be exploited to search the available health and social data and help construct a TSN protocol to mimic the perceptions and behaviors of a typical individual with chronic diabetes and hypertension. This protocol would be integrated with the TSN platform's caregiving and patient-safety functions and would be connected to the relevant cloud-based databases to ensure continuous updating and enrichment. With the addition of AI-supported speech capabilities, the enhanced TSN would be readily able to converse with chronically ill individuals and establish empathy while supporting self-management. A feedback channel can then be created and maintained to assist in the provision of immediate self-care guidance and initiate learning on the part of the TSN to enable future guidance and to change the patient's behavior and goals toward healthier living. In comparison with the prototype technology's *third-party* audio reminder function, which is described in the Prototype Technological Surrogate Nursing and Other Experimental Equipment subsection, an individual with irregular insulin administration habits would be more inclined to accept admonishment, persuasion, and corrective guidance from a "fellow diabetic" AI-enhanced TSN.

The above endeavors in paradigm development would fall, in part, under the ambit of technology augmentation. Given the absence of AI elements in the prototype technology and, hence, consistent applicability of base-level measurements to AI-based technology augmentation, the calibration presented in the previous section can be introduced to assist cost-benefit analysis in order to determine the best TSN in such cases. In addition to evaluating AI-enhanced TSN human-machine teams in the standard directions, the decision maker can seek to apply the cybernetic condition of team optimality [49] to guard against failure of value alignment between the subsystems [46].

Of interest to the business side is an approach to paradigm development suggested by consumer and especially patient acceptance of internet technology [50-54]. Paradigm development based on technology augmentation and its commercialization involve costly and difficult-to-reverse investment; therefore, it can be asked whether—by free-riding on this kind of consumer acceptance and with only nominal and, hence, readily reversible commitment to new technology—TSN can quickly and cheaply gain demand-side impact among chronically ill individuals. Such a question would be of interest to the health care business manager, such as suggesting informal paradigm development focused on marketing, benefits from free-riding on user experience with "old" but still-extant technology, and the capturing of economic rent implicit in knowledge-based external effects. Consider the following example: since shopping and banking are as much a part of everyday life as health care, consumer acceptance of internet e-shopping and e-banking [50,55] can be exploited in a marketing exercise of this type. By free-riding on well-established experience and familiarity with Technology Acceptance Model characteristics, such as usefulness, ease of use, and reliability, TSN can be advertised in these directions under health-oriented descriptions at lower cost-benefit ratios as compared to a *de novo* effort. Advertisements can suggest that TSN is easily programmable to unfailingly remind the chronically ill patient to take medication on time and in correct dosages. Additional free-riding would be possible by coupling TSN at behaviorally aware junctures to e-shopping or e-banking. For example, AI-generated friendly reminders to "watch the diet" can appear on monitors of the devices used by chronically ill individuals to access e-shops or e-banks. In this way, familiarity of use in one direction would breed familiarity of use and, hence, more patient compliance in the other.

Study Limitations

This study has several limitations. First, individuals recruited to the RCT were long-term patients with stably controlled conditions, on average. Therefore, our results may not be generalizable to higher-risk cases. Second, for reasons due to funding, the follow-up period was relatively short for the impact of TSN on chronic disease self-management to be more fully revealed. Finally, due to confidentiality concerns on the part of some patients, data regarding drug usage before and after the study were not collected.

Conclusions

Our research offers empirical support to the joint hypothesis that TSN would improve patient self-care while satisfying the *sine qua non* of patient self-safety. Practical utility is demonstrated by the derivation of an operational calibration of technology augmentation with applicability to health care management such as cost-benefit analysis. On a broader level, our findings suggest increased and wider use of technological surrogates in chronic disease management and help clarify the important question of whether introducing more technology in health care would be rational in an already technology-loaded society. Paradigm development in TSN is especially proposed with regard to multitasked technology augmentation aimed at inducing behavioral change along the lines of greater patient response in increasingly complex disease environments, AI enhancement, and synergy with internet technology and innovation.

Acknowledgments

This study was supported by the Health and Medical Research Fund of the Food and Health Bureau, The Government of Hong Kong Special Administrative Region (project No. 12133231; principal investigator: CKO).

Authors' Contributions

CKO was the principal investigator and was responsible for the study conception, design, and protocol development. CKO, BC, AT, and JT obtained the study funding. CKO, KL, and TL were responsible for the development of the study material. CKO, TL, LYCY, PSGH, HCC, and YFEL were responsible for patient recruitment. CKO and KL performed the data acquisition. KL and MKPS performed the statistical analysis. MTC, LYCY, and CKO were responsible for outcomes interpretation, analysis, and discussion. All authors contributed to the writing of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results from per-protocol analyses.
[DOCX File , 68 KB - jmir_v22i3e16769_app1.docx]

Multimedia Appendix 2
Results from sensitivity analyses.
[DOCX File , 68 KB - jmir_v22i3e16769_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2988 KB - jmir_v22i3e16769_app3.pdf]

References

1. 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]
2. Funnell MM, Anderson RM. Empowerment and self-management of diabetes. Clin Diabetes 2004 Jul 01;22(3):123-127. [doi: 10.2337/diaclin.22.3.123]
3. Hernandez-Tejada MA, Campbell JA, Walker RJ, Smalls BL, Davis KS, Egede LE. Diabetes empowerment, medication adherence and self-care behaviors in adults with type 2 diabetes. Diabetes Technol Ther 2012 Jul;14(7):630-634 [FREE Full text] [doi: 10.1089/dia.2011.0287] [Medline: 22524548]
4. World Health Organization (WHO) Regional Office for Europe. Geneva, Switzerland: World Health Organization; 2012 Apr 17. Empowering patients URL: http://www.euro.who.int/en/health-topics/Health-systems/patient-safety/news/news/2012/5/empowering-patients [accessed 2020-02-07]
5. Bergenstal R, Pearson J, Cembrowski GS, Bina D, Davidson J, List S. Identifying variables associated with inaccurate self-monitoring of blood glucose: Proposed guidelines to improve accuracy. Diabetes Educ 2000;26(6):981-989. [doi: 10.1177/014572170002600610] [Medline: 11912810]
6. Moen A, Brennan PF. Health@Home: The work of health information management in the household (HIMH): Implications for consumer health informatics (CHI) innovations. J Am Med Inform Assoc 2005;12(6):648-656 [FREE Full text] [doi: 10.1197/jamia.M1758] [Medline: 16049230]
7. 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]
8. Krousel-Wood M, Joyce C, Holt E, Muntner P, Webber LS, Morisky DE, et al. Predictors of decline in medication adherence: Results from the cohort study of medication adherence among older adults. Hypertension 2011 Nov;58(5):804-810. [doi: 10.1161/HYPERTENSIONAHA.111.176859] [Medline: 21968751]
9. Vervloet M, Linn AJ, van Weert JC, de Bakker DH, Bouvy ML, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: A systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
10. Wildevuur SE, Simonse LW. Information and communication technology-enabled person-centered care for the "big five" chronic conditions: Scoping review. J Med Internet Res 2015 Mar 27;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
11. Schooley B, San Nicolas-Rocca T, Burkhard R. Patient-provider communications in outpatient clinic settings: A clinic-based evaluation of mobile device and multimedia mediated communications for patient education. JMIR Mhealth Uhealth 2015 Jan 12;3(1):e2 [FREE Full text] [doi: 10.2196/mhealth.3732] [Medline: 25583145]
12. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An internet-based patient-provider communication system: Randomized controlled trial. J Med Internet Res 2005 Aug 05;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
13. Hsu J, Huang J, Fung V, Robertson N, Jimison H, Frankel R. Health information technology and physician-patient interactions: Impact of computers on communication during outpatient primary care visits. J Am Med Inform Assoc 2005;12(4):474-480 [FREE Full text] [doi: 10.1197/jamia.M1741] [Medline: 15802484]
14. Liederman EM, Morefield CS. Web messaging: A new tool for patient-physician communication. J Am Med Inform Assoc 2003;10(3):260-270 [FREE Full text] [doi: 10.1197/jamia.M1259] [Medline: 12626378]
15. Simon SR, Evans JS, Benjamin A, Delano D, Bates DW. Patients' attitudes toward electronic health information exchange: Qualitative study. J Med Internet Res 2009 Aug 06;11(3):e30 [FREE Full text] [doi: 10.2196/jmir.1164] [Medline: 19674960]
16. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: Qualitative study. J Med Internet Res 2013 Mar 27;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
17. 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]
18. 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 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
19. McLean G, Band R, Saunderson K, Hanlon P, Murray E, Little P, DIPSS co-investigators. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. J Hypertens 2016 Apr;34(4):600-612 [FREE Full text] [doi: 10.1097/HJH.0000000000000859] [Medline: 26845284]
20. Omboni S, Gazzola T, Carabelli G, Parati G. Clinical usefulness and cost effectiveness of home blood pressure telemonitoring: Meta-analysis of randomized controlled studies. J Hypertens 2013 Mar;31(3):455-468. [doi: 10.1097/HJH.0b013e32835ca8dd] [Medline: 23299557]

21. Or CK, Tao D. Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. Int J Med Inform 2014 May;83(5):320-329. [doi: 10.1016/j.ijmedinf.2014.01.009] [Medline: 24534118]
22. de Boer IH, Bangalore S, Benetos A, Davis AM, Michos ED, Muntner P, et al. Diabetes and hypertension: A position statement by the American Diabetes Association. Diabetes Care 2017 Sep;40(9):1273-1284. [doi: 10.2337/dci17-0026] [Medline: 28830958]
23. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: A randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]
24. McMahon GT, Gomes HE, Hickson Hohne S, Hu TM, Levine BA, Conlin PR. Web-based care management in patients with poorly controlled diabetes. Diabetes Care 2005 Jul;28(7):1624-1629 [FREE Full text] [doi: 10.2337/diacare.28.7.1624] [Medline: 15983311]
25. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 08;1(2):e13 [FREE Full text] [doi: 10.2196/resprot.2184] [Medline: 23612015]
26. Tao D, Or C. A paper prototype usability study of a chronic disease self-management system for older adults. In: Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management. 2012 Presented at: IEEE International Conference on IEEM; December 10-13, 2012; Hong Kong, China p. 1262-1266. [doi: 10.1109/ieem.2012.6837946]
27. 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]
28. Or CKL, Valdez R, Casper G, Carayon P, Burke LJ, Brennan PF, et al. Human factors and ergonomics in home care: Current concerns and future considerations for health information technology. Work 2009;33(2):201-209 [FREE Full text] [doi: 10.3233/WOR-2009-0867] [Medline: 19713630]
29. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986 Jan;24(1):67-74. [doi: 10.1097/00005650-198601000-00007] [Medline: 3945130]
30. Kravitz R, 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]
31. Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. Patient Educ Couns 2003 Nov;51(3):267-275. [doi: 10.1016/s0738-3991(02)00239-2] [Medline: 14630383]
32. Lorig KR, Ritter P, Stewart AL, Sobel DS, Brown BW, Bandura A, et al. Chronic disease self-management program: 2-year health status and health care utilization outcomes. Med Care 2001 Nov;39(11):1217-1223. [doi: 10.1097/00005650-200111000-00008] [Medline: 11606875]
33. Simon HA. Rational decision making in business organizations. Am Econ Rev 1979 Sep;69(4):493-513.
34. Simmons GF. Introduction to topology and modern analysis. New York, NY: McGraw-Hill; 1963.
35. Agency for Healthcare Research and Quality. 2019 Sep. Measurement of patient safety URL: https://psnet.ahrq.gov/primer/measurement-patient-safety [accessed 2020-01-01]
36. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): Prospective observational study. BMJ 2000 Aug 12;321(7258):405-412 [FREE Full text] [doi: 10.1136/bmj.321.7258.405] [Medline: 10938048]
37. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, et al. Global disparities of hypertension prevalence and control: A systematic analysis of population-based studies from 90 countries. Circulation 2016 Aug 09;134(6):441-450 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.018912] [Medline: 27502908]
38. Oliveria SA, Chen RS, McCarthy BD, Davis CC, Hill MN. Hypertension knowledge, awareness, and attitudes in a hypertensive population. J Gen Intern Med 2005 Mar;20(3):219-225 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.30353.x] [Medline: 15836524]
39. Brassington I. The ethics of reporting all the results of clinical trials. Br Med Bull 2017 Jan 01;121(1):19-29. [doi: 10.1093/bmb/ldw058] [Medline: 28104631]
40. Einstein A, Infeld L. The Evolution of Physics: The Growth of Ideas From Early Concepts to Relativity and Quanta. London, UK: Cambridge University Press; 1938.
41. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY: Basic Books; 2019.
42. Meskó B. The real era of the art of medicine begins with artificial intelligence. J Med Internet Res 2019 Nov 18;21(11):e16295 [FREE Full text] [doi: 10.2196/16295] [Medline: 31738169]
43. 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 [FREE Full text] [doi: 10.2196/12286] [Medline: 30950797]
44. 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]

45. Wiener N. The Human Use of Human Beings: Cybernetics and Society. Boston, MA: Da Capo Press; 1988.
46. Russell S. Human Compatible: Artificial Intelligence and the Problem of Control. New York, NY: Viking; 2019.
47. Martelaro N, Ju W. Cybernetics and the design of the user experience of AI systems. Interactions 2018 Oct 25;25(6):38-41. [doi: 10.1145/3274570]
48. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: Systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
49. Lai KK, Cheung MT, Fu Y. Resource allocation in public healthcare: A team-DEA model. J Syst Sci Complex 2017 Mar 21;31(2):463-472. [doi: 10.1007/s11424-017-6124-6]
50. Liao Z, Cheung MT. Internet-based e-banking and consumer attitudes: An empirical study. Inf Manage 2002 Jan;39(4):283-295. [doi: 10.1016/s0378-7206(01)00097-0]
51. Or CKL, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
52. Or CKL, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a Web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
53. Yan M, Or C. Factors in the 4-week acceptance of a computer-based, chronic disease self-monitoring system in patients with type 2 diabetes mellitus and/or hypertension. Telemed J E Health 2018 Feb;24(2):121-129. [doi: 10.1089/tmj.2017.0064] [Medline: 28737995]
54. Yan M, Or C. A 12-week pilot study of acceptance of a computer-based chronic disease self-monitoring system among patients with type 2 diabetes mellitus and/or hypertension. Health Informatics J 2019 Sep;25(3):828-843. [doi: 10.1177/1460458217724580] [Medline: 28820007]
55. Liao Z, Cheung MT. Internet-based e-shopping and consumer attitudes: An empirical study. Inf Manage 2001 Apr;38(5):299-306. [doi: 10.1016/s0378-7206(00)00072-0]

Abbreviations

AI: artificial intelligence
BG: blood glucose
BP: blood pressure
CG: control group
DBP: diastolic blood pressure
eHealth: electronic health
HbA_{1c}: hemoglobin A_{1c}
IG: intervention group
RCT: randomized controlled trial
SBP: systolic blood pressure
TSN: technological surrogate nursing

Edited by G Eysenbach; submitted 27.10.19; peer-reviewed by Q Cheng, M Dohan; comments to author 15.11.19; revised version received 14.02.20; accepted 01.03.20; published 27.03.20.

Please cite as:
Or CK, Liu K, So MKP, Cheung B, Yam LYC, Tiwari A, Lau YFE, Lau T, Hui PSG, Cheng HC, Tan J, Cheung MT
Improving Self-Care in Patients With Coexisting Type 2 Diabetes and Hypertension by Technological Surrogate Nursing: Randomized Controlled Trial
J Med Internet Res 2020;22(3):e16769
URL: http://www.jmir.org/2020/3/e16769/
doi:10.2196/16769
PMID:32217498

Original Paper

Experiences of Internet-Based Stepped Care in Individuals With Cancer and Concurrent Symptoms of Anxiety and Depression: Qualitative Exploration Conducted Alongside the U-CARE AdultCan Randomized Controlled Trial

Anna Hauffman[1], MSc; Sven Alfonsson[2,3], PhD; Helena Igelström[1,4], PhD; Birgitta Johansson[1,2], PhD

[1]Department of Immunology, Genetics and Pathology, Section of Experimental and Clinical Oncology, Uppsala University, Uppsala, Sweden
[2]Department of Women's and Children's Health, Section of Clinical Psychology in Healthcare, Uppsala University, Uppsala, Sweden
[3]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institute and Stockholm Health Care Services, Stockholm, Sweden
[4]Department of Neuroscience, Section of Physiotherapy, Uppsala University, Uppsala, Sweden

Corresponding Author:
Anna Hauffman, MSc
Department of Immunology, Genetics and Pathology
Section of Experimental and Clinical Oncology
Uppsala University
Akademiska University Hospital ing 100 D
Uppsala, SE-751 85
Sweden
Phone: 46 18 611 00 00
Fax: 46 18 611 10 27
Email: anna.hauffman@igp.uu.se

Abstract

Background: Individuals with newly diagnosed cancer may experience impaired health in several aspects and often have a large need for information and support. About 30% will experience symptoms of anxiety and depression, with varying needs of knowledge and support. Despite this, many of these patients lack appropriate support. Internet-based support programs may offer a supplement to standard care services, but must be carefully explored from a user perspective.

Objective: The purpose of this study was to explore the participants' perceptions of the relevance and benefits of an internet-based stepped care program (iCAN-DO) targeting individuals with cancer and concurrent symptoms of anxiety and depression.

Methods: We performed a qualitative study with an inductive approach, in which we used semistructured questions to interview 15 individuals using iCAN-DO. We analyzed the interviews using content analysis.

Results: The analysis found 17 subcategories regarding the stepped care intervention, resulting in 4 categories. Participants described the need for information as large and looked upon finding information almost as a survival strategy when receiving the cancer diagnosis. iCAN-DO was seen as a useful, reliable source of information and support. It was used as a complement to standard care and as a means to inform next of kin. Increased knowledge was a foundation for continued processing of participants' own feelings. The optimal time to gain access to iCAN-DO would have been when being informed of the diagnosis. The most common denominator was feeling acknowledged and supported, but with a desire for further adaptation of the system to each individual's own situation and needs.

Conclusions: Users saw the internet-based stepped care program as safe and reliable and used it as a complement to standard care. Similar interventions may gain from more personalized contents, being integrated into standard care, or using symptom tracking to adjust the contents. Offering this type of program close to diagnosis may provide benefits to users.

Trial Registration: ClincalTrials.gov NCT-01630681; https://clinicaltrials.gov/ct2/show/NCT01630681

(J Med Internet Res 2020;22(3):e16547) doi:10.2196/16547

KEYWORDS

internet-based stepped care; internet-based interactive health communication application; internet-based intervention; telemedicine; patient portals; oncology nursing; self care; psychoeducation

Introduction

Background

At the time of a cancer diagnosis and along the disease trajectory, individuals often have large information needs and may strive to get a fuller picture of what has befallen them and what awaits them [1]. The period around diagnosis is often described as frightening, with both new knowledge and support being needed to make the situation more predictable and manageable. Common reactions are shock and denial, followed by anger and symptoms of anxiety or depression [2]. Anxiety and depression may be part of the initial reaction and subside with time, but may also be persistent [3-5]. Targeted psychosocial interventions, especially those involving psychoeducation, may be helpful [6], as they aim to normalize, confirm, and explain feelings and symptoms, providing the individual with an explanatory model. Nevertheless, it is becoming increasingly common for people with cancer to be cared for as outpatients [7], thus spending more time at home during the period of illness. This may be perceived as positive, but it also leads to more limited possibilities of support from health professionals. One way of reaching people, regardless of the route of care delivery, is through electronic health (eHealth) interventions, including internet-based interactive health communication applications (IHCAs). An IHCA may be described as a set of components that offer information, support, and behavior change interventions. Further, internet-based IHCAs can provide relevant, quality-assured information and support when the individual feels ready for it or has time to access a computer or mobile device. IHCAs for people with cancer provide a way of offering support that may improve, for example, distress levels, social support, symptoms of fatigue, and health literacy [8,9].

The Uppsala University Psychosocial Care Program

The Uppsala University Psychosocial Care Program (U-CARE) [10] is a strategic research venture, in the fields of caring science, psychology, and computer science, supported by the Swedish government. Studies within U-CARE concern the psychosocial consequences of somatic diseases and the support that the affected individuals may need.

The U-CARE Portal

All interventions and the collection of patient-reported data in U-CARE take place via the U-CARE Portal, developed within U-CARE. The U-CARE Portal is an internet-based infrastructure that enables delivery of care such as self-care programs and psychological treatment within clinical studies and is not tethered to personal health records.

Objective

Using qualitative methods as part of the evaluation of complex interventions is increasingly common in order to gain further knowledge of aspects important to participants. The qualitative exploration may help explain the findings of the trial and aid understanding of the processes needed for change and implementation [11]. Since it can be difficult to predict how the intervention may work, as well as what users need and want, the qualitative exploration is an important measure to ensure that the intervention serves its purpose [12,13]. The aim of this study was to explore the participants' perceptions of the relevance and benefits of iCAN-DO.

Methods

U-CARE AdultCan

We recruited the informants in this study from AdultCan [14], a randomized controlled trial (RCT) (NCT-01630681) within U-CARE, targeting individuals with newly diagnosed breast cancer, colorectal cancer, or prostate cancer and with symptoms of anxiety or depression, or both. AdultCan aimed to investigate the effects of an internet-based IHCA (iCAN-DO), including a stepped care intervention for symptoms of anxiety and depression. iCAN-DO was developed in collaboration between the research group, staff in clinical cancer care, and individuals with lived experience of cancer. The main purpose of this collaboration, described in detail in a previous article [15], was to target iCAN-DO to the needs of individuals with cancer and concurrent symptoms of anxiety and depression. iCAN-DO is 1 intervention comprising 2 steps (Table 1): interactive support (step 1) based on psychoeducation and assumptions from Orem's self-care deficit nursing theory [16], social cognitive theory, and additional internet-delivered cognitive behavioral therapy (iCBT, step 2) for individuals with persistent symptoms of anxiety and depression despite the support provided in step 1. Individuals with newly diagnosed cancer were approached in a clinical setting and, after providing informed consent, were screened using the Hospital Anxiety and Depression Scale [17]. Participants with a score greater than 7 on either of the 2 subscales (indicating symptoms of anxiety or depression) were randomly assigned to iCAN-DO or standard care.

Table 1. Content and features of the internet-based stepped care in the iCAN-DO application, aimed at individuals with cancer and concurrent symptoms of anxiety and depression.

Step	Description	Who and when?	What?
1	Interactive support guided by a nurse. All communication was in the form of written messages via the U-CARE Portal.	Available for all participants in the intervention group from randomization. Duration 24 months.	The main feature of step 1 was a library containing psychoeducational information and self-care strategies in 16 modules concerning common problems surrounding cancer, such as anxiety, depression, pain, and sleeping issues. Information was delivered in texts, audiovisual presentations, slideshows, and video clips. Some information was directed at all users and some was directed at those with a specific diagnosis, but all contents were visible to all users. In addition to the library, there was a peer-support section and a frequently-asked-questions section. There was also a feature called “Ask an expert,” where users could pose questions to a nurse and read others’ anonymized questions with answers. The nurses were presented in brief, with photos.
2	iCBT[a] guided by a psychologist. All communication was in the form of written messages via the U-CARE Portal.	Offered only to participants with persistent symptoms of anxiety or depression, or both (>7 on either Hospital Anxiety and Depression Scale subscale) after using step 1. Duration 10 weeks.	Step 2 of the intervention provided a 10-week iCBT treatment program. The treatment contained 15 modules that comprised written texts, audiovisual presentations, and video clips. After completing an introductory module, all participants were free to choose the most relevant modules to work with over the course of 10 weeks. Each module included psychoeducational material, exercises, and assignments and spanned 2 to 4 weeks. Participants were guided by and received weekly feedback from a psychologist who monitored their work and answered any questions they had. A brief presentation of the psychologist, with a photo, was available in iCAN-DO.

[a] iCBT: internet-delivered cognitive behavioral therapy.

Study Design

This was a qualitative, descriptive study using semistructured interviews and was conducted alongside the AdultCan RCT.

Informants

Data from the RCT showed that most participants (105/124) used step 1, whereas only a few of those who were offered step 2 in addition to step 1 underwent or were interested in undergoing iCBT (7/82). Thus, through a purposeful selection we strived to include all participants who had used both step 1 and 2, as well as participants who had declined step 2. Since the amount of data necessary to answer the research question was dependent on the quality of data [18], we did not predefine the number of participants at the start of the study, but instead assessed the material for variation and consistencies in statements before ending data collection. We approached 20 individuals, of whom 2 declined to participate and 3 could not be reached. We interviewed 15 informants (Table 2) in 2016-2017. All had had access to iCAN-DO for at least seven months (since step 2 was offered at 1, 4, and 7 months). The informants’ general online activities varied, but all used the internet daily, and 2 informants described themselves as very inexperienced computer users.

Table 2. Informants’ characteristics (at time of diagnosis) (N=15).

Characteristics	Values
Age (years), mean (minimum-maximum)	59 (37-69)
Sex, n (%), and diagnosis	
Female	10 (70) (breast cancer)
Male	5 (30) (1 colorectal cancer, 4 prostate cancer)
Relationship status, n (%)	
Married or partner, living with someone	12 (80)
Married or partner, living alone	1 (7)
Widowed	1 (7)
Single	1 (7)
Level of education, n (%)	
Elementary or middle school	3 (20)
High school	1 (7)
University ≤3 years	6 (40)
University >3 years	5 (30)
Working situation, n (%)	
Working	10 (70)
Retired	4 (26)
Early retirement	1 (7)
General online activity outside iCAN-DO interactive health communication application, n (%)	
Daily, no social media	4 (26)
Daily, active on social media, lurking	7 (47)
Daily, active on social media, participating	4 (26)
Activity in step 1 (interactive support), n (%)	
Opening material (all sections)[a] >20 times	6 (40)
Opening material (some sections) >20 times	7 (47)
Opening material (some sections) <10 times	2 (13)
Step 2 (internet-delivered cognitive behavioral therapy), n (%)	
No	9 (60)
Yes	6 (40)

[a] Library, peer support, frequently asked questions, or ask an expert.

Procedure

Potential informants were sent a letter regarding the study and informed that the first author would call them within a few days to provide more information and ask them about their willingness to participate. If they did not want to get this phone call, they could email or call the principal investigator for AdultCan (no one used that option). If an informant consented to participate, a time and place for the interview were determined.

Ethics

All procedures were conducted in accordance with the ethical standards of the institutional and regional research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All informants were treated with confidentiality and had time to consider their participation after receiving written and verbal information. All provided informed consent before the interviews were conducted. The study was approved by the Swedish Ethics Review Authority (no. 2012/003/9).

Data Collection

Questionnaires and Log Data

We retrieved information on informants’ characteristics and activity in iCAN-DO from self-reported questionnaires and logs in the portal. We retrieved information about online activity beyond iCAN-DO at the time of the interviews.

Interviews

The first author (AH) conducted all the interviews. AH is a registered specialist nurse with experience in palliative cancer care. AH has been involved in parts of the development of the stepped care intervention, but has no overall responsibility in the program. AH strived to get as close to the users' experiences as possible to gain knowledge to facilitate further development of iCAN-DO. The interviews were tape recorded and an interview guide with open-ended questions was used. The first question in the interview guide was "Can you tell me about your experiences of using iCAN-DO?", then all parts of the system were listed as follow-up questions to ensure coverage, such as "Can you tell me about your experiences of using the library?" Follow-up questions were based on each informant's responses and used to get deeper descriptions of experiences important to the aim of the study [19]. Specifying and probing questions were used, such as "How did that feel?" "How did you use that?" or "Was there any part that you found helpful/problematic here?," as well as interpreting questions such as "Do you mean by that...?" One test interview was performed to explore the design and understanding of questions. Each interview was performed in a place chosen by the informant (eg, at home, at their workplace, or at the hospital) and lasted between 45 and 120 minutes. Informants could choose to log in to iCAN-DO during the interview.

Data Analysis

AH transcribed the interviews verbatim and used inductive content analysis according to Graneheim and Lundman [18] to analyze the manifest content of the interviews. The text was read several times and each interview was divided into meaning units and condensed (reduced while preserving the core). Each condensed meaning unit was then labelled with a code to describe the key message. Codes with similar content were allocated to subcategories and categories (Table 3 shows an example). The analysis was then discussed and assessed within the group of authors several times before reaching the final categories. The main purpose of this validation by the coauthors was to determine that data were labelled and sorted in a way that corresponded not only to the research question, but also to what the informants conveyed in the interviews. The excerpts presented in the paper were translated by AH and Linnea Holmén at Calyptic.

Table 3. Steps in the content analysis illustrated by a sample meaning unit.

Meaning unit	Condensed meaning unit	Code	Subcategory	Category
And then I could also show it to my husband, read this, and that was also good because we both got the same information.	I could show it to my husband, read this, that was good because we both got the same information.	My husband and I could get the same information.	iCAN-DO as a source of information for others	A complement to standard care

Results

Content Categories

The analysis resulted in 4 categories in the stepped care intervention: (1) gaining knowledge and support but wanting more personalization, (2) a feeling of safety that was needed earlier, (3) own situation, preferences, and timing determine the use of peer support, and (4) a complement to standard care (Table 4). We clarify the context or subject in quotes below in brackets and indicate the 17 subcategories in italic text.

Table 4. Categories and subcategories.

Category	Subcategories
Gaining knowledge and support but wanting more personalization	• The importance of information and support • Confirmation, recognition, and being taken seriously • A wish for more specifically tailored contents • Positive, supportive contacts with the psychologist • Turning down the offer of iCBT[a] • Limitations of iCBT
A feeling of safety that was needed earlier	• Reliable and safe to trust • Wanted to have access to step 1 earlier • Seeking information online started early • Information needs vary over time
Own situation, preferences, and timing determine the use of peer support	• Not interested in peer support at all, besides lurking • Higher presence of health care professionals in the forum • Facebook provided a better environment for online peer support
A complement to standard care	• Information given at the hospital was insufficient • iCAN-DO as a source of information for others • Standard care did not offer any support for emotional problems • Fit into everyday life

[a] iCBT: internet-delivered cognitive behavioral therapy.

Gaining Knowledge and Support But Wanting More Personalization

Informants used iCAN-DO to gain understanding of what was going on, both physically and emotionally. The informants described *the importance of information and support* and talked about information gathering and increased knowledge as a way of coping with their new situation. Some of the informants, who had used both steps 1 and 2, discussed how the 2 steps complemented each other.

> *The first thing you need is information, a foundation knowing that this is correct information that I can trust, this is how it works. Then you need different tools based on that, like for example CBT [cognitive behavioral therapy]. But this thing, the information, it's the core to understanding the disease. It has to start with that, in order for you to, like, move on.* [Female, breast cancer]

> *These modules [step 2] suit me rather well, I have worked with several of them and this first part that is more about self-care [step 1] also suits me quite well....I had pretty big problems with pain, so those parts were important to me; I could get help with that both in iCBT and in the information section [step 1].* [Male, colorectal cancer]

The content in step 1 was perceived as helpful, calming, and confirmatory. Informants described *confirmation, recognition, and being taken seriously* as factors that could mitigate troublesome thoughts and feelings. Several informants expressed a wish to have access to as much information as possible, even if the information was unpleasant. Symptoms of the disease and side effects of treatment had an impact on daily life and participants considered it a relief to recognize descriptions of their own symptoms or side effects in step 1. This could be achieved through studying information in the library, asking questions, or reading questions from others in the frequently asked questions section. The important thing was to get confirmation that the experiences were well known and real.

> *Nothing that I read made me sad. It was more like gaining freedom, when reading the answers to questions, for example...to be able to read that this is normal, this is a known side effect, it's nothing new. And I have been able to use that knowledge many times later on: calm down, this is a common experience.* [Female, breast cancer]

The possibility of posing questions in step 1 was seen as providing a sense of security and of being taken seriously. The feedback to questions was described as educating and encouraging, although the function would have been more valuable if the nurse who answered questions had access to medical records to provide more specific answers. Furthermore, participants suggested that such a function would be suitable for regular health care.

> *It's fantastic, I have gotten such good answers all the time. I must say that "Ask an expert" is the best thing [in step 1] because it feels like someone takes me seriously, someone is giving me a scientifically based answer in words that I can understand. You never have time for those things at the clinic.* [Female, breast cancer]

> *I wish that you didn't have to be put on hold on the phone to make appointments to get advice [in regular health care]. I think that "Ask an expert" should be developed further, as a part of regular health care. You should be able to ask your questions online like this, I think.* [Male, prostate cancer]

Participants expressed *a wish for more specifically tailored contents* regarding diagnosis, age, the person's sex, treatment, and symptoms regarding parts of step 1. For example, feeling much younger than other users could create a sense of loneliness. One informant explained that she chose a Facebook forum instead of the forum addressing women with breast cancer in step 1, as she had a need to talk to those with exactly the same diagnosis, prognosis, and treatment.

> *I can't identify with a man who has colorectal cancer, for example. We don't have a lot in common. Not with those who have just breast cancer either. I want to talk to women who have exactly my type of breast cancer [human epidermal growth factor receptor 2-positive]. We discuss things based on our diagnosis and situation, because it is different, like survival and side effects of medicines and how we feel ahead of the surgery, things like that....* [Female, breast cancer]

The informants who underwent the iCBT program (step 2) in addition to step 1 experienced *positive, supportive contacts with the psychologist* and materials that were useful for managing symptoms. The feedback was described as encouraging and genuine and was also a trigger to log in and participate further. Undergoing iCBT was described as demanding, but worth putting effort into. Informants felt that they got help and that, while symptoms such as pain, anxiety, or insomnia would still be present, iCBT gave them tools to manage the situation.

> *My anxiety was relieved, absolutely; it felt like...when working with the [cognitive behavioral therapy] program...going from constant anxiety, I could use what I had learned. I'd think, that's right, now I'll do this...and then the anxiety faded. Then it might come back, but then I can use what I have learned again. I now have the tools. I still think that way.* [Female, breast cancer]

A feeling of sufficient support from relatives and friends was a reason for *turning down the offer of iCBT* provided in step 2. A fear of "making it worse" by focusing on the negative feelings was also mentioned as a reason, as were past experiences affecting confidence in the method or already having a professional contact. Among the informants who participated in step 2, experiences of its helpfulness varied and several *limitations of iCBT* were described. A common reflection was that the time limit set for therapy (10 weeks) was too short in an already strained situation. Choosing among modules was sometimes difficult and a feeling of "needing them all" was experienced as stressful. The division into modules could also be perceived as difficult, as all the symptoms interacted. Symptoms and needs were described as dynamic and changing during the period of illness. Informants understood the purpose of the set time frame and modules, but still found it hard to deal with. One informant described choosing to work with self-help apps instead, because then time was not a problem.

> *When I realized that I was supposed to finish all that in those weeks I felt, "no"...because that's a very intense schedule, oh I felt such pressure. I just felt I wouldn't manage, so I didn't finish it. Since I got my diagnosis, my sensitivity to stress has been, well...I can't take any stress, zero.* [Female, breast cancer]

Some experienced the online delivery of the treatment and interaction with the psychologist by written messages as positive because they felt that they could form the level of relationship they needed. The use of written communication was described as a kind of processing in itself and as a way to get a grip on one's own thoughts. One informant stated that cognitive behavioral therapy in a real-life context would not be as appealing. Communicating with the psychologist in writing could also be experienced as limiting and seen as an obstacle by some, who felt that the psychologist was absent and uncaring. A video meeting at some point during treatment was one suggested facilitator. Informants further suggested that the contacts with the psychologist should be visible in medical records, to facilitate continued work with a psychologist in a real-life context.

> *It was almost fascinating that it worked so well, but also a bit strange that someone that you never meet knows so much about your life...it's strange and astounding...it was a new kind of relationship that I have never experienced before.* [Female, breast cancer]
>
> *It wasn't at all the relationship I wanted; this was faceless. I felt like the line of text that I got from the psychologist had probably been sent to 10 other people that day. Like...it wasn't about me.* [Female, breast cancer]

A Feeling of Safety That Was Needed Earlier

The information and self-care advice provided in step 1 was perceived as *reliable and safe to trust*. Informants highlighted it as positive that they could be sure there were no hidden agendas or profits involved, as that was something perceived to make other online information difficult to evaluate. Evaluating information on the internet could be strenuous and time consuming. Informants described looking for confirmation in step 1, using it to judge whether something they had been told or read elsewhere was accurate. Another aspect of security was that of knowing that there was easily available and reliable information in step 1, in case it was needed in the future, even if it was not actively used at the time. Having access to information was experienced as a privilege.

> *Because there is such a wealth of information on the internet, I think it has been good to have information from an official page, so to speak; it felt good. It toned down the whole thing a bit. There are horrible stories on the internet about the side effects of the treatment. On various forums on the internet, medical knowledge is questioned.... So, therefore, I stopped searching the internet. I haven't used the internet all that much; it's mostly the portal I've used.* [Female, breast cancer]

Many of the informants described that they would have *wanted to have access to step 1 earlier*, preferably when informed of the diagnosis. That occasion had been followed by large information needs, and informants looked for information that could give them a sense of control. *Seeking information online started early* in close proximity to gaining knowledge about the diagnosis during the medical investigation, or even before. Many had already searched for information online when gaining access to step 1 and some had bad experiences from such online information. The informants who reported that they used online forums for support described that they had already found forums elsewhere when they got access to the portal. Informants also described how their *information needs would vary over time*, and it was seen as important to have access over a period of time, as the material could be used at different times during the illness, depending on individual information needs.

> *You should get access at once, at the point when your world falls apart, because that's the time when you need the answers. To me, it didn't matter that the doctor said that I would most likely get well, I still needed to read the facts myself and I needed to do that right away.* [Female, breast cancer]

> *Sometimes you may not get full insight into your own situation at once, it takes time...and then you should have access to this when the right time comes...it takes time, it comes later...and I am the kind of person who...I don't want to bother others, some can argue for themselves but no, I can't, since I'm not like that. So for me, it's actually better that it is available like this.* [Male, prostate cancer]

Own Situation, Preferences, and Timing Determine the Use of Peer Support

When asked about the section containing peer support, most of the informants said they were *not interested in peer support at all, besides lurking*. This could have several different causes, such as preferring professional support or peer support face-to-face, not being the "social type," feeling that peer support was just opinions often not based on facts, or fearing that discussions would degenerate into being nonsupportive. Others said that they might have used peer support more if they did not have family and friends to talk to. Some of those describing themselves as not interested declared that they sometimes read forums, both in step 1 and elsewhere online, but that they would never write something themselves. Some would consider active participation if health professionals were more involved, and a *higher presence of health care professionals in the forum* in step 1 was suggested. Some informants found peer support in general to be highly important and helpful but expressed that *Facebook provided a better environment for online peer support*. The informants had all already found Facebook groups when they got access to step 1 and described both positive and negative experiences from this. They also stated that the forum in step 1 could not compete with the Web environment, easy access, and high specificity and activity in Facebook groups. However, some informants described a lack of moderators and a nonsupportive environment in Facebook groups.

> *Well, I didn't look in here so often. I think it's because I have a lot of people around me to talk to. If I was alone it would have been different, because you need to have something...but since I can choose, I'd rather sit down and talk to someone.* [Male, prostate cancer]

A Complement to Standard Care

Informants described step 1 as a complement to standard health care in several ways. It was useful in the cases when *information given at the hospital was insufficient*, for example, that it was brief or difficult to understand. Being in a state of shock when getting the news of the diagnosis also made the information at the hospital hard to remember, and there was a need for repetition in step 1. Informants were relieved to see that the information given in the hospital was consistent with the information in step 1. The information could also be used to prepare for hospital visits. All informants stated that factual information and self-care advice were highly appropriate to provide in a Web-based form, but some stressed that they preferred emotional support in a real-life context.

> *You should have a system like this with the opportunity to get in touch with a person in health care. It doesn't have to be a physician, but a person connected to the clinic, because you feel a lot of anxiety in between medical examinations and such.* [Male, prostate cancer]

> *I booked an extra meeting with the physician because I was so scared about the radiation, and, sure, I got answers from the physician, but then I could get it confirmed through the material here, and the information here was consistent with what he said. That fact made me feel safer.* [Female, breast cancer]

Informants further described both step 1 and step 2 in *iCAN-DO as a source of information for others*. Relatives were described as being left out and it could be burdensome to manage their worries and questions. Informants described how relatives could have trouble trusting information from informants and that it was an advantage to be able to read the same information together.

> *Well, I let her read all this. She thinks it's good, and good that I have something to do as well. We also discussed things with the material as a starting point; I think it's good. Some of the material describes...well, relations and such. This is something that we must handle, between us, with those close to us.* [Male, colorectal cancer]

> *I felt a confirmation, a sense of security regarding the disease, and I felt that I could trust what was said. Since it was published here in writing, then I could also show it to my husband, "read this," and that was also good. We both got the same information.* [Female, breast cancer]

When talking about both steps in iCAN-DO, several informants mentioned that *standard care did not offer any support for emotional problems*, and some said that even if they thought that they could get support from a counselor, they did not know whether it would be helpful, or did not want to burden health care services. The fact that iCAN-DO included emotional, practical, and bodily issues was noted by some informants who described standard health care as separating the body from the emotions, even if emotions emanated from a bodily problem.

> *No, this doesn't exist in regular health care and sometimes when I was in the portal I felt at least I am not completely alone, there are others whose situations are like mine. I also saw those who felt even worse than I did...I don't see it like...well, this was*

> *not a complement, this was the only, the one support that I got.* [Female, breast cancer]
>
> *It was important to be able to read this [in the step 1 library] at my own pace, but I also wanted to pose my questions to a real person. When I got the relapse, I told the physician that it felt quite stressful and the answer I got was, well, there are counselors...and I think that is so...well, it's the worst thing I know! You can't divide yourself up like that...a lot of worry is connected to the body.* [Female, breast cancer]

Further, it was difficult to fit standard health care into everyday life. Informants stated that life did not stop or slow down when they got ill: their work had to be done, families had to be taken care of, and so on. Working informants experienced a lot of stress, and they appreciated being able to use the stepped care intervention when it *fit into everyday life*. Both working informants and those who were retired or on sick leave described difficulties in getting in touch with regular health care services during office hours and having limited time to get questions answered by regular health care providers. iCAN-DO was a possibility to get support and answers close to when a need for it occurred and when the informants had time. Informants also talked about features in relation to standard health care and some suggested possible implementations.

> *You are not bound to a set time—you can do it anytime you wish in the evening, for example—that's the biggest gain...if you have time in the middle of the night, you can sit down then.* [Male, colorectal cancer]
>
> *It should be easy, because you can't call the clinic all the time, and they may not be able to answer when you call...it's hard to get in touch, so it was here in the library I found the answers and that was good.* [Female, breast cancer]

Discussion

Principal Findings

Participants experienced iCAN-DO as a useful and reliable source of information and support and used it as a complement to standard care. They described the need for information as being large and looked upon the gathering of information almost as a survival strategy when cancer was newly diagnosed. To be able to understand and handle the entire situation, informants expressed a need for increased knowledge, a result that corresponds well with the theoretical basis of the intervention, where psychoeducation is a prominent feature. They suggested that knowledge was the foundation for continued processing, implying that the stepped care model in iCAN-DO would be suitable. In addition, informants often talked about the stepped care intervention in relation to standard health care, where they often felt that they were not seen as whole individuals, but rather as a set of separate symptoms.

The overall picture was that standard health care did not seem to suit their needs, leaving them with a lot of unanswered questions and a lack of emotional support focused on their cancer type and its treatment. The high, often unmet needs for information and psychosocial support among individuals with cancer is well known [20]. Before gaining access to step 1, informants sought answers on the internet, an action previously described as the first choice in the absence of contact with health care professionals [21,22]. Here iCAN-DO seemed to fill a void, providing easy access to information that could be trusted. It is essentially positive that individuals look for information that can facilitate the handling of disease, but the quality of this information could be better ensured if integrated as a natural part of routine care, which informants also suggested. Most informants described lurking (nonactive participation) in forums, both in iCAN-DO and elsewhere online, and this is by far the most common behavior in internet forums, including those regarding behavior change and health [23]. By reading about other people's experiences, the lurker may learn and benefit on a personal level, but still a forum needs active participants to stay alive and healthy. Informants in this study suggested that they would be encouraged to participate by more active participation from health care providers, and this is in line with a literature review [24] suggesting that moderators should create opportunities for delurking by providing new forum members with supportive, encouraging information, as well as highlighting the value of contribution. Such attitudes could be further developed within iCAN-DO forums.

The need for information and support is the highest at the time of diagnosis [20], and the timing of the intervention was also highlighted by most informants, who would have wanted access to step 1 immediately at the time of diagnosis. Information seeking was sometimes described as having started in the early investigation phase. During the development of iCAN-DO, the clinical staff and those with lived experience of cancer agreed that it was best not to approach presumptive study participants immediately at the time of diagnosis. This was decided because shock and denial might be present [2] and the question posed would be regarding participating in the AdultCan RCT, with no promises of additional psychosocial support. Also, the time of diagnosis and start of treatment imply an intensive period with diagnostic procedures, loads of information about the treatment, and possible inclusion in clinical trials. Thus, we regarded the time of diagnosis as inappropriate for inclusion in AdultCan. However, the findings of this study suggest that when implementing evaluated internet-based support in routine clinical care, the time of diagnosis could be a suitable occasion for introduction.

iCBT was fitting and helpful for some, but not all. For some, the ability to handle symptoms seemed to increase, while others seemed to require the face-to-face presence and guidance of a psychologist. The burden and strain of the illness and treatment itself seemed to negatively affect the possibilities of engaging in a fixed, time-limited therapy like iCBT. Adjustments to a set time frame during iCBT have previously been described as inefficient [25], suggesting that iCBT treatment should be agreed upon and adapted to the individual at the initial stage.

Besides the specific burden of illness and treatment, participants' perseverance may be affected by the fact that they did not actively seek support. Individuals undergoing iCBT both in trials and in clinical settings are often recruited via the internet and by self-referral [26,27]. Since our sample consisted of individuals with a recent diagnosis of cancer, recruited in a clinical setting, the program would likely have gained from being more adaptable to the individual situation. A previous evaluation of a similar iCBT intervention for individuals who had myocardial infarction [28] suggested that individuals' technical skills, personal preferences, and life context must be considered, and other results suggested that individuals with comorbidity benefit more from a tailored iCBT program than from one-size-fits-all protocols [29]. Another study on the feasibility of iCBT for individuals with a new diagnosis of cancer [30] reported that some of the 13 informants interviewed found it difficult to complete the program. Further, participants in that study appreciated the flexibility and private nature of the iCBT program that was provided but suggested that they wanted supplementary content about the side effects of treatment. This might further support the stepped care approach in iCAN-DO, which offered information and self-care strategies for side effects in step 1, in addition to iCBT. Again, support for this group may be more relevant if aiming to address the totality of their situation following the cancer rather than focusing on individual symptoms.

Informants appreciated the possibility to ask an expert in step 1 but stated that it would have been even more useful if the nurse had access to medical records, and they also wanted the psychologist's treatment to be visible in the medical records. This suggests that the intervention would have gained from being integrated into standard care. There were thoughts of a system tethered to electronic health records in the development stage, but unfortunately the lack of interoperability and lengthy legal processes would most likely have delayed the project for years. Thus, further development and cooperation between stakeholders is much needed. These needs and possible solutions have recently been described in depth by Signorelli et al [31], who also suggest that the integration of electronic health records can increase personalization.

Further, the informants used iCAN-DO to inform next of kin, who were described as being left out. Earlier investigations have shown that next of kin are at risk of developing symptoms of anxiety or depression [32] and, as eHealth resources have shown the ability to decrease both perceived burden and negative mood symptoms among next of kin [33], we plan to include information and support specifically directed at them in our forthcoming projects.

Despite positive experiences of iCAN-DO, the common denominator was the desire for individual tailoring to one's own situation and needs. Informants suggested that the intervention should target them more specifically, not only on the basis of having cancer and concurrent symptoms of anxiety or depression. The results suggest that, for example, diagnosis, type of treatment, symptoms, age, and the person's sex could be taken in consideration when tailoring the support. We based the design of iCAN-DO on an evidence-based triad [15] of clinical expertise, best scientific evidence, and patient involvement. In retrospect, we could have aimed at a more heterogeneous group of individuals with lived experience of cancer. As Bandura described, attempting to tailor health communications is no guarantee of a positive outcome, since the benefits depend on the value of the tailored factors [34]. Future interventions could probably gain from being integrated into standard care, as it may be easier to achieve more relevant personalization there than in a separate research context. Further, some previous studies have used tracking of symptoms as a means to tailor the content by, for example, suggesting material to the user [35-37]. In addition to these possible adjustments, the delivery mode and design of the program and its included features must be considered in regard to the entire user experience. We have analyzed these aspects of iCAN-DO separately (HI, unpublished data, 2020).

Methodological Considerations

The purposeful sample strategy aiming to include participants who used both step 1 and step 2 and participants who had declined step 2 resulted in a predominance of female informants with breast cancer. This reflects the total sample in the AdultCan RCT, in which this group was the largest, followed by those with a diagnosis of prostate or colorectal cancer. Informants still varied in other aspects such as general online activity, level of education, and age. The informants in this study had access to a specified internet-based stepped care intervention and may therefore be considered a "specific sample" according to Malterud et al [38]. Their description of sample sizes in qualitative studies suggested that a smaller sample size is sufficient when the anticipated specificity in informants' experiences is high. Thus, when we had interviewed 15 informants, we assessed the material and, as statements seemed to be both varied and consistent, we terminated data collection. However, we interviewed only 2 informants with a low activity level in iCAN-DO, since they had difficulties discussing the topics covered in the interview guide, which might have caused them some discomfort. The purpose of the interview guide was to provide dependability and stability over time and under varying conditions [19], thus strengthening credibility. The low-activity users could be interviewed separately, using a different research question, to further explore possible customizations of the program.

Regarding transferability, both the content of iCAN-DO and the sample of informants were specific, but the needs for information and support may be similar in individuals in other diagnostic groups. Also, we have strived to describe iCAN-DO and the surrounding conditions in a way that allows those interested to determine whether results are usable in another context, such as when developing an internet-based support program. Issues of confirmability must be taken into consideration in that informants might have been affected by the fact that someone working within the project conducted the interviews. Even though the interviewer assured informants that she was focused on program development, they might have expressed their opinions in more positive terms than they would have if someone completely independent had conducted the interviews. However, an independent interviewer might have affected the quality of the dialogue due to a lack of knowledge about the contents and features of iCAN-DO. Further, regarding confirmability, the analysis described in the Methods section was conducted mainly by the first author but then assessed and discussed within the group of coauthors at several occasions during the process. The group of authors consisted of individuals with expertise in nursing, psychology, and physiotherapy, although 1 of the coauthors contributing to analysis had nothing to do with iCAN-DO.

Conclusion

iCAN-DO was experienced as a safe and reliable source of information and support, but informants highlighted the importance of individualized information, support, and delivery in both step 1 and step 2. Further, internet-based support should be offered close to the time of diagnosis and would gain from being integrated into standard care. Future trials should focus on how to personalize and integrate internet-based information and support into clinical everyday life and, meanwhile, the clinic should assist patients by guiding them to evidence-based, reliable sources on the internet.

Acknowledgments

The first author/interviewer was struck by the informants’ friendliness and willingness to contribute and would like to say a warm thank you to all of them.

This work was supported by grants from Forte and the Swedish Cancer Society.

Authors' Contributions

AH participated in research design, data collection, data analysis, and writing the paper. BJ participated in research design, data analysis, and writing the paper. HI and SA participated in data analysis and writing the paper.

Conflicts of Interest

None declared.

References

1. Fletcher C, Flight I, Chapman J, Fennell K, Wilson C. The information needs of adult cancer survivors across the cancer continuum: a scoping review. Patient Educ Couns 2017 Mar;100(3):383-410. [doi: 10.1016/j.pec.2016.10.008] [Medline: 27765377]
2. Singer S. Psychosocial impact of cancer. Recent Results Cancer Res 2018;210:1-11. [doi: 10.1007/978-3-319-64310-6_1] [Medline: 28924676]
3. Fröjd C, Larsson G, Lampic C, von Essen L. Health related quality of life and psychosocial function among patients with carcinoid tumours. A longitudinal, prospective, and comparative study. Health Qual Life Outcomes 2007 Apr 11;5:18 [FREE Full text] [doi: 10.1186/1477-7525-5-18] [Medline: 17428340]
4. Maass SWMC, Roorda C, Berendsen AJ, Verhaak PFM, de Bock GH. The prevalence of long-term symptoms of depression and anxiety after breast cancer treatment: a systematic review. Maturitas 2015 Sep;82(1):100-108. [doi: 10.1016/j.maturitas.2015.04.010] [Medline: 25998574]
5. 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]
6. Galway K, Black A, Cantwell M, Cardwell CR, Mills M, Donnelly M. Psychosocial interventions to improve quality of life and emotional wellbeing for recently diagnosed cancer patients. Cochrane Database Syst Rev 2012 Nov 14;11:CD007064 [FREE Full text] [doi: 10.1002/14651858.CD007064.pub2] [Medline: 23152241]
7. 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]
8. Bouma G, Admiraal JM, de Vries EGE, Schröder CP, Walenkamp AME, Reyners AKL. Internet-based support programs to alleviate psychosocial and physical symptoms in cancer patients: a literature analysis. Crit Rev Oncol Hematol 2015 Jul;95(1):26-37. [doi: 10.1016/j.critrevonc.2015.01.011] [Medline: 25701515]
9. Ventura F, Ohlén J, Koinberg I. An integrative review of supportive e-health programs in cancer care. Eur J Oncol Nurs 2013 Aug;17(4):498-507. [doi: 10.1016/j.ejon.2012.10.007] [Medline: 23158437]
10. 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]
11. Malterud K, Siersma VD, Guassora AD, Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. BMJ 2009 Sep 10;339:b3496 [FREE Full text] [doi: 10.1136/bmj.b3496] [Medline: 19744976]
12. Sainio C, Lauri S, Eriksson E. Cancer patients' views and experiences of participation in care and decision making. Nurs Ethics 2001 Mar;8(2):97-113. [doi: 10.1177/096973300100800203] [Medline: 16010885]
13. Stoop AP, van’t Riet A, Berg M. Using information technology for patient education: realizing surplus value? Patient Education and Counseling 2004 Aug;54(2):187-195. [doi: 10.1016/S0738-3991(03)00211-8] [Medline: 15288913]

14. 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]
15. 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]
16. Orem DE, Taylor SG, Renpenning KM. Nursing: Concepts of Practice. St Louis, MO: Mosby; 2001.
17. 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]
18. 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]
19. Brinkmann S, Kvale S. Interviews: Learning The Craft Of Qualitative Research Interviewing. Los Angeles: Sage Publications, Inc; 2020.
20. Puts MTE, Papoutsis A, Springall E, Tourangeau AE. A systematic review of unmet needs of newly diagnosed older cancer patients undergoing active cancer treatment. Support Care Cancer 2012 Jul;20(7):1377-1394. [doi: 10.1007/s00520-012-1450-7] [Medline: 22476399]
21. Birkler J, Dahl M. Den Digitala Patienten. Stockholm, Sweden: Liber; 2014.
22. 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] [Medline: 15062905]
23. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014 Feb 04;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
24. Sun N, Rau PP, Ma L. Understanding lurkers in online communities: A literature review. Computers in Human Behavior 2014 Sep;38:110-117. [doi: 10.1016/j.chb.2014.05.022]
25. Paxling B, Lundgren S, Norman A, Almlöv J, Carlbring P, Cuijpers P, et al. Therapist behaviours in internet-delivered cognitive behaviour therapy: analyses of e-mail correspondence in the treatment of generalized anxiety disorder. Behav Cogn Psychother 2013 May;41(3):280-289. [doi: 10.1017/S1352465812000240] [Medline: 22717145]
26. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
27. Murphy MJ, Newby JM, Butow P, Loughnan SA, Joubert AE, Kirsten L, et al. Randomised controlled trial of internet-delivered cognitive behaviour therapy for clinical depression and/or anxiety in cancer survivors (iCanADAPT Early). Psychooncology 2020 Jan;29(1):76-85. [doi: 10.1002/pon.5267] [Medline: 31659822]
28. Wallin E, Norlund F, Olsson EMG, Burell G, Held C, Carlsson T. Treatment Activity, User Satisfaction, and Experienced Usability of Internet-Based Cognitive Behavioral Therapy for Adults With Depression and Anxiety After a Myocardial Infarction: Mixed-Methods Study. J Med Internet Res 2018 Mar 16;20(3):e87 [FREE Full text] [doi: 10.2196/jmir.9690] [Medline: 29549067]
29. Webb CA, Rosso IM, Rauch SL. Internet-Based Cognitive-Behavioral Therapy for Depression: Current Progress and Future Directions. Harv Rev Psychiatry 2017;25(3):114-122 [FREE Full text] [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
30. Alberts NM, Hadjistavropoulos HD, Titov N, Dear BF. Patient and provider perceptions of Internet-delivered cognitive behavior therapy for recent cancer survivors. Support Care Cancer 2018 Feb;26(2):597-603. [doi: 10.1007/s00520-017-3872-8] [Medline: 28913732]
31. Signorelli GR, Lehocki F, Mora Fernández M, O'Neill G, O'Connor D, Brennan L, et al. A Research Roadmap: Connected Health as an Enabler of Cancer Patient Support. J Med Internet Res 2019 Oct 29;21(10):e14360 [FREE Full text] [doi: 10.2196/14360] [Medline: 31663861]
32. Mitchell AJ, Chan M, Bhatti H, Halton M, Grassi L, Johansen C, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. The Lancet Oncology 2011 Feb;12(2):160-174. [doi: 10.1016/S1470-2045(11)70002-X] [Medline: 21251875]
33. DuBenske LL, Gustafson DH, Namkoong K, Hawkins RP, Atwood AK, Brown RL, et al. CHESS improves cancer caregivers' burden and mood: results of an eHealth RCT. Health Psychol 2014 Oct;33(10):1261-1272 [FREE Full text] [doi: 10.1037/a0034216] [Medline: 24245838]
34. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
35. 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]

36. 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]
37. 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]
38. Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]

Abbreviations

eHealth: electronic health
iCBT: internet-delivered cognitive behavioral therapy
IHCA: interactive health communication application
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 08.10.19; peer-reviewed by K Day, S March; comments to author 24.11.19; accepted 26.01.20; published 30.03.20.

Please cite as:
Hauffman A, Alfonsson S, Igelström H, Johansson B
Experiences of Internet-Based Stepped Care in Individuals With Cancer and Concurrent Symptoms of Anxiety and Depression: Qualitative Exploration Conducted Alongside the U-CARE AdultCan Randomized Controlled Trial
J Med Internet Res 2020;22(3):e16547
URL: http://www.jmir.org/2020/3/e16547/
doi:10.2196/16547
PMID:32224483

Original Paper

Assessing Real-Time Moderation for Developing Adaptive Mobile Health Interventions for Medical Interns: Micro-Randomized Trial

Timothy NeCamp[1], PhD; Srijan Sen[2,3], MD, PhD; Elena Frank[2], PhD; Maureen A Walton[4,5], MPH, PhD; Edward L Ionides[1], PhD; Yu Fang[2], MSE; Ambuj Tewari[1], PhD; Zhenke Wu[6,7], PhD

[1]Department of Statistics, University of Michigan, Ann Arbor, MI, United States
[2]Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, United States
[3]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
[4]Addiction Center, Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
[5]Injury Prevention Center, University of Michigan, Ann Arbor, MI, United States
[6]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States
[7]Michigan Institute for Data Science, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Timothy NeCamp, PhD
Department of Statistics
University of Michigan
1085 South University Ave
Ann Arbor, MI, 48109
United States
Phone: 1 7346474820
Email: tnecamp@umich.edu

Abstract

Background: Individuals in stressful work environments often experience mental health issues, such as depression. Reducing depression rates is difficult because of persistently stressful work environments and inadequate time or resources to access traditional mental health care services. Mobile health (mHealth) interventions provide an opportunity to deliver real-time interventions in the real world. In addition, the delivery times of interventions can be based on real-time data collected with a mobile device. To date, data and analyses informing the timing of delivery of mHealth interventions are generally lacking.

Objective: This study aimed to investigate when to provide mHealth interventions to individuals in stressful work environments to improve their behavior and mental health. The mHealth interventions targeted 3 categories of behavior: mood, activity, and sleep. The interventions aimed to improve 3 different outcomes: weekly mood (assessed through a daily survey), weekly step count, and weekly sleep time. We explored when these interventions were most effective, based on previous mood, step, and sleep scores.

Methods: We conducted a 6-month micro-randomized trial on 1565 medical interns. Medical internship, during the first year of physician residency training, is highly stressful, resulting in depression rates several folds higher than those of the general population. Every week, interns were randomly assigned to receive push notifications related to a particular category (mood, activity, sleep, or no notifications). Every day, we collected interns' daily mood valence, sleep, and step data. We assessed the causal effect moderation by the previous week's mood, steps, and sleep. Specifically, we examined changes in the effect of notifications containing mood, activity, and sleep messages based on the previous week's mood, step, and sleep scores. Moderation was assessed with a weighted and centered least-squares estimator.

Results: We found that the previous week's mood negatively moderated the effect of notifications on the current week's mood with an estimated moderation of −0.052 (*P*=.001). That is, notifications had a better impact on mood when the studied interns had a low mood in the previous week. Similarly, we found that the previous week's step count negatively moderated the effect of activity notifications on the current week's step count, with an estimated moderation of −0.039 (*P*=.01) and that the previous week's sleep negatively moderated the effect of sleep notifications on the current week's sleep with an estimated moderation of −0.075 (*P*<.001). For all three of these moderators, we estimated that the treatment effect was positive (beneficial) when the moderator was low, and negative (harmful) when the moderator was high.

Conclusions: These findings suggest that an individual’s current state meaningfully influences their receptivity to mHealth interventions for mental health. Timing interventions to match an individual’s state may be critical to maximizing the efficacy of interventions.

Trial Registration: ClinicalTrials.gov NCT03972293; http://clinicaltrials.gov/ct2/show/NCT03972293

(J Med Internet Res 2020;22(3):e15033) doi:10.2196/15033

KEYWORDS

mobile health; digital health; smartphone; mobile phone; wearable devices; ecological momentary assessment; depression; mood; physical activity; sleep; moderator variables

Introduction

Background

According to the World Health Organization, depression is the leading cause of disease-associated disability in the world [1]. In the United States, the burden of depression, including suicide, has continued to grow [2]. In populations at high risk, prevention of depression may be an effective strategy. The US National Academy of Medicine has highlighted the need to develop, evaluate, and implement prevention interventions for depression and other mental, emotional, and behavioral disorders [3].

Prevention interventions for depression are critical for individuals in stressful work environments because these environments can lead to increased rates of depression [4]. However, individuals in these work environments may have inadequate time or resources to access traditional mental health care services. High stress can also make individuals less receptive to interventions and behavior change [5,6].

Unlike other recent advances, mobile technology has the potential to transform the delivery and timing of depression prevention interventions to meet the needs of highly stressed individuals. In contrast to more intensive treatments (such as therapeutic appointments), mobile health (mHealth) interventions (such as push notifications) can be delivered at low burden, which may be critical given the individuals’ high stress workloads. Mobile devices hold the power to deliver just-in-time adaptive interventions (JITAIs) [7] to individuals during times when they are able to receive and respond to them. Mobile devices also collect objective measurements of an individual’s context and behavior with minimal burden (eg, step counts, sleep duration). These data may, in turn, be used to determine when to deliver interventions, and evaluate intervention efficacy, without bothering the individuals.

When initially designing a JITAI, these states of opportunity [7]—times when individuals are receptive to positive behavior change—are not known. Timing is critical because poorly timed interventions can lead to loss of engagement with the intervention [8]. Timing interventions is also particularly important for individuals in stressful work environments because poorly timed interventions could cause disengagement and treatment fatigue [9].

Current behavioral theories lack the granularity and adaptivity necessary to inform the timing of the delivery of mHealth interventions [10,11]. Many theoretical models are nondynamic—they only consider treatment adaptation based on baseline characteristics, such as sex and depression history [12]. Timing and adapting treatments based on real-time variables is essential for developing high-quality JITAIs [7].

This study follows a data-driven approach to inform the dynamic timing of intervention delivery. Experimentation and data collection were used to provide empirical evidence for determining states of opportunity—the data illustrate when interventions cause positive behavior change in individuals and when they do not.

There have been other empirical studies showing the promise of JITAIs to improve mental health [13]. Those studies are either focused on feasibility and acceptability of the JITAI [14-17] or use a randomized controlled trial (RCT) to demonstrate the impact of the JITAI on a distal outcome [18-20]. They do not focus on the timing of intervention delivery. In two studies [21,22], the authors demonstrated the benefits of timing mHealth intervention delivery based on real-time variables. In that work, the timing of intervention delivery is specified before the study. In contrast, because we did not know *a priori* how to time our interventions, our work used empirical evidence to learn how to dynamically time intervention delivery.

In statistical terms, we formulated the task of empirically learning how to dynamically time interventions as *discovering time-varying moderators of causal treatment effects* [23]. Time-varying moderators are *moderators* because they change—or moderate—the efficacy of subsequent treatments and are *time-varying* because the moderators’ values vary throughout the study (such as daily mood). For example, if push notifications containing sleep messages cause a larger increase in sleep when individuals had little sleep in the previous night compared with when individuals had high sleep, then the previous night’s sleep moderates the effect of sleep notifications. Discovering time-varying moderators informs treatment timing because treatment delivery can now be based on the observed values of these moderators. In the example, it may be better to send sleep notifications only after individuals have insufficient sleep. Note that time-varying moderators should have meaningful variability to allow the possibility of sending different interventions at different times.

We assessed time-varying moderators of mHealth interventions targeting 3 categories: mood, activity, and sleep. Stressful work environments can lead to sleep deprivation and physical inactivity [24-26], two behaviors directly associated with depression [24,27,28]. To prevent depression among individuals experiencing high stress, it is critical to develop high-quality interventions that can help them maintain and improve their mood, either through targeting mood directly or by indirectly improving activity and sleep [24,28].

Our study population comprised medical interns. Medical internship, the first year of physician residency training, is highly stressful, causing the depression rates of interns to be several folds higher than those of the general population [29]. Focusing on physician training, a rare situation in which a dramatic increase in stress can be anticipated, provides an ideal experimental model to develop interventions for maintaining mental wellness during life and work stressors.

Our study, the 2018 Intern Health Study (IHS) [30], was a 6-month-long mHealth cohort study that tracked medical interns using phones and wearables. During the internship year, we conducted a micro-randomized trial (MRT) [23]. Standard single–time point RCTs only inform moderation by baseline variables [31] and do not permit the discovery of time-varying moderators. The MRT was advantageous because it allowed us to discover time-varying moderators of causal treatment effects [23].

During each week in the 6-month study, an intern was randomized to 1 of 4 possible treatments: a week of mood notifications, activity notifications, sleep notifications, or no notifications. The outcomes were average daily self-reported mood valence (measured through a one-question survey), average daily steps (as a proxy for activity), and average daily sleep duration, where averages were 7-day averages of data collected during the week of treatment. The strongest moderators were hypothesized to be the previous week's average daily mood, steps, and sleep, as these were the strongest predictors of the outcomes (based on previous years' IHS data [30]). We were only interested in a subset of combinations of outcomes, treatments, and moderators.

Study Aims

Here, we have highlighted the primary and secondary aims of this paper. Below, the *effect* (for which we are assessing moderation) corresponds to how a week of a certain notification category causally changes an outcome *compared with weeks of no notifications*.

The moderator aims listed below were not the only aims of the 2018 IHS. Analyses of main effects analyses were conducted before the analysis of moderator effects (see Additional Analyses in Multimedia Appendix 1). This paper has focused on moderator analyses as those were the most interesting findings.

Primary Aim

Our primary aim focused on discovering how an intern's previous mood moderates the effect of notifications in general. Specifically, we examined the following: Is the effect of a week of notifications (of *any* category) on the average daily mood moderated by the previous week's mood? Here, *Outcome*=mood; *Treatment*=any (mood, activity, or sleep); and *Moderator*=mood.

Exploratory Subaim

If we do find that mood moderates the effect of notifications, generally, we will assess if this moderation is consistent across all intervention categories. Specifically, we will examine the following: Is the effect of *each* category of notification on the average daily mood moderated by the previous week's mood? Here, *Outcome*=mood; *Treatment*=mood, activity, and sleep separately; and *Moderator*=mood.

Secondary Aim 1

Secondary aim 1 focused on discovering how an intern's previous activity moderates the effect of notifications containing activity messages. Specifically, we examined the following: Is the effect of a week of activity notifications on the average daily step count moderated by the previous week's step count? Here, *Outcome*=steps, *Treatment*=activity, and *Moderator*=steps.

Secondary Aim 2

Secondary aim 2 focused on discovering how an intern's previous sleep moderates the effect of notifications containing sleep messages. Specifically, we examined the following: Is the effect of a week of sleep notifications on the average daily sleep moderated by the previous week's sleep? Here, *Outcome*=sleep; *Treatment*=sleep; and *Moderator*=sleep.

Methods

The Study App

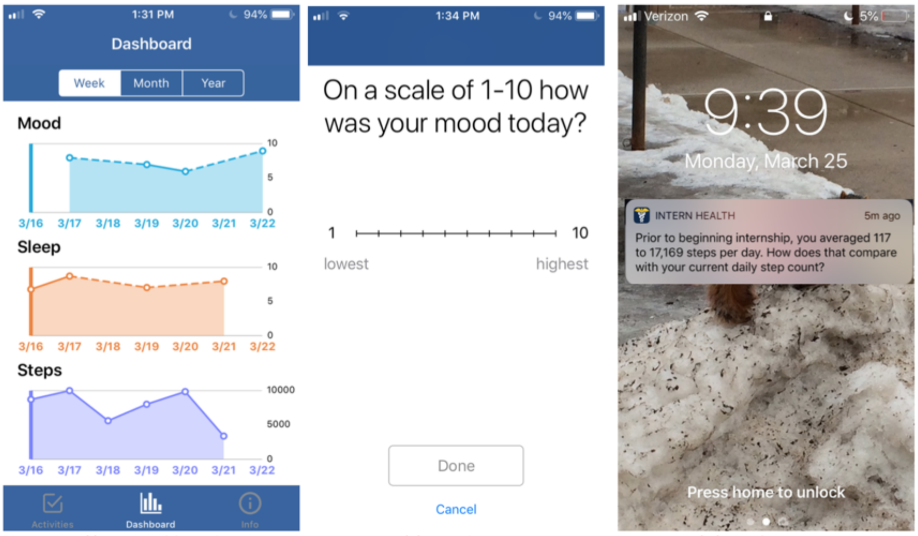
Study participants were provided a Fitbit Charge 2 [32] to collect sleep and activity data, and a mobile app was downloaded to the intern's phone. The app can conduct ecological momentary assessments (EMAs) [33], aggregate and visualize data, and deliver push notifications. The app was designed for iOS using Apple ResearchKit [34].

As the primary aim of the study was focused on understanding the effects of interventions on the mental health of interns, we employed a daily EMA to measure mood valence (see Figure 1, Mood EMA). Daily mood is one of the 2 cardinal symptoms of depression [35]. This daily mood EMA is used widely to track the mood in patients with depression [36]. There are more widely used measurements of mental health other than mood valence (such as the Patient Health Questionnaire-9, PHQ-9 [37]). However, these questionnaires are more time-intensive and their assessment may cause higher nonresponse rates. Participants are prompted to enter their daily mood every day at a user-specified time between 5 PM and 10 PM.

In addition to collecting EMA data, the app aggregates and displays visual summaries of interns' historical data, including step and sleep counts (collected through the Fitbit) and mood EMA data (Figure 1, App Dashboard). The data are integrated with the app using Fitbit's publicly available application programming interface [38]. Displaying historical trends to the intern helps them self-monitor their mood, activity, and sleep trajectories and could potentially lead to positive reactive behavior change [39]. These displays are a type of *pull* intervention, that is, interventions that are available at all times but only accessed upon user request. The *pull* component was available to all participants, and assessing its effects was not the focus of this study.

The IHS app also delivers *push* interventions, that is, interventions delivered without user prompting. Evaluating and improving the delivery timing of the push notification intervention was the focus of this study.

Figure 1. Screenshots of the app dashboard, mood ecological momentary assessment, and lock screen notifications.



On a scale of 1-10 how was your mood today?

Prior to beginning internship, you averaged 117 to 17,169 steps per day. How does that compare with your current daily step count?

(i) App Dashboard (ii) Mood EMA (iii) Notifications

Push Notification Intervention

As applied to mHealth, theoretically, behavior change comprises an individual's motivation and ability to change, combined with a trigger to elicit change [40]. Push notifications are such a trigger, potentially providing motivational messages for change (eg, to spark change), strategies for change (eg, to facilitate change), and/or reminders to engage with the app (eg, to signal change) [40]. Importantly, research supports the potential of push notifications for behavior change [41,42].

Push notifications are particularly advantageous for medical interns because they are delivered as needed with minimal burden to the user [42-44]. However, poorly timed push notifications can lead to loss of engagement and treatment fatigue [9,45], demonstrating the importance of evaluating and improving the delivery timing of the push notifications.

Push notifications were provided to the interns through the app, with the goal of improving healthy behavior in a target category of interest: mood, activity, and sleep (ie, mood notifications improve mood, activity notifications increase physical activity, and sleep notifications increase sleep duration). For all 3 categories, there were 2 types of notifications: tips and life insights. Consistent with theory [40] and motivational interviewing approaches [46-48], tips are non–data-based notifications that provide autonomy support (eg, motivational focused messages on why change) and tools (eg, ability-focused messages on how to change) to promote healthy mood, activity, or sleep. Next, consistent with theory [40,49,50] and research showing that interventions that enhance self-monitoring promote behavior change [51], life insight notifications summarize an individual's data, to provide reminders (eg, signals) and/or reduce the burden of accessing the app to view visualizations. Table 1 contains examples of different push notifications used in the study.

Table 1. Examples of 6 different groups of notifications.

Notification groups	Life insight	Tip
Mood	Your mood has ranges from 7 to 9 over the past 2 weeks. The average intern's daily mood goes down by 7.5% after intern year begins.	Treat yourself to your favorite meal. You've earned it!
Activity	Prior to beginning internship, you averaged 117 to 17,169 steps per day. How does that compare with your current daily step count?	Exercising releases endorphins which may improve mood. Staying fit and healthy can help increase your energy level.
Sleep	The average nightly sleep duration for an intern is 6 hours 42 minutes. Your average since starting internship is 7 hours 47 minutes.	Try to get 6 to 8 hours of sleep each night if possible. Notice how even small increases in sleep may help you to function at peak capacity & better manage the stresses of internship.

The Intern Health Study Micro-Randomized Trial Design

To discover time-varying moderators for informing the timing of notification delivery, we conducted an MRT. The MRT design is shown in Figure 2. The MRT design and protocol were approved by the University of Michigan Institutional Review Board (Protocol #HUM00033029).

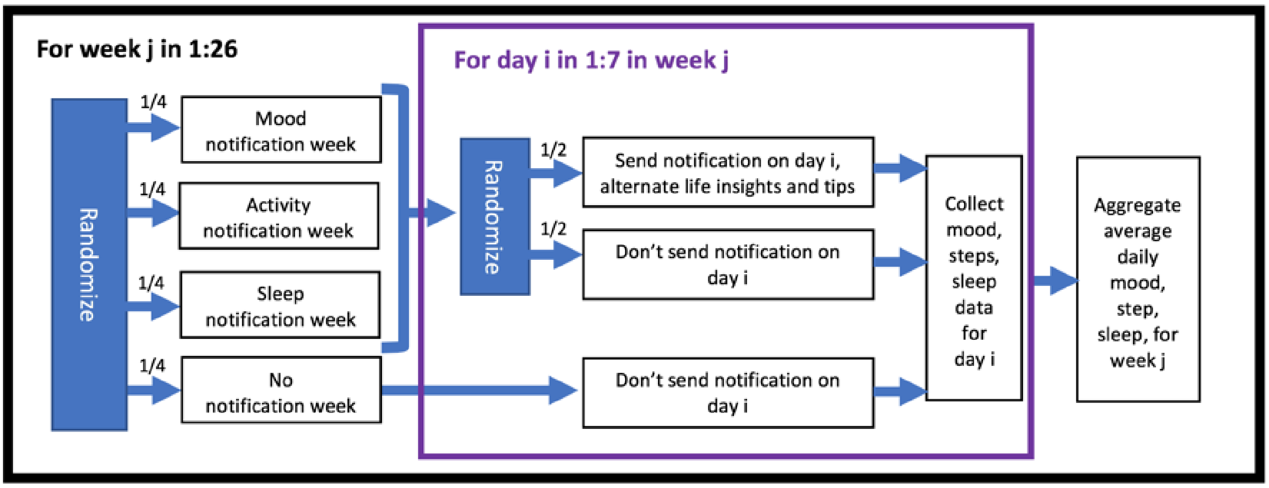
The main randomization was the weekly randomization to a specific notification category (mood, activity, sleep) or to no notification. Thus, we were able to compare how a week of a certain notification category changed intern behavior when compared with a week of no notifications.

The randomization—and the ensuing analysis of effects—occurred at the weekly level for two reasons. First, the notifications are not intended to change the interns' behavior in the next few hours, but over the next few days. Randomizing and analyzing effects at the weekly level, as opposed to a daily or minute level, permitted the discovery of longer-term effects. Second, as interns are quite busy, their behavior may not change significantly after receiving a single notification. Instead, interns received several notifications related to the same category and had a consistent reminder about improving that category.

Given a week when a user was randomized to receiving notifications, every day they were further randomized (with 50% probability) to receive a notification on that day. Hence, for a mood notification week, the user received, on average, 3.5 mood notifications that week. The purpose of this randomization was to balance delivering enough notifications to be noticeable and cause behavior change but not too often that it leads to treatment fatigue [9]. Treatment fatigue is pervasive in mHealth [7] and for individuals with heavy workloads [9]. Additional Analyses in Multimedia Appendix 1 contains a summary of how many notifications users received in a given week.

Another way to prevent treatment fatigue is through increased variability in notifications and the order in which they are received [52]. For each notification category, the notifications alternated between life insights and tips. In addition, given a type and category, each notification was drawn randomly, without replacement, from a corresponding bucket of notifications. The bucket refilled once it was completely emptied. Alternating between life insights and tips increased the day-to-day variability of the notification framing. Drawing notifications without replacement ensured that users were not receiving repeats of the same notification. Under this scheme, on average, a user did not receive a repeat notification for 16 weeks. Weekly and daily notification randomization and notification delivery were implemented using the Firebase Cloud Messaging platform [53].

Figure 2. Randomization scheme of the Intern Health Study micro-randomized trial.



Participants

Medical doctors starting their year-long internship in the summer of 2018 were eligible to participate in the study. Interns were onboarded before the start of their internship (between April 2018 and June 2018), in which they were instructed to download the study app, were provided Fitbits, completed a baseline survey, and were able to begin entering mood scores. Baseline and follow-up surveys were administered through the app using Qualtrics survey software [54]. Data collection began when an intern enrolled in the study and continued until the end of the trial. Collecting data before the start of the internship provided baseline measurements of mood, step counts, and sleep, which are valuable control variables in the analysis. The weekly randomizations and notification delivery began on June 30, 2018, 1 day before the start of interns' clinical duties. Interns were rerandomized every 7 days thereafter. During the study, notifications were sent at 3 PM, mood EMAs were collected daily between 5 PM and 10 PM, and sleep and step data were collected every minute. Data were transferred directly from the subjects' phones to a secure, Health Insurance Portability and Accountability Act–compliant server managed by the University of Michigan Health Information and Technology Services. The interns received notifications for 6 months (26 weeks), and the trial ended on December 28, 2018.

Statistical Analysis

Overview

To analyze the primary and secondary aims, we performed a moderator analysis for each of the outcomes, treatments, and moderators specified in Study Aims. More details on the methods can be found in Further Details on the Statistical Methods in Multimedia Appendix 1.

In the analysis, there were 4 sets of variables:

1. The *treatment outcome* variable of interest, Y_t.
2. The *treatment indicator*, Z_t. For now, Z_t is a binary indicator, where Z_t=1 implies it is a notification week (of any category) and Z_t=0 is a no-notification week. The case with multicategorical treatments—mood, activity, and sleep notifications—will be described under the secondary aims.

3. The *moderator*, M_t, corresponding to the causal effect moderator of interest.
4. The last set of variables, X_t, are the *control variables*. The control variables are variables measured before each weekly randomization (eg, baseline data and previous weeks' data) and are included in the model to reduce variation in the outcome, Y_t.

The outcomes, treatment, and moderators correspond exactly to the outcomes, treatments, and moderators described in Study Aims. As interns were randomized to different treatments each week, the outcomes, treatments, moderators, and control variables were aggregated at the weekly level and were indexed by time, *t*, corresponding to each week of the study (t=1,…,26).

To perform the moderator analysis, we used a linear model with an interaction term. The outcome of interest (such as average daily mood), Y_t, was regressed on X_t, M_t, Z_t and the interaction between M_t and Z_t, Z_tM_t, giving the model the following form:

$$E(Y_t|X_t,M_t,Z_t) = a_0X_t + a_1M_t + b_0Z_t + b_1Z_tM_t$$

The moderation effect of interest is the coefficient b_1 for the interaction of Z_t and M_t. This coefficient is interpreted as the change in treatment effect of treatment Z_t on Y_t per unit change in M_t. A positive value for b_1 indicates that the treatment works better after weeks when M_t is high, whereas a negative value indicates that the treatment works better after time points when M_t is low.

For the primary and secondary aims, to evaluate if the moderator effect is statistically significant, we performed a hypothesis test comparing the coefficient b_1 to 0, with a 0.05 type I error rate. We reported the estimate of b_1, the standard error, and *P* value of this test. Though estimating and testing the moderation effect is useful, it does not demonstrate whether the notifications had a positive or negative effect on the outcome. Hence, in addition to a hypothesis test, we also plotted the estimated treatment effect at various values of the moderator. We did this by using both the estimate of the slope, b_1, and intercept, b_0, of the moderation effect. The plots also included histograms of the moderator to illustrate the distribution of treatment effects.

Estimation Techniques

To estimate the coefficients, we used a multicategorical extension of the weighted and centered least-squares estimator described in Boruvka et al [55]. The estimation method provides asymptotically unbiased estimates of the causal effect moderation of interest. The method also protects against potential misspecification of terms not interacted with treatment ($a_0X_t + a_1M_t$). The method assesses the uncertainty of the coefficient estimates using robust standard error estimation—the *sandwich* estimator [56]—to account for correlation between outcomes over time. The method was implemented in R using the package geepack [57]. The code is available on the first author's website.

Missing Data

Missing data occurred throughout the trial because of interns not completing the self-reported mood survey or not wearing Fitbits. Multiple imputation [58], a robust method for dealing with missing data, was used to impute missing data at the daily level. Due to the complexity of the trial design and data structure, our imputation method combines imputation methods for longitudinal data [59] and sequentially randomized trials [60]. Results were aggregated across 20 imputed datasets using Rubin's rules [58,61]. We also assessed the sensitivity of the conclusions to the imputation method. See Missing Data and Sensitivity Analyses in Multimedia Appendix 1 for further details on the missingness and sensitivity analysis results.

Primary Aim

The primary aim assessed the previous week's average daily self-reported mood valence as a moderator of the effect of notifications on the average daily self-reported mood valence. For this analysis, the interpretation, b_1, was *the change in treatment effect (for delivering a week of notifications compared with a week of no notifications) on the average daily mood when the previous week's average daily mood increased by 1*.

Secondary Aim 1

The first secondary aim assessed the previous week's average daily step count as a moderator of the effect of activity notifications on the average daily step count. For this aim, the treatment variable (Z_t) and corresponding coefficients (b_0 and b_1) were no longer binary because there were 4 possible notification categories. See Further Details on the Statistical Methods in Multimedia Appendix 1 for further details on the multicategorical treatment model. The focus of inference for secondary aim 1 was on the first dimension of the moderation effect, b_{11}, which corresponds to the comparison between activity notification weeks and no-notification weeks. In addition, to reduce right skew and decrease outliers, the outcome and moderator used average daily *square root* step count. After the square root transformation, the average daily step counts more closely resembled a Gaussian distribution.

The interpretation, b_{11}, was *the change in treatment effect (for delivering a week of activity notifications compared with a week of no notifications) on the average daily square root step count when the previous week's average daily square root step count increased by 1*. Hypothesis testing was performed on b_{11} and plots were made using estimates of b_{01} and b_{11}.

Secondary Aim 2

Secondary aim 2 assessed the previous week's average daily sleep count as a moderator of the effect of sleep notifications on the average daily sleep count. Similar to secondary aim 1, the treatment here was no longer binary, and we encoded the treatment vector the same way as secondary aim 1. For this analysis, the focus of inference was on the second dimension, b_{12}, which compared sleep notification weeks with no-notification weeks. Again, to reduce right skew and decrease outliers, the outcome and moderator used average daily *square root* sleep minutes.

The interpretation, b_{12}, was *the change in treatment effect (for delivering a week of sleep notifications compared with a week of no notifications) on average daily square root sleep minutes* *when the previous week's average daily square root sleep minutes increased by 1*. Hypothesis testing was performed on b_{12}, and plots were made using estimates of b_{02} and b_{12}.

Exploratory Subaim

The exploratory aim assessed the previous week's mood as a moderator of the effect of each notification category on the average daily mood. For the exploratory aim, the outcome and moderator were the same as the primary aim, except that the treatment was separated into 4 treatment categories (as in the secondary aims). As this aim was only exploratory, we did not calculate P values. Instead, for each notification category, we plotted the estimated treatment effect at various values of the moderator. This required making 3 separate lines using each dimension, with estimates of b_{0i} providing the intercept and estimates of b_{1i} providing the slope.

Results

Participants

Participants were recruited through emails, which were sent to future interns from 47 different recruitment institutions between April 1 and June 25, 2018. The recruitment institutions comprised both medical schools, where emails were sent to all graduates, and residency locations, where emails were sent to all incoming interns. A total of 5233 future interns received the initial email inviting them to participate in the study. In all, 40.78% (2134/5233) of interns downloaded the study app, completed the consent form, and filled out the baseline survey sometime before June 25, 2018. The study app and study participation were restricted to interns using an iPhone, the phone brand used by most interns. A total of 2134 interns received a Fitbit Charge 2. Of the 2134 interns, 1565 (73.34%) were randomly selected to participate in the MRT (see Additional Analyses in Multimedia Appendix 1 for an explanation of this initial randomization). These 1565 interns were randomized according to Figure 2. Interns were incentivized to participate in the study by receiving the Fitbit wearable and up to US $125, distributed 5 times throughout the year (US $25 each time) based on continued participation.

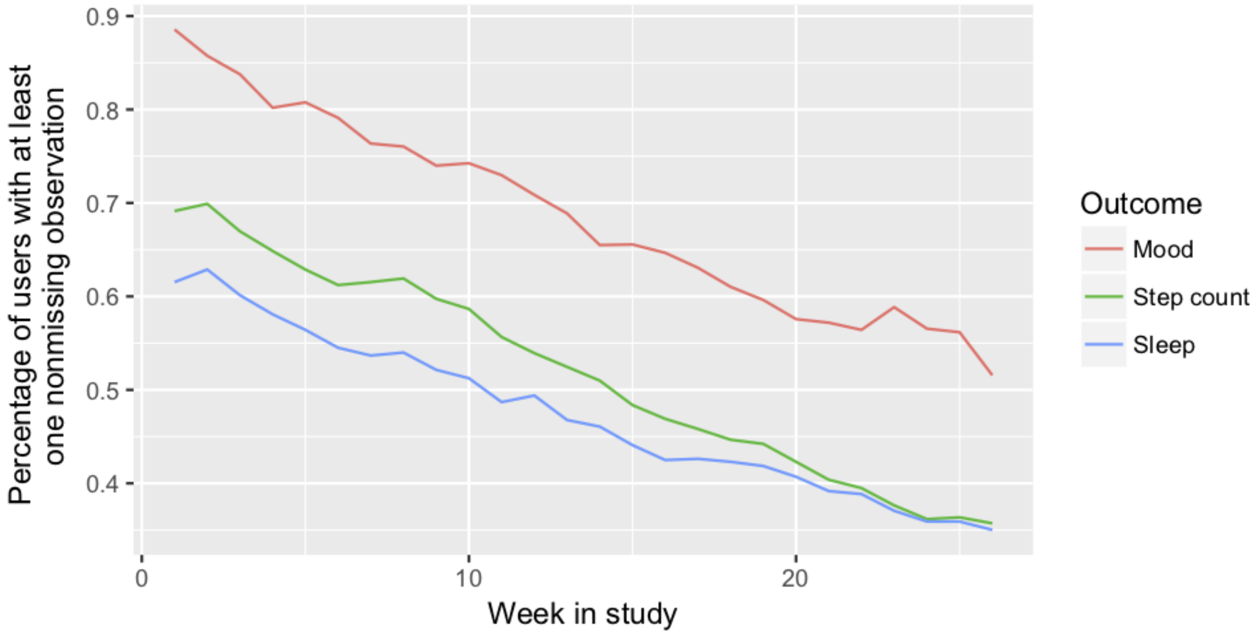
Of the 1565 interns in the MRT, 875 (55.91%) were female, and 774 (49.45%) had previously experienced an episode of depression. The interns represented 321 different residency locations and 42 specialties. The study interns' baseline information closely resembled the known characteristics of the general medical intern population [29]. Throughout the trial, we measured intern mood valence, steps, and nightly sleep. Summaries of the weekly averages of those data can be found in Table 2.

Missing data occurred throughout the study. Figure 3 displays the percentage of interns with at least one nonmissing sleep, step, or mood observation for each week in the study. See Missing Data and Sensitivity Analyses in Multimedia Appendix 1 for further details on the missingness and sensitivity analyses.

Table 2. Summary statistics of daily mood, activity, and sleep during the study, averaged over each week of the study. These are the primary outcomes and moderators used in the analyses of all study aims.

Daily measure	First quartile	Median	Mean (SD)	Third quartile
Average daily mood	6.50	7.33	7.21 (1.43)	8.00
Average daily step count	6193	7983	8274 (3285)	10,050
Average daily hours of sleep	6.02	6.65	6.54 (1.25)	7.25

Figure 3. Percentage of interns with at least one nonmissing sleep, step, or mood observation for each week in the study.



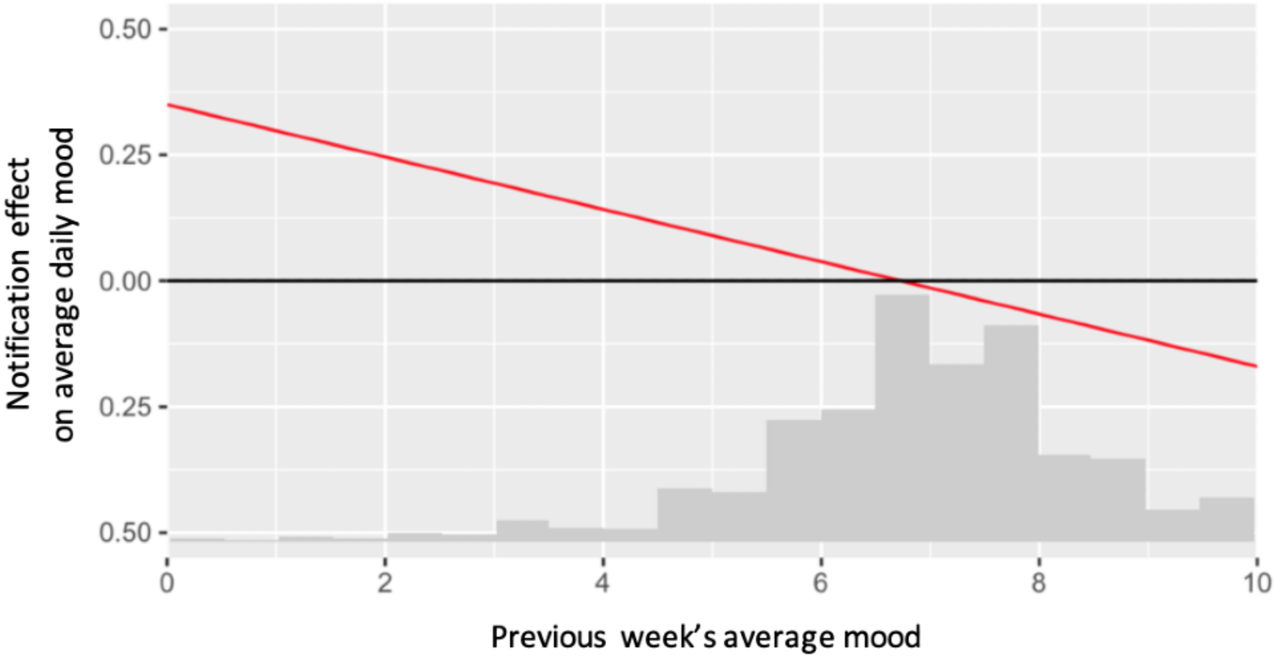
Main Findings

Primary Aim

We conclude that the previous week's average daily self-reported mood valence is a statistically significant negative moderator of the effect of notifications on the average daily self-reported mood valence. The estimate for the moderation is −0.052 (SE 0.014; 95% CI −0.081 to −0.023; *P*=.001).

Figure 4 plots the estimated treatment effect at various values of the moderator. Figure 4 shows that the effect of notifications (compared with no notifications) was positive for weeks when the previous mood was low, but negative for weeks when the previous mood was high. For example, when the previous week's average daily mood was 3, we estimated that a week of notifications *increased* an intern's average daily mood by 0.19 (effect size=0.14). However, when the previous week's average daily mood was 9, we estimated that a week of notifications *decreased* an intern's average daily mood by 0.12 (effect size=−0.08).

Figure 4. Estimated treatment effects (compared with no notifications) of notifications on average daily mood, at various values of previous week's mood. The x-axis also contains a scaled histogram of previous week's average mood.

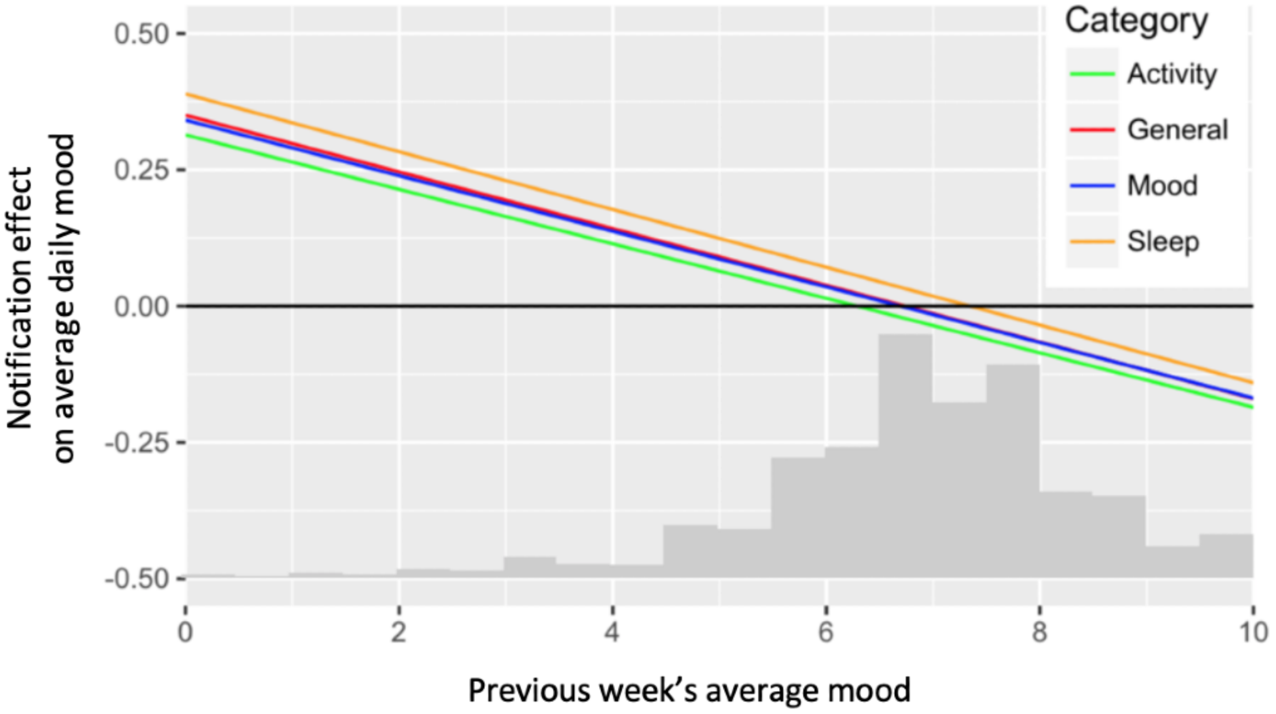


Exploratory Subaim

For each notification category, we plotted the estimated treatment effect at various values of the moderator. Essentially, we broke apart the moderation effect in Figure 4 into 3 categories of notifications. The result is shown in Figure 5. We included the line for general notifications from Figure 4 for reference. Figure 5 demonstrates that the moderation by the previous week's average daily mood was similar for all 3 notification categories.

When the previous week's average daily mood was 3, we estimated that a week of mood, activity, and sleep notifications *increased* an intern's average daily mood by 0.19, 0.16, and 0.23 (effect sizes=0.13, 0.11, and 0.16), respectively. When the previous week's average daily mood was 9, we estimated that a week of mood, activity, and sleep notifications *decreased* an intern's average daily mood by 0.12, 0.14, and 0.09 (effect sizes=−0.08, −0.10, and −0.06), respectively.

Figure 5. Estimated treatment effects (compared with no notifications) of different notification categories on average daily mood, at various values of previous week's mood. The x-axis also contains a scaled histogram of previous week's average mood.

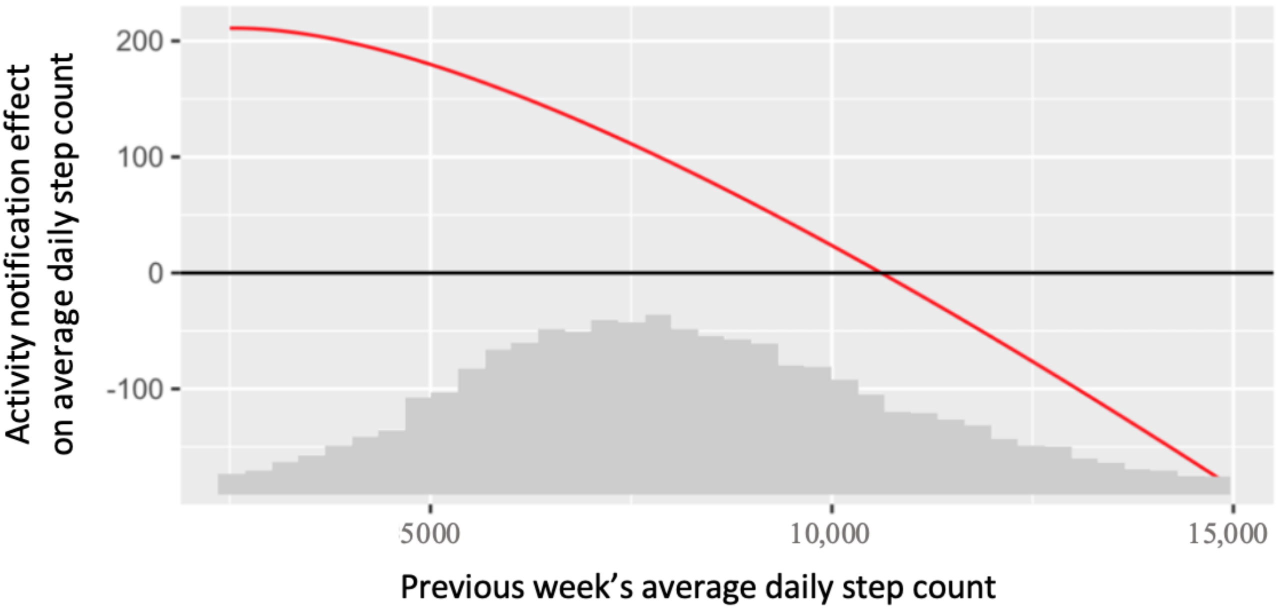


Secondary Aim 1

We conclude that the previous week's average daily step count is a statistically significant negative moderator of the effect of activity notifications on average daily steps. The estimate for the moderation is −0.039 (SE 0.015; 95% CI −0.069 to −0.008; *P*=.01).

Figure 6 plots the estimated treatment effect at various values of the moderator. In Figure 6, for interpretability, we retransformed the moderation effect back from the analysis scale (square root) to the original scale. We see from Figure 6 that the effect of activity notifications (compared with no notifications) was positive for weeks when previous steps were low, but negative for weeks when previous steps were high. For example, when the previous week's average daily step count was 5625, we estimated that a week of activity notifications *increased* an intern's average daily step count by 165 steps (effect size=0.05). However, when the previous week's average daily step count was 12,100, we estimated that a week of activity notifications *decreased* an intern's average daily step count by 60 steps (effect size=−0.02).

Figure 6. Estimated treatment effects (compared with no notifications) of activity notifications on average daily steps, at various values of previous week's step counts. The x-axis also contains a scaled histogram of previous week's average daily step count.

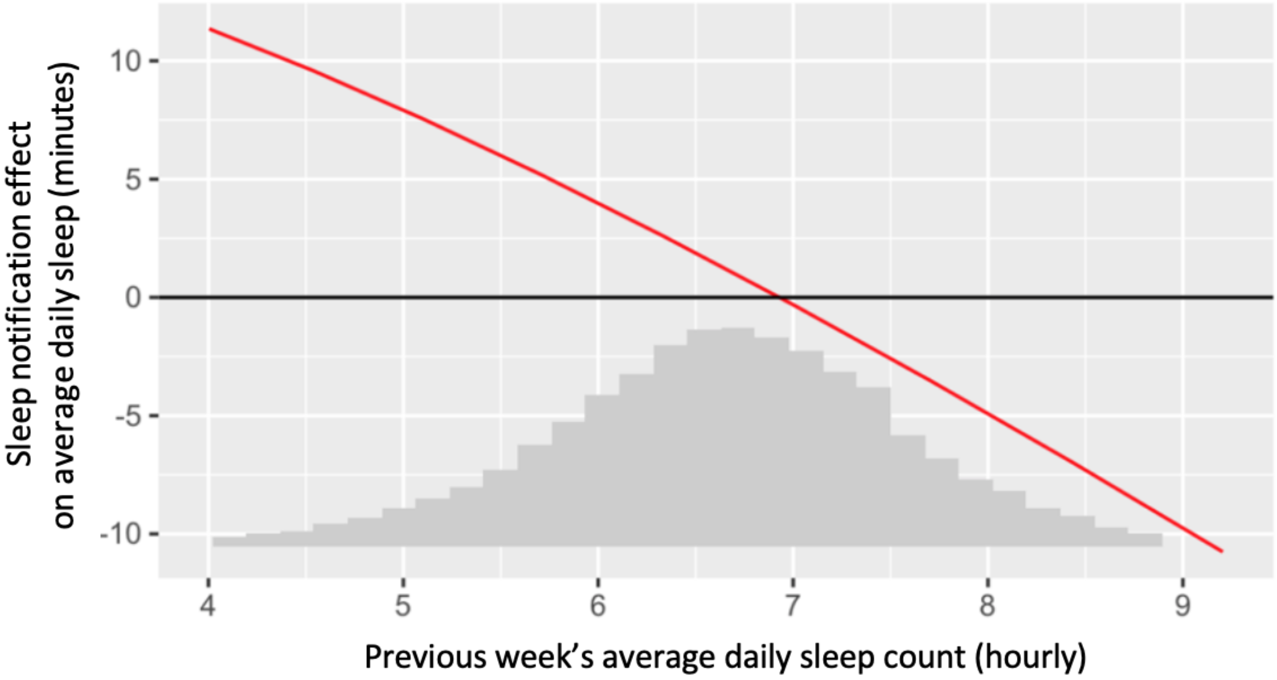


Secondary Aim 2

We conclude that the previous week's average daily sleep is a statistically significant negative moderator of the effect of sleep notifications on average daily sleep. The estimate for the moderation is −0.075 (SE 0.018; 95% CI −0.111 to −0.038; *P*<.001).

Figure 7 plots the estimated treatment effect at various values of the moderator. Again, we retransformed the moderation effect back from the analysis scale (square root) to the original scale. In addition, for interpretability, the x-axis is on the hourly scale, whereas the y-axis is on the minute scale. We see from Figure 7 that the effect of sleep notifications (compared with no notifications) was positive for weeks when previous sleep was low, but negative for weeks when previous sleep was high. For example, when the previous week's average daily sleep was 5 hours, we estimated that a week of sleep notifications *increased* an intern's average daily sleep by 8 min (effect size=0.11). However, when the previous week's average daily sleep was 8 hours, we estimated that a week of sleep notifications *decreased* an intern's average daily sleep by 5 min (effect size=−0.07).

Figure 7. Estimated treatment effects (compared with no notifications) of sleep notifications on average daily sleep minutes, at various values of previous week's hourly sleep. The x-axis also contains a scaled histogram of previous week's average daily sleep count.



Additional Analyses

The Additional Analyses section of Multimedia Appendix 1 contains detailed results on other analyses, including an analysis of nonmoderated main effects, changes in effects over time, the effects of life insights and tips, the effects on long-term PHQ-9 scores, and an analysis of baseline moderators. There is evidence of a negative effect of (general) notifications on mood. There is also evidence of a positive effect of activity notifications on step count and a positive effect of sleep notifications on sleep duration. All of these effect sizes, however, are small. There is no strong evidence that these effects change over time. There is minor evidence that tips perform better than life insights in improving step count and sleep duration. We did not see any effects on long-term mental health outcomes. We saw some evidence of nonlinear moderation for the primary and secondary aims. The nonlinear moderator analysis suggested that when the moderators are high, the treatment effect on sleep hours and step count is close to 0 (as opposed to negative). Finally, we found that baseline variables, such as gender and depression history, were weak moderators of notification effects, demonstrating the value of personalizing intervention delivery on real-time data.

Discussion

Principal Findings

Through this MRT of an mHealth push notification intervention, we found that the effects of notifications were negatively moderated by the subject's previous measurement of the outcome of interest. Specifically, we found that previous mood negatively moderated the effect of notifications on mood, previous step count negatively moderated the effect of activity notifications on step count, and previous sleep duration negatively moderated the effect of sleep notifications on sleep duration.

Comparison With Other Studies

A few previous studies explored using real-time variables to determine the timing of mHealth interventions for mental health. These studies postulated that messages would be most effective when self-reported mood was outside the typical range [21], or when self-reported stress or negative affect was high [22]. The studies found that such timing does improve efficacy. Our work differs from these studies because we did not assume, beforehand, that interventions would be most effective during a predetermined time. Instead, we used the MRT design to *learn* opportune times to send interventions, based on real-time objective and self-reported data.

Outside of mental health, there have been studies that have sought to learn opportune times to send interventions. Much of that work is focused on assessing in-the-moment interruptibility, namely times when a user is open to interruption and willing to engage with a notification. For example, in one study [62], the authors found that phone usage, time of day, and location were strong predictors of a user's willingness to engage with content provided via a push notification. Another study [63] found that location, affect, current activity, time of day, day of week, and current stress are significant predictors of a user's willingness to respond to an EMA prompt. Another study [64] used an MRT to causally demonstrate that notifications (which ask users to self-monitor) are more effective when sent mid-day and on weekends. Our study differs from this work. In our study, the outcome was not focused on short-term engagement with the notification but rather longer-term behavior change, such as improved *weekly* mood, activity, or sleep.

Most standardized effect sizes within this paper fell within the 0.05 to 0.15 range. According to the suggested definitions of *small* and *large* [65], the effect sizes for our interventions are small. However, these definitions of *small* and *large* may not directly apply to the causal effects assessed in MRTs [66]. As MRTs are a relatively new trial design, there are currently no accepted definitions of *large* and *small* [66]. For the 3 MRTs with published effect sizes, the effects sizes were 0.074 [67], 0.2 [66], and 0.1 [66]. The effect sizes within this paper are similar in magnitude to these other works.

Implications

Our principal findings demonstrate that the study interns' current state meaningfully influences their receptivity to mHealth interventions for mental health. Effective mHealth interventions for individuals in stressful work environments must consider timing notification delivery based on recent real-time data. Delivering notifications when previous measurements of mood, sleep, and activity are low—when improvement is needed—benefits mood and behavior. However, delivering notifications when those variables are high, negatively impacts mood and behavior.

mHealth interventions aiming to increase mood, activity, and sleep can be improved based on these findings. An improved mHealth intervention for increasing mood would deliver notifications (of any type) only when the user's previous week's average daily mood is below 7 and sends nothing when previous mood is at or above 7. Similarly, for activity, an improved intervention would deliver activity notifications only when the user's previous week's average activity is below 10,614 steps and delivers nothing otherwise. For sleep, an improved intervention would deliver sleep notifications only when the user's previous week's average daily sleep duration is below 6.9 hours. These improved interventions are based upon our single trial, with small effect sizes. There is potential for larger effects through further intervention optimization and using different intervention groups in conjunction with each other. Consistent with the multiphase optimization strategy (MOST) framework [68,69], these suggested interventions should be further refined and evaluated in additional studies and confirmatory trials before being used broadly.

Study Strengths

Through the MRT design and repeatedly randomizing interns throughout the trial, we were able to assess causal effect moderation by time-varying measurements. Our large sample size (1565 interns) allowed us to detect the moderators of interest. The relatively long duration of the study (6 months) demonstrated that our conclusions are valid beyond the first few weeks and months of the study (we analyzed how treatment effects vary over time in the Additional Analyses of Multimedia Appendix 1). Our study focused on medical interns, which provided a unique opportunity to assess the efficacy of mHealth interventions on wellness during life and work stressors. There were also advantages of our analytic approach. First, the use of the multicategorical extension of the weighted and centered least-squares estimator allowed us to unbiasedly assess the causal effect moderation. Second, our imputation method allowed us to cope with missing data without requiring strong assumptions.

Limitations

The primary outcome for the study, mood valence, was self-reported. Self-reported outcomes may be less reliable and valid compared with objective measurements [70,71]. In addition, because of user nonresponse, missing data is a common issue with self-reported outcomes collected over an extended period [59]. In future studies, developing and using a passively collected objective measurement of depression could be beneficial for improving objectivity and reducing missing data.

The main findings of the IHS MRT are partially sensitive to the imputation method used for overcoming missing data (see Missing Data and Sensitivity Analyses in Multimedia Appendix 1). The conclusions of the primary aim and secondary aim 2 are not sensitive. The conclusion of secondary aim 1 (the negative moderation of the activity notification effect by previous step count), however, is sensitive to the imputation method.

The results of the IHS MRT may not extrapolate to other populations because medical interns are different from the general population in the average education level and socioeconomic status. Within the population of medical interns, sampling bias may still exist as the study's interns self-selected into the study, as opposed to being randomly sampled. This self-selection bias may cause the study interns to be different from the general population of interns. For example, because they were motivated to participate in the study, they may also be motivated to change their behavior. Although it is difficult to show self-selection unbiasedness, the bias may be mitigated because a large percentage of interns agreed to participate in the study (40.78%), and the study interns' baseline information closely resembles that of the general medical intern population [29].

Daily work schedules were not reliably measured in this study. Previous studies [63,64] have found that mHealth message effectiveness varies between weekdays and weekends, suggesting that future studies should assess work schedule as a potential moderator.

Measurements of app engagement could provide further insights into how these notifications are promoting behavior change. For example, after receiving a notification, users may have an increased rate of opening the app and viewing their historical data displays. Unfortunately, the app does not currently collect data on app access and app clicks. It also does not measure a user's interactions with the notification messages. Including these capabilities in future versions of the app would be useful.

We did not have message tailoring in this study. Currently, the message framing and wording was the same, no matter the intern's current behavior. The messages (see Table 1) are framed toward improving mood, sleep, and activity. This framing may be frustrating to an intern who already has a high mood or sufficient sleep or activity. Tailoring the wording of the messages [72,73] could potentially eliminate the negative effect of messages when previous mood, sleep, or activity is high (eg, providing a reminder message as opposed to an unnecessary ability-focused message).

There were also a couple technological errors that occurred throughout the trial. There were 8 days (of the 182 total days) when, because of server issues, no notifications were sent to any subject. In addition, the weekly randomization to a notification category occurred without replacement, as opposed to with replacement as originally intended.

Future Iterations of the Intern Health Study

The IHS is an annual study that continues each year with a new cohort of interns [30]. Consistent with the MOST framework [68,69], this provides multiple trial phases to continually update, optimize, and test interventions and provides confirmation of findings from previous cohorts. Starting in the fall of 2019, we will run another study to test new hypotheses with improved interventions. Using the results and conclusions drawn from this study, in 2019, we plan to introduce tailored messages that are tailored based on an intern's previous mood, activity, and sleep [72,73]. For people with high previous measurements, the messages will be framed toward maintenance of healthy behavior, not improvement. The cutoffs that define *high* and *low* scores will be based on data collected from the 2018 study. We also plan to improve the missing data protocol and incentive structure to reduce the frequency of missing data. We will collect work schedule information to compare message efficacy between work days and days off. Finally, in addition to providing notifications on the phone lock screen, we will also show the notifications within the app to give interns more opportunities to read them.

Conclusions

Overall, our study demonstrates the importance of real-time moderators for the development of high-quality mHealth interventions, especially for individuals in stressful work environments. There were times when the notifications were beneficial and times when the notifications were harmful to the study participants. Developers of mHealth interventions are encouraged to think deeply about the delivery of interventions and how real-time variables can be used to inform the timing of intervention delivery. The MRT design allowed us to discover real-time moderators and may be useful for other app developers also aiming to learn when to deliver notification messages.

In addition to the research aims for future iterations of the IHS, assessing the value of mHealth interventions and delivery timing in other highly stressed populations is beneficial for understanding the generality of these results. Future MRTs should also examine the efficacy of mHealth content (eg, content focused on motivation, ability, or triggers) incorporated into other app features for behavior change. In this regard, developing mHealth intervention features beyond push notifications (eg, integrating ability-focused mindfulness exercises) could provide a greater overall benefit.

Acknowledgments

The authors would like to thank Dr Inbal Nahum-Shani, Dr Daniel Almirall, Jessica Liu, Mengjie Shi, and Dr Stephanie Carpenter for valuable feedback on this work. ZW and SS acknowledge support by the National Institutes of Health (NIH) grant R01MH101459 and an investigator award from the Michigan Precision Health Initiative. ZW acknowledges the support of the NIH National Cancer Institute award P30CA046592 through the Cancer Center Support Grant Development Funds from the Rogel Cancer Center and start-up funds from Michigan Institute for Data Science. AT acknowledges the support of the National Science Foundation via CAREER grant IIS-1452099. SS acknowledges the support of the American Foundation for Suicide Prevention (LSRG-0-059-16).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional analyses and statistical details.
[PDF File (Adobe PDF File), 555 KB - jmir_v22i3e15033_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 665 KB - jmir_v22i3e15033_app2.pdf]

References

1. World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2020-01-31]
2. Greenberg PE, Fournier A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry 2015 Mar;76(2):155-162 [FREE Full text] [doi: 10.4088/JCP.14m09298] [Medline: 25742202]
3. Hawkins JD, Jenson JM, Catalano R, Fraser MW, Botvin GJ, Shapiro V, et al. Unleashing the power of prevention. NAM Perspectives 2015;5(6):1343-1344. [doi: 10.31478/201506c]
4. Tennant C. Work-related stress and depressive disorders. J Psychosom Res 2001 Dec;51(5):697-704. [doi: 10.1016/s0022-3999(01)00255-0] [Medline: 11728512]
5. Baucom KJ, Queen TL, Wiebe DJ, Turner SL, Wolfe KL, Godbey EI, et al. Depressive symptoms, daily stress, and adherence in late adolescents with type 1 diabetes. Health Psychol 2015 May;34(5):522-530 [FREE Full text] [doi: 10.1037/hea0000219] [Medline: 25798545]
6. Everett KD, Brantley PJ, Sletten C, Jones GN, McKnight GT. The relation of stress and depression to interdialytic weight gain in hemodialysis patients. Behav Med 1995;21(1):25-30. [doi: 10.1080/08964289.1995.9933739] [Medline: 7579772]
7. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
8. Zhang S, Elhadad N. Factors contributing to dropping-out in an online health community: static and longitudinal analyses. AMIA Annu Symp Proc 2016;2016:2090-2099 [FREE Full text] [Medline: 28269969]
9. Heckman BW, Mathew AR, Carpenter MJ. Treatment burden and treatment fatigue as barriers to health. Curr Opin Psychol 2015 Oct 1;5:31-36 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.004] [Medline: 26086031]
10. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
11. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 2014;13(3):13-17. [doi: 10.1109/MPRV.2014.46]
12. Riley WT. Theoretical models to inform technology-based health behavior interventions. In: Marsch LA, Lord SE, Dallery J, editors. Behavioral Healthcare and Technology: Using Science-Based Innovations to Transform Practice. Oxford, United Kingdom: Oxford University Press; 2014:13-24.
13. Wang L, Miller LC. Just-in-the-moment adaptive interventions (JITAI): a meta-analytical review. Health Commun 2019 Oct 5:1-14. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
14. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
15. Ahmedani BK, Crotty N, Abdulhak MM, Ondersma SJ. Pilot feasibility study of a brief, tailored mobile health intervention for depression among patients with chronic pain. Behav Med 2015;41(1):25-32. [doi: 10.1080/08964289.2013.867827] [Medline: 24313728]
16. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
17. Rizvi SL, Dimeff LA, Skutch J, Carroll D, Linehan MM. A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. Behav Ther 2011 Dec;42(4):589-600. [doi: 10.1016/j.beth.2011.01.003] [Medline: 22035988]
18. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013 Jan 7;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
19. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Dec 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
20. Newman MG, Przeworski A, Consoli AJ, Taylor CB. A randomized controlled trial of ecological momentary intervention plus brief group therapy for generalized anxiety disorder. Psychotherapy (Chic) 2014 Jul;51(2):198-206 [FREE Full text] [doi: 10.1037/a0032519] [Medline: 24059730]
21. 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]

22. Smyth JM, Heron KE. Is Providing Mobile Interventions 'Just-in-time' Helpful? An Experimental Proof of Concept Study of Just-in-time Intervention for Stress Management. In: Proceedings of the 2016 IEEE Wireless Health. 2016 Presented at: WH'16; October 25-27, 2016; Bethesda, MD, USA p. 1-7. [doi: 10.1109/WH.2016.7764561]
23. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
24. Kalmbach DA, Fang Y, Arnedt JT, Cochran AL, Deldin PJ, Kaplin AI, et al. Effects of sleep, physical activity, and shift work on daily mood: a prospective mobile monitoring study of medical interns. J Gen Intern Med 2018 Jun;33(6):914-920 [FREE Full text] [doi: 10.1007/s11606-018-4373-2] [Medline: 29542006]
25. Akerstedt T. Psychosocial stress and impaired sleep. Scand J Work Environ Health 2006 Dec;32(6):493-501 [FREE Full text] [Medline: 17173205]
26. Lallukka T, Sarlio-Lähteenkorva S, Roos E, Laaksonen M, Rahkonen O, Lahelma E. Working conditions and health behaviours among employed women and men: the Helsinki Health Study. Prev Med 2004 Jan;38(1):48-56. [doi: 10.1016/j.ypmed.2003.09.027] [Medline: 14672641]
27. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
28. Ströhle A. Physical activity, exercise, depression and anxiety disorders. J Neural Transm (Vienna) 2009 Jul;116(6):777-784. [doi: 10.1007/s00702-008-0092-x] [Medline: 18726137]
29. Mata DA, Ramos MA, Bansal N, Khan R, Guille C, di Angelantonio E, et al. Prevalence of depression and depressive symptoms among resident physicians: a systematic review and meta-analysis. J Am Med Assoc 2015 Dec 8;314(22):2373-2383 [FREE Full text] [doi: 10.1001/jama.2015.15845] [Medline: 26647259]
30. The Sen Lab. Intern Health Study URL: https://www.srijan-sen-lab.com/intern-health-study [accessed 2020-01-31]
31. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002 Oct;59(10):877-883. [doi: 10.1001/archpsyc.59.10.877] [Medline: 12365874]
32. Fitbit. Fitbit Charge 2 User Manual URL: https://staticcs.fitbit.com/content/assets/help/manuals/manual_charge_2_en_US.pdf [accessed 2020-01-31]
33. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
34. Apple Developer. ResearchKit URL: https://developer.apple.com/researchkit/ [accessed 2020-01-31]
35. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Mar;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
36. Foreman AC, Hall C, Bone K, Cheng J, Kaplin A. Just text me: using SMS technology for collaborative patient mood charting. J Particip Med 2011;3:e45 [FREE Full text]
37. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Oct;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
38. Fitbit. Welcome. It's an API URL: https://www.fitbit.com/dev [accessed 2020-01-31]
39. Korotitsch WJ, Nelson-Gray RO. An overview of self-monitoring research in assessment and treatment. Psychol Assess 1999;11(4):415-425. [doi: 10.1037/1040-3590.11.4.415]
40. Fogg BJ. A Behavior Model for Persuasive Design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive'09; April 26 - 29, 2009; Claremont, California, USA p. 1-7. [doi: 10.1145/1541948.1541999]
41. de Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: systematic review. J Med Internet Res 2014 Mar 25;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]
42. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS One 2017;12(1):e0169162 [FREE Full text] [doi: 10.1371/journal.pone.0169162] [Medline: 28046034]
43. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167 [FREE Full text] [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]
44. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 1;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
45. Suggs S, Blake H, Bardus M, Lloyd S. Effects of text messaging in addition to emails on physical activity among university and college employees in the UK. J Health Serv Res Policy 2013 May;18(1 Suppl):56-64. [doi: 10.1177/1355819613478001] [Medline: 27552780]
46. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change. Second Edition. New York, US: The Guilford Press; 2002.

47. Miller WR, Rose GS. Toward a theory of motivational interviewing. Am Psychol 2009 Oct;64(6):527-537 [FREE Full text] [doi: 10.1037/a0016830] [Medline: 19739882]
48. Resnicow K, McMaster F. Motivational Interviewing: moving from why to how with autonomy support. Int J Behav Nutr Phys Act 2012 Mar 2;9:19 [FREE Full text] [doi: 10.1186/1479-5868-9-19] [Medline: 22385702]
49. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Upper Saddle River, New Jersey: Prentice Hall; 1986.
50. DiClemente CC, Marinilli AS, Singh M, Bellino LE. The role of feedback in the process of health behavior change. Am J Health Behav 2001;25(3):217-227. [doi: 10.5993/ajhb.25.3.8] [Medline: 11322620]
51. Harkin B, Webb TL, Chang BP, Prestwich A, Conner M, Kellar I, et al. Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. Psychol Bull 2016 Mar;142(2):198-229. [doi: 10.1037/bul0000025] [Medline: 26479070]
52. Hockey R. The Psychology of Fatigue: Work, Effort and Control. Cambridge, UK: Cambridge University Press; 2013.
53. Firebase - Google. Firebase Cloud Messaging URL: https://firebase.google.com/docs/cloud-messaging [accessed 2020-01-31]
54. Qualtrics XM // The Leading Experience Management Software. URL: https://www.qualtrics.com [accessed 2020-01-31]
55. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. J Am Stat Assoc 2018;113(523):1112-1121 [FREE Full text] [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]
56. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. Berkeley Symp Math Stat Probab 1967;1:221-233 [FREE Full text]
57. Halekoh U, Højsgaard S, Yan J. The R package GEEPACK for generalized estimating equations. J Stat Soft 2006;15(2):1-11. [doi: 10.18637/jss.v015.i02]
58. Little RJ, Rubin DB. Statistical Analysis with Missing Data. Second Edition. Hoboken, New Jersey: Wiley-interscience; 2002.
59. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Stat Methods Med Res 2014 Oct;23(5):440-459. [doi: 10.1177/0962280213476378] [Medline: 23427225]
60. Shortreed SM, Laber E, Scott Stroup T, Pineau J, Murphy SA. A multiple imputation strategy for sequential multiple assignment randomized trials. Stat Med 2014 Oct 30;33(24):4202-4214 [FREE Full text] [doi: 10.1002/sim.6223] [Medline: 24919867]
61. Grund S, Robitzsch A, Luedtke O. mitml: Tools for Multiple Imputation in Multilevel Modeling. URL: https://rdrr.io/cran/mitml/
62. Pielot M, Cardoso B, Katevas K, Serrà J, Matic A, Oliver N. Beyond interruptibility: predicting opportune moments to engage mobile phone users. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017;1(3):1-25. [doi: 10.1145/3130956]
63. Sarker H, Sharmin M, Ali AA, Rahman MM, Bari R, Hossain SM, et al. Assessing the availability of users to engage in just-in-time intervention in the natural environment. Proc ACM Int Conf Ubiquitous Comput 2014;2014:909-920 [FREE Full text] [doi: 10.1145/2632048.2636082] [Medline: 25798455]
64. 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]
65. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. Mahwah, New Jersey, United States: Lawrence Erlbaum Associates; 1988.
66. Luers B, Klasnja P, Murphy S. Standardized effect sizes for preventive mobile health interventions in micro-randomized trials. Prev Sci 2019 Jan;20(1):100-109 [FREE Full text] [doi: 10.1007/s11121-017-0862-5] [Medline: 29318443]
67. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 3;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
68. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
69. The Methodology Center. Overview of MOST URL: http://www.methodology.psu.edu/ra/most/research/ [accessed 2020-01-31]
70. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Dec 6;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
71. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health 2007;10(Suppl 2):S94-105 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00272.x] [Medline: 17995479]
72. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
73. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jul;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]

Abbreviations

EMA: ecological momentary assessment
IHS: Intern Health Study
JITAI: just-in-time adaptive intervention
MRT: micro-randomized trial
mHealth: mobile health
MOST: multiphase optimization strategy
NIH: National Institutes of Health
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 13.06.19; peer-reviewed by J Kim, P Chow, E Yom-Tov; comments to author 01.09.19; revised version received 26.10.19; accepted 15.12.19; published 31.03.20.

Please cite as:
NeCamp T, Sen S, Frank E, Walton MA, Ionides EL, Fang Y, Tewari A, Wu Z
Assessing Real-Time Moderation for Developing Adaptive Mobile Health Interventions for Medical Interns: Micro-Randomized Trial
J Med Internet Res 2020;22(3):e15033
URL: http://www.jmir.org/2020/3/e15033/
doi:10.2196/15033
PMID:32229469

Original Paper

Adherence to Prescribed E-Diary Recording by Patients With Seasonal Allergic Rhinitis: Observational Study

Marco Di Fraia[1,2], MD; Salvatore Tripodi[3,4,5], MD; Stefania Arasi[1,6], MD, PhD; Stephanie Dramburg[1], MD; Sveva Castelli[1], MD; Danilo Villalta[7], MD; Francesca Buzzulini[7], MD; Ifigenia Sfika[3], MD; Valeria Villella[3], MD; Ekaterina Potapova[1], MSc; Serena Perna[1], MSc; Maria Antonia Brighetti[8], PhD; Alessandro Travaglini[8], MA; Pierluigi Verardo[9], MA; Simone Pelosi[5], MA; Anna Maria Zicari[2], MD; Paolo Maria Matricardi[1], MD

[1]Department of Pediatric Pulmonology, Immunology and Intensive Care Medicine, Charité University Medicine, Berlin, Germany
[2]Department of Pediatrics, Sapienza University of Rome, Rome, Italy
[3]Pediatric Allergology Unit, Sandro Pertini Hospital, Rome, Italy
[4]Allergology Service, Policlinico Casilino, Rome, Italy
[5]TPS Production, Rome, Italy
[6]Pediatric Allergology Unit, Department of Pediatric Medicine, Bambino Gesù Children's Research Hospital, Rome, Italy
[7]Department of Immunology-Allergy, Santa Maria degli Angeli Hospital, Pordenone, Italy
[8]Department of Biology, University of Rome Tor Vergata, Rome, Italy
[9]Center of Aerobiology, Agenzia Regionale per la Protezione Ambientale, Pordenone, Italy

Corresponding Author:
Paolo Maria Matricardi, MD
Department of Pediatric Pulmonology, Immunology and Intensive Care Medicine
Charité University Medicine
Augustenburger Platz 1
Berlin, 13353
Germany
Phone: 49 30 450 566 406
Email: paolo.matricardi@charite.de

Abstract

Background: Complete diagnosis and therapy of seasonal allergic rhinoconjunctivitis require evidence that exposure to the sensitizing pollen triggers allergic symptoms. Electronic clinical diaries, by recording disease severity scores and pollen exposure, can demonstrate this association. However, patients who spontaneously download an e-diary app show very low adherence to their recording.

Objective: The objective of our study was to assess adherence of patients with seasonal allergic rhinitis to symptom recording via e-diary explicitly prescribed by an allergist within a blended care approach.

Methods: The @IT-2020 project is investigating the diagnostic synergy of mobile health and molecular allergology in patients with seasonal allergic rhinitis. In the pilot phase of the study, we recruited Italian children (Rome, Italy) and adults (Pordenone, Italy) with seasonal allergic rhinitis and instructed them to record their symptoms, medication intake, and general conditions daily through a mobile app (Allergy.Monitor) during the relevant pollen season.

Results: Overall, we recruited 101 Italian children (Rome) and 93 adults (Pordenone) with seasonal allergic rhinitis. Adherence to device use slowly declined during monitoring in 3 phases: phase A: first week, ≥1267/1358, 90%; phase B: second to sixth week, 4992/5884, 80% to 90%; and phase C: seventh week onward, 2063/2606, 70% to 80%. At the individual level, the adherence assessed in the second and third weeks of recording predicted with enough confidence (Rome: Spearman $\rho=0.75$; $P<.001$; Pordenone: $\rho=0.81$; $P<.001$) the overall patient adherence to recording and was inversely related to postponed reporting ($\rho=-0.55$; $P<.001$; in both centers). Recording adherence was significantly higher during the peak grass pollen season in Rome, but not in Pordenone.

Conclusions: Adherence to daily recording in an e-diary, prescribed and motivated by an allergist in a blended care setting, was very high. This observation supports the use of e-diaries in addition to face-to-face visits for diagnosis and treatment of seasonal allergic rhinitis and deserves further investigation in real-life contexts.

(J Med Internet Res 2020;22(3):e16642) doi:10.2196/16642

KEYWORDS

mobile health; e-Diary; precision medicine; pollen; seasonal allergic rhinitis; blended care

Introduction

Background

Seasonal allergic rhinoconjunctivitis (SAR) affects patients exposed to pollens to which they are sensitized. The etiological diagnosis and therapy of SAR require a demonstration that exposure to the sensitizing pollen triggers allergic symptoms [1]. Objectively, this link is established by a positive outcome to nasal allergen provocation tests [2] or allergen exposure in pollen chambers [3]. Unfortunately, both these tests are costly and time consuming and are mostly used in clinical trials [4]. In clinical daily life, a causality between pollen exposure and symptoms is often assessed by a careful *retrospective* clinical history [5]. However, recall biases make the diagnosis based on retrospective data somewhat imprecise, especially in patients apparently sensitized to multiple pollens that share the same pollination periods [6], which is a frequent setting in Mediterranean countries [7].

This diagnostic problem can be partially solved through a *prospective* clinical history, based on the patient's daily recording of symptoms and medication intake in a clinical diary [8]. Indeed, the trajectories of daily symptom scores or a combined symptom and medication score (CSMS) are free from recall bias and can be matched with daily concentration counts, obtained in parallel, of the potentially eliciting pollen sources [9,10]. While traditional and time-consuming clinical diaries on paper records are rarely used, electronic clinical diaries (e-diaries) have become increasingly prevalent [11,12]. E-diaries are apps consisting of short questionnaires filled in daily by the patient, usually on his or her mobile phone or tablet computer [11-13]. Recording e-diaries is easy and quick, and the software automatically provides daily scores, time trajectories, and descriptive reports [8,10-16].

Several e-diaries are available for pollen allergies in European countries, and some of them have also been used in trials or observational studies [8,10,12,14-19]. In most of the study settings, the app was directly downloaded by the patients, with no or only occasional intervention by their allergist [14-17]. The observational studies were characterized by large population size (more than 9000 participants) and big datasets (112,054 registered visual analog scale [VAS] data) [14], balanced by a poor mean adherence (<10%) to daily recording [14,15].

Objective

We hypothesized that the patients' adherence to recording of e-diaries would be significantly increased if the rationale and the use of the e-diary were personally explained by an allergist to the patient (blended approach). To test this hypothesis, we examined the rate and cofactors of adherence to recording of an e-diary among Italian patients with SAR participating in the @IT-2020 project, a study of combined molecular diagnostics and mobile health for allergen immunotherapy in patients with SAR.

Methods

@IT-2020 Project

The pilot study of the @IT-2020 project was carried out in 2 Italian centers differing significantly in terms of environmental setting and patient characteristics.

Climate and Study Area

Pordenone, Italy, is a city with about 50,000 inhabitants, which extends over an area of 38 km^2 [20]. Pordenone is 600 km north of Rome and the territory is located in northeastern Italy, about 50 km from the Adriatic Sea, in the Po-Veneto plain south of the Carnic Pre-Alps, in the continental biogeographical region [21]. It has a mean annual temperature of 13.1°C and mean rainfall of 1292 mm [22].

Rome, Italy, has 3 million inhabitants in an area of almost 1300 km^2 [20] and is 20 km from the Tyrrhenian Sea. Rome is located in the Mediterranean biogeographical region [21] with mean annual temperature of 15.7°C and mean rainfall of 798 mm (Rome Monte Mario) [22].

Study Population

Between November 2016 and February 2017, we recruited 101 children aged 10 to 18 years at Ospedale Sandro Pertini in Rome and 93 adults aged over 18 years at Ospedale Santa Maria degli Angeli in Pordenone. Criteria for eligibility were (1) being followed up for at least one year for allergic rhinoconjunctivitis (objectively confirmed by skin prick tests or in vitro immunoglobulin E tests, or both) due to outdoor aeroallergens (pollen or spores), (2) residing within 30 km of the aerobiological station of the study center, (3) having no intention to change residence in the 6 months after recruitment, and (4) being able to use a mobile phone (by the patient or the patient's parents). Exclusion criteria were (1) previous allergen immunotherapy for any outdoor allergen, and (2) any other severe nonatopic chronic disease. All participants (in the case of children, their parents or guardians) provided informed written consent to the clinical investigations.

Study Design

Recruited patients underwent a first clinical assessment (T0), including clinical questionnaires, during which they were instructed on the use of the Allergy.Monitor (Technology Project and Software [TPS] Production, Rome, Italy) mobile app to monitor their symptoms and medication intake during the following study period. According to the timing of retrospective symptoms and skin prick test results, participants were assigned an individual monitoring period during the suspected high season of the putative eliciting pollen. During this period, participants were asked to monitor their eye, nose, and lung symptoms, as well as their effect on daily activities and daily medication intake, and report them via Allergy.Monitor. After the monitoring period, all participants underwent a second clinical assessment (T1), including a repetition of the initial clinical questionnaires focused on the past pollen season, internationally validated by the International Study of Asthma and Allergies in Childhood [23], the Allergic Rhinitis and its Impact on Asthma (ARIA) initiative [24,25], and the Global Initiative for Asthma [26]. The study design and procedures had been approved by the ethics committee of each participating center.

Skin Prick Tests

Skin prick tests were performed using a standard panel of commercial extracts (ALK-Abelló, Milan, Italy) of outdoor and indoor aeroallergens (*Alternaria*, Bermuda grass, birch, cat dander, cypress, dog dander, hazel, house dust mite, mugwort, olive tree, plane tree, ragweed, Russian thistle, timothy grass, and pellitory-of-the-wall). Histamine 0.1 mg/mL and glycerol solution were used as positive and negative controls, respectively. Morrow Brown needles were used to prick the skin and the wheal reactions were read after 15 minutes. A wheal equal to or greater than 3 mm after subtraction of the negative control was regarded as positive.

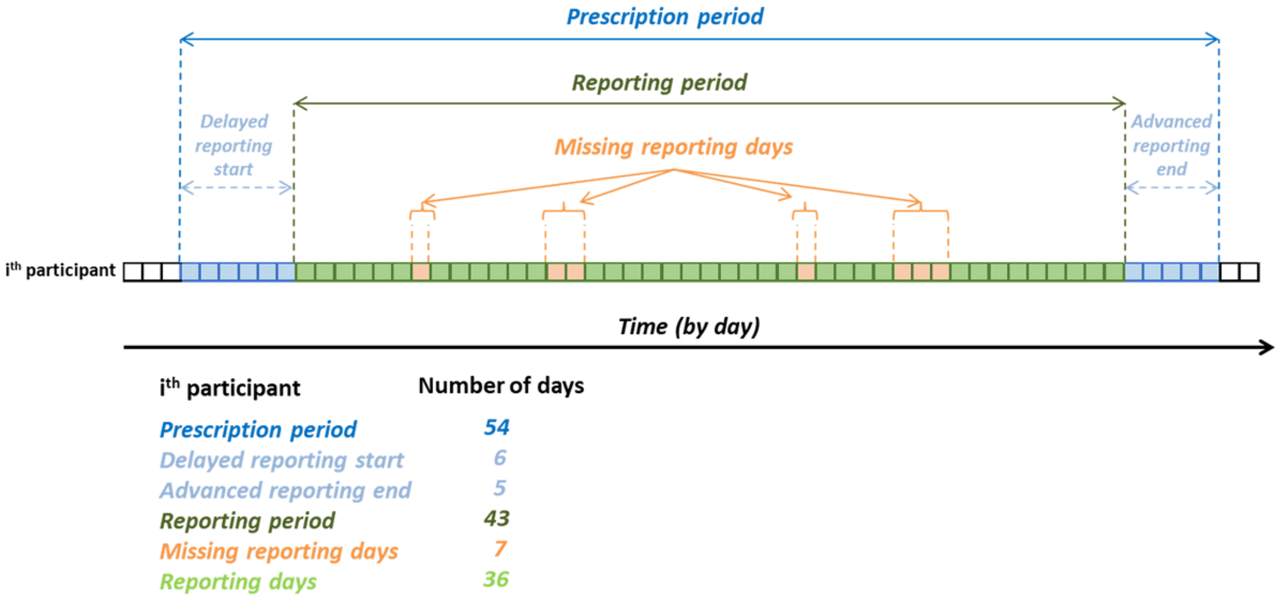
Pollen Counts and Pollen Periods

The pollen count data, acquired from March 1 to September 30, 2016, were provided by the pollen stations of Rome (Tor Vergata University) and Pordenone (Agenzia Regionale per la Protezione dell’Ambiente del Friuli Venezia Giulia). Pollen was collected using a VPPS 2000 pollen sampler (Lanzoni srl, Bologna, Italy), and data were acquired as reported in Standard UNI CEN/TS 16868:2015 [27]. Pollen periods were determined (1) according to the 2017 European Academy of Allergy and Clinical Immunology (EAACI) position paper on pollen exposure times [28] (EAACI criteria), and (2) by adapting these criteria to the pollen situation in Italy (local criteria).

Allergy.Monitor

Allergy.Monitor is a mobile app designed for daily reporting of symptoms and medication intake related to allergic rhinitis or asthma. In this study, medical doctors, on the basis of clinical history, defined a time frame (*prescription period*; Figure 1) for each patient, in which he or she was encouraged to fill in a daily questionnaire regarding his or her symptoms and medication intake. The system offers a bidirectional interaction between physician and patient via email, chat, and text messaging. Patients not entering their data for 2 consecutive days received an automatic alert message on their mobile phone or by email; after 4 days without reporting, the alert was followed by a phone call from the physician or nurse. The patient could insert data referring to 1 day only on the same day or on the following one (*postponed reporting*). For each participant, *adherence to prescription* was calculated as the number of actual reporting days / prescription period ×100; *adherence* was calculated as the number of actual reporting days / reporting period ×100; and postponed reporting was calculated as the number of postponed reporting days / actual reporting days ×100.

Figure 1. Graphical representation of definitions used in this study. The box line represents the monitored period (each box is a specific day) of a hypothetical participant. In this example the medical doctor, according to the individual participant’s clinical history, invited the patient to fill in the e-diary questionnaire for 54 days (prescription period). The patient started to record symptoms 6 days after the prescribed beginning day (delayed reporting start) and finished recording symptoms 5 days before the prescribed ending day (advanced reporting end). Thus, the reporting period lasted 43 days, during which the participant did not fill in the e-diary questionnaire for 7 days (missing reporting days). Overall, the participant filled in the e-diary questionnaire for 36 days (reporting days).



Symptom and Medication Scores

We used the following symptom and medication scores in this study: Rhinoconjunctivitis Total Symptom Score (RTSS; score 0-18) [29]; CSMS (score 0-6) [30]; and VAS (score 0-10) [31].

RTSS and CSMS were calculated automatically by the Allergy.Monitor app, for every reporting day, on the basis of 4 questions on nasal symptoms (sneezing, rhinorrhea, nasal pruritus, nasal congestion), 2 on ocular symptoms (itchy eyes, watery eyes), and 3 questions on medication intake (antihistaminic drugs, local corticosteroids, systemic corticosteroids). The severity of each of the symptoms was also measured by the patient using 4 different emoticons, each representing a distinct severity grade (no symptoms, mild, moderate, or severe). Overall severity was also measured by a VAS in response to the question “How do you feel in relation to your allergic symptoms today?”

Statistics

We summarized data as numbers (n) and frequencies (%) if they were categorical and as mean or median and standard deviation or interquartile range if quantitative. We examined all described analyses for each of the study centers (Rome and Pordenone). We evaluated the prevalence of atopic sensitization (skin prick test ≥3 mm) to airborne allergens. For every pollen period considered, we calculated adherence values (see above for definition) for each participant and compared their means using a nonparametric Friedman test for repeated measures. We adjusted the *P* of multiple comparison by the Bonferroni correction. We studied adherence trends over time considering the time (in days) that had passed since the first day of the reporting period. We used the Spearman rank correlation coefficient to investigate the relationship between total adherence (%), postponed reporting (%), and adherence achieved between the seventh and the 21st reporting day (%). Mean CSMS scores by time were computed for the local whole season. We considered $P<.05$ to be statistically significant. Statistical analyses were performed with R version 3.2.3 (R Foundation).

Results

Study Population

Overall, 101 children (Rome) and 93 adults (Pordenone) with mean (SD) ages of 13.7 (SD 2.8) and 34.3 (14.4), respectively, met the inclusion criteria (Table 1). Male sex was slightly more frequent in both populations: 62.4% (63/101) for Rome and 56% (52/93) for Pordenone. At T0, according to the ARIA questionnaire, the population in Pordenone was characterized by a higher prevalence of moderate to severe (intermittent and persistent) allergic rhinitis than in Rome (90/93, 97% vs 51/101, 50.5%, respectively). At T1, this difference was less evident (64/75, 85% vs 68/91, 75%). The prevalence of allergic asthma was similar in both groups (Rome: 28/101, 27.7%; Pordenone: 24/93, 26%), whereas the Rome population seemed to be more affected by oral allergy syndrome, urticaria, atopic dermatitis, and anaphylaxis (Table 1). Grass pollen was the most relevant allergen in both study populations. Positive skin prick test reactions to olive tree and cypress were more frequent in Rome, while sensitization to birch was more prevalent in Pordenone. Sensitization to indoor allergens was equally prevalent in both populations.

Table 1. Characteristics of the study population.

Characteristic	Rome (n=101)	Pordenone (n=93)
Males, n (%)	63 (62.4)	52 (56)
Age (years), mean (SD)	13.7 (2.8)	34.3 (14)
Allergic rhinitis, n (%)		
Age at onset (years), median (IQR[a])	6 (4-8)	15 (8-22)
ARIA[b] classification at T0		
Mild intermittent	19 (18.8)	1 (1)
Mild persistent	31 (30.7)	2 (2)
Moderate to severe intermittent	11 (10.9)	17 (18)
Moderate to severe persistent	40 (39.6)	73 (79)
ARIA classification at T1[c], n (%)		
Mild intermittent	6 (7)	2 (3)
Mild persistent	17 (19)	9 (12)
Moderate to severe intermittent	4 (4)	13 (17)
Moderate to severe persistent	64 (70)	51 (68)
Other allergic comorbidities, n (%)		
Allergic asthma	28 (27.7)	24 (26)
Oral allergic syndrome	32 (32.3)	23 (25)
Urticaria or angioedema	19 (19.2)	8 (9)
Atopic dermatitis	28 (28.3)	11 (12)
Gastrointestinal disorders	4 (4.0)	1 (1)
Anaphylaxis episode	10 (10.1)	1 (1)
Other	5 (5.1)	2 (2)

[a] IQR: interquartile range.

[b] ARIA: Allergic Rhinitis and its Impact on Asthma.

[c] Study population at T1: Rome, n=91; Pordenone, n=75.

Pollen Periods

The graphical representation of grass pollen counts (grains/m^3) highlighted differences between the 2 cities. The maximum grass pollen count in Rome (199 grains/m^3) was higher than in Pordenone (145 grains/m^3), and the grass pollination period was longer in Rome. Grass pollen periods in 2016 differed significantly if calculated according to EAACI criteria or local criteria (Table 2). While we used EAACI criteria for their reproducibility and standardization, the application of locally adapted criteria resulted in shorter and less fragmented periods.

Table 2. Grass pollen period criteria and duration, by study center.

Pollen period	Criteria	Rome (n=101)			Pordenone (n=93)		
		No. of time intervals[a]	Cumulative duration (days), n	Adherence mean (95% CI) (%)	No. of time intervals[a]	Cumulative duration (days), n	Adherence mean (95% CI) (%)
EAACI whole season[b]	5 days (out of 7 consecutive days) each with ≥3 pollen grains/m^3 and with a sum of ≥30 pollen grains/m^3	2	132	82.1 (79.3-84.9)	2	132	86.3 (83.5-89.2)
Local whole season[c]	5 days (out of 7 consecutive days) each with ≥10 pollen grains/m^3 and with a sum of ≥100 pollen grains/m^3	1	97	81.9 (79.1-84.8)	2	72	86.2 (83.1-89.2)
EAACI peak season[b]	3 consecutive days, each with ≥50 pollen grains/m^3	5	28	81.0 (77.6-84.4)	2	13	90.4 (87.5-93.3)
Local peak season[c]	3 days (out of 5 consecutive days) each with ≥50 pollen grains/m^3	1	55	81.0 (78.0-84.1)	2	24	90.3 (87.6-92.9)
EAACI high days[b]	Days with at least 50 pollen grains/m^3	18	45	80.9 (77.9-84.0)	9	23	89.3 (86.4-92.1)

[a] See Figure 1 for specifications of time periods.

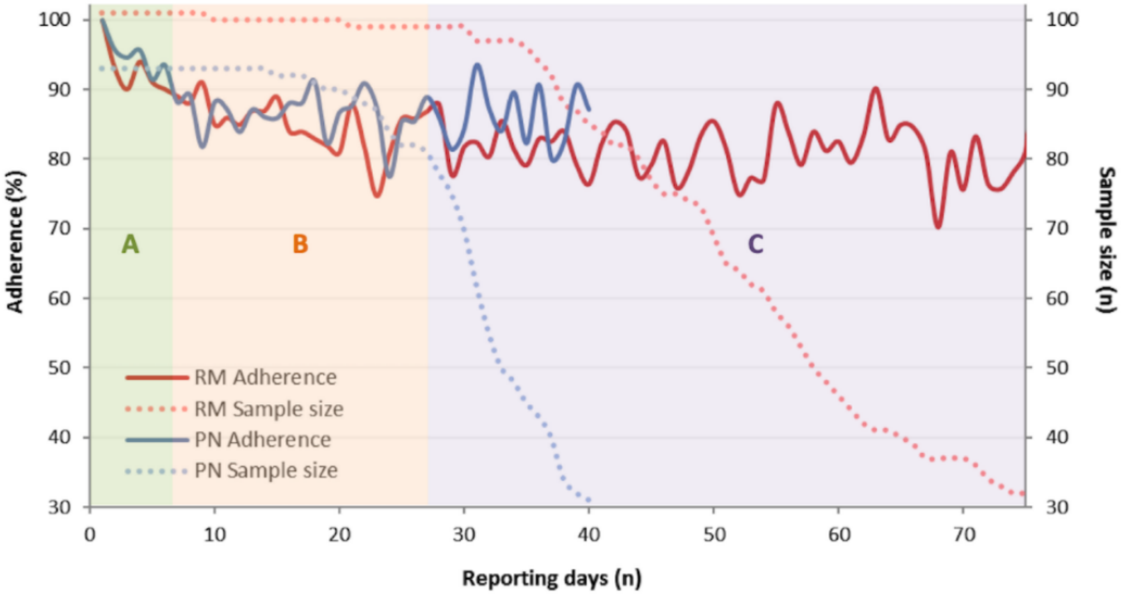
[b] European Academy of Allergy and Clinical Immunology (EAACI) criteria [28].

[c] Adaptation of EAACI criteria to the local scenario.

Adherence to E-Diary Recording

The mean prescription period was longer in Rome than in Pordenone (76.2, 95% CI 70.4-82.0 vs 53.9, 95% CI 50.1-57.7 days, respectively). The pattern was similar for the mean reporting period (Rome: 70.6, 95% CI 64.9-74.4 vs Pordenone: 48.2, 95% CI 44.6-51.7 days) (Figure 2). Mean adherence levels were 85.7% (SD 13.9) in Pordenone and 82.3% (SD 13.7) in Rome. The analysis of mean adherence values by reporting day showed a similar trend for both participating study centers. In Rome, the adherence trend by reporting day displayed 3 different phases: phase A, a first phase of 6 days with an adherence 93.1% (564/606); phase B, a second phase of approximately 40 days, during which the adherence fluctuated around 83.65% (2834/3388); and phase C, a final phase of slowly declining adherence, oscillating around 78.55% (1952/2485). Pordenone's adherence trend by reporting day followed the same evolution for phases A and B. Due to a shorter pollen season and mean prescribed period, we did not investigate phase C in Pordenone (Figure 2).

Figure 2. Adherence (%) by reporting day and study center. It is possible to describe three phases (indicated by light background color): the first phase (A), lasting 6 days, during which adherence fell from 100% to 90%; the second phase (B), lasting approximately 20 days, during which adherence fluctuated until reaching 88%; and the final phase (C), during which it slowly declined to 80%. RM: Rome; PN, Pordenone.



Interestingly, the total adherence was directly proportional to the adherence assessed between the seventh and 21st reporting days (Spearman ρ=0.75; *P*<.001 and ρ=0.81; *P*<.001 for Rome and Pordenone, respectively) (Figure 3) and inversely related, although with less intensity, to postponed reporting (ρ=–0.55; *P*<.001 for both Rome and Pordenone) (Figure 4). In both populations, the mean RTSS, evaluating symptoms of the eyes and nose, showed a parallel trend with the mean VAS scores assessing the general disease-related impairment. Also, the CSMS followed a similar trend but with less distinct variance (Figure 5).

Mean adherence values differed only slightly in Rome during the different pollen periods (Figure 6, part A). By contrast, adherence values were significantly higher in Pordenone during the peak pollen season and the high day (Figure 6, part B).

Figure 3. Correlation between adherence achieved between the seventh and the 21st reporting days and total reporting period adherence, by study center: (A) Rome (n=101); (B) Pordenone (n=93).

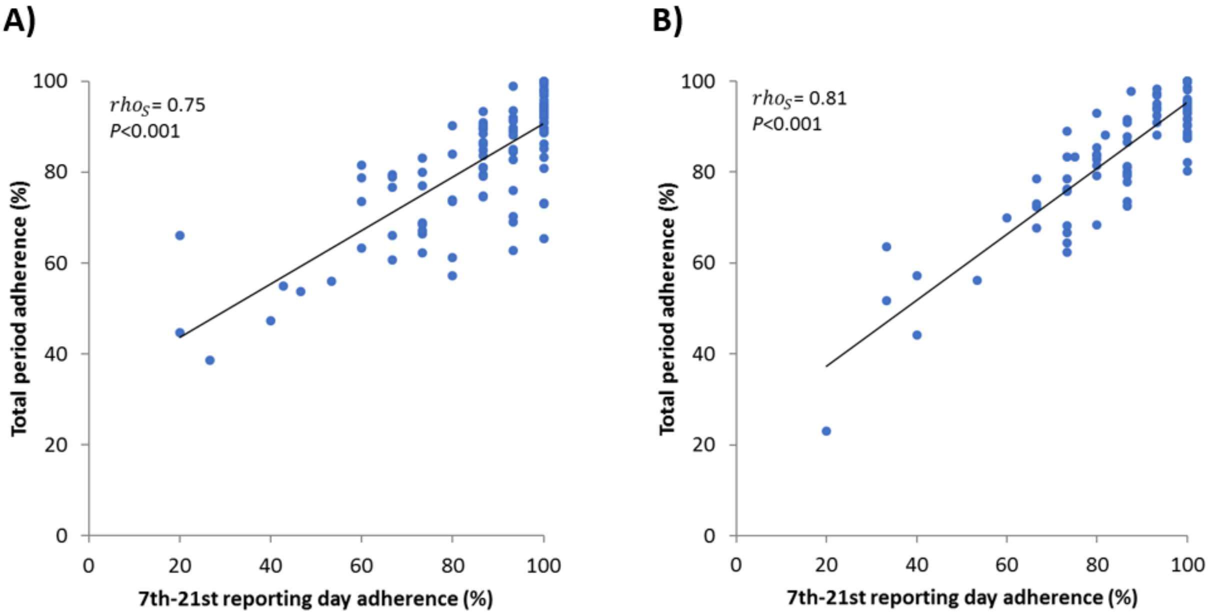


Figure 4. Correlation between postponed reporting (%) and total reporting period adherence (%) by study center: (A) Rome (n=101); (B) Pordenone (n=93).

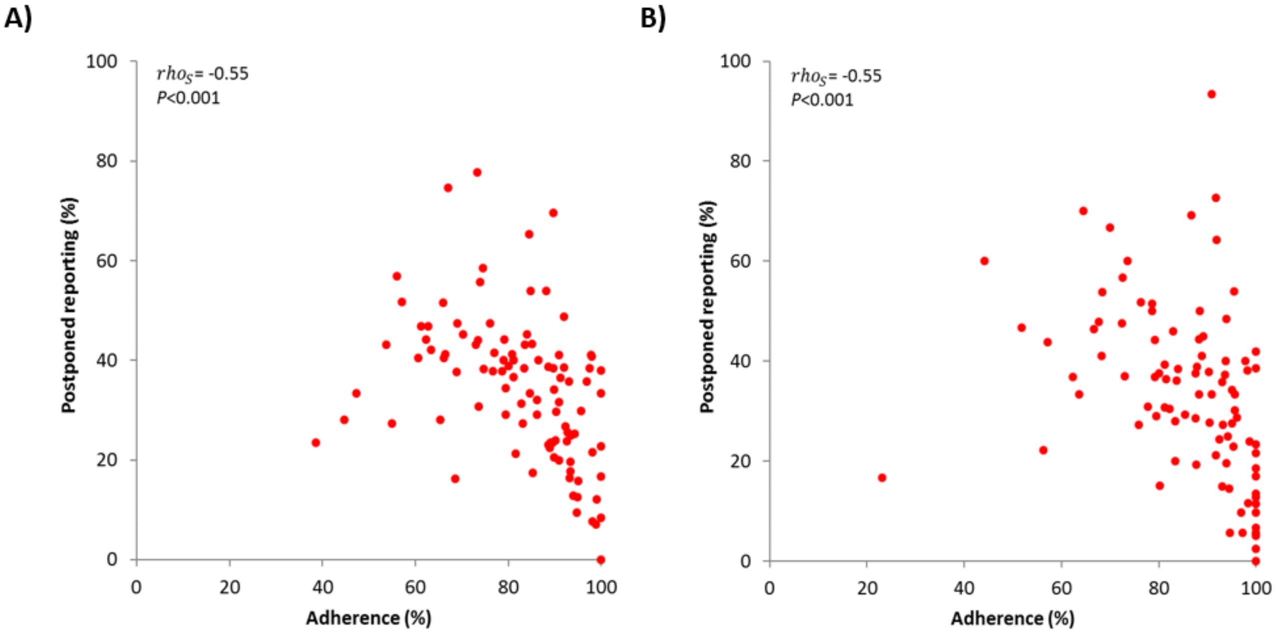


Figure 5. Mean visual analog scale (VAS) score, Rhinoconjunctivitis Total Symptom Score (RTSS), and combined symptom and medication score (CSMS) by time considering the local whole season of grass pollen in (A) Rome (n=101) and (B) Pordenone (n=93; see Figure 2 and Table 2 for definitions).

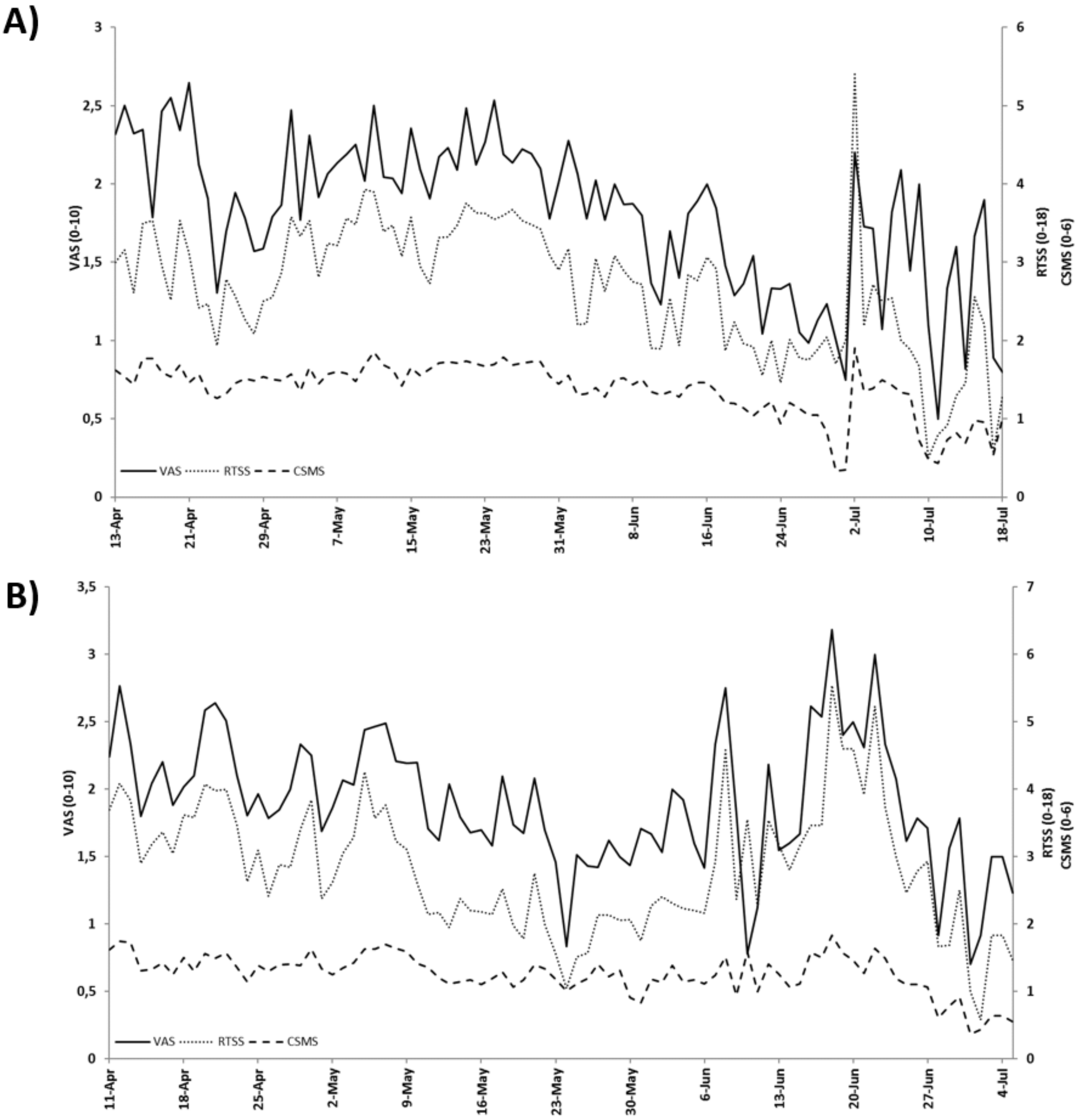
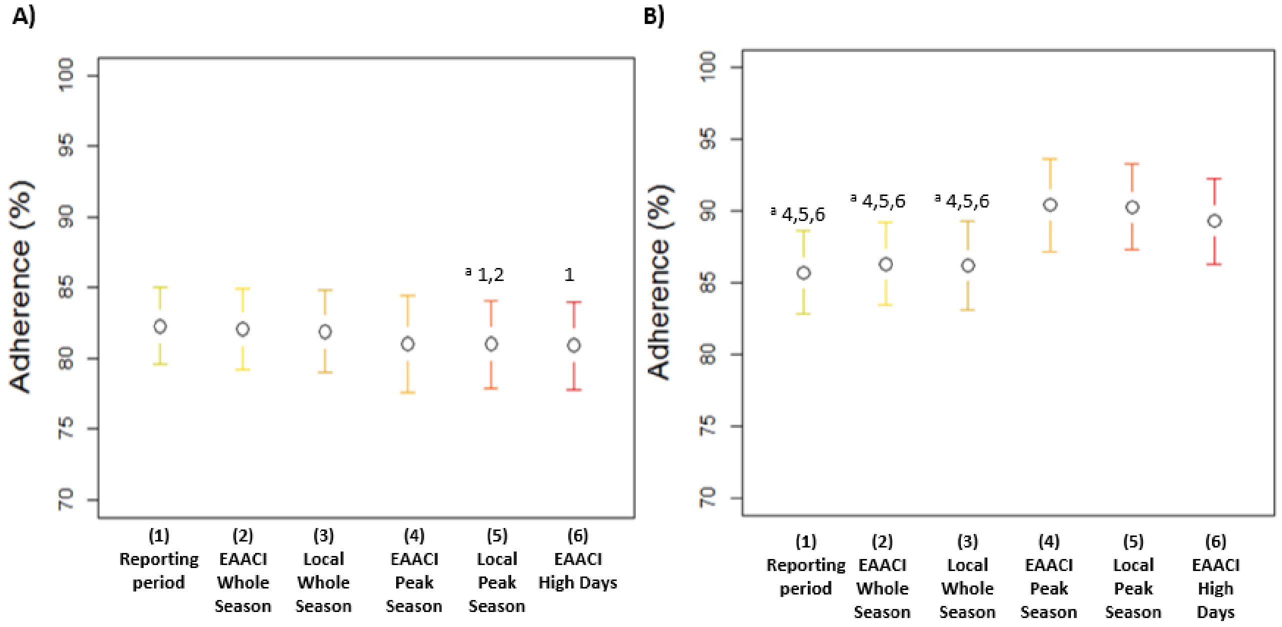


Figure 6. Mean (95% CI) adherence to recording of the e-diary during the pollen season in children affected by seasonal allergic rhinitis in (A) Rome and (B) Pordenone. Adherence was calculated for each patient considering the total reporting period and according to whole, peak, and high days of pollen periods defined by the European Academy of Allergy and Clinical Immunology (EAACI) and local criteria. See Figure 1 and Table 2 for criteria. "a" indicates that nonparametric Friedman test for repeated measures was applied and only statistically significant *P* values of multiple comparisons, adjusted by Bonferroni correction, are highlighted.



Discussion

Principal Findings

In this bicenter study, we investigated the adherence of Italian patients with SAR to symptom and medication monitoring via an e-diary prescribed by their doctor in the context of an observational study. We found that adherence to recording was (1) very high (>80%) in the first 7 weeks of monitoring, (2) predicted by the adherence in the first 3 weeks of the monitoring period, (3) inversely associated with the frequency of delayed e-diary compilation, and (4) higher during the peak pollen season.

The trajectories of the mean adherence to recording were highly similar in both study populations, notwithstanding their differences in geographical location (northern vs central Italy) and age (adults vs children). Moreover, we observed only a few patients with very low adherence to e-diary recording, that is, failing to register their symptoms during more than 60% of days within their monitoring period (not shown).

This level of adherence is at great variance from levels published in previous studies on e-diaries in patients with allergic rhinitis who had not been specifically instructed and advised by a doctor to use an app. With this approach, the Mobile Airways Sentinel Network observational pilot study among 2871 allergic users from 15 countries reported an adherence to symptom recording of only 9.5% after 14 days of recording [14]. A follow-up project among 9122 users from 22 countries showed that only 16.4% of the users were still recording their symptoms after 14 days [15].

Digital technologies have been shown to be a very useful tool for the assessment of real-life data among big patient groups [14-17]. While the patient-initiated use of an e-diary app may be very helpful in highly motivated patients looking for self-management opportunities, it seems that this scenario applies to only a minority of the users spontaneously downloading, installing, and using an e-diary app for allergic rhinitis [14,15]. However, our results showed that in a blended care approach combining face-to-face visits with internet-based support technologies, patients are keen and able to correctly use an e-diary when contacted and instructed to do so by their allergist. It has to be underlined, though, that our patients were participating in an observational clinical study and we do not know whether their high adherence would have been also maintained in the context of routine clinical practice. This hypothesis deserves to be tested in a real-life or surveillance study.

The adherence to e-diary recording of the patients in Rome was slightly, but significantly, higher during the grass pollen peak season, when allergic symptoms were also more severe. This observation may be easily explained by increased awareness and motivation linked to symptom severity. This outcome should be taken into account when considering the use of e-diaries outside the pollen season or in patients with very mild symptoms. With regard to monitoring scores, we demonstrated that the overall VAS score reliably reflected the results of the RTSS and CSMS, which confirms the usefulness of VASs for digital symptom assessment as previously shown in other studies [31,32].

Of great relevance is, in our opinion, that the overall adherence of a patient to e-diary recording over a period of 2 or more months can already be predicted with enough confidence in the second and third weeks of monitoring. Patients at risk of poor adherence could therefore be identified and receive supplementary information and education, thus facilitating a higher compliance.

Limitations

First, our study population consisted of Italian patients only, so that our results now require further evaluation in different cultural contexts. To this end, we are examining the outcomes of a similar study performed in 7 southern European and Mediterranean countries. Second, we cannot comment on possible outcome improvements, as the study did not include any control group. Third, we limited our monitoring period to a maximum of 90 days; we do not know whether the patients' adherence to recording would have remained high enough beyond this time frame. Fourth, our results and proposal cannot be applied to SAR patients not using a mobile phone. Fifth, we did not evaluate potential adverse effects of the use of an e-diary, such as excessive attention to disease or even facilitation of anxiety and obsessive disturbances.

Conclusion

Our study showed that adherence to the daily symptom and medication monitoring via an e-diary was maintained at a high level up to 2 months by SAR patients properly informed and educated by their allergist. This outcome underlines the strength of a blended care approach and needs now to be confirmed in a real-life clinical allergy setting. Our results contribute to reinforce positive expectations for a proper use of mHealth technology in monitoring patients with SAR for diagnostic and therapeutic purposes.

Acknowledgments

We thank Ms Theresa Lipp for English language revision.

SA was supported by the EAACI Fellowship Award of the European Academy of Allergy and Clinical Immunology. The study was supported by an unrestricted grant from EUROIMMUN (no. 118583). The Informatics Platform AllergyCARD and the app Allergy.Monitor were kindly provided by TPS Production, Rome, Italy. We acknowledge support from the German Research Foundation and the Open Access Publication Funds of Charité – Universitätsmedizin Berlin.

Conflicts of Interest

PMM reports grants and personal fees from EUROIMMUN AG, during the conduct of the study; and grants and personal fees from Thermo Fisher Scientific, and personal fees from Hycor Biomedical Inc, outside the submitted work. ST is cofounder of TPS Production. S Pelosi reports personal fees from TPS Production. FB reports personal fees from Abbvie. The remaining authors declare no conflict of interest.

References

1. van Cauwenberge P, Bachert C, Passalacqua G, Bousquet J, Canonica GW, Durham SR, et al. Consensus statement on the treatment of allergic rhinitis. European Academy of Allergology and Clinical Immunology. Allergy 2000 Feb;55(2):116-134. [doi: 10.1034/j.1398-9995.2000.00526.x] [Medline: 10726726]
2. Agache I, Bilò M, Braunstahl G, Delgado L, Demoly P, Eigenmann P, et al. In vivo diagnosis of allergic diseases--allergen provocation tests. Allergy 2015 Apr;70(4):355-365. [doi: 10.1111/all.12586] [Medline: 25640808]
3. Krzych-Fałta E, Sowa J, Wojas O, Piekarska B, Sybilski A, Samolinski B. Allergen Challenge Chamber: an innovative solution in allergic rhinitis diagnosis. Postepy Dermatol Alergol 2015 Dec;32(6):421-425 [FREE Full text] [doi: 10.5114/pdia.2015.56096] [Medline: 26755904]
4. Boelke G, Berger U, Bergmann K, Bindslev-Jensen C, Bousquet J, Gildemeister J, et al. Peak nasal inspiratory flow as outcome for provocation studies in allergen exposure chambers: a GALEN study. Clin Transl Allergy 2017;7:33 [FREE Full text] [doi: 10.1186/s13601-017-0169-4] [Medline: 28932387]
5. Scadding G, Hellings P, Alobid I, Bachert C, Fokkens W, van Wijk RG, et al. Diagnostic tools in Rhinology EAACI position paper. Clin Transl Allergy 2011 Jun 10;1(1):2 [FREE Full text] [doi: 10.1186/2045-7022-1-2] [Medline: 22410181]
6. Assing K, Bodtger U, Poulsen LK, Malling HJ. Grass pollen symptoms interfere with the recollection of birch pollen symptoms - a prospective study of suspected, asymptomatic skin sensitization. Allergy 2007 Apr;62(4):373-377. [doi: 10.1111/j.1398-9995.2006.01280.x] [Medline: 17362247]
7. Dondi A, Tripodi S, Panetta V, Asero R, Businco ADR, Bianchi A, Italian Pediatric Allergy Network (I-PAN). Pollen-induced allergic rhinitis in 1360 Italian children: comorbidities and determinants of severity. Pediatr Allergy Immunol 2013 Dec;24(8):742-751. [doi: 10.1111/pai.12136] [Medline: 24236692]
8. Florack J, Brighetti MA, Perna S, Pizzulli A, Pizzulli A, Tripodi S, et al. Comparison of six disease severity scores for allergic rhinitis against pollen counts a prospective analysis at population and individual level. Pediatr Allergy Immunol 2016 Jun;27(4):382-390. [doi: 10.1111/pai.12562] [Medline: 26992008]
9. Bastl K, Kmenta M, Pessi AM, Prank M, Saarto A, Sofiev M, et al. First comparison of symptom data with allergen content (Bet v 1 and Phl p 5 measurements) and pollen data from four European regions during 2009-2011. Sci Total Environ 2016 Apr 01;548-549:229-235. [doi: 10.1016/j.scitotenv.2016.01.014] [Medline: 26802351]
10. Bianchi A, Tsilochristou O, Gabrielli F, Tripodi S, Matricardi PM. The smartphone: a novel diagnostic tool in pollen allergy? J Investig Allergol Clin Immunol 2016;26(3):204-207 [FREE Full text] [doi: 10.18176/jiaci.0060] [Medline: 27326995]

11. Matricardi PM, Dramburg S, Alvarez-Perea A, Antolín-Amérigo D, Apfelbacher C, Atanaskovic-Markovic M, et al. The role of mobile health technologies in allergy care: An EAACI position paper. Allergy 2019 Jun 22. [doi: 10.1111/all.13953] [Medline: 31230373]
12. Bousquet J, Caimmi DP, Bedbrook A, Bewick M, Hellings PW, Devillier P, et al. Pilot study of mobile phone technology in allergic rhinitis in European countries: the MASK-rhinitis study. Allergy 2017 Jun;72(6):857-865. [doi: 10.1111/all.13125] [Medline: 28072463]
13. Huang X, Matricardi PM. Allergy and asthma care in the mobile phone era. Clin Rev Allergy Immunol 2016 May 21;56:161-173. [doi: 10.1007/s12016-016-8542-y] [Medline: 27209270]
14. Bédard A, Basagaña X, Anto JM, Garcia-Aymerich J, Devillier P, Arnavielhe S, MASK study group. Mobile technology offers novel insights into the control and treatment of allergic rhinitis: The MASK study. J Allergy Clin Immunol 2019 Jul;144(1):135-143.e6. [doi: 10.1016/j.jaci.2019.01.053] [Medline: 30951790]
15. Bousquet J, Bedbrook A, Czarlewski W, Onorato GL, Arnavielhe S, Laune D, The MASK study group. Guidance to 2018 good practice: ARIA digitally-enabled, integrated, person-centred care for rhinitis and asthma. Clin Transl Allergy 2019;9:16 [FREE Full text] [doi: 10.1186/s13601-019-0252-0] [Medline: 30911372]
16. Bousquet J, Ansotegui IJ, Anto JM, Arnavielhe S, Bachert C, Basagaña X, et al. Mobile technology in allergic rhinitis: evolution in management or revolution in health and care? J Allergy Clin Immunol Pract 2019;7(8):2511-2523. [doi: 10.1016/j.jaip.2019.07.044] [Medline: 31445223]
17. Menditto E, Costa E, Midão L, Bosnic-Anticevich S, Novellino E, Bialek S, MASK group. Adherence to treatment in allergic rhinitis using mobile technology. The MASK Study. Clin Exp Allergy 2019 Apr;49(4):442-460. [doi: 10.1111/cea.13333] [Medline: 30597673]
18. Pizzulli A, Perna S, Florack J, Pizzulli A, Giordani P, Tripodi S, et al. The impact of telemonitoring on adherence to nasal corticosteroid treatment in children with seasonal allergic rhinoconjunctivitis. Clin Exp Allergy 2014 Oct;44(10):1246-1254. [doi: 10.1111/cea.12386] [Medline: 25109375]
19. Karatzas K, Katsifarakis N, Riga M, Werchan B, Werchan M, Berger U, et al. New European Academy of Allergy and Clinical Immunology definition on pollen season mirrors symptom load for grass and birch pollen-induced allergic rhinitis. Allergy 2018 Sep;73(9):1851-1859. [doi: 10.1111/all.13487] [Medline: 29791010]
20. I.Stat. URL: http://dati.istat.it/
21. European Environment Agency. Copenhagen, Denmark; 2019. URL: https://www.eea.europa.eu
22. Ente per le Nuove tecnologie, l'Energia e l'Ambiante. Archivio Climatico DBT. URL: http://clisun.casaccia.enea.it/
23. Asher MI, Keil U, Anderson HR, Beasley R, Crane J, Martinez F, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. Eur Respir J 1995 Mar;8(3):483-491 [FREE Full text] [doi: 10.1183/09031936.95.08030483] [Medline: 7789502]
24. Brozek JL, Bousquet J, Baena-Cagnani CE, Bonini S, Canonica GW, Casale TB, Global Allergy and Asthma European Network, Grading of Recommendations Assessment, Development and Evaluation Working Group. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision. J Allergy Clin Immunol 2010 Sep;126(3):466-476. [doi: 10.1016/j.jaci.2010.06.047] [Medline: 20816182]
25. Jáuregui I, Dávila I, Sastre J, Bartra J, del Cuvillo A, Ferrer M, et al. Validation of ARIA (Allergic Rhinitis and its Impact on Asthma) classification in a pediatric population: the PEDRIAL study. Pediatr Allergy Immunol 2011 Jun;22(4):388-392. [doi: 10.1111/j.1399-3038.2010.01108.x] [Medline: 21261745]
26. Global Initiative for Asthma Report: Update 2019. URL: https://ginasthma.org/wp-content/uploads/2019/06/GINA-2019-main-report-June-2019-wms.pdf [accessed 2019-11-13]
27. CEN/TS 16868: 2015. Ambient air - sampling and analysis of airborne pollen grains and fungal spores for allergy networks - volumetric Hirst method. Brussels, Belgium: European Committee for Standardization; 2015.
28. Pfaar O, Bastl K, Berger U, Buters J, Calderon MA, Clot B, et al. Defining pollen exposure times for clinical trials of allergen immunotherapy for pollen-induced rhinoconjunctivitis - an EAACI position paper. Allergy 2017 May;72(5):713-722. [doi: 10.1111/all.13092] [Medline: 27874202]
29. Canonica GW, Baena-Cagnani CE, Bousquet J, Bousquet PJ, Lockey RF, Malling H, et al. Recommendations for standardization of clinical trials with allergen specific immunotherapy for respiratory allergy. A statement of a World Allergy Organization (WAO) taskforce. Allergy 2007 Mar;62(3):317-324 [FREE Full text] [doi: 10.1111/j.1398-9995.2006.01312.x] [Medline: 17298350]
30. Pfaar O, Demoly P, Gerth van Wijk R, Bonini S, Bousquet J, Canonica GW, European Academy of Allergy and Clinical Immunology. Recommendations for the standardization of clinical outcomes used in allergen immunotherapy trials for allergic rhinoconjunctivitis: an EAACI Position Paper. Allergy 2014 Jul;69(7):854-867. [doi: 10.1111/all.12383] [Medline: 24761804]
31. Bousquet PJ, Combescure C, Neukirch F, Klossek JM, Méchin H, Daures J, et al. Visual analog scales can assess the severity of rhinitis graded according to ARIA guidelines. Allergy 2007 Apr;62(4):367-372. [doi: 10.1111/j.1398-9995.2006.01276.x] [Medline: 17362246]

32. Caimmi D, Baiz N, Tanno LK, Demoly P, Arnavielhe S, Murray R, MASK Study Group. Validation of the MASK-rhinitis visual analogue scale on smartphone screens to assess allergic rhinitis control. Clin Exp Allergy 2017 Dec;47(12):1526-1533. [doi: 10.1111/cea.13025] [Medline: 28886234]

Abbreviations

ARIA: Allergic Rhinitis and its Impact on Asthma
CSMS: combined symptom and medication score
EAACI: European Academy of Allergy and Clinical Immunology
RTSS: Rhinoconjunctivitis Total Symptoms Score
SAR: seasonal allergic rhinoconjunctivitis
VAS: visual analog scale

Edited by G Eysenbach; submitted 10.10.19; peer-reviewed by D Ryan, J Ivancevich; comments to author 30.10.19; revised version received 02.12.19; accepted 16.12.19; published 16.03.20.

Please cite as:
Di Fraia M, Tripodi S, Arasi S, Dramburg S, Castelli S, Villalta D, Buzzulini F, Sfika I, Villella V, Potapova E, Perna S, Brighetti MA, Travaglini A, Verardo P, Pelosi S, Zicari AM, Matricardi PM
Adherence to Prescribed E-Diary Recording by Patients With Seasonal Allergic Rhinitis: Observational Study
J Med Internet Res 2020;22(3):e16642
URL: https://www.jmir.org/2020/3/e16642
doi:10.2196/16642
PMID:32175909

Original Paper

Text Messaging as a Screening Tool for Depression and Related Conditions in Underserved, Predominantly Minority Safety Net Primary Care Patients: Validity Study

Haomiao Jin[1,2], PhD; Shinyi Wu[1,2,3], PhD

[1]Suzanne Dworak-Peck School of Social Work, University of Southern California, Los Angeles, CA, United States

[2]Edward R Roybal Institute on Aging, University of Southern California, Los Angeles, CA, United States

[3]Daniel J Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Haomiao Jin, PhD
Suzanne Dworak-Peck School of Social Work
University of Southern California
1150 S Olive Street
Suite 1400
Los Angeles, CA, 90015
United States
Phone: 1 2138216441
Email: haomiaoj@usc.edu

Abstract

Background: SMS text messaging is an inexpensive, private, and scalable technology-mediated assessment mode that can alleviate many barriers faced by the safety net population to receive depression screening. Some existing studies suggest that technology-mediated assessment encourages self-disclosure of sensitive health information such as depressive symptoms while other studies show the opposite effect.

Objective: This study aimed to evaluate the validity of using SMS text messaging to screen depression and related conditions, including anxiety and functional disability, in a low-income, culturally diverse safety net primary care population.

Methods: This study used a randomized design with 4 study groups that permuted the order of SMS text messaging and the gold standard interview (INTW) assessment. The participants for this study were recruited from the participants of the prior Diabetes-Depression Care-management Adoption Trial (DCAT). Depression was screened by using the 2-item and 8-item Patient Health Questionnaire (PHQ-2 and PHQ-8, respectively). Anxiety was screened by using the 2-item Generalized Anxiety Disorder scale (GAD-2), and functional disability was assessed by using the Sheehan Disability Scale (SDS). Participants chose to take up the assessment in English or Spanish. Internal consistency and test-retest reliability were evaluated by using Cronbach alpha and intraclass correlation coefficient (ICC), respectively. Concordance was evaluated by using an ICC, a kappa statistic, an area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity. A regression analysis was conducted to examine the association between the participant characteristics and the differences in the scores between the SMS text messaging and INTW assessment modes.

Results: Overall, 206 participants (average age 57.1 [SD 9.18] years; females: 119/206, 57.8%) were enrolled. All measurements except the SMS text messaging–assessed PHQ-2 showed Cronbach alpha values ≥.70, indicating acceptable to good internal consistency. All measurements except the INTW-assessed SDS had ICC values ≥0.75, indicating good to excellent test-retest reliability. For concordance, the PHQ-8 had an ICC of 0.73 and AUROC of 0.93, indicating good concordance. The kappa statistic, sensitivity, and specificity for major depression (PHQ-8 ≥8) were 0.43, 0.60, and 0.86, respectively. The concordance of the shorter PHQ-2, GAD-2, and SDS scales was poor to fair. The regression analysis revealed that a higher level of personal depression stigma was associated with reporting higher SMS text messaging–assessed PHQ-8 and GAD-2 scores than the INTW-assessed scores. The analysis also determined that the differences in the scores were associated with marital status and personality traits.

Conclusions: Depression screening conducted using the longer PHQ-8 scale via SMS text messaging demonstrated good internal consistency, test-retest reliability, and concordance with the gold standard INTW assessment mode. However, care must be taken when deploying shorter scales via SMS text messaging. Further regression analysis supported that a technology-mediated assessment, such as SMS text messaging, may create a private space with less pressure from the personal depression stigma and therefore encourage self-disclosure of depressive symptoms.

Trial Registration: ClinicalTrials.gov NCT01781013; https://clinicaltrials.gov/ct2/show/NCT01781013

International Registered Report Identifier (IRRID): RR2-10.2196/12392

(J Med Internet Res 2020;22(3):e17282) doi:10.2196/17282

KEYWORDS

depression; diabetes mellitus; comorbidity; screening; primary care; health information technology; mobile health; text messaging; patient reported outcome measures

Introduction

Depression is an underdiagnosed comorbidity that can negatively affect functional status, morbidity/mortality, and cost for the treatment of chronic illnesses, such as diabetes [1-5]. Depression screening is an effective approach to reduce the rate of undiagnosed depression and provide timely treatment for patients [6]. On the basis of the growing evidence for the benefits of depression screening, the US Preventive Services Task Force recommends depression screening for every adult in the 2016 update of the clinical guidelines [6].

Nevertheless, there are significant barriers for adopting mass depression screening, particularly in underserved, predominantly minority patients with chronic illnesses. This patient population has an increased risk of depression and often prefers safety net primary care over specialty psychiatric care when seeking mental health care [7-9]. However, safety net primary care providers often find themselves lacking time and resources to address mental health issues on top of managing other medical conditions such as diabetes [10-13]. In addition, minority patients are less likely to voluntarily report depressive symptoms. They may view depression as a moral weakness or character flaw rather than an illness and may be more likely to ascribe symptoms of depression to a physical illness [14]. Therefore, underserved minority patients in safety net care systems often miss out on screening and are less than half as likely as non-Hispanic whites to receive any depression care or guideline-level depression care [11,15]

The increasing usage of mobile services, particularly SMS text messaging, provides opportunities to overcome the barriers for adopting universal depression screening in underserved populations. The use of SMS text messaging is highly prevalent globally; among the 4 billion mobile phones in use, 3.05 billion (75%) are SMS text messaging–enabled [16]. In the United States, texting among adult mobile users is higher among minorities such as Hispanics/Latinos (83%) than non-Hispanic whites (70%) [17]. SMS text messaging is also inexpensive, private, and can be scaled to large populations [16,17]. Thus, SMS text messaging could be an ideal approach for conducting mass depression screening for underserved, predominantly minority patients in safety net primary care systems.

Previous studies have tested the validity of conducting standardized depression screening, such as the Patient Health Questionnaire (PHQ), by using paper-based self-reported assessment [18-20], in-person interviewer (INTW) assessment [21,22], and telephone INTW assessment [11,21]. Patients with depression are at a higher risk of comorbid anxiety and functional disability; as many as 50% of depressed patients in the primary care setting suffer from anxiety and/or functional disability [8,23]. INTW-administered anxiety screening using the 2-item Generalized Anxiety Disorder (GAD-2) scale has been validated in 6 studies (reviewed by Plummer et al [24]). The INTW-administered functional disability assessment using the Sheehan Disability Scale (SDS) has been validated by Leon et al [25]. Few studies have examined the validity of technology-mediated assessment for depression and its related conditions such as anxiety and functional disability. Depression screening requires participants to self-disclose sensitive health information such as a sad mood, anhedonia, and eating and sleeping problems. Current evidence on the effect of technology being used to disclose such information is weak and inconsistent. It has been suggested that technology-mediated assessments, such as SMS text messaging, may help to create an idealized perception of the information collector and thus reduce social desirability bias [26]. This effect may encourage disclosure of sensitive health information [27,28]. In contrast, there is also evidence suggesting that technology-mediated assessments discourage disclosure of sensitive information as the distance and private space created by technology may discourage patients to seek help [29].

To fill in this knowledge gap, this study examined the validity of using standardized tools to assess depression and its related conditions via SMS text messaging vs the gold standard INTW assessment in underserved, predominantly minority patients from a large safety net primary care system. This study examined the internal consistency, test-retest reliability, and concordance of the 2 modes of assessment. Patient characteristics, including demographics such as age, gender, race/ethnicity, and marital status; technology use; and psychological traits such as personality, cognitive vulnerability of depression, and depression stigma were further examined in a regression analysis to explore their correlations with the differences in the 2 modes of assessment.

Methods

Study Design

This study protocol was approved by the Institutional Review Board of the University of Southern California and has been published in *JMIR Research Protocol* [30]. Underserved, predominantly minority safety net primary care patients were recruited and randomly assigned using a simple randomization method to 1 of the 4 study groups: SMS text messaging/INTW,

INTW/SMS text messaging, SMS text messaging/SMS text messaging, or INTW/INTW. Participants in the SMS text messaging/INTW group received the SMS text messaging assessment in English or Spanish as chosen by the participant for depression and its related comorbid conditions, including anxiety and functional disability. Within 7 to 10 days following the SMS text messaging assessment, a bilingual INTW contacted the participant over telephone to repeat the same assessment. Participants in the INTW/SMS text messaging group first answered the INTW assessment over telephone; then, they replied to the SMS text messaging assessment within 7 to 10 days following the INTW assessment. Participants in the SMS text messaging/SMS text messaging and INTW/INTW groups received 2 SMS text messages and 2 INTW assessments each, respectively. The interval between the 2 assessments was 7 to 10 days. The choice of the interval between the 2 assessments was based on a widely cited study that examined the validity of INTW assessments conducted by telephone vs in-person assessments of depression [21]. A shorter interval could increase the likelihood of repeating the answer from the first assessment in the subsequent assessment, whereas a longer interval could increase the probability of change in the actual severity of depression.

The SMS text messaging/INTW and INTW/SMS text messaging groups were used to examine the concordance between the SMS text messaging and INTW assessments. The SMS text messaging/SMS text messaging and INTW/INTW groups were used to evaluate test-retest reliability. Validity of the INTW assessment has been established in prior studies [21,31]; thus, the INTW assessment served as the gold standard in this study. The participants for this study were recruited from the participants of the prior Diabetes-Depression Care-management Adoption Trial (DCAT), a large, US Department of Health and Human Services–funded translational study, in partnership with the Los Angeles County Department of Health Services, the second largest safety net system in the United States [11,32-39]. These patients were chosen from the DCAT due to prior contact and rapport built in the DCAT, and the study fit the timeline for the funding requirement. The inclusion criteria were as follows: (1) patients were DCAT participants, (2) possessed an SMS text messaging–capable phone, (3) knew how to send and receive SMS text messages, and (4) could speak and read English or Spanish. Patients unable to provide consent were excluded from the study.

As described in the study protocol paper [30], there is no consensus in the method to determine the sample size *a priori* for a validity study. Well-received published studies that evaluated the PHQ and the SDS in primary care using an INTW assessment typically had a sample size that ranged from 100 to more than 3000 [21,33,40,41]. Using the method developed by Walter et al [42], the sample size needed to evaluate the concordance using an intraclass correlation coefficient (ICC) was 80 to attain a type I error of .05 and a type II error of .20 based on the assumptions that the minimum acceptable concordance was 0.6 (ie, threshold of good concordance as suggested by Cicchetti [43]) and the expected concordance was 0.75. The sample size needed to evaluate test-retest reliability using ICC was 40 to attain a type I error of 0.05 and a type II error of 0.20 based on the assumptions that the minimum acceptable test-retest reliability was 0.6 and the expected reliability was 0.8. The targeted sample size of this study was set to 200 (ie, 50 participants in each of the 4 groups). This led to a total of 100 subjects (ie, 50 in the SMS text messaging/INTW group and 50 in the INTW/SMS text messaging group) to evaluate concordance and 50 subjects per mode of assessment to evaluate test-retest reliability.

Measurements

The depression screening was conducted using the 2-item and 8-item PHQ (PHQ-2 and PHQ-8, respectively), which are widely used depression screening tools in primary care and general populations [27]. The PHQ-8 has 8 questions; each question uses a score of 0 to 3 to assess the frequency of a depressive symptom in the past 2 weeks. The total PHQ-8 score ranges from 0 to 24, with a higher score indicating severe depressive symptoms. A cutoff score of 8 has been suggested to identify major depression using the PHQ-8 [27]. The PHQ-2 comprises the first 2 questions of the PHQ-8. The PHQ-2 score ranges from 0 to 6, with PHQ-2≥3 indicating major depression [19]. Anxiety was assessed by the GAD-2 [44]. Each GAD-2 question uses a score of 0 to 3 to assess the frequency of an anxiety symptom in the past 2 weeks. The total GAD-2 score ranges from 0 to 6, with a higher score indicating severe anxiety symptoms. Functional disability was assessed by the SDS, which includes 3 questions to assess the degree of disruption (scored from 0 to 10) caused by health problems to work/school work, social life, and family life/home responsibilities [25]. The total SDS score ranges from 0 to 30, with a higher score indicating severe functional disability.

Participant characteristics included demographics (such as age, gender, race/ethnicity, language, marital status, and education), personality, cognitive diathesis to depression, depression stigma, and mobile phone use. Personality was measured by using the Ten-Item Personality measure of the Big Five personality scale: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience [45]. Cognitive diathesis to depression was measured by using the 9-item Dysfunctional Attitudes Scale (DAS)–Short Form [46]. The DAS measures 2 depression diatheses, ie, perfectionism and dependency, and has a score ranging from 0 to 3, with a higher score indicating higher depression diathesis. Depression stigma was measured by the Depression Stigma Scale (DSS) [47], which assesses both personal and perceived depression stigma. Both the personal and perceived DSS have a score range of 0 to 4, with a higher score indicating a higher stigma. Mobile phone usage was measured by recall questions for using the phone at least once per day during the past 2 weeks for the following functions: making a telephone call, sending or reading an SMS text message, using the internet, and using a mobile app. The number of mobile functions used daily by the participants was counted to generate a dichotomous variable indicating the use of three or more mobile functions. Using a mobile phone for health care was measured by recall questions asking if the mobile device was ever used for the following health care purposes: contacting a doctor, getting health information, and assistance with self-care. A dichotomous variable was generated to indicate whether the participant ever used a mobile phone for multiple health care purposes.

Statistical Analysis

The participant characteristics were summarized using mean and standard deviation for continuous variables and frequency and percentage for dichotomous variables. The internal consistency was evaluated by using Cronbach alpha. The test-retest reliability of the SMS text messaging and INTW assessments was evaluated by using ICC. The concordance between the SMS text messaging and INTW assessments was evaluated by using ICC, a kappa statistic, an area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity. ICC was used to measure the consistency or reproducibility of the SMS text messaging and INTW assessments. AUROC, sensitivity, and specificity were used to measure discriminative validity. The kappa statistic was used to measure interrater agreement. The kappa statistic, sensitivity, and specificity were computed using the threshold levels of PHQ-2 ≥3, PHQ-8 ≥8, GAD-2 ≥3, and SDS ≥12. The differences in the scores between the SMS text messaging and INTW assessments were summarized by using means and standard deviations. The differences were detected using a paired 2-tailed *t* test.

A regression analysis was conducted to further examine the associations between the participant characteristics and the differences in the scores between the SMS text messaging and INTW assessments. To identify the most predictive variables, all patient characteristics, as summarized in Table 1, were entered into a least absolute shrinkage and selection operator (LASSO) variable selection procedure [39,48]. LASSO is a regression-based variable selection method that introduces a penalization parameter, lambda, to a standard regression to penalize the size of the coefficient estimate. As the lambda value increases, the coefficient estimate shrinks toward 0 but at varying speeds. The shrinkage speed provides a way to rank the predictive power of each variable, as variables with a slower shrinkage speed are ranked with stronger predictive power. The top 4 predictive variables selected by LASSO were included in the linear regression models to estimate their associations with the differences in the scores between the SMS text messaging and INTW assessments. The goodness of fit of the linear regression models was evaluated using the original and adjusted measures.

All statistical analyses were conducted using R, version 3.5.2 (R Core team) [49]. Cronbach alpha and ICC were calculated using the *alpha* and *ICC* functions, respectively, in the R *psych* package [50]. The kappa statistic was evaluated using the *Kappa.test* function in the R *fmsb* package [51]. The AUROC was evaluated using the *roc* function in the R *pROC* package [52]. LASSO variable selection was conducted using the *glmnet* function in the R *glmnet* package [53]. Finally, the linear regression analysis was performed using the R *lm* function.

Table 1. Summary of the participant characteristics.

Variable	All (N=206)	SMS text messaging/INTW[a] (n=52)	SMS text messaging/SMS text messaging (n=53)	INTW/SMS text messaging (n=49)	INTW/INTW (n=52)
Age (years), mean (SD)	57.11 (9.18)	58.54 (8.60)	55.35 (10.06)	57.24 (8.08)	57.33 (9.76)
Female, n (%)	119 (57.8)	33 (63.5)	34 (64.2)	26 (53.1)	26 (50.0)
Latino, n (%)	192 (93.2)	50 (96.2)	51 (96.2)	44 (91.7)	47 (92.2)
Preferred Spanish language, n (%)	160 (77.7)	39 (75.0)	47 (88.7)	38 (77.6)	36 (69.2)
Less than high-school level education, n (%)	131 (63.6)	31 (59.6)	33 (62.3)	35 (71.4)	32 (61.5)
Extraversion score, mean (SD)	3.84 (1.15)	3.84 (1.23)	4.03 (1.00)	3.68 (1.05)	3.81 (1.28)
Agreeableness score, mean (SD)	6.43 (0.87)	6.45 (0.84)	6.75 (0.53)	6.17 (0.95)	6.31 (1.01)
Conscientiousness score, mean (SD)	5.51 (1.48)	5.60 (1.24)	5.52 (1.73)	5.36 (1.56)	5.55 (1.37)
Emotional stability score, mean (SD)	5.47 (1.44)	5.37 (1.47)	5.47 (1.48)	5.43 (1.42)	5.61 (1.43)
Openness to experience score, mean (SD)	3.77 (1.23)	3.61 (1.38)	3.90 (1.15)	3.72 (1.24)	3.83 (1.17)
Dysfunctional attitude scale, mean (SD)	0.58 (0.70)	0.55 (0.65)	0.65 (0.76)	0.50 (0.66)	0.61 (0.74)
Personal depression stigma, mean (SD)	2.16 (1.08)	2.00 (1.12)	2.31 (1.11)	2.13 (0.96)	2.21 (1.13)
Perceived depression stigma, mean (SD)	3.20 (0.91)	3.07 (1.02)	3.40 (0.73)	3.14 (0.99)	3.16 (0.86)
Mobile phone functions used every day (≥3), n (%)	131 (63.6)	33 (63.5)	39 (73.6)	33 (67.3)	26 (50.0)
0	3 (1.5)	0 (0.0)	0 (0.0)	1 (2.1)	2 (3.8)
1	13 (6.4)	3 (5.8)	1 (1.9)	1 (2.1)	8 (15.4)
2	56 (27.6)	16 (30.8)	12 (23.1)	12 (25.5)	16 (30.8)
3	5 (2.5)	0 (0.0)	4 (7.7)	0 (0.0)	1 (1.9)
4	126 (62.1)	33 (63.5)	35 (67.3)	33 (70.2)	25 (48.1)
Mobile phone used for multiple health care purposes, n (%)	86 (41.7)	22 (42.3)	26 (49.1)	22 (44.9)	16 (30.8)
0 purposes	22 (10.8)	6 (11.5)	2 (3.8)	4 (8.3)	10 (19.2)
1 purpose	96 (47.1)	24 (46.2)	24 (46.2)	22 (45.8)	26 (50.0)
2 purposes	54 (26.5)	14 (26.9)	17 (32.7)	14 (29.2)	9 (17.3)
3 purposes	32 (15.7)	8 (15.4)	9 (17.3)	8 (16.7)	7 (13.5)

[a] INTW: interviewer.

Results

Participant Characteristics

Participants were recruited from June 2017 to November 2017, which led to the enrollment of 206 participants: 52 in the SMS text messaging/INTW, 53 in the SMS text messaging/SMS text messaging, 49 in the INTW/SMS text messaging, and 52 in the INTW/INTW groups. The average age of the participants was 57.1 years, 57.8% (119/206) were females, and 93.2% (192/206) were Latinos. In addition, 77.7% (160/206) chose Spanish as their preferred language. Compared with the personality norms from a large sample [54], participants in this study were more agreeable (mean: this study=6.43; norm for males aged 51-60 years=4.89; and norm for females aged 51-60 years=5.43), more emotionally stable (mean: this study=5.47; norm for males aged 51-60 years=4.80; and norm for females aged 51-60 years=4.66), less open to new experiences (mean: this study=3.77; norm for males aged 51-60 years=5.39; and norm for females aged 51-60 years=5.42), similar in extraversion (mean: this study=3.84;

norm for males aged 51-60 years=3.87; and norm for females aged 51-60 years=4.18), and similar in conscientiousness (this study=5.51; norm for males aged 51-60 years=5.11; and norm for females aged 51-60 years=5.35). Overall, 63.6% (131/206) of the participants used three or more mobile phone functions every day; only 41.7% (86/206) of the participants ever used a mobile phone for multiple health care purposes. Table 1 summarizes the participant characteristics.

Internal Consistency and Test-Retest Reliability

The internal consistency and test-retest reliability of the INTW and SMS text messaging assessments were evaluated by using Cronbach alpha and ICC, respectively. As shown in Table 2, all measurements except the SMS text messaging–assessed PHQ-2 had Cronbach alpha values ≥.70. Following the guidelines [55,56], a Cronbach alpha value ≥.70 indicates greater than acceptable internal consistency. Both the INTW and SMS text messaging assessments for the PHQ-8 and SDS had Cronbach alpha values ≥.80, indicating good internal consistency [55,56]. All measurements except the INTW-assessed SDS had ICC values ≥0.75. Following the guidelines given by Cicchetti [43], these values indicate good to excellent test-retest reliability. The INTW-assessed SDS had an ICC value of 0.47, indicating fair test-retest reliability [43].

Table 2. Internal consistency and test-retest reliability of the interviewer and SMS text messaging assessments.

Assessment mode	Internal consistency (Cronbach alpha)	Test-retest reliability (intraclass correlation coefficient)
Interviewer assessment		
Depression		
PHQ-2[a]	.71	0.76
PHQ-8[b]	.86	0.78
Anxiety (GAD-2[c])	.82	0.75
Functional disability (SDS[d])	.80	0.47
SMS text messaging assessment		
Depression		
PHQ-2	.68	0.74
PHQ-8	.86	0.81
Anxiety (GAD-2)	.71	0.73
Functional disability (SDS)	.86	0.82

[a] PHQ-2: 2-item Patient Health Questionnaire.

[b] PHQ-8: 8-item Patient Health Questionnaire.

[c] GAD-2: 2-item Generalized Anxiety Disorder scale.

[d] SDS: Sheehan Disability Scale.

Concordance

Table 3 summarizes the results of evaluating the concordance between the INTW and SMS text messaging assessments. The results show that the INTW-assessed depression and anxiety scores were lower on average than their paired SMS text messaging–assessed scores, indicating that people reported fewer symptoms of depression and anxiety via the INTW assessment than the SMS text messaging assessment. The INTW-assessed SDS scores were higher on average than their paired SMS text messaging–assessed scores, indicating that people reported more functional disability in the INTW assessment than the SMS text messaging assessment. Although paired *t* tests showed no significant differences in the mean scores, ICC and kappa statistic evaluations revealed some scales with poor concordance. ICC of the PHQ-2 was 0.32, indicating poor concordance between the INTW and SMS text messaging assessments [43]. ICC values of both the GAD-2 and the SDS were 0.54, suggesting fair concordance [43]. The PHQ-8 assessments had an ICC value of 0.73, indicating good concordance [43]. The kappa statistic suggested that the categorical agreements between the INTW and SMS text messaging assessments were poor for PHQ-2 ≥3 (kappa=0.19) and SDS ≥12 (kappa=0.13), following Landis and Koch [57]. The kappa statistic for GAD-2 ≥3 was 0.35, indicating fair agreement [57]. The kappa statistic for PHQ-8 ≥8 was 0.43, indicating moderate agreement [57]. The AUROC values were 0.84, 0.93, 0.76, and 0.94 for the PHQ-2, PHQ-8, GAD-2, and SDS, respectively. The sensitivity for the 3 shorter scales, ie, PHQ-2, GAD-2, and SDS, was <0.60, whereas the sensitivity for the PHQ-8 was 0.60. The specificity for all 4 scales was >0.85.

Table 3. Concordance between the interviewer and SMS text messaging assessments.

Measurement	Interviewer assessment, mean (SD)	SMS text messaging assessment, mean (SD)	P[a] value	Intraclass correlation coefficient	Kappa value[b]	Area under the receiver operating characteristic curve	Sensitivity[b]	Specificity[b]
Depression								
Patient Health Questionnaire (2-item)	0.67 (1.27)	1.23 (1.79)	.13	0.32	0.19	0.84	0.34	0.89
Patient Health Questionnaire (8-item)	3.29 (4.47)	3.89 (4.20)	.39	0.73	0.43	0.93	0.60	0.86
Anxiety (2-item Generalized Anxiety Disorder scale)	0.97 (1.49)	1.16 (1.63)	.64	0.54	0.35	0.76	0.50	0.89
Functional disability (Sheehan Disability Scale)	8.09 (6.40)	6.83 (8.03)	.16	0.54	0.13	0.94	0.59	1.00

[a] *P* value was calculated by using a paired *t* test.

[b] The kappa statistic, sensitivity, and specificity were evaluated using a cutoff point of 3 for the 2-item Patient Health Questionnaire and 2-item Generalized Anxiety Disorder scale, 8 for the 8-item Patient Health Questionnaire, and 12 for the Sheehan Disability Scale.

Associations Between the Participant Characteristics and the Differences in the Interviewer and SMS Text Messaging Assessment Scores

A regression analysis was performed to further examine the associations between the participant characteristics and the differences in the INTW and SMS text messaging assessment scores. Table 4 summarizes the results. Compared with the participants who were more conscientious, the less-conscientious participants were significantly associated with reporting more symptoms of depression (as assessed by the PHQ-2 and PHQ-8) and anxiety (as assessed by the GAD-2) in the INTW assessment than the SMS text messaging assessment. Compared with the more emotionally stable participants, the less emotionally stable participants were significantly associated with reporting fewer symptoms of depression (as assessed by the PHQ-2) and anxiety (as assessed by the GAD-2) in the INTW assessment than the SMS text messaging assessment. Compared with the participants who were not extremely agreeable, the extremely agreeable participants were significantly associated with reporting more depression symptoms (as assessed by the PHQ-2 and PHQ-8) in the INTW assessment than the SMS text messaging assessment. Compared with the participants who were open to new experiences, those who were less open to new experiences were significantly associated with reporting more functional disability (as assessed by the SDS) in the INTW assessment than the SMS text messaging assessment. All personality-related differences were non-negligible as the differences were >1 point for the PHQ-2 and GAD-2 (both have scores ranging from 0 to 6) and >2 points for the PHQ-8 (with a score ranging from 0 to 24). A personal depression stigma was significantly associated with reporting less depression (as assessed by the PHQ-8) and anxiety (as assessed by the GAD-2) in the INTW assessment than the SMS text messaging assessment. The only significant demographic variable was being married, which was significantly associated with reporting less depression (as assessed by the PHQ-8) in the INTW assessment than the SMS text messaging assessment. The R^2 goodness of fit evaluation model showed that all regression models explained at least 40% of the variance in the data. The adjusted R^2 values were >0.30 for all models. The regression model for the difference in PHQ-8 had the best goodness of fit, with R^2=0.56 and adjusted R^2=0.48. Diagnostic plots of the regression did not reveal any violation of the underlying assumptions of the model.

Table 4. Linear regression analysis using the top 4 predictors selected by least absolute shrinkage and selection operator to predict the differences between the interviewer and SMS text messaging assessments.

Predictors	Difference between interviewer and SMS text messaging assessments, estimate of coefficient (95% CI)			
	Patient Health Questionnaire (2-item)[a]	Patient Health Questionnaire (8-item)[b]	Generalized Anxiety Disorder scale (2-item)[c]	Sheehan Disability Scale[d]
Conscientiousness score ≤4.5	1.76 (0.58 to 2.94)[e]	2.39 (0.27 to 4.51)[e]	1.09 (0.09 to 2.05)[e]	−3.75 (−8.57 to 1.07)
Emotional stability score ≤4.5	−1.45 (−2.54 to −0.36)[e]	—[f]	−1.09 (−2.04 to −0.14)[e]	—
Agreeable score=7	1.33 (0.17 to 2.49)[e]	2.35 (0.38 to 0.32)[e]	—	2.74 (−1.88 to 7.36)
Openness to experience score ≥4.5	—	—	—	5.51 (0.50 to 10.51)[e]
Personal depression stigma	—	−0.94 (−1.87 to −0.10)[e]	−0.50 (−0.98 to −0.02)[e]	—
Dysfunctional attitude score	—	—	−0.36 (−1.14 to 0.42)	—
Married	—	−2.37 (−4.39 to −0.34)[e]	—	—
Gender	0.62 (−0.50 to 1.74)	—	—	1.76 (−2.75 to 6.26)

[a] R^2 value=0.46, adjusted; R^2 value=0.38

[b] R^2 value=0.56, adjusted; R^2 value=0.48

[c] R^2 value=0.44, adjusted; R^2 value=0.36

[d] R^2 value=0.40, adjusted; R^2 value=0.31

[e] $P<.05$.

[f] Some cells are empty because the corresponding variables are not selected into the regression model.

Discussion

Principal Findings

This study examined the validity of screening depression and related comorbid conditions, including anxiety and functional disability via the SMS text messaging and INTW assessments for underserved, predominantly minority safety net primary care patients. Although the longer PHQ-8 depression screening scale had good internal consistency, test-retest reliability, and concordance, the 3 shorter scales, ie, the PHQ-2, GAD-2, and SDS, had poor-to-moderate levels of concordance between the SMS text messaging and INTW assessments. In particular, the PHQ-2 depression screening scale had poor concordance, as measured by ICC and Cohen kappa, between the SMS text messaging and INTW assessments. The kappa value of the SDS also indicated poor agreement. The interrater agreement as measured using Cohen kappa would improve if different cutoff points were assigned based on the modes of assessment. The kappa value for the PHQ-2 depression screening scale would improve from 0.19 (indicating poor agreement) to 0.52 (indicating moderate agreement) if the cutoff points were changed from 3 for both modes of assessment to 2 for the INTW assessment and to 3 for the SMS text messaging assessment. Similarly, the kappa value for the SDS would improve from 0.13 (indicating poor agreement) to 0.49 (indicating moderate agreement) if the cutoff points were changed from 12 for both modes of assessment to a cutoff point of 12 for the INTW assessment and 9 for the SMS text messaging assessment.

This study found that participants reported more symptoms of depression and anxiety via the SMS text messaging assessment than the INTW assessment. In contrast, less functional disability was reported via the SMS text messaging assessment than the INTW assessment. The regression analysis revealed that a higher level of personal depression stigma was associated with reporting more symptoms of depression and anxiety via the SMS text messaging assessment than the INTW assessment. This finding supports the hypothesis that SMS text messaging creates a private and secure environment with less social desirability bias and therefore encourages people to self-report stigmatized symptoms of depression and anxiety [26-28]. The analysis also identified that the differences in the scores between the SMS text messaging and INTW assessments were associated with personality traits. Personality traits refer to habitual patterns of behavior, thoughts, and emotions that are relatively stable over time, are relatively consistent among situations, and influence behavior [58]. As few studies have examined the relationship between personality traits and self-reporting of sensitive health information, it is difficult to judge whether these findings imply causal relationships or merely a statistical association. The finding that a higher level of conscientiousness was related to reporting more symptoms of depression and anxiety via the SMS text messaging assessment than the INTW assessment may be explained by the nature of conscientious participants, who tended to be careful, diligent, and deliberate and who were better at retrieving and reporting symptoms of depression and anxiety in the more private, less time-pressured and less socially biased environment created by the SMS text messaging assessment. A lower level of emotional stability appeared to be associated with reporting more symptoms of depression and anxiety via the SMS text messaging assessment than the INTW assessment. This may be explained by the fact that people with a lower level of emotional stability tend to have a higher level of negative emotional experience [59] such as sadness and fear, which are core symptoms of depression and anxiety measured by the PHQ-2 and GAD-2. The less-pressured environment of the SMS text messaging assessment may facilitate better retrieval and reporting of these negative emotional experiences. To the best of our knowledge, no study has reported a reasonable explanation for the correlations between agreeableness and openness to new experience and the differences in reporting via the SMS text messaging and INTW assessments. Those correlations may be caused by some unknown mechanism or merely randomness in the data. Finally, the regression analysis identified that unmarried participants reported more depressive symptoms via the INTW assessment than the SMS text messaging assessment. This phenomenon may be explained by unmarried participants who may be more likely to use the INTW assessment to seek help by reporting more symptoms [29], whereas married participants were less likely to do so as they may have stronger social network support. Overall, the regression analysis suggested that people may self-report sensitive health information differently to technology-mediated assessment than INTW assessment modes based on their personality traits, stigma with depression, and certain demographic characteristics.

Limitations

This study had a few limitations that should be discussed. First, the study participants' experience built in the prior DCAT study may have made those participants more familiar with technology-mediated assessments than the average person in the targeted study population. Nevertheless, the 4-year interval between the DCAT study (conducted during 2010-2013) and this study (conducted in 2017) was not short and thus is likely to decrease the potential influence of the DCAT assessment. Second, the study participants were predominantly Latinos, which may limit the generalizability of the results to other safety net primary care populations, particularly those of African American patients. Finally, the statistical associations revealed by the regression analysis need further exploration for the causal mechanism underlying self-reporting sensitive health information via different modes of assessment.

Conclusions

This study examined the validity of screening depression and related conditions via an SMS text messaging vs interview assessment for underserved, predominantly minority safety net primary care patients. The depression screening conducted using the longer PHQ-8 scale via SMS text messaging demonstrated good internal consistency, test-retest reliability, and concordance with the gold standard INTW assessment mode. Deploying shorter scales via SMS text messaging should be done cautiously. A further regression analysis supported that technology-mediated assessments, such as SMS text messaging, may create a private space with less pressure from personal depression stigma and therefore encourage self-disclosure of depressive symptoms. Other characteristics such as personality traits and certain demographic characteristics were also associated with the difference between technology-mediated and INTW assessment modes.

Acknowledgments

The Suzanne Dworak-Peck School of Social Work at the University of Southern California funded this study.

Conflicts of Interest

None declared.

References

1. Li C, Ford ES, Zhao G, Ahluwalia IB, Pearson WS, Mokdad AH. Prevalence and correlates of undiagnosed depression among US adults with diabetes: the Behavioral Risk Factor Surveillance System, 2006. Diabetes Res Clin Pract 2009 Feb;83(2):268-279. [doi: 10.1016/j.diabres.2008.11.006] [Medline: 19111364]
2. van Dooren FE, Nefs G, Schram MT, Verhey FR, Denollet J, Pouwer F. Depression and risk of mortality in people with diabetes mellitus: a systematic review and meta-analysis. PLoS One 2013;8(3):e57058 [FREE Full text] [doi: 10.1371/journal.pone.0057058] [Medline: 23472075]
3. Cuijpers P, Vogelzangs N, Twisk J, Kleiboer A, Li J, Penninx BW. Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. Am J Psychiatry 2014 Apr;171(4):453-462. [doi: 10.1176/appi.ajp.2013.13030325] [Medline: 24434956]
4. Park M, Katon WJ, Wolf FM. Depression and risk of mortality in individuals with diabetes: a meta-analysis and systematic review. Gen Hosp Psychiatry 2013;35(3):217-225 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.01.006] [Medline: 23415577]
5. Egede LE, Walker RJ, Bishu K, Dismuke CE. Trends in costs of depression in adults with diabetes in the United States: Medical Expenditure Panel Survey, 2004-2011. J Gen Intern Med 2016 Jun;31(6):615-622 [FREE Full text] [doi: 10.1007/s11606-016-3650-1] [Medline: 26969312]
6. Siu AL, US Preventive Services Task Force (USPSTF), Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, et al. Screening for depression in adults: US Preventive Services Task Force Recommendation Statement. J Am Med Assoc 2016 Jan 26;315(4):380-387. [doi: 10.1001/jama.2015.18392] [Medline: 26813211]

7. Katon WJ. The comorbidity of diabetes mellitus and depression. Am J Med 2008 Nov;121(11 Suppl 2):S8-15 [FREE Full text] [doi: 10.1016/j.amjmed.2008.09.008] [Medline: 18954592]
8. Egede LE. Diabetes, major depression, and functional disability among US adults. Diabetes Care 2004 Feb;27(2):421-428. [doi: 10.2337/diacare.27.2.421] [Medline: 14747223]
9. Black SA, Markides KS, Ray LA. Depression predicts increased incidence of adverse health outcomes in older Mexican Americans with type 2 diabetes. Diabetes Care 2003 Oct;26(10):2822-2828. [doi: 10.2337/diacare.26.10.2822] [Medline: 14514586]
10. Alegría M, Chatterji P, Wells K, Cao Z, Chen C, Takeuchi D, et al. Disparity in depression treatment among racial and ethnic minority populations in the United States. Psychiatr Serv 2008 Nov;59(11):1264-1272 [FREE Full text] [doi: 10.1176/appi.ps.59.11.1264] [Medline: 18971402]
11. Wu S, Ell K, Gross-Schulman SG, Sklaroff LM, Katon WJ, Nezu AM, et al. Technology-facilitated depression care management among predominantly Latino diabetes patients within a public safety net care system: comparative effectiveness trial design. Contemp Clin Trials 2014 Mar;37(2):342-354. [doi: 10.1016/j.cct.2013.11.002] [Medline: 24215775]
12. Katon W. Collaborative depression care models: from development to dissemination. Am J Prev Med 2012 May;42(5):550-552. [doi: 10.1016/j.amepre.2012.01.017] [Medline: 22516497]
13. Katon W, Unützer J, Wells K, Jones L. Collaborative depression care: history, evolution and ways to enhance dissemination and sustainability. Gen Hosp Psychiatry 2010;32(5):456-464 [FREE Full text] [doi: 10.1016/j.genhosppsych.2010.04.001] [Medline: 20851265]
14. Lyness JM, Cox C, Curry J, Conwell Y, King DA, Caine ED. Older age and the underreporting of depressive symptoms. J Am Geriatr Soc 1995 Mar;43(3):216-221. [doi: 10.1111/j.1532-5415.1995.tb07325.x] [Medline: 7884106]
15. Lagomasino IT, Dwight-Johnson M, Miranda J, Zhang L, Liao D, Duan N, et al. Disparities in depression treatment for Latinos and site of care. Psychiatr Serv 2005 Dec;56(12):1517-1523. [doi: 10.1176/appi.ps.56.12.1517] [Medline: 16339612]
16. Jones KR, Lekhak N, Kaewluang N. Using mobile phones and short message service to deliver self-management interventions for chronic conditions: a meta-review. Worldviews Evid Based Nurs 2014 Apr;11(2):81-88. [doi: 10.1111/wvn.12030] [Medline: 24597522]
17. 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]
18. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. J Am Med Assoc 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
19. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
20. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br J Gen Pract 2008 Jan;58(546):32-36 [FREE Full text] [doi: 10.3399/bjgp08X263794] [Medline: 18186994]
21. Pinto-Meza A, Serrano-Blanco A, Peñarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? J Gen Intern Med 2005 Aug;20(8):738-742 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0144.x] [Medline: 16050884]
22. Wulsin L, Somoza E, Heck J. The feasibility of using the Spanish PHQ-9 to screen for depression in primary care in Honduras. Prim Care Companion J Clin Psychiatry 2002 Oct;4(5):191-195 [FREE Full text] [doi: 10.4088/pcc.v04n0504] [Medline: 15014707]
23. Hirschfeld RM. The comorbidity of major depression and anxiety disorders: recognition and management in primary care. Prim Care Companion J Clin Psychiatry 2001 Dec;3(6):244-254 [FREE Full text] [doi: 10.4088/pcc.v03n0609] [Medline: 15014592]
24. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
25. Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. Int J Psychiatry Med 1997;27(2):93-105. [doi: 10.2190/T8EM-C8YH-373N-1UWD] [Medline: 9565717]
26. Nguyen M, Bin YS, Campbell A. Comparing online and offline self-disclosure: a systematic review. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):103-111. [doi: 10.1089/cyber.2011.0277] [Medline: 22032794]
27. Gribble JN, Miller HG, Cooley PC, Catania JA, Pollack L, Turner CF. The impact of T-ACASI interviewing on reported drug use among men who have sex with men. Subst Use Misuse 2000;35(6-8):869-890. [doi: 10.3109/10826080009148425] [Medline: 10847215]
28. Li C, Ford ES, Zhao G, Tsai J, Balluz LS. A comparison of depression prevalence estimates measured by the Patient Health Questionnaire with two administration modes: computer-assisted telephone interviewing versus computer-assisted personal interviewing. Int J Public Health 2012 Feb;57(1):225-233. [doi: 10.1007/s00038-011-0253-9] [Medline: 21523617]

29. Newman JC, Des Jarlais DC, Turner CF, Gribble J, Cooley P, Paone D. The differential effects of face-to-face and computer interview modes. Am J Public Health 2002 Feb;92(2):294-297. [doi: 10.2105/ajph.92.2.294] [Medline: 11818309]
30. Jin H, Wu S. Screening depression and related conditions via text messaging versus interview assessment: protocol for a randomized study. JMIR Res Protoc 2019 Mar 29;8(3):e12392 [FREE Full text] [doi: 10.2196/12392] [Medline: 30924787]
31. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
32. Wu S, Vidyanti I, Liu P, Hawkins C, Ramirez M, Guterman J, et al. Patient-centered technological assessment and monitoring of depression for low-income patients. J Ambul Care Manage 2014;37(2):138-147 [FREE Full text] [doi: 10.1097/JAC.0000000000000027] [Medline: 24525531]
33. Ramirez M, Wu S, Jin H, Ell K, Gross-Schulman S, Sklaroff LM, et al. Automated remote monitoring of depression: acceptance among low-income patients in diabetes disease management. JMIR Ment Health 2016 Jan 25;3(1):e6 [FREE Full text] [doi: 10.2196/mental.4823] [Medline: 26810139]
34. Jin H, Wu S, Di Capua P. Development of a clinical forecasting model to predict comorbid depression among diabetes patients and an application in depression screening policy making. Prev Chronic Dis 2015 Sep 3;12:E142 [FREE Full text] [doi: 10.5888/pcd12.150047] [Medline: 26334714]
35. Jin H, Wu S, Vidyanti I, Di Capua P, Wu B. Predicting depression among patients with diabetes using longitudinal data. A multilevel regression model. Methods Inf Med 2015;54(6):553-559. [doi: 10.3414/ME14-02-0009] [Medline: 26577265]
36. Wu S, Ell K, Jin H, Vidyanti I, Chou C, Lee P, et al. Comparative effectiveness of a technology-facilitated depression care management model in safety-net primary care patients with type 2 diabetes: 6-month outcomes of a large clinical trial. J Med Internet Res 2018 Apr 23;20(4):e147 [FREE Full text] [doi: 10.2196/jmir.7692] [Medline: 29685872]
37. Hay JW, Lee P, Jin H, Guterman JJ, Gross-Schulman S, Ell K, et al. Cost-effectiveness of a technology-facilitated depression care management adoption model in safety-net primary care patients with type 2 diabetes. Value Health 2018 May;21(5):561-568 [FREE Full text] [doi: 10.1016/j.jval.2017.11.005] [Medline: 29753353]
38. Jin H, Wu S. Developing Depression Symptoms Prediction Models to Improve Depression Care Outcomes: Preliminary Results. In: Proceedings of the 2nd International Conference on Big Data and Analytics in Healthcare. 2014 Presented at: ICBDAH'14; July 22-24, 2014; Singapore URL: https://www.researchgate.net/publication/278668625_Developing_Depression_Symptoms_Prediction_Models_to_Improve_Depression_Care_Outcomes_Preliminary_Results
39. Jin H, Wu S. Use of patient-reported data to match depression screening intervals with depression risk profiles in primary care patients with diabetes: development and validation of prediction models for major depression. JMIR Form Res 2019 Oct 1;3(4):e13610 [FREE Full text] [doi: 10.2196/13610] [Medline: 31573900]
40. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
41. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
42. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998 Jan 15;17(1):101-110. [doi: 10.1002/(sici)1097-0258(19980115)17:1<101::aid-sim727>3.0.co;2-e] [Medline: 9463853]
43. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994 Dec;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
44. Skapinakis P. The 2-item Generalized Anxiety Disorder scale had high sensitivity and specificity for detecting GAD in primary care. Evid Based Med 2007 Oct;12(5):149. [doi: 10.1136/ebm.12.5.149] [Medline: 17909240]
45. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Personal 2003;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
46. Beevers CG, Strong DR, Meyer B, Pilkonis PA, Miller IR. Efficiently assessing negative cognition in depression: an item response theory analysis of the Dysfunctional Attitude Scale. Psychol Assess 2007 Jun;19(2):199-209. [doi: 10.1037/1040-3590.19.2.199] [Medline: 17563201]
47. Griffiths KM, Christensen H, Jorm AF. Predictors of depression stigma. BMC Psychiatry 2008 Apr 18;8:25 [FREE Full text] [doi: 10.1186/1471-244X-8-25] [Medline: 18423003]
48. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc Series B Stat Methodol 2011 Jun;73(3):273-282. [doi: 10.1111/j.1467-9868.2011.00771.x]
49. R: The R Project for Statistical Computing. URL: https://www.R-project.org/ [accessed 2020-03-11]
50. Revelle W. The Comprehensive R Archive Network. 2020. psych: Procedures for Psychological, Psychometric, and Personality Research URL: https://cran.r-project.org/web/packages/psych/index.html [accessed 2020-03-11]
51. Nakazawa M. The Comprehensive R Archive Network. 2019. fmsb: Functions for Medical Statistics Book with some Demographic Data URL: https://CRAN.R-project.org/package=fmsb [accessed 2020-03-12]
52. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011 Mar 17;12:77 [FREE Full text] [doi: 10.1186/1471-2105-12-77] [Medline: 21414208]
53. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Soft 2010;33(1):1-22 [FREE Full text] [doi: 10.18637/jss.v033.i01]

54. Gosling SD, Rentfrow PJ, Potter J. Goz Lab. Ten Item Personality Measure (TIPI) URL: https://gosling.psy.utexas.edu/scales-weve-developed/ten-item-personality-measure-tipi/ [accessed 2020-03-12]
55. Kline P. Handbook of Psychological Testing. London: Routledge; 2013.
56. DeVellis RF. Scale Development: Theory And Applications. Thousand Oaks, CA: Sage; 2016.
57. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
58. Kassin S. Psychology. Upper Saddle River, NJ: Prentice Hall; 2003.
59. Larsen RJ, Ketelaar T. Personality and susceptibility to positive and negative emotional states. J Pers Soc Psychol 1991 Jul;61(1):132-140. [doi: 10.1037//0022-3514.61.1.132] [Medline: 1890584]

Abbreviations

AUROC: area under the receiver operating characteristic curve
DAS: Dysfunctional Attitudes Scale
DCAT: Diabetes-Depression Care-management Adoption Trial
DSS: Depression Stigma Scale
GAD: Generalized Anxiety Disorder
ICC: intraclass correlation coefficient
INTW: interviewer
LASSO: least absolute shrinkage and selection operator
PHQ: Patient Health Questionnaire
SDS: Sheehan Disability Scale

Edited by G Eysenbach; submitted 02.12.19; peer-reviewed by L Sun, S Buchholz, W Gong; comments to author 23.12.19; revised version received 11.02.20; accepted 27.02.20; published 26.03.20.

Please cite as:
Jin H, Wu S
Text Messaging as a Screening Tool for Depression and Related Conditions in Underserved, Predominantly Minority Safety Net Primary Care Patients: Validity Study
J Med Internet Res 2020;22(3):e17282
URL: http://www.jmir.org/2020/3/e17282/
doi:10.2196/17282
PMID:32213473

Review

Social Comparison Features in Physical Activity Promotion Apps: Scoping Meta-Review

Danielle Arigo[1], PhD; Megan M Brown[1], BSc; Kristen Pasko[1], BSc; Jerry Suls[2], PhD

[1]Department of Psychology, Rowan University, Glassboro, NJ, United States

[2]Center for Personalized Health, Feinstein Institute for Medical Research, New York, NY, United States

Corresponding Author:
Danielle Arigo, PhD
Department of Psychology
Rowan University
201 Mullica Hill Road
Robinson 116G
Glassboro, NJ, 08028
United States
Phone: 1 8562564500 ext 53775
Email: Arigo@Rowan.edu

Abstract

Background: Smartphone apps promoting physical activity (PA) are abundant, but few produce substantial and sustained behavior change. Although many PA apps purport to induce users to compare themselves with others (by invoking social comparison processes), improvements in PA and other health behaviors are inconsistent. Existing literature suggests that social comparison may motivate PA for some people under some circumstances. However, 2 aspects of work that apply social comparison theory to PA apps remain unclear: (1) how comparison processes have been operationalized or harnessed in existing PA apps and (2) whether incorporating sources of variability in response to comparison have been used to tailor comparison features of apps, which could improve their effectiveness for promoting PA.

Objective: The aim of this meta-review was to summarize existing systematic, quantitative, and narrative reviews of behavior change techniques in PA apps, with an emphasis on social comparison features, to examine how social comparison is operationalized and implemented.

Methods: We searched PubMed, Web of Science, and PsycINFO for reviews of PA smartphone apps. Of the 3743 initial articles returned, 26 reviews met the inclusion criteria. Two independent raters extracted the data from these reviews, including the definition of social comparison used to categorize app features, the percentage of apps categorized as inducing comparison, specific features intended to induce comparison, and any mention of tailoring comparison features. For reference, these data were also extracted for related processes (such as behavioral modeling, norm referencing, and social networking).

Results: Of the included review articles, 31% (8/26) categorized app features as prompting social comparison. The majority of these employed Abraham and Michie's earliest definition of comparison, which differs from versions in later iterations of the same taxonomy. Very few reviews specified what dimension users were expected to compare (eg, steps, physical fitness) or which features of the apps were used to induce comparison (eg, leaderboards, message boards). No review referenced tailoring of comparison features. In contrast, 54% (14/26) reviews categorized features for prompting behavioral modeling and 31% (8/26) referenced tailoring app features for users' personal goals or preferences.

Conclusions: The heterogeneity across reviews of PA apps and the absence of relevant information (eg, about dimensions or features relevant for comparison) create confusion about how to best harness social comparison to increase PA and its effectiveness in future research. No evidence was found that important findings from the broader social comparison literature (eg, that people have differing preferences for and responses to social comparison information) have been incorporated in the design of existing PA apps. Greater integration of the mobile health (mHealth) and social comparison literatures may improve the effectiveness of PA apps, thereby increasing the public health impact of these mHealth tools.

International Registered Report Identifier (IRRID): RR2-https://osf.io/nh4td/

(J Med Internet Res 2020;22(3):e15642) doi:10.2196/15642

KEYWORDS
smartphone app; physical activity; mHealth; social comparison; behavior change technique

Introduction

Despite decades of intervention efforts by several health care disciplines, physical inactivity remains a leading cause of morbidity and mortality in the United States [1]. Many emerging digital health interventions focus on promoting physical activity (PA) [2], delivered via mobile health (mHealth) applications or smartphone apps. For example, more than 5000 apps available from the iTunes and Google Play app stores are designed to promote PA (alone or in the context of weight loss) [3]. Although many of these apps are user-friendly and elicit high user engagement [4], most are designed without input from behavioral scientists or other health professionals and reach the market without rigorous scientific evaluation [5,6]. Conversely, evidence-based PA apps have been developed by researchers, but these apps rarely reach the commercialization stage (due to a lack of resources) and research participants show modest engagement with them [7]. These limitations may contribute to the low efficacy of existing PA apps; those that have been tested in randomized controlled trials produce only short-term increases in activity [8].

Thus, few existing PA apps are simultaneously grounded in behavior change science, engaging for potential users, and effective over the long term. Efforts are needed to improve PA app design to optimize both user engagement and intervention effectiveness. Currently, both commercial and researcher-developed PA apps vary in the extent to which they employ specific behavior change techniques (BCTs) [9]. In fact, considerable research effort has been devoted to determining the number and type of BCTs in existing apps. *Social comparison* (ie, evaluating one's standing relative to others) [10] is a BCT used in several commercial and researcher-developed apps [6]. Comparison has also been identified as one of the most effective techniques for promoting PA in face-to-face behavioral interventions [11,12]. In PA apps, social comparison is activated when a user's information is listed alongside that of other users, for example, via activity engagement rankings (leaderboards). Comparison may also be activated by any feature that exposes app users to information about other users (eg, message boards or other social networking features). However, PA app developers have not always recognized that social comparison is a complex process; it can be activated by various factors and has several possible outcomes. A comprehensive assessment of how social comparison is being currently used in PA apps and whether current methods capitalize fully on the theoretical and empirical social comparison literature has not been available. Such a review could begin to suggest how to optimize an app's social comparison features and, potentially, improve its efficacy.

To illustrate the complexities of social comparison processes, consider that PA is a multifaceted concept; there are various *dimensions* of PA (eg, steps per day, minutes of intense aerobic activity per week, appearance of muscularity, overall physical fitness), and app users may focus on any or all of these as the subject of social comparison. In addition, BCTs such as behavioral modeling (ie, providing examples of behavior engagement to encourage others to engage) and norm referencing (ie, providing information about group norms or averages) often are differentiated from social comparison as mechanisms of behavior change [9]. However, these mechanisms can explicitly or implicitly prompt a comparison of an aspect of the self to another person (or persons). Furthermore, modeling and norm referencing are assumed to prompt social comparisons in some classification systems [13]. An additional complication is that although research has found that social comparisons (via leaderboards or through these other processes) may promote PA [14,15], some experiments find that social comparisons can have negative consequences, such as worsened mood and decreased motivation for or engagement in healthy behavior [16-19]. Exposing users to others who have engaged in more PA than they have might be either inspiring (by learning what they might achieve [20]) or discouraging (by seeing themselves as inferior or incapable of achieving activity goals [16,21,22]). Conversely, exposing users to others who have engaged in less PA than they have may be satisfying (because they are outperforming their peers) or stressful (because they might also become more sedentary) [23,24].

Moreover, existing literature on social comparison processes shows that people's responses to comparison, as well as their preferences for the comparison information they receive, differ at 2 levels. At the between-person (or dispositional) level, different users may show different responses or preferences that are consistent over time [25]. At the within-person level, the same user may show variability in their responses and preferences over time [26,27]. Devising apps to modify social comparison features to match the general preferences of individual users or contextual preferences over time might be more effective for promoting PA, versus exposing everyone to the same comparison information. Such personalization or tailoring may prevent users from disengaging from social comparison or from PA apps altogether, especially if they repeatedly receive (potentially) discouraging comparison information [16,28].

To what extent distinct dimensions and possible outcomes of social comparison are considered in existing PA apps remains an open question. A search of available literature reveals more than 100 published reviews about PA apps, surveying thousands of individual app-based programs. A number of these reviews intentionally categorize app features, including social comparison (using the BCT taxonomy [9] and other frameworks). These summaries are intended to inform future app design and evaluation [29,30]. However, to our knowledge, no review or synthesis of reviews has focused on social comparison or considered whether findings from the mainstream comparison literature have been incorporated.

This scoping review had the following objectives: (1) to determine how social comparison is currently defined and categorized in existing systematic, meta-analytic, and narrative reviews of commercially available and researcher-developed

PA apps, (2) to examine the methods for activating and facilitating social comparison in the PA apps identified in these reviews, and (3) to determine to what extent different elements of social comparison are included as design features in the PA apps. This review represents an initial step for the formulation of best practice recommendations for including social comparison features in PA apps.

Methods

Guidelines

This review followed the initial guidelines delineated by Arksey and O'Malley [31] and the recent Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Extension for Scoping Reviews (PRISMA-ScR) [32]. A description of the protocol for this review is registered with the Open Science Framework.

Research Questions

This scoping review was guided by the following a priori research questions:

1. How often does social comparison appear as a key behavior change mechanism in published reviews of PA smartphone apps?
2. How is social comparison defined in published reviews of PA smartphone apps?
3. How are app features categorized as social comparison (vs other behavior change processes) in published reviews of PA apps?
4. What methods by which social comparison is activated or facilitated in PA apps are included in published reviews?
5. To what extent (and how) have PA apps included in published reviews addressed between- and within-person variability in responses to social comparison (eg, via tailoring)?
6. To what extent (and how) is social comparison differentiated from related processes, such as modeling and norm referencing, in published reviews of PA apps?

How effective social comparison features of apps are in changing PA behavior is also an important question. It is not included in the preceding list because we did not find any randomized controlled trials, narrative reviews, meta-analyses, or dismantling studies focused on social comparison app features or directly comparing the effects of different app features. We elaborate on this point in the *Discussion* section.

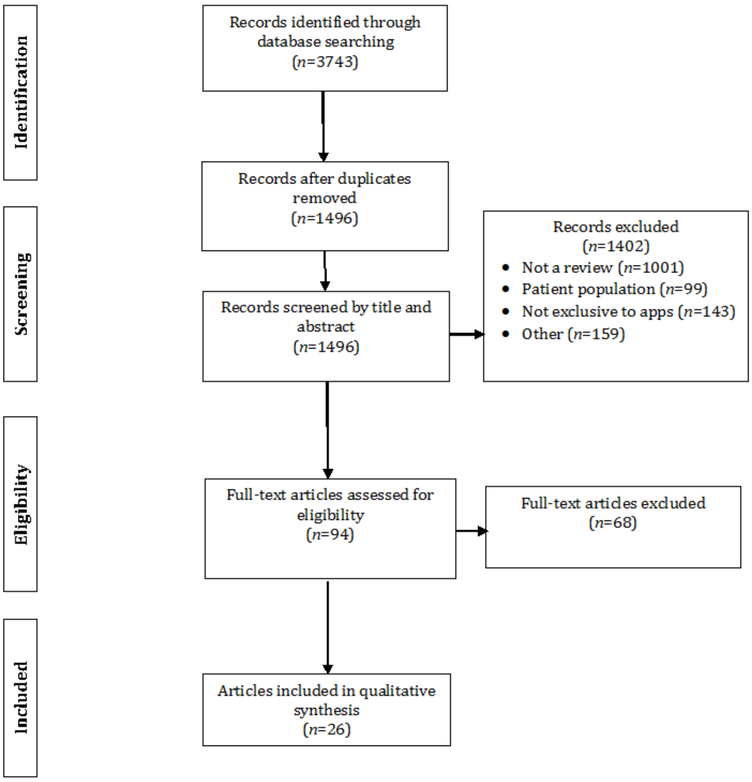
Identification and Selection of Relevant Reviews

Inclusion and exclusion criteria were chosen by the first and last authors (DA and JS). Review articles were eligible for inclusion if they met the following criteria: (1) available in English; (2) published on or before May 31, 2019; (3) conducted a systematic or narrative review, or meta-analysis; (4) reviewed the features of commercially available smartphone apps or included formal intervention programs delivered via smartphone apps; and (5) used increasing PA or reducing sedentary time as a key behavioral outcome. An initial examination of the literature revealed that many reviews in the domain of mHealth combine PA with related weight control outcomes. Consequently, reviews that met the first 4 criteria and used weight loss or PA plus other behaviors (eg, diet and weight loss) as outcomes were included.

Reviews were excluded if they considered interventions that combined an app modality with other modalities (eg, websites, text messages, etc) because they might obscure conclusions specific to apps. Apps geared toward particular medical populations also were excluded because these interventions tend to promote multiple behavior changes and set illness-specific PA targets, rather than focusing on broad-based PA increases. Finally, apps involving gamified interventions were excluded because they typically engage a variety of social processes, in addition to social comparison, to create competition with other users or teams of users. Dissecting comparison features from features intended to induce other processes in this context can be very challenging.

We searched PubMed, PsycINFO, and Web of Science for publications related to the use of smartphone apps for increasing PA. Search terms were combinations of "physical activity" or "exercise" and "smartphone app(lication)," "mobile app(lication)," or "mHealth." Resulting titles and abstracts were reviewed to determine relevance to our 6 research questions. Initial database and hand searches returned 3743 individual articles of which 2247 were duplicates, leaving 1496 unique articles. A PRISMA-ScR flowchart, shown in Figure 1, details the evaluation of each article for inclusion in this review. The majority of articles that were identified described empirical studies. Initial reviews were conducted by the first 3 authors (DA, MB, and KP) who were responsible for determining inclusion for an equal subset of identified articles. Final review and inclusion decisions were made by the first author (DA). The final set of 26 review articles were coded for the characteristics described in the following section.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Extension for Scoping Reviews flowchart.



Data Extraction and Article Coding

The first and last authors (DA and JS) determined the types of data to be extracted from each article. The second and third authors (MB and KP, respectively) each independently read and extracted the following data from the 26 reviews: authors; year of publication; review of commercially available versus researcher-developed apps (or combination); number of apps reviewed; specific behavior change outcome targeted by the app (eg, overall PA, sedentary behavior, weight loss); percentage of apps that included social comparison features; the definition of social comparison; the specific features for inducing social comparison (eg, leaderboards); the social comparison dimension (eg, steps, physical fitness); and the presence (vs absence) and type or types of social comparison tailoring. Additional data extracted included the percentage of apps categorized as modeling/demonstrating a behavior, providing normative information about others' behavior, and social networking (eg, message boards). These features are associated with the opportunity to make comparisons, even if comparison is not considered the primary BCT induced.

For reviews that explicitly categorized features based on social comparison or other types of social influence (eg, modeling), the percentages attributed to social comparison processes were taken directly from the original published review. For reviews that did not use these terms, the percentages were calculated manually by reviewing the details available in the original published review, where possible (eg, references to social networking features or exposure to information about other users). As for all other data extraction, the second and third authors (MB and KP, respectively) independently determined the percentages of apps that categorized features as inducing social comparison or other social processes. The first author (DA) then calculated the interrater agreement (91%) and independently rated a subset of included reviews to verify the accuracy; the remaining discrepancies were resolved by consensus.

Results

Types of Reviews

Among the 26 articles reviewed, the number of apps identified as promoting PA or weight control ranged from 12 [33] to more than 28,000 [34]. Of these 26 articles, 10 (38%) focused exclusively on apps intended to increase PA and 10 (38%) focused on weight loss, weight management, or obesity intervention (the largest subsets; see Table 1). The remaining reviews (6/26, 23%) considered a combination of diet, PA, and/or weight control/obesity prevention. The majority of articles reviewed only commercially available apps (19/26, 73%), primarily those available through the iTunes App Store (Apple operating system). Only 27% (7/26) of reviews appeared to include apps developed or empirically tested by researchers. The popularity of these reviews appeared to increase through 2014 (peaking in 2014-2016) and then decrease through 2019.

Table 1. Descriptive information for each included review of physical activity and related apps.

Author and year	Type(s) of apps	Apps reviewed, n	Operating system(s)	Behavioral target or targets
Azar et al (2013) [35]	Commercial	23	Apple	Weight management
Bardus et al (2016) [4]	Commercial	23	Apple and Android	Weight management
Bondaronek et al (2018) [36]	Commercial	65	Apple and Android	Increase physical activity
Brannon and Cushing (2015) [37]	Both[a]	200 (number of PA apps only)	Apple	Improve diet or physical activity
Breton et al (2011) [38]	Commercial	204	Apple	Weight management
Conroy et al (2014) [29]	Commercial	200	Apple and Android	Increase physical activity
Direito et al (2014) [39]	Commercial	40	Apple	Weight management
Dute et al (2016) [33]	Researcher developed	12	Apple	Healthy nutrition, or physical activity, or overweight prevention
Jee (2017) [40]	Both	200	Apple and Android	Increase physical activity
Jeon et al (2014) [41]	Commercial	104	Apple	Weight management
Knight et al (2015) [30]	Commercial	379	Apple and Android	Increase physical activity
Middelweerd et al (2014) [3]	Commercial	64	Apple and Android	Increase physical activity
Modave et al (2015) [42]	Commercial	30	Apple	Increase physical activity
Nikolaou and Lean (2017) [34]	Commercial	28,905	Apple, Android, Amazon, Windows, Blackberry	Weight management
Payne et al (2015) [43]	Commercial	52	Apple	Increase physical activity
Quelly et al (2016) [44]	Researcher developed	9	Not applicable	Improve weight-related behaviors, psychosocial factors, and/or anthropometric outcomes
Rivera et al (2016) [5]	Commercial	393	Apple, Android, Blackberry, Windows	Weight management
Schoeppe et al (2016) [45]	Both	27	Not mentioned	Improve diet, physical activity, and/or sedentary behavior
Schoeppe et al (2017) [46]	Commercial	25	Apple and Android	Improve diet, physical activity, and/or sedentary behavior
Schoffman et al (2013) [47]	Commercial	57	Apple	Pediatric weight management
Stuckey et al (2017) [48]	Both	18	Not mentioned	Increase physical activity
Vlahu-Gjorgievska et al (2018) [49]	Researcher tested[b]	6	Not mentioned	Weight management
Wang et al (2015) [50]	Commercial	10	Android	Increase physical activity
Wearing et al (2014) [51]	Commercial	62	Apple	Pediatric weight management
West et al (2012) [52]	Commercial	3336	Apple	Increase physical activity
Yang et al (2015) [53]	Commercial	100	Apple and Android	Increase physical activity

[a] Both: the article reviewed both commercially available and researcher-developed apps.

[b] Researcher tested: commercially available apps evaluated in formal research studies.

Social Comparison in Physical Activity Apps

Reference to Social Comparison as a Behavior Change Mechanism

Of the included review articles, 31% (8/26) categorized app features as inducing social comparison (see Table 2). The percentages of apps with social comparison features ranged from 8% (2/27) [45] to 66% (43/65) [36], with an average of 30% across reviews that used social comparison as a category (see Table 3).

Table 2. Summary of social comparison data extracted from each review of physical activity apps.

Author and year	SC[a] definition	Apps with SC as a category, n (%)	SC features	Apps specifying the SC dimension, n (%)	Apps with SC tailoring, n (%)
Azar et al (2013) [35]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Bardus et al (2016) [4]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Bondaronek et al (2018) [36]	Michie et al (2013) [13]	43 (66)[b]	Not mentioned	43 (66)[b] (comparison of behavior)	Not mentioned
Brannon and Cushing (2015) [37]	Abraham and Michie (2008) [9]	9 (14)[c]	"Most commonly be seen in the case of group practice but could also be employed using detailed case studies in text or video or by pairing people as supports."	Not mentioned	Not mentioned
Breton et al (2011) [38]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Conroy et al (2014) [29]	Michie et al (2011) [54]	25 (15.0)[d]	Not mentioned	Not mentioned	Not mentioned
Direito et al (2014) [39]	Abraham and Michie (2008) [9]	22 (55)[e]	Not mentioned	Not mentioned	Not mentioned
Dute et al (2016) [33]	Abraham and Michie (2008) [9]	1 (8)[f]	Activity sharing and connection to a partner whose activities are visible	4 (33)[f] (sharing activities)	Not mentioned
Jee (2017) [40]	Not mentioned	Not mentioned	Competitions	Not specified	Not mentioned
Jeon et al (2014) [41]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Knight et al (2015) [30]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Middelweerd et al (2014) [3]	Abraham and Michie (2008) [9]	10 (16)[g]	Not mentioned	Not mentioned	Not mentioned
Modave et al (2015) [42]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Nikolaou and Lean (2017) [34]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Payne et al (2015) [43]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Quelly et al (2016) [44]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Rivera et al (2016) [5]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Schoeppe et al (2016) [45]	Not mentioned	Not mentioned	Social comparison with friends via leaderboards	Not specified (comparison of *physical activity behavior*)	Not mentioned
Schoeppe et al (2017) [46]	Abraham and Michie (2008) [9]	10 (40)[h]	Competitions, leaderboards	Not specified	Not mentioned
Schoffman et al (2013) [47]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Stuckey et al (2017) [48]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Vlahu-Gjorgievska et al (2018) [49]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Wang et al (2015) [50]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Wearing et al (2014) [51]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
West et al (2012) [52]	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Yang et al (2015) [53]	Michie et al (2013) [13]	25 (25.0)[i]	Not specified	Not specified	Not mentioned

[a] SC: social comparison.

[b] N=65.
[c] N=66.
[d] N=167.
[e] N=40.
[f] N=12.
[g] N=64.
[h] N=25.
[i] N=100.

Table 3. Percentages of articles reviewed (26 articles) that included each behavior change technique (BCT) category, and average percentages of apps identified by these articles as including features that belong to each BCT category.

Behavior change technique (BCT)	Percentage of articles reviewed (N=26)	Average percent of apps with designated features
Social comparison	31	30
Modeling	54	35
Normative feedback	12	13
Social networking	38	32
Tailoring (general)	31	40

Definitions of Social Comparison

The majority of articles that referenced social comparison (5/8, 63%) employed Abraham and Michie's [9] BCT definition of social comparison—"facilitate[ing] observation of nonexpert others' performance for example, in a group class or using video or case study." Other definitions included those proposed by Michie et al's [54] revised Coventry, Aberdeen & London – Refined CALO-RE BCT or Michie et al's [13] hierarchy of BCTs; see Table 4 for the full text and frequencies of these definitions. Of note, Abraham and Michie's [9] definition specifies that comparison targets are *nonexperts*, and Michie et al's [54] definition explicitly states that merely exposing users to others using group settings does not constitute social comparison, as several other processes could be engaged (eg, modeling, social support).

Table 4. Definitions of social comparison used in existing reviews of physical activity apps.

Author and year	Definition	Reviews using this definition, n (%)[a]
Abraham and Michie (2008) [9]	"Facilitate observation of nonexpert others' performance for example, in a group class or using video or case study."	6 (67)
Michie et al (2011) [54]	"Facilitate social comparison Involves explicitly drawing attention to others' performance to elicit comparisons. NB: The fact the intervention takes place in a group setting, or have been placed in groups on the basis of shared characteristics, does not necessarily mean social comparison is actually taking place. Social support may also be encouraged in such settings. Group classes may also involve instruction, demonstration, and practice."	1 (11)
Michie et al (2013) [13]	"Draw attention to others' performance to allow comparison with the person's own performance. *Note:* being in a group setting does not necessarily mean that social comparison is actually taking place."	2 (22)

[a] Percentages above use a denominator of N=8, the number of reviews that categorized app features as social comparison.

Social Comparison App Features

Across definitions, only some of the articles that categorized social comparison (5/8, 63%) specified or implied which features they considered to induce comparison. These reviews referenced *leaderboards* [46], *competitions* [40], *sharing information with other users* [33], and *connections between users* [30]. One article described social comparison as features such as "group practice… [and] detailed case studies in text or video or by pairing people as supports" [37]. Another review indicated that *friendly competitions* were available in some apps but *did not* include them as features that prompt social comparison [45].

Dimension of Comparison

Of the 8 articles that categorized features inducing social comparison, 3 (38%) referenced the specific dimension. One review indicated that users could share/compare *their activities* (33% of apps reviewed) [33]; the other distinguished between apps that allowed for *comparison of behavior* (66% of apps reviewed) and *comparison of outcomes* (13% of apps reviewed) [36]. *Comparison of behavior* was most often described as a demonstration of particular exercises (ie, modeling), whereas *comparison of outcomes* referred to potential consequences of a behavior, rather than to social comparison [13]. The third review described apps that allowed sharing/comparing *PA information* [46], although without specifying the percentages of apps with such features.

Acknowledgment of Between- and Within-Person Variability or Tailoring of Comparison Features

None of the articles reviewed referred to individual (between-person) differences in social comparison responses or preferences, a change in these responses or preferences (within-person) over time, or tailoring social comparison features to address either level of variability. In contrast, 8 of the 26 included articles (31%) described tailoring or personalization with respect to feedback on user progress toward behavioral goals (92% of apps reviewed; see Table 5) [36]. For example, users who did not meet the PA guidelines for a given period were given a visual comparison of their PA to the recommended level of PA (vs reinforcement for those who met the guidelines), with PA information matched to users' demographic characteristics (eg, PA and aging for those over 45, PA and weight loss for those with BMIs greater than 25) [55]. Reviews also referenced tailoring with respect to matching *motivational cueing* (28% of apps reviewed) [48], *exercise prescriptions* (11% of apps reviewed) [48], and *encouraging messages* (33% of apps reviewed) [49] to users' progress and/or preferences.

Table 5. Summary of tailoring, modeling, norm referencing, and social networking data extracted from each review of physical activity apps.

Author and year	Apps with any tailoring, n/N (%)	Apps with modeling as a category, n/N (%)	Apps with normative feedback, n/N (%)	Apps with social networking capabilities, n/N (%)
Azar et al (2013) [35]	Not mentioned	Not specified	Not mentioned	Not mentioned
Bardus et al (2016) [4]	Not mentioned	Not mentioned	Not mentioned	3.45/23 (15)
Bondaronek et al (2018) [36]	60/65 (92)	31/65 (47)	Not mentioned	Not mentioned
Brannon and Cushing (2015) [37]	Not mentioned	124/200 (62.0)	Not mentioned	Not mentioned
Breton et al (2011) [38]	Not mentioned	Not mentioned	Not mentioned	7/204 (3.4)
Conroy et al (2014) [29]	Not mentioned	106/200 (53.0)	2/200 (1.0)	Not mentioned
Direito et al (2014) [39]	Not mentioned	21/40 (53)	Not mentioned	Not mentioned
Dute et al (2016) [33]	4/12 (33)	4/12 (33)	Not mentioned	3/12 (25)
Jee (2017) [40]	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Jeon et al (2014) [41]	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Knight et al (2015) [30]	Not mentioned	Not mentioned	Not mentioned	209/379 (55.1)
Middelweerd et al (2014) [3]	64/64 (100)	7/64 (11)	Not mentioned	Not mentioned
Modave et al (2015) [42]	Not mentioned	Not mentioned	Not mentioned	1/30 (3)
Nikolaou and Lean (2017) [34]	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Payne et al (2015) [43]	Not mentioned	16/52 (31)	3/52 (6)	Not mentioned
Quelly et al (2016) [44]	3/9 (33)	1/9 (11)	3/9 (33)	Not mentioned
Rivera et al (2016) [5]	7/393 (1.8)	Not mentioned	Not mentioned	Not mentioned
Schoeppe et al (2016) [45]	7/27 (26)	2/27 (7)	Not mentioned	4/27 (15)
Schoeppe et al (2017) [46]	Not mentioned	6.75/25 (27)	Not mentioned	Not mentioned
Schoffman et al (2013) [47]	5/57 (9)	Not mentioned	Not mentioned	9/57 (16)
Stuckey et al (2017) [48]	5/18 (28)	Not mentioned	Not mentioned	Not mentioned
Vlahu-Gjorgievska et al (2018) [49]	Not mentioned	Not mentioned	Not mentioned	Not mentioned
Wang et al (2015) [50]	Not mentioned	3.7/10 (37)	Not mentioned	7.8/10 (78)
Wearing et al (2014) [51]	Not mentioned	18/62 (29)	Not mentioned	Not mentioned
West et al (2012) [52]	Not mentioned	1235/3336 (37.0)	Not mentioned	1535/3336 (46.0)
Yang et al (2015) [53]	Not mentioned	47/100 (47.0)	Not mentioned	32/100 (32.0)

Other App Processes Relevant to Social Comparison

Modeling/Demonstrating Behavior

Of the 26 articles, 14 (54%) classified app features as modeling or demonstrating particular behaviors (eg, proper exercise form; see Table 5). The percentage of app features categorized as modeling in each review ranged in size from 7% [45] to 53% [29,38], with an average of 35%. One review indicated that modeling was a popular BCT but did not specify the percentage of apps with this feature [35]. Behavioral models were either fitness professionals (coaches) or app users who appeared via a photo or video. Although these features were not counted as inducing comparison, modeling represents an attempt to increase similarity (or decrease the perceived difference) between the app user's behavior and a comparison target's behavior. Consequently, modeling features may facilitate social comparison.

Normative Feedback

Providing normative information about others' behavior is intended to give an individual user a sense of how they compare to the average for a relevant group. Although social comparison often refers to comparisons against individual targets, comparison to a group average is a related process [56]. Of the 26 articles, 3 (12%) evaluated whether apps provided normative information to users. These articles reported that normative information appeared in 1% [29] to 33% [44] of the apps reviewed, with an average of 13%.

Social Networking

Of the 26 articles, 10 (38%) referenced social networking features via app-specific communities or connections to existing social media platforms. Percentages of apps designated as offering these features ranged from 3% [38,42] to 78% [50], with an average of 32%. Although social networking platforms can facilitate several social influence processes (eg, social reinforcement or support), social comparisons between users of these platforms are common (based on shared text, objective data, or images) and are associated with a range of affective and behavioral responses [57,58].

Discussion

Reviewing Evidence of Social Comparison in Physical Activity Apps

Social comparison is known to influence motivation and health behavior and is frequently manipulated in health behavior change interventions [9]. Comparison processes may be particularly useful for promoting PA with technology such as smartphone apps; objective measures of PA can be visualized and shared between app users, and users can see evidence of change in their relative standing by increasing their PA behavior over short time frames. Despite the interest in social comparison as a motivator of PA change and the exponential increase in publications about digital health interventions [59], no review to date has attempted to summarize existing literature on the social comparison features of PA apps. We undertook the present scoping review to address this gap and provide recommendations for future research in this area.

Defining and Classifying Social Comparison

A modest proportion of the 26 available and eligible reviews of PA promotion apps categorized app features as eliciting social comparison (31%). Comparison fell behind modeling as a popular intervention process (54%) but was as common as social networking (38%; which also may facilitate comparison) and was more common than related processes such as norm referencing (12%). All the articles that included social comparison as a category used versions of the BCT taxonomy [9,13,54]. However, the versions differ in their definitions of social comparison. The original BCT taxonomy specifies that the potential target of comparison must be a *nonexpert* [9]; exposure to an expert is classified as modeling. Although modeling appeared more frequently in apps than did social comparison, the percentages of apps with features in each category differed modestly (ie, 35% vs 30%; see Table 3). Later iterations of the BCT taxonomy removed the requirement that only social comparisons with nonexperts would qualify [13,54]. Visual inspection of the percentage of apps classified as having social comparison features suggests that using the broader definition, ie, including experts, slightly increases the average proportions of apps that receive a social comparison designation (ie, 27% to 35%). The broader definition also is consistent with definitions of social comparison used in the mainstream comparison literature, where targets often include media figures or fashion models, in addition to peers [60].

Abraham and Michie's [9] initial taxonomy also defined comparison as simply *observation* of another's performance, which may occur in a variety of contexts (eg, group classes). Using this definition, PA app features such as social networking or message boards (where users can report on their performance) may count as social comparison [30,33]. In contrast, later versions explicitly state that attention must be drawn to the other's performance and that contexts such as group classes do not necessarily induce comparison (vs other social processes) [13,54]. This definition implies that social networking and message boards would not count as social comparison, whereas leaderboards or competitions would [40,46].

The majority of reviews did not include any mention of specific dimensions of social comparison, and those that did made only vague references to dimensions (eg, *comparison of behavior* without specifying which behavior, eg, steps, etc). A recent meta-analysis suggests that comparison dimension provides information about the target's relevance to the self; if relevance to the self is not clear, the individual might reflect on their target's performance but not engage in comparative self-evaluation [61]. Owing to the many dimensions potentially relevant to PA promotion (eg, steps, calories burned, minutes of activity, and overall fitness) and the likelihood that these dimensions are not relevant for all app users [62], this review highlights the need for increased specificity in future work that describes social comparison features of apps.

As very few articles included descriptions of the specific features eliciting comparison, the exact degree of heterogeneity is unclear. What can be concluded is that existing reviews of PA apps show considerable variability in their approaches to defining and classifying social comparison. Specifically, comparison, modeling, and information sharing are not consistently differentiated. The heterogeneity associated with which features activate social comparison represents a challenge for future research to evaluate the unique effect of comparison as a mechanism of app-based behavior change, or its efficacy relative to other mechanisms [15]. Inconsistency in the definition of comparison also creates challenges for optimizing app-based interventions to address comparison preferences and needs between users, which may be either stable or dynamic. In this vein, PA app development has not yet integrated theoretical and empirical advances that the mainstream social comparison literature has made.

Social Comparison Theory and Evidence Relevant to Physical Activity App Design

Interest in and responsiveness to social comparison information vary across individuals. This construct, called social comparison orientation (SCO) [63], has been positively associated with engagement in PA [64]. PA app users with strong SCOs may engage in comparison in response to a wide variety of social features in PA apps, including social networking and message boards, and they may find this information motivating. Here, comparison information is available, but the comparison process itself is not intentionally activated. In contrast, users with weaker SCOs may engage in comparison only when the comparison process is deliberately induced, such as by competitive challenges or leaderboards that display PA data ranked from most to least [65]. Social comparison features also may be ineffective for users with weaker SCOs. These hypotheses imply that PA app effectiveness might be improved by guiding users toward the types of social features that match their level of SCO or away from social comparison features at particularly low levels of SCO.

Additional variability may exist with respect to users' social comparison preferences and their affective and behavioral responses to comparisons. As noted, users may find comparisons to targets who are *doing better* with respect to PA (ie, upward comparisons) either inspiring or disheartening and may find comparisons to targets who are *doing worse* (ie, downward comparisons) either comforting or anxiety-provoking [18,23]. Which combinations lead to the greatest increases in PA (or lead to increases vs decreases) and for whom are significant empirical questions [25,66,67]. Basic research indicates that the opportunity to select a comparison target does not always lead to optimal affective or health-relevant outcomes, nor does it always fulfill comparers' goals (eg, *to feel better*) [18,68,69]. Thus, providing information about only the targets that a PA app user wants may not lead to benefits. Providing only the targets that they do *not* want may create an aversive experience, however, and may lead users to discontinue engagement with the app [28].

The optimal combination of comparison target and affective response for increasing PA may differ between people. The best combination may also vary within the same person over time, as a function context (eg, precomparison mood), shift over the course of behavior change (eg, as users experience progress and setbacks) [56,70], and differ from users' stated preferences, depending on whether users are just starting with the app or have been engaged for some time. The degree of within-person variability in social comparison preference and response (either affective or behavioral) remains unclear. The quantification of within-person variability and its responsiveness to social comparison interventions (eg, using N-of-1 designs) represent important next steps for PA app development and a broader understanding of social comparison processes [71].

Future Directions for Social Comparison Features of Physical Activity Apps: Social Comparison Tailoring

Despite gaps in the social comparison literature, evidence suggests that the effects of social comparison and preferences for a comparison type differ between people and within people over time. This review, however, detected no reference to between- or within-person variability in comparison response/preference or to tailoring social comparison features of PA apps. In contrast, this review indicates that tailoring in PA apps is common with respect to goals and feedback, which suggests that technology for such tailoring is currently in use. Tailoring the PA app experience to match user characteristics such as SCO or user-relevant PA comparison dimensions might improve the app's acceptability and engagement and, in turn, enhance PA outcomes [28]. Indeed, tailoring has been shown to outperform generic messaging in PA interventions across a range of modalities, including apps [48,72]. Tailoring also might discourage negative consequences of comparison (eg, giving up in response to a failure to match another user's achievements) by matching a user's comparison preferences with the types of comparisons that optimize engagement in PA. Such tailoring will require nuanced assessment of the effect of factors such as SCO, dimensions of relevance, comparison preferences, affective response to comparison, and PA engagement. The adaptive capabilities of many existing apps and those under development may lend themselves to such tailoring [73].

Strengths, Limitations, and Additional Future Directions

Strengths of this scoping review include its use of preregistered methods, adherence to PRISMA-ScR guidelines, and a comprehensive search for relevant reviews to provide insights into how social comparison is currently applied in existing PA apps. A subset of pertinent articles may have been overlooked, but the extensive and systematic search increases confidence in the overall conclusions. Additional app comparison features (eg, specific dimensions and tailoring) may have not been described in the reviews or missed by our coders. As a check, we examined several primary sources of empirical data and failed to find these additional details. One exception, an empirical study by Mollee and Klein [28], demonstrated PA benefits of matching (tailoring) versus not matching comparison targets to user preferences. There is need for additional work of this kind to inform best practices for tailoring social comparison features of PA apps.

Although social comparison has been shown as effective for increasing PA in other types of interventions (eg, team-based competitions) [26], there are very few studies of the effectiveness of social comparison as a mechanism of change in PA apps (eg, randomized controlled trials, meta-analyses, and dismantling studies) to answer the question of whether, for whom, or under what circumstances social comparison features of apps produce positive changes in PA. Such research is critical to advance our basic understanding of comparison processes and their utility as BCTs, as is further information about within-person variability in comparison preferences and responses. This information would inform the necessary or sufficient social comparison features of PA apps needed for a successful intervention. To what extent our findings and conclusions apply beyond PA promotion (alone or in the context of weight control) to such health behaviors as smoking cessation or skin cancer prevention [74,75] remains to be addressed in future research.

Conclusions

This review documents that social comparison is frequently identified as a potential mechanism of behavior change in smartphone apps designed to promote PA, on par with mechanisms such as social networking (broadly defined). Behavioral modeling, which is considered in some reviews as a means of inducing social comparison, was the only comparison-related mechanism to appear in more reviews of PA apps than social comparison (as explicitly differentiated from other processes). Our findings highlight the need for careful consideration of social processes as behavior change mechanisms in app design and evaluation. Considerable gaps currently exist between theory and evidence relevant to social comparison and its implementation in PA apps. Greater attention to individual differences, dynamic responses, relevant PA dimensions, and comparison preferences and the potential to tailor apps on the basis of these characteristics may meaningfully improve the effectiveness of existing PA promotion apps.

Acknowledgments

DA was supported by NIH (grant no. K23HL136657; PI: Arigo).

Conflicts of Interest

None declared.

References

1. Borrell LN. The effects of smoking and physical inactivity on advancing mortality in US adults. Ann Epidemiol 2014 Jun;24(6):484-487. [doi: 10.1016/j.annepidem.2014.02.016] [Medline: 24842763]
2. Lewis BA, Napolitano MA, Buman MP, Williams DM, Nigg CR. Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. J Behav Med 2017 Feb;40(1):112-126 [FREE Full text] [doi: 10.1007/s10865-016-9797-8] [Medline: 27722907]
3. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
4. 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]
5. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile apps for weight management: A scoping review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]
6. Simões P, Silva AG, Amaral J, Queirós A, Rocha NP, Rodrigues M. Features, behavioral change techniques, and quality of the most popular mobile apps to measure physical activity: systematic search in app stores. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11281 [FREE Full text] [doi: 10.2196/11281] [Medline: 30368438]
7. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83 [FREE Full text] [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
8. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
9. 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]
10. Festinger L. A theory of social comparison processes. Hum Relat 1954;7(2):117-140. [doi: 10.1177/001872675400700202]
11. Olander EK, Fletcher H, Williams S, Atkinson L, Turner A, French DP. What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2013 Mar 3;10:29 [FREE Full text] [doi: 10.1186/1479-5868-10-29] [Medline: 23452345]
12. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour--and are they the same? Health Educ Res 2011 Apr;26(2):308-322. [doi: 10.1093/her/cyr005] [Medline: 21321008]
13. 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]
14. Leahey TM, Crane MM, Pinto AM, Weinberg B, Kumar R, Wing RR. Effect of teammates on changes in physical activity in a statewide campaign. Prev Med 2010 Jul;51(1):45-49 [FREE Full text] [doi: 10.1016/j.ypmed.2010.04.004] [Medline: 20394768]

15. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: A randomized controlled trial. Prev Med Rep 2016 Dec;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
16. Arigo D, Smyth JM, Suls JM. Perceptions of similarity and response to selected comparison targets in type 2 diabetes. Psychol Health 2015;30(10):1206-1220. [doi: 10.1080/08870446.2015.1040018] [Medline: 25871344]
17. Patel MS, Volpp KG, Rosin R, Bellamy SL, Small DS, Fletcher MA, et al. A randomized trial of social comparison feedback and financial incentives to increase physical activity. Am J Health Promot 2016 Jul;30(6):416-424 [FREE Full text] [doi: 10.1177/0890117116658195] [Medline: 27422252]
18. Schokker MC, Keers JC, Bouma J, Links TP, Sanderman R, Wolffenbuttel BHR, et al. The impact of social comparison information on motivation in patients with diabetes as a function of regulatory focus and self-efficacy. Health Psychol 2010 Jul;29(4):438-445. [doi: 10.1037/a0019878] [Medline: 20658832]
19. Tiggemann M, Polivy J. Upward and downward: Social comparison processing of thin idealized media images. Psychol Women Q 2010;34(3):356-364. [doi: 10.1111/j.1471-6402.2010.01581.x]
20. Sabiston CM, Chandler K. Effects of fitness advertising on weight and body shape dissatisfaction, social physique anxiety, and exercise motives in a sample of healthy-weight females. J Appl Biobehav Res 2009;14(4):165-180. [doi: 10.1111/j.1751-9861.2010.00047.x]
21. Pila E, Barlow MA, Wrosch C, Sabiston CM. Comparing the body to superior others: Associations with daily exercise and body evaluation in men and women. Psychol Sport Exerc 2016;27:120-127. [doi: 10.1016/j.psychsport.2016.08.001]
22. Zuckerman O, Gal-Oz A. Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Pers Ubiquit Comput 2014;18(7):1705-1719. [doi: 10.1007/s00779-014-0783-2]
23. Buunk BP, Ybema JF. Social comparisons and occupational stress: The identification-contrast model. In: Buunk BP, Gibbons FX, editors. Health, Coping, and Well-being: Perspectives From Social Comparison Theory. New Jersey: Lawrence Erlbaum Associates Inc; 1997:359-388.
24. Wills TA. Downward comparison principles in social psychology. Psychol Bull 1981;90(2):245-271. [doi: 10.1037/0033-2909.90.2.245]
25. van der Zee K, Buunk B, Sanderman R, Botke G, van den Bergh F. Social comparison and coping with cancer treatment. Pers Indiv Differ 2000;28(1):17-34. [doi: 10.1016/s0191-8869(99)00045-8]
26. Leahey TM, Crowther JH, Ciesla JA. An ecological momentary assessment of the effects of weight and shape social comparisons on women with eating pathology, high body dissatisfaction, and low body dissatisfaction. Behav Ther 2011 Jun;42(2):197-210 [FREE Full text] [doi: 10.1016/j.beth.2010.07.003] [Medline: 21496506]
27. Wheeler L, Miyake K. Social comparison in everyday life. J Pers Soc Psychol 1992;62(5):760-773. [doi: 10.1037/0022-3514.62.5.760]
28. Mollee JS, Klein MC. The Effectiveness of Upward and Downward Social Comparison of Physical Activity in an Online Intervention. In: Proceedings of the 2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security.: IEEE; 2016 Presented at: IUCC-CSS'16; December 14-16, 2016; Granada, Spain p. 109-115. [doi: 10.1109/iucc-css.2016.023]
29. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
30. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. JMIR Mhealth Uhealth 2015 May 21;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
31. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Method 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
32. 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]
33. Dute DJ, Bemelmans WJ, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: A review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 5;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
34. Nikolaou CK, Lean ME. Mobile applications for obesity and weight management: current market characteristics. Int J Obes (Lond) 2017 Jan;41(1):200-202. [doi: 10.1038/ijo.2016.186] [Medline: 27780974]
35. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
36. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: Review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]

37. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
38. Breton ER, Fuemmeler BF, Abroms LC. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? Transl Behav Med 2011 Dec;1(4):523-529 [FREE Full text] [doi: 10.1007/s13142-011-0076-5] [Medline: 24073074]
39. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
40. Jee H. Review of researches on smartphone applications for physical activity promotion in healthy adults. J Exerc Rehabil 2017 Feb;13(1):3-11 [FREE Full text] [doi: 10.12965/jer.1732928.464] [Medline: 28349027]
41. Jeon E, Park H, Min YH, Kim H. Analysis of the information quality of Korean obesity-management smartphone applications. Healthc Inform Res 2014 Jan;20(1):23-29 [FREE Full text] [doi: 10.4258/hir.2014.20.1.23] [Medline: 24627815]
42. Modave F, Bian J, Leavitt T, Bromwell J, Iii CH, Vincent H. Low quality of free coaching apps with respect to The American College of Sports Medicine Guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015 Jul 24;3(3):e77 [FREE Full text] [doi: 10.2196/mhealth.4669] [Medline: 26209109]
43. Payne HE, Moxley VB, MacDonald E. Health behavior theory in physical activity game apps: a content analysis. JMIR Serious Games 2015 Jul 13;3(2):e4 [FREE Full text] [doi: 10.2196/games.4187] [Medline: 26168926]
44. Quelly SB, Norris AE, DiPietro JL. Impact of mobile apps to combat obesity in children and adolescents: A systematic literature review. J Spec Pediatr Nurs 2016 Jan;21(1):5-17. [doi: 10.1111/jspn.12134] [Medline: 26494019]
45. 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]
46. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
47. Schoffman DE, Turner-McGrievy G, Jones SJ, Wilcox S. Mobile apps for pediatric obesity prevention and treatment, healthy eating, and physical activity promotion: just fun and games? Transl Behav Med 2013 Sep;3(3):320-325 [FREE Full text] [doi: 10.1007/s13142-013-0206-3] [Medline: 24073184]
48. Stuckey MI, Carter SW, Knight E. The role of smartphones in encouraging physical activity in adults. Int J Gen Med 2017;10:293-303 [FREE Full text] [doi: 10.2147/IJGM.S134095] [Medline: 28979157]
49. Vlahu-Gjorgievska E, Mulakaparambil US, Win K. mHealth applications: A tool for behaviour change in weight management. In: Studies in Health Technology and Informatics. Amsterdam, Netherlands: IOS Press; 2018:158-163.
50. Wang Y, Sun L, Xu Y, Xiao Q, Chang P, Wu Y. Comparision and analysis of top 10 exercise android apps in mainland China. Stud Health Technol Inform 2015;216:1024. [Medline: 26262324]
51. Wearing JR, Nollen N, Befort C, Davis AM, Agemy CK. iPhone app adherence to expert-recommended guidelines for pediatric obesity prevention. Child Obes 2014 Apr;10(2):132-144 [FREE Full text] [doi: 10.1089/chi.2013.0084] [Medline: 24655230]
52. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012 May 14;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
53. Yang C, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. Am J Prev Med 2015 Apr;48(4):452-455. [doi: 10.1016/j.amepre.2014.10.010] [Medline: 25576494]
54. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
55. Soetens KCM, Vandelanotte C, de Vries H, Mummery KW. Using online computer tailoring to promote physical activity: a randomized trial of text, video, and combined intervention delivery modes. J Health Commun 2014 Dec;19(12):1377-1392. [doi: 10.1080/10810730.2014.894597] [Medline: 24749983]
56. Zell E, Alicke MD. The local dominance effect in self-evaluation: evidence and explanations. Pers Soc Psychol Rev 2010 Nov;14(4):368-384. [doi: 10.1177/1088868310366144] [Medline: 20435806]
57. Fardouly J, Diedrichs PC, Vartanian LR, Halliwell E. Social comparisons on social media: the impact of Facebook on young women's body image concerns and mood. Body Image 2015 Mar;13:38-45. [doi: 10.1016/j.bodyim.2014.12.002] [Medline: 25615425]
58. Vogel EA, Rose JP, Okdie BM, Eckles K, Franz B. Who compares and despairs? The effect of social comparison orientation on social media use and its outcomes. Pers Indiv Differ 2015;86:249-256. [doi: 10.1016/j.paid.2015.06.026]
59. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical activity, sedentary behavior, and diet-related eHealth and mHealth research: Bibliometric analysis. J Med Internet Res 2018 Apr 18;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]

60. Leahey TM, Crowther JH. An ecological momentary assessment of comparison target as a moderator of the effects of appearance-focused social comparisons. Body Image 2008 Sep;5(3):307-311. [doi: 10.1016/j.bodyim.2008.03.002] [Medline: 18585108]
61. Gerber JP, Wheeler L, Suls J. A social comparison theory meta-analysis 60+ years on. Psychol Bull 2018 Feb;144(2):177-197. [doi: 10.1037/bul0000127] [Medline: 29144145]
62. Harrison D, Marshall P, Bianchi-Berthouze N, Bird J. Activity Tracking: Barriers, Workarounds and Customisation. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing.: Association for Computing Machinery; 2015 Presented at: UbiComp'15; September 7 - 11, 2015; Osaka, Japan p. 617-621. [doi: 10.1145/2750858.2805832]
63. Gibbons FX, Buunk BP. Individual differences in social comparison: Development of a scale of social comparison orientation. J Pers Soc Psychol 1999;76(1):129-142. [doi: 10.1037/0022-3514.76.1.129]
64. Luszczynska A, Gibbons FX, Piko BF, Tekozel M. Self-regulatory cognitions, social comparison, and perceived peers' behaviors as predictors of nutrition and physical activity: a comparison among adolescents in Hungary, Poland, Turkey, and USA. Psychol Health 2004;19(5):577-593. [doi: 10.1080/0887044042000205844]
65. Arigo D, Suls JM. Smartphone apps providing social comparison for health behavior change: a need for better tailoring to person and context. Mhealth 2018;4:46 [FREE Full text] [doi: 10.21037/mhealth.2018.09.10] [Medline: 30505844]
66. Bennenbroek FT, Buunk BP, van der Zee KI, Grol B. Social comparison and patient information: what do cancer patients want? Patient Educ Couns 2002 May;47(1):5-12. [doi: 10.1016/s0738-3991(02)00018-6] [Medline: 12023095]
67. DeVellis RF, Holt K, Renner BR, Blabock SJ, Blanchard LW, Cook HL, et al. The relationship of social comparison to rheumatoid arthritis symptoms and affect. Basic Appl Soc Psych 1990;11(1):1-18. [doi: 10.1207/s15324834basp1101_1]
68. Arigo D, Cornell M, Smyth JM. Social comparisons in adults with type 2 diabetes: Patients' reasons for target selection. Psychol Health 2018 Jul;33(7):906-924. [doi: 10.1080/08870446.2018.1432759] [Medline: 29424244]
69. French D, Sutton S, Marteau T, Kinmonth A. The impact of personal and social comparison information about health risk. Br J Health Psychol 2004 May;9(Pt 2):187-200. [doi: 10.1348/135910704773891041] [Medline: 15125804]
70. Johnson BK, Knobloch-Westerwick S. Glancing up or down: Mood management and selective social comparisons on social networking sites. Comput Hum Behav 2014;41:33-39. [doi: 10.1016/j.chb.2014.09.009]
71. Arigo D, Mogle JA, Brown MM, Pasko K, Travers L, Sweeder L, et al. Methods to assess social comparison processes within persons in daily life: A scoping review. Front Psychol 2019;10:2909 [FREE Full text] [doi: 10.3389/fpsyg.2019.02909] [Medline: 32038352]
72. Direito A, Walsh D, Hinbarji M, Albatal R, Tooley M, Whittaker R, et al. Using the intervention mapping and behavioral intervention technology frameworks: Development of an mHealth intervention for physical activity and sedentary behavior change. Health Educ Behav 2018 Jun;45(3):331-348. [doi: 10.1177/1090198117742438] [Medline: 29216765]
73. Nahum-Shani I, Smith S, Spring B, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
74. Gerrard M, Gibbons FX, Lane DJ, Stock ML. Smoking cessation: social comparison level predicts success for adult smokers. Health Psychol 2005 Nov;24(6):623-629. [doi: 10.1037/0278-6133.24.6.623] [Medline: 16287409]
75. Mahler HI, Kulik JA, Gerrard M, Gibbons FX. Effects of upward and downward social comparison information on the efficacy of an appearance-based sun protection intervention: a randomized, controlled experiment. J Behav Med 2010 Dec;33(6):496-507 [FREE Full text] [doi: 10.1007/s10865-010-9279-3] [Medline: 20652391]

Abbreviations

BCT: behavior change technique
mHealth: mobile health
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SC: social comparison

Edited by G Eysenbach; submitted 30.07.19; peer-reviewed by FJ Yang, R Ciptaningtyas, E Kozasa; comments to author 07.12.19; revised version received 06.01.20; accepted 24.01.20; published 27.03.20.

Please cite as:
Arigo D, Brown MM, Pasko K, Suls J
Social Comparison Features in Physical Activity Promotion Apps: Scoping Meta-Review
J Med Internet Res 2020;22(3):e15642
URL: http://www.jmir.org/2020/3/e15642/
doi:10.2196/15642
PMID:32217499

Original Paper

Technological, Organizational, and Environmental Factors Influencing Social Media Adoption by Hospitals in Switzerland: Cross-Sectional Study

Michael Beier[1], Dr rer pol; Sebastian Früh[1], MSc

[1]University of Applied Sciences of the Grisons, Chur, Switzerland

Corresponding Author:
Michael Beier, Dr rer pol
University of Applied Sciences of the Grisons
Pulvermühlestrasse 57
Chur, 7000
Switzerland
Phone: 41 81 286 3755
Email: michael.beier@fhgr.ch

Abstract

Background: Social media platforms are important tools for hospitals. These platforms offer many potential benefits in various areas of application for hospitals to connect and interact with their stakeholders. However, hospitals differ immensely in their social media adoption. There are studies that provide initial findings on individual factors influencing social media adoption by hospitals, but there is no comprehensive and integrated model.

Objective: This study aimed to develop a comprehensive model of social media adoption by hospitals in the context of the Swiss health care system and to test the model with empirical data from Switzerland.

Methods: To develop our model, we applied the general technology-organization-environment framework of organizational technology adoption and adapted it to the specific context of social media adoption by hospitals in Switzerland. To test our model, we collected empirical data on all 283 hospitals in Switzerland and identified the accounts they operate on 7 different social media platforms (Facebook, Google+, Twitter, Instagram, LinkedIn, XING, and YouTube). We tested the hypotheses of our model by means of binary logistic regression (dependent variable: platform adoption) and negative binomial regression (dependent variable: number of different platforms adopted).

Results: Our general model on social media adoption received broad support. Overall, hospitals in Switzerland are more likely to adopt social media if they have a higher share of patients with voluntary health insurance or have a higher patient volume. In contrast, they are less likely to operate their own social media accounts if they are associated with a hospital network. However, some hypotheses of our model received only partial support for specific social media platforms; for instance, hospitals in Switzerland are more likely to adopt XING if they provide an educational program and are more likely to adopt LinkedIn if they are located in regions with higher competition intensity.

Conclusions: Our study provides a comprehensive model of social media adoption by hospitals in Switzerland. This model shows, in detail, the factors that influence hospitals in Switzerland in their social media adoption. In addition, it provides a basic framework that might be helpful in systematically developing and testing comprehensive models of social media adoption by hospitals in other countries.

(J Med Internet Res 2020;22(3):e16995) doi:10.2196/16995

KEYWORDS

social media; social media adoption; hospitals; Switzerland; organizational technology adoption; TOE framework

Introduction

Background

Social media has changed the way actors in health care interact and relate with each other in numerous ways [1]. One key actor in this field is hospitals, which have a central position in regional and national health care systems. Social media provides hospitals with many potential benefits in various areas of application but also forces them to adapt the ways in which they connect with their stakeholders. In this regard, hospitals might use social media platforms as a communication and marketing tool to reach new patients, provide information about current health topics, establish their presence in the general public, or present their service offerings [2-5]. In addition, hospitals can apply their social media channels to enhance service delivery to their patients [6-8]. Hospitals often use social media channels for educational purposes in the teaching and training of medical students and doctors and apply these channels for employer branding and recruiting [9-11]. Finally, as Web content on health care topics often is of questionable quality, hospitals' social media channels might also act as trustworthy curators and reliable sources for online health information in the society [1,6,8].

In contrast to the outlined importance of the topic, research on social media adoption by hospitals is still in its infancy. Initial research in this field focused on the nationwide adoption rates of hospitals with regard to specific social media platforms [8,12-16]. Comparisons of the findings of these studies show remarkable differences in the hospitals' rates of social media adoption among different countries, regions, hospital types, and social media platforms. However, deeper insights into the interplay of factors influencing social media adoption by hospitals, which could explain these differences, are scarce so far. Exploratory studies provide initial findings on individual factors influencing the adoption of social media platforms by hospitals [8,14,15]. Nevertheless, there is no comprehensive and integrated model of social media adoption by hospitals. Such a model would allow us to better understand the differences in the adoption rates described earlier. In addition, a deeper understanding of the quantitative patterns of social media adoption by hospitals would support and supplement further qualitative and quantitative research on the strategies, processes, and content of social media usage by hospitals.

In this study, we developed a comprehensive model on social media adoption by hospitals in the context of the Swiss health care system and tested the model with empirical data from Switzerland. To develop the model, we applied the general technology-organization-environment (TOE) framework of organizational technology adoption [17] and adapted it to the specific context of social media adoption by hospitals. Furthermore, we considered relevant aspects of the Swiss health care system as we assumed that some factors of influence on social media adoption by hospitals can be understood only by considering specific characteristics of the surrounding health care system [18,19].

The Swiss health care system ranks high in many indicators and is highly valued by patients. Nevertheless, it is challenged by high and rising costs and faces physician shortage [20]. In 2012, Switzerland introduced a case-based remuneration scheme (Swiss Diagnosis Related Groups, DRGs) for hospital inpatient services. In this scheme, hospitals are reimbursed a certain fee depending on the patients' diagnoses and region (canton) of residence [19]. Before this change, most patients could only choose a hospital in their home region (canton). With the new system, patients have more freedom to choose a hospital nationwide. However, there are differences between the 2 main types of insurance in Switzerland [20,21]. The first insurance type is mandatory for everybody (Mandatory Health Insurance) and covers all general services of the health care system in accordance with the regional DRG rate. The second type is complementary and voluntary (Voluntary Health Insurance, VHI) and provides patients who pay an additional insurance premium enhanced reimbursements and advanced services, such as free choice of hospitals nationwide, single rooms with a higher level of hospital accommodation, and free choice of hospital doctors. Most hospitals in Switzerland provide services to patients with both types of health insurance. However, patients with VHI allow hospitals to generate revenues in addition to the DRG rates of their medical diagnoses.

Research Model and Hypotheses

A hospital's decision to officially run its own account on a specific social media platform can be seen as a specialization of organizational technology adoption. Within the general field of organizational technology adoption, Tornatzky and Fleischer [17] developed the TOE framework, which is a generic framework that comprehensively covers potential areas of relevant influence on organizational technology adoption structured by 3 different contexts: a technological context, an organizational context, and an environmental context [22]. The TOE framework has been extensively applied to study hospitals' adoption of various information technologies [23-25]. A strength (and weakness) of the TOE framework is its general applicability [22]. On the one hand, it can be adapted to all technologies used by organizations. On the other hand, its high level of abstraction requires sufficient adjustments and specifications to allow meaningful application to a specific technology in a specific industry.

Therefore, in this study, we developed an integrated research model on social media adoption by hospitals by following the TOE approach. We developed hypotheses on the factors influencing social media adoption with regard to the specific technological, organizational, and environmental contexts of hospitals. For this purpose, we systematically adapted the fundamental concepts of the TOE approach to social media as the technology to be adopted and to the specific context of hospitals. At certain points, we also relied on specific characteristics of the Swiss hospital system.

Technological Context

The technological context of the TOE framework addresses questions on the benefits and costs related to the adoption of a new technology [26,27]. The perceived benefits of new information technology can comprise assumed possibilities for the generation of a relative competitive advantage by its application [24,28] or enhanced processes because of internal improvements or better collaboration with external partners [29,30]. In contrast, perceived costs mainly arise from expected integration costs and barriers in accordance with existing internal or external technologies in use [26,27]. Therefore, a lack of relevant competencies or resources leads to higher perceived costs and barriers [31,32].

With regard to the outlined calculus of organizational technology adoption, social media platforms are a unique technology. From a technological cost perspective, social media appears rather simple as the underlying third-party platforms are readily available at no cost and are user friendly [33]. Therefore, the technological affordances of social media platforms for organizations and their employees may be perceived as low compared with those of more complex and expensive information systems. However, significant integration costs and barriers are mainly based on the internal processes of communication, marketing, and compliance, which must be adapted to the affordances of social media communication. For hospitals, the potential benefits of social media adoption primarily lie in the enhanced interactions and relationships with relevant stakeholders [34,35]. Correspondingly, questions of internal compatibility and capacities for social media adoption by hospitals are mainly strategic and are operational questions regarding marketing and communication [36]. Therefore, hospitals might decide to run their own accounts on social media platforms based on their expectations of the extent to which doing so will improve the effectiveness or efficiency of relevant stakeholder interactions and relationships [37-39].

Patients with additional VHI represent one external stakeholder group that is very important for many hospitals (in Switzerland) and is accessible via social media platforms. This group is very attractive for hospitals as health insurance compensates hospitals beyond the DRG rate for advanced accommodation and other amenities [20]. Patients with VHI have more options in their hospital choice and are more strongly influenced by the nonmedical characteristics of hospitals, eg, ambience, accommodation, and comfort [40,41]. Similar to the hotel sector, these characteristics are particularly suitable for an effective and efficient presentation in social media [42,43]. Correspondingly, we hypothesized that the more important patients with VHI are for a hospital, the higher is its propensity to adopt social media. We thus proposed the following hypothesis:

> *H1a: The higher the share of patients with VHI in hospitals, the more likely they are to adopt social media.*

The education and training sector is another important field in which hospitals might seek to improve their interactions and relations with their stakeholders via social media [14]. Social media is particularly suited for health care organizations to relate and interact with communities of medical students and practitioners [9,44]. Social media platforms provide effective channels for hospitals to communicate with potential, actual, and former participants of their educational programs [45]. Therefore, it can be expected that hospitals that provide such programs are more likely to run their own social media accounts. In alignment with this assumption, findings from studies in the United States and China show that hospitals that are involved in graduate medical education or affiliated with a university are more likely to adopt social media [8,14,15]. Thus, we proposed the following hypothesis:

> *H1b: Hospitals that provide an educational program are more likely to adopt social media.*

Organizational Context

The organizational context of the TOE framework refers to the organizations’ internal structures and processes that may facilitate or constrain the adoption of new information technology [18,46]. One basic organizational factor included in many TOE studies on the adoption of technology in hospitals is the hospital size [23,24,47,48]. The findings of these studies are mainly similar, showing that larger hospitals adopt new technologies faster. Similarly, initial exploratory findings in the United States show that larger hospitals (measured by the number of beds) are more likely to adopt social media than smaller ones [14,15]. However, in organizational technology adoption, hospital size covers two different arguments. On the one hand, hospital size is seen as an indicator of the extent of internal infrastructure. In this regard, a certain size means that a hospital has all necessary resources and capabilities as well as the ability to set up additional assets to apply new technology adequately [23,24,48]. On the other hand, hospital size is seen as an indicator for the operational volume of patients served by the hospital [47,49,50]. In these cases, some technologies can be applied effectively or efficiently only if a hospital has a certain volume of patients. The adoption of social media platforms does not necessitate any internal infrastructure for hospitals but mainly shows benefits of enhanced effectiveness or efficiency depending on the size of the respective stakeholder groups with which a hospital wants to connect [51-53]. Therefore, regarding social media adoption by hospitals, the organization size might be more a matter of operational patient volume than of infrastructure capacities. Consequently, we proposed the following hypothesis:

> *H2a: The higher the patient volume in hospitals, the more likely they are to adopt social media.*

Another organizational factor of technology adoption refers to internal structures [54]. One important aspect of hospital structure in this context is centralization [18,49]. Early studies argue that technology adoption is easier with decentralized structures when the technology is only locally applied within a hospital (for instance, in a specific department). In contrast, adoption decisions for technologies that affect a whole hospital benefit from centralized decision-making rights [50]. As social media adoption raises questions of ownership within an organization, centralized decision making should be beneficial for social media adoption decisions [55]. Accordingly, exploratory findings from outside the hospital sector show that centralized leadership fosters social media adoption in organizations as it makes resource allocations and decision-making processes easier [56]. In Switzerland, there is a tendency to form consolidated hospital organizations that run several hospital sites [20]. Currently, approximately 30% of the hospitals in Switzerland run at more than one site [57]. For a hospital running at several sites, each with its own directorates, it might be more difficult or take longer to reach an agreement on the adoption of social media. On the one hand, more individual interests of different parties and differences in local needs must be considered [18]. On the other hand, more decision makers have the opportunity to block a decision on social media adoption because of risk concerns [33]. Overall, we hypothesized that a more complex and dispersed structure of a hospital organization is related to a lower propensity for social media adoption. Thus, we proposed the following hypothesis:

> *H2b: The more sites hospitals run, the less likely they are to adopt social media.*

Environmental Context

The environmental context of the TOE framework refers to all factors outside an organization that may facilitate or constrain the adoption of new information technology [22]. One important factor in this context is external partners and their influence on the organizations' decision making regarding technology adoption [37]. In this regard, the hospitals' decision to adopt new information technology can be influenced by their association with a parent group or hospital network (eg, Hirslanden or the Swiss Medical Network in Switzerland). Such groups are seen as effective communicators and facilitators of innovations for their member hospitals, and they often provide technological support and shared services to implement new technologies [23]. Similarly, exploratory studies have shown that hospitals that are affiliated with a health or hospital system are more likely to be present on social media [14,15]. However, these exploratory studies do not explicitly focus on the organizational social media adoption by individual hospitals. Thus, they do not explicitly differentiate between cases where a hospital runs its own social media account on a platform versus cases where a hospital is only co-represented in the social media account of a parent group, hospital network, or health system. With regard to organizational social media adoption, however, this difference is significant. Correspondingly, regarding individual hospitals and their social media adoption, we hypothesized a negative influence of their affiliation to a hospital network. On the one hand, in some cases, parent networks run social media accounts as shared services representing their member hospitals with social media accounts of the whole network [58]. Therefore, for a focal hospital, the perceived benefits of running its own account (in addition to the network account) on a social media platform should be diminished compared with the situation of an independent hospital. On the other hand, some hospital networks might apply a centralized communication strategy that does not allow member hospitals to run their own accounts in specific social media platforms. Both patterns reduce the propensity of an individual hospital to run its own social media accounts if it is affiliated with such a network. Thus, we proposed the following hypothesis:

> *H3a: Hospitals that are affiliated with a hospital network are less likely to adopt social media.*

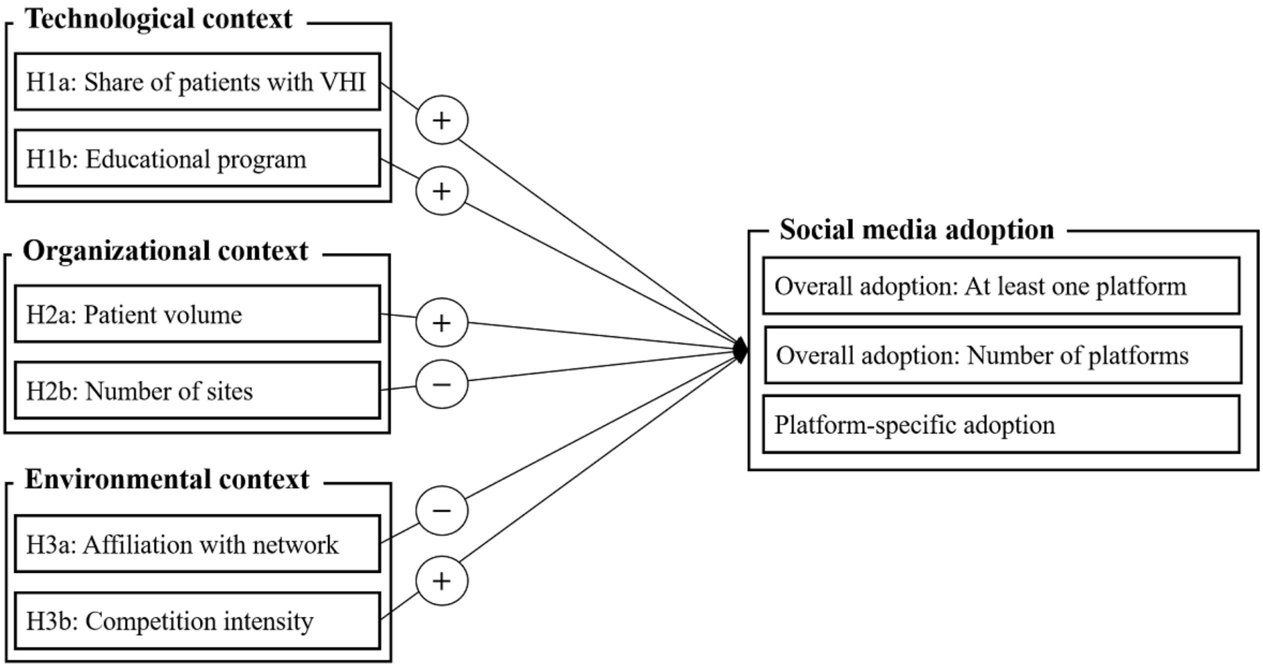
Another important environmental factor of technology adoption is competition intensity [59]. Several studies on information technology adoption by hospitals have observed a positive relationship between the level of competition intensity that hospitals face and their propensity to adopt new information technology [47,60,61]. On the one hand, in an environment with high competition intensity, organizations try to change the rules of the competition by adopting new technology and try to decrease the threat of competitors by leveraging new ways to outperform their rivals [31,62]. On the other hand, under higher competition intensity, managers perceive higher peer pressure to adopt new technologies [18,63]. Initial research on general social media adoption proposes a similar relationship between the competition intensity and social media adoption of organizations [64]. Thus, we proposed the following hypothesis:

> *H3b: The higher the competition intensity that hospitals face, the more likely they are to adopt social media.*

In our research model, we analyzed social media adoption at the organizational level of hospitals. However, in the model, we also covered two other organizational levels as attributes of these hospitals. On the one hand, we incorporated lower-level units of these hospitals by analyzing the number of sites they run (H2b) as an attribute of their internal structure (organizational context). On the other hand, we included higher-level entities by analyzing if the hospitals are affiliated with a hospital network or parent group (H3a) as an attribute of their external partner relationships (environmental context).

Our research model covered social media adoption as the dependent variable of our hypotheses in 3 different ways: first, the binary overall adoption of at least one social media platform; second, the total number of different social media platforms adopted, covering the breadth of overall adoption; and third, the binary adoption of specific social media platforms. Figure 1 presents an overview of our integrated research model.

Figure 1. The research model. VHI = voluntary health insurance. The + sign indicates a hypothesized positive relation. The − sign indicates a hypothesized negative relation.



Methods

Data Collection

To test our research model, we used data on all Swiss hospitals from several different sources. We collected data on the overall population of hospitals in Switzerland. To achieve this, we applied the official list of key figures for all Swiss hospitals from the Federal Office of Public Health (FOPH) in Switzerland [57]. The list provided us with all relevant data on the hospitals' characteristics needed for our study. Additional data on the main languages in different regions (cantons) of Switzerland were gathered from the Federal Statistical Office (FSO) of Switzerland [65]. Finally, we used the FOPH list to collect data on the social media presence of all hospitals in Switzerland on different social media platforms.

In line with previous studies [14-16], we collected the social media data for all hospitals in 3 steps. In the first step, we performed a Google search for the official homepages of all hospitals on the FOPH list. During this step, we had to reduce our sample from 283 hospitals on the FOPH list to 279 hospitals. Two of the hospitals on the list had closed. Another hospital organization is listed as 2 suborganizations on the FOPH list. As it was not possible to merge the data of these 2 suborganizations with the social media data, we had to exclude these 2 entries from our sample. In the second step, we gathered all social media accounts that were linked from the official homepages of hospitals in our sample. On the basis of these links, we identified the relevant social media platforms for this study. We included all social media platforms where at least 10 hospitals in our sample provided a link on their official homepage to their own social media account on the platform. Correspondingly, in further data collection, we included 7 platforms: Facebook (126 links from individual hospital homepages to their accounts on the platform), YouTube (89 links), LinkedIn (62 links), Twitter (53 links), XING (36 links), Instagram (25 links), and Google+ (15 links). In the third step, we completed our systematic search for social media accounts of the hospitals in our sample. To achieve this, we performed a Google search for the name of each hospital in combination with the name of each platform. In addition, we searched each social media platform for the accounts of each hospital. To exclude informal social media accounts operated under the name of hospitals, we applied a low-threshold minimal criterion. Therefore, in our data collection, we did not include unofficial Facebook pages (pages for public places automatically generated by the platform itself) or accounts on the other platforms where no background picture or profile picture was uploaded and no *About Us* text was provided. The final social media data were gathered from September 17 to 19, 2018. We collected only publicly available data from the social media channels of hospitals. No persons were directly involved. The dataset generated and analyzed in this study is available from the corresponding author on reasonable request.

Measures

With regard to the independent variables of our research model, we measured the hospitals' share of patients with VHI by the percentage of patients with VHI among all hospital patients, operationalized as a value between 0 and 100. We used a dummy variable to indicate whether a hospital ran an educational program. This variable took the value of 1 if a hospital was listed on the FOPH list as a provider of educational programs for medical students or doctors and 0 if not. We measured the patient volume of hospitals in values of 1000 patients per year. The number of sites a hospital ran was operationalized as a hospital's number of locations listed in the FOPH list. We operationalized the hospitals' affiliation to a network with a dummy variable. This variable took the value of 1 for all hospitals whose online communication included any indication that the hospital was associated with a parent group or hospital network (shared homepage, any shared social media channel, hyperlink to the parent group from any online channel, or listed as a member on the group website). In all cases without such an indication, the dummy variable took the value of 0. As competition in the Swiss hospital market is mainly focused within regions (cantons) [20], competition intensity was considered as the number of hospitals on the FOPH list that were active in the same region (canton).

We operationalized social media adoption as the dependent variable in our research model in 3 different ways. First, we used 7 dummy variables (one for each social media platform) to measure platform-specific social media adoption by hospitals. For each platform, the dummy variable took the value of 1 if a hospital operated its own account on the platform (not via a hospital group as a representative of the hospital) and 0 if this was not the case. Second, we operationalized the binary overall adoption of social media by hospitals with an additional dummy variable if a hospital had a social media presence in at least one of the platforms included in our study. Third, we measured the total number of different platforms wherein a hospital operated its own accounts as the sum of the dummy variables of the specific platforms with a numerical variable.

We applied additional measures to control for further influences on the hospitals' social media adoption. Prior results indicate that the ownership of hospitals (especially privately vs publicly owned hospitals) might influence their social media adoption [14,15]. Therefore, we controlled for public hospitals with a dummy variable that took the value of 1 if a hospital was listed with the legal status of a public company in the FOPH list and 0 in all other cases. As Switzerland is a multilingual country with different language regions (German, French, and Italian as main languages in specific cantons), we also used dummy variables to control for possible language-related effects. With one dummy variable each, we indicated whether a hospital was located in a canton where the main language was German, French, or Italian, as classified by the FSO of Switzerland [65].

Analytical Approach

As described earlier, the dependent variables of our research model are operationalized on the basis of 8 binary dummy variables (7 dummies for specific platforms and 1 for any platform) and 1 numeric variable (the total number of different platforms). Correspondingly, for each binary dependent variable, we computed a separate binary logistic regression model. The total number of different social media platforms used by a hospital shows the characteristics of count data. Furthermore, we found significant evidence of overdispersion in the data [66]. Therefore, we tested our hypotheses with regard to the total number of different social media platforms as the dependent variable by means of a negative binomial regression. All analyses were performed in R (The R Foundation for Statistical Computing, Vienna, Austria). Details are provided in Multimedia Appendix 1.

We also tested for multicollinearity issues by computing correlations (Kendall tau) between all explanatory variables in our analyses. As all correlations were less than 0.45, no multicollinearity issues were indicated regarding our regression analyses [67].

Results

Descriptive Results

In Table 1, we present the respective adoption rates of all binary dependent variables of our research model. First, we display the overall social media adoption rate of hospitals using at least one social media platform (Any). In the other rows, we present the adoption rates for the specific social media platforms included in our data collection. The data show that 74.6% (208/279) of all hospitals in Switzerland run their own social media accounts on at least one platform. The adoption rates of specific platforms range from a maximum of 58.4% (163/279) (Facebook) to a minimum of 13.3% (37/279) (Instagram).

Table 1. Rates of social media adoption by Swiss hospitals (N=279).

Platform	Adoption rate, n (%)
Any	208 (74.6)
Facebook	163 (58.4)
Google+	79 (28.3)
Twitter	57 (20.4)
Instagram	37 (13.3)
LinkedIn	114 (40.9)
XING	57 (20.4)
YouTube	98 (35.1)

Furthermore, 59.5% (166/279) of the hospitals in our sample run an educational program, 23.3% (65/279) of them are affiliated with a hospital network, and 20.4% (57/279) are public hospitals. Most hospitals in our sample are located in a German-speaking region (207/279, 74.2%), 21.1% (59/279) are located in a French-speaking region, and 4.7% (13/279) are located in an Italian-speaking region.

For all other nonbinary variables, we provide descriptive statistics in Table 2. The hospitals' share of patients with additional VHI range between 0% and 100%, with a mean of 25.7%. The patient volumes per year for the hospitals in our sample show a maximum of 55,200 patients and a mean of 4940 patients. Hospitals run at a maximum of 22 sites, with a mean of 2.01. Competition intensity varies from no competition to 42 hospitals in the same region (canton). The total number of different social media platforms used by a hospital (dependent variable) ranges from 0 to 7, with a mean of 2.17.

Table 2. Descriptive results of nonbinary variables (N=279).

Variable	Mean (SD)	Median	Minimum value	Maximum value
H1a: Share of patients with voluntary health insurance (%)	25.70 (26.400)	19.22	0.00	100.00
H2a: Patient volume (in 1000)	4.94 (8.445)	1.56	0.01	55.20
H2b: Number of sites	2.01 (2.663)	1.00	1.00	22.00
H3b: Competition intensity	20.22 (12.160)	16.00	1.00	42.00
Dependent variable: Number of platforms	2.17 (1.935)	2.00	0.00	7.00

Regression Results

We have presented our regression results in 9 models (results of models 1 to 3 are provided in Table 3, results of models 4 to 6 in Table 4, and results of models 7 to 9 in Table 5). Model 1 covers the logistic regression results of the binary overall adoption of at least one social media platform (any platform) by the hospitals. Model 2 displays the negative binomial regression results for the total number of different social media platforms adopted by hospitals (number of platforms). Models 3 to 9 show a binary logistic regression model of platform-specific social media adoption for each individual social media platform.

The results for the binary logistic regression models (all models except model 2) are presented as odds ratios (ORs). The ORs indicate the expected changes in the hospitals' odds of social media adoption when the respective explanatory variable changes by one unit. Correspondingly, we have displayed the results for the negative binomial regression in model 2 (on the number of different platforms adopted) as incidence rate ratios (IRRs). Similar to the ORs, the IRRs indicate the expected changes in the number of different social media platforms that the hospitals adopt when the respective explanatory variable changes by one unit. For all regression models, we displayed the 95% confidence intervals of the ORs or IRRs and *P* values for each explanatory variable as well as the Nagelkerke R^2 for the overall regression model. For the categorical data on language regions, we applied German-speaking regions as a reference category.

Table 3. Regression results of models 1 to 3.

Variable	Model 1: Any platform[a]		Model 2: Number of platforms[b]		Model 3: Facebook[c]	
	OR[d] (95% CI)	*P* value	IRR[e] (95% CI)	*P* value	OR (95% CI)	*P* value
H1a: Share of patients with voluntary health insurance	1.01 (1.00[f]-1.03)	.02	1.01 (1.00[f]-1.01)	<.001	1.01 (1.00[f]-1.02)	.03
H1b: Educational program	1.37 (0.72-2.60)	.33	1.13 (0.90-1.40)	.30	0.95 (0.54-1.67)	.85
H2a: Patient volume	1.08 (1.01-1.15)	.02	1.03 (1.02-1.04)	<.001	1.06 (1.01-1.11)	.009
H2b: Number of sites	1.00 (0.89-1.13)	.97	0.96 (0.92-1.00)	.06	0.87 (0.77-0.99)	.03
H3a: Affiliation with network	0.28 (0.14-0.53)	<.001	0.47 (0.36-0.61)	<.001	0.54 (0.30-0.98)	.04
H3b: Competition intensity	1.00 (0.97-1.02)	.80	1.00 (1.00-1.01)	.48	1.00 (0.98-1.03)	.76
Public hospital	0.88 (0.38-2.06)	.77	1.25 (0.97-1.61)	.09	1.38 (0.66-2.90)	.39
French-speaking region	0.65 (0.32-1.34)	.25	0.81 (0.63-1.04)	.09	0.66 (0.35-1.26)	.21
Italian-speaking region	0.36 (0.10-1.30)	.12	0.66 (0.40-1.08)	.09	0.95 (0.29-3.15)	.93
Constant	2.31 (1.06-5.03)	.04	1.59 (1.21-2.10)	.001	1.20 (0.60-2.41)	.61

[a] Nagelkerke R^2=0.151.
[b] Nagelkerke R^2=0.248.
[c] Nagelkerke R^2=0.112.
[d] OR: odds ratio.
[e] IRR: incidence rate ratio.
[f] The value is greater than 1 but rounds to 1.00.

Table 4. Regression results of models 4 to 6.

Variable	Model 4: Google+[a]		Model 5: Twitter[b]		Model 6: Instagram[c]	
	OR[d] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
H1a: Share of patients with voluntary health insurance	1.02 (1.00[e]-1.03)	.01	1.02 (1.01-1.04)	.002	1.02 (1.01-1.04)	.002
H1b: Educational program	0.85 (0.45-1.61)	.61	1.16 (0.51-2.67)	.72	0.74 (0.29-1.89)	.53
H2a: Patient volume	1.04 (1.00[e]-1.08)	.03	1.11 (1.05-1.16)	<.001	1.08 (1.03-1.13)	<.001
H2b: Number of sites	0.96 (0.85-1.08)	.46	0.85 (0.68-1.06)	.14	0.83 (0.60-1.14)	.25
H3a: Affiliation with network	0.50 (0.25-1.03)	.06	0.07 (0.02-0.28)	<.001	0.21 (0.07-0.70)	.01
H3b: Competition intensity	1.00 (0.98-1.03)	.84	0.99 (0.96-1.01)	.30	0.99 (0.96-1.02)	.63
Public hospital	2.06 (0.96-4.42)	.06	1.97 (0.81-4.76)	.13	1.46 (0.50-4.28)	.49
French-speaking region	0.34 (0.15-0.78)	.01	0.70 (0.28-1.74)	.44	0.82 (0.30-2.19)	.69
Italian-speaking region	0.57 (0.14-2.34)	.44	0.38 (0.06-2.31)	.29	0.00 (0.00-∞)	.99
Constant	0.30 (0.14-0.66)	.003	0.18 (0.07-0.50)	<.001	0.12 (0.04-0.37)	<.001

[a] Nagelkerke R^2=0.125.

[b] Nagelkerke R^2=0.305.

[c] Nagelkerke R^2=0.225.

[d] OR: odds ratio.

[e] The value is greater than 1 but rounds to 1.00.

Table 5. Regression results of models 7 to 9.

Variable	Model 7: LinkedIn[a]		Model 8: XING[b]		Model 9: YouTube[c]	
	OR[d] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
H1a: Share of patients with voluntary health insurance	1.01 (1.00[e]-1.02)	.03	1.01 (0.99-1.03)	.26	1.01 (1.00[e]-1.02)	.04
H1b: Educational program	1.41 (0.77-2.58)	.26	2.69 (1.11-6.53)	.03	1.16 (0.60-2.28)	.66
H2a: Patient volume	1.06 (1.02-1.10)	.004	1.10 (1.04-1.16)	<.001	1.10 (1.05-1.15)	<.001
H2b: Number of sites	0.94 (0.84-1.05)	.27	0.96 (0.85-1.08)	.48	0.97 (0.87-1.08)	.55
H3a: Affiliation with network	0.44 (0.23-0.85)	.01	0.13 (0.04-0.45)	.001	0.05 (0.01-0.16)	<.001
H3b: Competition intensity	1.03 (1.01-1.05)	.01	1.01 (0.98-1.04)	.44	1.00 (0.98-1.02)	.95
Public hospital	1.23 (0.59-2.57)	.58	1.76 (0.73-4.25)	.21	1.50 (0.67-3.34)	.32
French-speaking region	2.48 (1.28-4.80)	.007	0.02 (0.00-0.25)	.003	0.44 (0.20-0.99)	.05
Italian-speaking region	0.75 (0.20-2.73)	.66	0.00 (0.00-∞)	.98	0.51 (0.13-2.09)	.35
Constant	0.19 (0.09-0.41)	<.001	0.09 (0.03-0.27)	<.001	0.43 (0.19-0.95)	.04

[a] Nagelkerke R^2=0.173.

[b] Nagelkerke R^2=0.405.

[c] Nagelkerke R^2=0.316.

[d] OR: odds ratio.

[e] The value is greater than 1 but rounds to 1.00.

Technological Context

Hypotheses H1a and H1b of our research model addressed the technological context of social media adoption by hospitals. The general proposition of our model in this context was that hospitals are more likely to run their own account on a social media platform when they have higher expectations that this platform provides benefits for their organizational communication with stakeholder groups that are individually important for them. In this regard, hypothesis H1a suggested that hospitals with a higher share of patients with VHI are more likely to adopt social media. We found broad support for hypothesis H1a in our regression results that showed significant positive effects of the hospitals' share of patients with VHI on their social media adoption. More specifically, we found significant positive effects for the adoption of at least one social media platform (model 1), the total number of different platforms adopted (model 2), and the platform-specific adoption of all platforms included in our study except the online business-related social network, XING (model 8). This result seemed plausible as XING is a regional business platform focused on German-speaking countries [68] and may be less helpful for relationship marketing and community building with patients. In contrast to the result for XING, the regression result for hospitals' adoption of LinkedIn (the other business platform included in our study) showed a significant positive effect of their share of patients with VHI (model 7). This difference might be explained by 2 factors. First, LinkedIn is a global business platform that facilitates connections to wealthy and internationally oriented patients all over the world, whereas XING is mainly focused only on German-speaking countries. Second, hospitals might use LinkedIn as an indirect way to attract patients. Instead of reaching patients directly via social media, hospitals might aim to reach national and international physicians and other health experts in their role as referring doctors or influencers in the field [69,70]. In Switzerland, patients do not need a referral from their attending physician to choose a hospital in all cases [20], but these physicians still have an important influence on the patients' hospital choices [70,71].

Another important sector in which hospitals can establish and maintain beneficial relationships with external parties via social media is the education and training sector [14]. Correspondingly, hypothesis H1b suggested that hospitals providing an educational program are more likely to adopt social media. We found only partial support for hypothesis H1b as our regression results showed only a significant positive effect of the hospitals' provision of an educational program on their propensity to run an own XING account (OR 2.69, 95% CI 1.11-6.53; *P*=.03). However, this result seems convincing, as one main focus of XING is the market for coaching and training [72]. In contrast, reaching business-related stakeholders in a private context on general purpose social media platforms might seem less beneficial for hospitals. Similar patterns have already been observed regarding social media applications in business-to-business marketing [73].

Overall, the results support the general perspective we developed in our research model regarding the technological context of hospitals' social media adoption. The main drivers in this context are the potential benefits that hospitals can derive from their social media presence. However, the concrete benefits that hospitals perceive as relevant are influenced by their individual market situation and positioning strategies. Therefore, hospitals with a stronger market focus on patients with VHI show higher adoption propensities on social media platforms for private individuals, physicians, and health professionals. In contrast, only hospitals offering education or training on the market show higher propensities to adopt a business-related social network that specializes in the education and training market.

Organizational Context

Hypotheses H2a and H2b of our research model addressed the organizational context of the hospitals' social media adoption. Hypothesis H2a stated that hospitals with a higher patient volume are more likely to run their own social media accounts. This hypothesis was fully supported by the results of all regression models we have tested. For hospitals serving more patients, additional efforts for social media communication are more cost-efficient as the costs per patient decrease with higher patient numbers. Such hospitals also tend to have more resources for communication and marketing activities. Furthermore, hospitals with more patients might also attract greater public attention and therefore be under more institutional pressure to have a social media presence [74,75].

Another aspect of the organizational context of technology adoption is the centralization and distribution of decision rights. In this regard, hypothesis H2b claimed that hospitals that have more locations and therefore have more local directorates show lower propensities for social media adoption. This hypothesis received only partial support. The results of our regression models showed only a significant negative effect of the number of hospital sites on the hospitals' propensity to run their own Facebook account (OR 0.87, 95% CI 0.77-0.99; *P*=.03). Facebook is by far the most widely used online social network among individuals in Switzerland [76] and is well known for facilitating negative word of mouth [77]. In recent years, Facebook has been involved in many scandals on data misuse and privacy breaches [78,79]. Therefore, local directorates of hospital sites might have higher risk aversion and greater fear of a potential loss of control with regard to the communication of internal and external stakeholders on this platform [80]. Correspondingly, local directorates of hospital sites may show higher propensities to veto an organization-wide adoption of Facebook. Hospitals might also be less likely to operate a Facebook account when they run more sites as they do not want to provide a public space for employees and other stakeholders to communicate on issues related to the hospital, especially among different sites [81].

Environmental Context

Hypotheses H3a and H3b of our research model covered the environmental context of social media adoption by hospitals. In this regard, hypothesis H3a proposed that hospitals that are associated with a hospital network are less likely to have their own social media accounts. We found broad support for hypothesis H3a in our regression results, which showed significant negative effects of the hospitals' affiliation with a network on their social media adoption. More specifically, we found significant negative effects for the adoption of at least one social media platform (model 1), the total number of different platforms adopted (model 2), and the platform-specific adoption of all platforms included in our study except Google+ (model 4). Our results extend previous research by explaining in more detail how affiliation with a group of hospitals influences the social media adoption of individual hospitals. Previous research showed that hospitals that are affiliated with a health or hospital system are more likely to be represented on social media by their own social media account or the account of their affiliated group [14,15]. Our study showed an opposite effect as we only considered the social media presence of accounts run by a focal hospital itself. As a concretization of previous research, we found strong support for our hypothesis that group affiliations reduce hospitals' propensity to run their own social media accounts. Indeed, affiliation with a hospital network might increase hospitals' propensity to be present on social media in some way (be it directly with their own account or indirectly via a network account). However, such an affiliation also decreases their propensity to run accounts on social media by themselves. On the one hand, social media accounts run by a hospital network might already provide sufficient benefits to their member hospitals that some of them perceive that there is no need to run individual accounts in addition. On the other hand, in cases where a hospital network applies a centralized communication strategy, member hospitals might also be encouraged by the network not to run individual accounts in addition to the collective accounts of the network. As described earlier, to address research questions on organizational social media adoption, it is essential to include only social media accounts run by a focal organization as the dependent variable. If not, by also including accounts run by an associated network, the analysis levels of an organization and its environment get mixed up.

Regional competition intensity is another aspect of the environmental context of the hospitals' social media adoption. In this regard, hypothesis H3b suggested that hospitals located in regions with higher competition intensity are more likely to run their own social media accounts. This hypothesis was supported only by the results of our binary logistic regression with regard to LinkedIn accounts (OR 1.03, 95% CI 1.01-1.05; *P*=.01). On the one hand, the explanation for this finding may be similar to that for the hospitals' orientation toward patients with VHI. Correspondingly, hospitals under more competitive pressure might not address patients directly via social media. Instead, they may establish and maintain contacts with physicians and other health experts (as referring doctors or influencers in the field) via the business-related social network, LinkedIn [69]. On the other hand, as Switzerland faces significant physician shortage, another plausible explanation for the observed pattern is that higher competition intensity among hospitals might be less a question of patients and more a question of the regional labor market for physicians [20]. In both cases, LinkedIn would be the most appropriate platform in our sample.

Discussion

Principal Findings

Previous studies provide initial empirical evidence of individual factors influencing hospitals' adoption of social media [8,14,15]. However, a comprehensive and integrated model of the factors influencing the adoption of social media by hospitals was missing. To our knowledge, this study is the first to provide such a model. To develop our model, we applied the general TOE framework [17] and adapted it systematically to the context of our study. We derived specific hypotheses on technological, organizational, and environmental factors influencing the adoption of social media platforms by hospitals. We tested our hypotheses with regard to the hospitals' overall adoption (adoption of at least one platform and the total number of platforms adopted) and platform-specific adoption with regard to the 7 social media platforms most commonly used by hospitals in Switzerland. Table 6 displays our final model with an overview of the empirical evidence we found in our data.

Table 6. Final model and empirical evidence.

Explanatory variables	Overall adoption		Platform-specific adoption							Hypothesis support
	At least one platform	Number of platforms	Facebook	Google+	Twitter	Instagram	LinkedIn	XING	YouTube	
Technological context										
H1a: Share of patients with voluntary health insurance	+[a]	+	+	+	+	+	+	NS[b]	+	Broad
H1b: Educational program	NS	NS	NS	NS	NS	NS	NS	+	NS	Partial
Organizational context										
H2a: Patient volume	+	+	+	+	+	+	+	+	+	Full
H2b: Number of sites	NS	NS	–[c]	NS	NS	NS	NS	NS	NS	Partial
Environmental context										
H3a: Affiliation with network	–	–	–	NS	–	–	–	–	–	Broad
H3b: Competition intensity	NS	NS	NS	NS	NS	NS	+	NS	NS	Partial

[a] The + sign indicates a significant positive relation (*P*<.05).

[b] NS: no significant relation (*P*>.05).

[c] The − sign indicates a significant negative relation (*P*<.05).

Our overall research model received comprehensive support. All hypotheses received at least some empirical support in the data analyses, as expected. However, our findings also allowed us to derive more detailed patterns of the platform-specific adoption of social media by hospitals. In particular, our findings regarding the technological context indicated that social media platforms should be perceived not as homogeneous technology but as a heterogeneous set of specific tools for different communication purposes with different stakeholder groups. In this regard, our results showed, for instance, that hospitals with a stronger market focus on patients with VHI tend to adopt social media platforms for private individuals (eg, Facebook, Twitter, Instagram, or YouTube), physicians, and health professionals (LinkedIn). In contrast, hospitals offering education or training on the market showed higher propensities to adopt a business-related social network that specializes in the education and training market (XING). Overall, this finding indicates that hospitals choose the social media platforms on which they should be present based on their relevant stakeholder groups. Our findings also show that hospitals with higher patient volumes generally tend to adopt social media more, independent of specific social media platforms. Finally, we found broad support for our hypothesis that individual hospitals affiliated with a hospital network or group are less likely to adopt any social media.

Limitations and Opportunities for Future Research

Our study has 3 main limitations that also indicate opportunities for future research. The first limitation is that we applied only abstract data on hospital characteristics and binary data on social media adoption. Therefore, we observed quantitative patterns of social media adoption but were not able to directly address the logic really applied during the underlying adoption decisions and adoption processes. Further (especially qualitative) research is necessary to understand in more detail how hospitals make adoption decisions and actually adopt social media. To achieve this, future studies could apply interview data of key informants who are directly involved in decision-making and implementation processes to gather more concrete information on the considerations, structures, and processes leading to social media adoption by hospitals.

The second limitation of this study is the low explanatory power of some of the regression models. Although our overall model is supported comprehensively by the data, some of the regression models have only low model strength (eg, model 3 on Facebook adoption with a Nagelkerke R^2 of 0.112), indicating that in these models our explanatory variables can only explain a small part of the variance of the dependent variable. Therefore, future research might search for further factors influencing the adoption of social media by hospitals and add them to our model. However, to this end, it might be useful (or necessary) to develop specific models for individual social media platforms.

The third limitation of this study is its limited generalizability as we focused on Switzerland, a country with a rather unique hospital sector. Indeed, it is also a strength of this study that we adapted the model to the specific context of the Swiss health care system as this allowed us to develop more concrete and context-specific hypotheses. However, this national specialization of the model also reduces the direct generalizability of our findings. Therefore, future research should adapt the model to the relevant characteristics and empirically test it in other national health care systems. In this regard, our model, as a context-specific adaptation of the general TOE approach, allows future studies to integrate further hypotheses adapted to other specific national contexts. For instance, in the technical context of our model, we identified patients with additional VHI and participants of educational programs as relevant stakeholder groups that Swiss hospitals can reach effectively and efficiently via social media platforms. For studies in other countries, other relevant stakeholder groups (eg, private or charitable fundraising, recruiting, and societal or political legitimacy) could be integrated into the model. Furthermore, the organizational and environmental contexts of the model can be adapted to specific characteristics of the hospitals and their relevant settings in other countries. In this regard, our model provides a starting point as a basic framework for the further development of more comprehensive and detailed models on the social media adoption of hospitals in all countries.

Acknowledgments

This study was funded by Innosuisse - Swiss Innovation Agency.

Authors' Contributions

MB designed the study, carried out the statistical analysis, and wrote the initial draft of the manuscript. Both authors conceptualized and performed the collection of social media data and revised the manuscript in collaboration. SF made substantial contributions to the interpretation of data and to the revision of the manuscript. Both authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Specifications R statistics.
[PDF File (Adobe PDF File), 49 KB - jmir_v22i3e16995_app1.pdf]

References

1. 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 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
2. Diddi P, Lundy LK. Organizational Twitter use: Content analysis of tweets during breast cancer awareness month. J Health Commun 2017 Mar;22(3):243-253. [doi: 10.1080/10810730.2016.1266716] [Medline: 28248621]
3. Kordzadeh N, Young DK. Exploring hospitals' use of Facebook: Thematic analysis. J Med Internet Res 2018 May 16;20(5):e190 [FREE Full text] [doi: 10.2196/jmir.9549] [Medline: 29769173]
4. Smith KT. Hospital marketing and communications via social media. Serv Mar Q 2017;38(3):187-201. [doi: 10.1080/15332969.2017.1363518]
5. Wong CA, Ostapovich G, Kramer-Golinkoff E, Griffis H, Asch DA, Merchant RM. How US children's hospitals use social media: A mixed methods study. Healthc (Amst) 2016 Mar;4(1):15-21. [doi: 10.1016/j.hjdsi.2015.12.004] [Medline: 27001094]
6. Beier M, Früh S. Social media teams of hospitals as mediators in digital health ecosystems. In: Song Y, Grippa F, Gloor P, Leitão J, editors. Collaborative Innovation Networks. Cham, Switzerland: Springer; 2019:115-124.
7. O'Connor MI, Brennan K, Kazmerchak S, Pratt J. YouTube videos to create a ‘virtual hospital experience’ for hip and knee replacement patients to decrease preoperative anxiety: A randomized trial. Interact J Med Res 2016 Apr 18;5(2):e10 [FREE Full text] [doi: 10.2196/ijmr.4295] [Medline: 27091674]
8. Zhang W, Deng Z, Evans R, Xiang F, Ye Q, Zeng R. Social media landscape of the tertiary referral hospitals in China: Observational descriptive study. J Med Internet Res 2018 Aug 9;20(8):e249 [FREE Full text] [doi: 10.2196/jmir.9607] [Medline: 30093370]
9. Carpentier M, Van Hoye G, Stockman S, Schollaert E, Van Theemsche B, Jacobs G. Recruiting nurses through social media: Effects on employer brand and attractiveness. J Adv Nurs 2017 Nov;73(11):2696-2708. [doi: 10.1111/jan.13336] [Medline: 28494103]
10. Cartledge P, Miller M, Phillips B. The use of social-networking sites in medical education. Med Teach 2013 Oct;35(10):847-857. [doi: 10.3109/0142159X.2013.804909] [Medline: 23841681]
11. Widmer RJ, Engler NB, Geske JB, Klarich KW, Timimi FK. An academic healthcare Twitter account: The Mayo Clinic experience. Cyberpsychol Behav Soc Netw 2016 Jun;19(6):360-366. [doi: 10.1089/cyber.2015.0607] [Medline: 27327062]
12. Griffis HM, Kilaru AS, Werner RM, Asch DA, Hershey JC, Hill S, et al. Use of social media across US hospitals: Descriptive analysis of adoption and utilization. J Med Internet Res 2014 Nov 27;16(11):e264 [FREE Full text] [doi: 10.2196/jmir.3758] [Medline: 25431831]
13. Martinez-Millana A, Fernandez-Llatas C, Bilbao IB, Salcedo MT, Salcedo VT. Evaluating the social media performance of hospitals in Spain: A longitudinal and comparative study. J Med Internet Res 2017 May 23;19(5):e181 [FREE Full text] [doi: 10.2196/jmir.6763] [Medline: 28536091]
14. Richter JP, Muhlestein DB, Wilks CE. Social media: How hospitals use it, and opportunities for future use. J Healthc Manag 2014;59(6):447-460. [Medline: 25647968]
15. Thaker SI, Nowacki AS, Mehta NB, Edwards AR. How US hospitals use social media. Ann Intern Med 2011 May 17;154(10):707-708. [doi: 10.7326/0003-4819-154-10-201105170-00021] [Medline: 21576547]
16. van de Belt TH, Berben SA, Samsom M, Engelen LJ, Schoonhoven L. Use of social media by Western European hospitals: Longitudinal study. J Med Internet Res 2012 May 1;14(3):e61 [FREE Full text] [doi: 10.2196/jmir.1992] [Medline: 22549016]
17. Tornatzky LG, Fleischer M. Processes of Technological Innovation. Issues in Organization and Management Series. Lanham, Maryland: Lexington Books; 1990.
18. Cao Q, Jones DR, Sheng H. Contained nomadic information environments: Technology, organization, and environment influences on adoption of hospital RFID patient tracking. Inf Manag 2014;51(2):225-239. [doi: 10.1016/j.im.2013.11.007]
19. Cleven A, Mettler T, Rohner P, Winter R. Healthcare quality innovation and performance through process orientation: Evidence from general hospitals in Switzerland. Technol Forecast Soc Change 2016 Dec;113:386-395. [doi: 10.1016/j.techfore.2016.07.007]
20. de Pietro C, Camenzind P, Sturny I, Crivelli L, Edwards-Garavoglia S, Spranger A, et al. Switzerland: Health system review. Health Syst Transit 2015;17(4):1-288, xix [FREE Full text] [Medline: 26766626]
21. Von Wyl V, Beck K. Switzerland. In: Sagan A, Thomson S, editors. Voluntary Health Insurance in Europe: Country Experience (number 42). Copenhagen, Denmark: European Observatory on Health Systems and Policies; 2016:147-150.
22. Baker J. The technology–organization–environment framework. In: Dwivedi YK, Wade MR, Schneberger SL, editors. Information Systems Theory. New York: Springer; 2012:231-245.
23. Asagbra OE, Burke D, Liang H. Why hospitals adopt patient engagement functionalities at different speeds? A moderated trend analysis. Int J Med Inform 2018 Mar;111:123-130. [doi: 10.1016/j.ijmedinf.2017.12.023] [Medline: 29425623]
24. Hung SY, Hung WH, Tsai CA, Jiang SC. Critical factors of hospital adoption on CRM system: Organizational and information system perspectives. Decis Support Syst 2010;48(4):592-603. [doi: 10.1016/j.dss.2009.11.009]
25. Lian JW, Yen DC, Wang YT. An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in Taiwan hospital. Int J Inf Manag 2014;34(1):28-36. [doi: 10.1016/j.ijinfomgt.2013.09.004]

26. Chau PY, Tam KY. Factors affecting the adoption of open systems: An exploratory study. Manag Inf Syst Q 1997;21(1):1-24. [doi: 10.2307/249740]
27. Khoumbati K, Themistocleous M, Irani Z. Evaluating the adoption of enterprise application integration in health-care organizations. J Manag Inf Syst 2006;22(4):69-108. [doi: 10.2753/mis0742-1222220404]
28. Ramdani B, Kawalek P, Lorenzo O. Predicting SMEs' adoption of enterprise systems. J Enterp Inf Manag 2009;22(1/2):10-24. [doi: 10.1108/17410390910922796]
29. Lee CP, Shim JP. An exploratory study of radio frequency identification (RFID) adoption in the healthcare industry. Eur J Inf Syst 2007;16(6):712-724. [doi: 10.1057/palgrave.ejis.3000716]
30. Venkatesh V, Bala H. Adoption and impacts of interorganizational business process standards: Role of partnering synergy. Inf Syst Res 2012;23(4):1131-1157. [doi: 10.1287/isre.1110.0404]
31. Zhu K, Kraemer KL, Dedrick J. Information technology payoff in e-business environments: An international perspective on value creation of e-business in the financial services industry. J Manag Inf Syst 2004;21(1):17-54. [doi: 10.1080/07421222.2004.11045797]
32. Zhu K, Kraemer KL, Xu S. The process of innovation assimilation by firms in different countries: a technology diffusion perspective on e-business. Manag Sci 2006;52(10):1557-1576. [doi: 10.1287/mnsc.1050.0487]
33. Beier M, Wagner K. Social media adoption: Barriers to the strategic use of social media in SMEs. In: Proceedings of the European Conference on Information Systems. USA: AIS; 2016 Presented at: ECIS'16; June 12-15, 2016; Istanbul, Turkey URL: https://www.researchgate.net/publication/301357654_Social_Media_Adoption_Barriers_to_the_Strategic_Use_of_Social_Media_in_SMEs
34. Abeza G, O'Reilly N. Social media platforms' use in building stakeholder relationships. J Appl Sport Manag 2014;6(3):103-126 [FREE Full text]
35. Michaelidou N, Siamagka NT, Christodoulides G. Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands. Ind Mark Manage 2011 Oct;40(7):1153-1159. [doi: 10.1016/j.indmarman.2011.09.009]
36. Omosigho OE, Abeysinghe G. Evaluating Readiness of Organizations to Adopt Social Media for Competitive Advantage. In: Proceedings of the International Conference on Information Society. Piscataway Township, NJ, USA: IEEE; 2012 Presented at: i-Society'12; June 25-28, 2012; London, UK.
37. Panagiotopoulos P, Barnett J. Social media in union communications: An international study with UNI global union affiliates. Br J Ind Relat 2015;53(3):508-532. [doi: 10.1111/bjir.12060]
38. Schaupp LC, Bélanger F. The value of social media for small businesses. J Inf Syst 2014 Jun;28(1):187-207. [doi: 10.2308/isys-50674]
39. Sharif MH, Troshani I, Davidson R. Public sector adoption of social media. J Comput Inf Syst 2015;55(4):53-61. [doi: 10.1080/08874417.2015.11645787]
40. Tountas Y, Karnaki P, Pavi E, Souliotis K. The 'unexpected' growth of the private health sector in Greece. Health Policy 2005 Oct;74(2):167-180. [doi: 10.1016/j.healthpol.2005.01.013] [Medline: 16153477]
41. White K, Collyer F. Health care markets in Australia: ownership of the private hospital sector. Int J Health Serv 1998;28(3):487-510. [doi: 10.2190/A9U4-JXGX-87Y7-5B34] [Medline: 9711477]
42. Kim WH, Chae B. Understanding the relationship among resources, social media use and hotel performance: The case of Twitter use by hotels. Int J Contemp Hospitality Manag 2018;30(9):2888-2907. [doi: 10.1108/ijchm-02-2017-0085]
43. Leung XY, Bai B, Stahura KA. The marketing effectiveness of social media in the hotel industry. J Hospitality Tour Res 2015;39(2):147-169. [doi: 10.1177/1096348012471381]
44. Kind T, Genrich G, Sodhi A, Chretien KC. Social media policies at US medical schools. Med Educ Online 2010 Sep 15;15:- [FREE Full text] [doi: 10.3402/meo.v15i0.5324] [Medline: 20859533]
45. Tuckett A, Turner C. Do you use social media? A study into new nursing and midwifery graduates' uptake of social media. Int J Nurs Pract 2016 Apr;22(2):197-204. [doi: 10.1111/ijn.12411] [Medline: 26531296]
46. Zhu K, Kraemer K, Xu S. Electronic business adoption by European firms: a cross-country assessment of the facilitators and inhibitors. Eur J Inf Syst 2003;12(4):251-268. [doi: 10.1057/palgrave.ejis.3000475]
47. Chang IC, Hwang HG, Hung MC, Lin MH, Yen DC. Factors affecting the adoption of electronic signature: Executives' perspective of hospital information department. Decis Support Syst 2007 Nov;44(1):350-359. [doi: 10.1016/j.dss.2007.04.006]
48. Marques A, Oliveira T, Dias SS, Martins MFO. Medical records system adoption in European hospitals. Electron J Inform Syst Eval 2011;14(1):89-99 [FREE Full text]
49. Kimberly JR, Evanisko MJ. Organizational innovation: The influence of individual, organizational, and contextual factors on hospital adoption of technological and administrative innovations. Acad Manag J 1981;24(4):689-713. [doi: 10.5465/256170]
50. Moch MK, Morse EV. Size, centralization and organizational adoption of innovations. Am Sociol Rev 1977;42(5):716-725. [doi: 10.2307/2094861]
51. Brech FM, Messer U, Vander Schee BA, Rauschnabel PA, Ivens BS. Engaging fans and the community in social media: interaction with institutions of higher education on Facebook. J Mark High Educ 2017;27(1):112-130. [doi: 10.1080/08841241.2016.1219803]

52. Cvijikj IP, Spiegler ED, Michahelles F. Evaluation framework for social media brand presence. Soc Netw Anal Min 2013;3(4):1325-1349. [doi: 10.1007/s13278-013-0131-y]
53. Dunbar RI. Do online social media cut through the constraints that limit the size of offline social networks? R Soc Open Sci 2016 Jan;3(1):150292 [FREE Full text] [doi: 10.1098/rsos.150292] [Medline: 26909163]
54. Orlikowski WJ. CASE tools as organizational change: Investigating incremental and radical changes in systems development. Manag Inf Syst Q 1993;17(3):309-340. [doi: 10.2307/249774]
55. Kuikka M, Äkkinen M. Determining the Challenges of Organizational Social Media Adoption and Use. In: Proceedings of the European Conference on Information Systems. USA: AIS; 2011 Presented at: ECIS'11; June 9-11, 2011; Helsinki, Finland URL: https://pdfs.semanticscholar.org/a85b/f571d6d707b59b752a7d77b1581ae2842295.pdf
56. Brink T. B2B SME management of antecedents to the application of social media. Ind Mark Manag 2017;64:57-65. [doi: 10.1016/j.indmarman.2017.02.007]
57. Bundesamt für Gesundheit BAG. Kennzahlen der Schweizer Spitäler 2016 [Key figures for Swiss hospitals] URL: https://www.bag.admin.ch/bag/de/home/zahlen-und-statistiken/zahlen-fakten-zu-spitaelern/kennzahlen-der-schweizer-spitaeler.html [accessed 2019-11-06]
58. Buro JA, Töpfer A, Leffler P. Marketing und Markenstrategie der Privatklinikgruppe Hirslanden [Marketing and brand strategies at the Hirslanden Group]. In: Albert DM, Töpfer A, editors. Handbuch Changemanagement im Krankenhaus [Change management in Hospitals]. Berlin, Germany: Springer; 2017:475-496.
59. Oliveira T, Martins MF. Literature review of information technology adoption models at firm level. Electro J Inf Syst Eval 2011;14(1):110-121.
60. Ahmadi H, Nilashi M, Ibrahim O. Organizational decision to adopt hospital information system: An empirical investigation in the case of Malaysian public hospitals. Int J Med Inform 2015 Mar;84(3):166-188. [doi: 10.1016/j.ijmedinf.2014.12.004] [Medline: 25612792]
61. Hsiao S, Li YC, Chen YL, Ko HC. Critical factors for the adoption of mobile nursing information systems in Taiwan: The nursing department administrators' perspective. J Med Syst 2009 Oct;33(5):369-377. [doi: 10.1007/s10916-008-9199-8] [Medline: 19827263]
62. Porter M, Millar V. How information gives you competitive advantage. Harv Bus Rev 1985;63(4):149-160 [FREE Full text]
63. Jia Q, Guo Y, Barnes SJ. Enterprise 2.0 post-adoption: Extending the information system continuance model based on the technology-organization-environment framework. Comput Hum Behav 2017;67:95-105. [doi: 10.1016/j.chb.2016.10.022]
64. Ahmad SZ, Bakar AR, Ahmad N. Social media adoption and its impact on firm performance: the case of the UAE. Int J Entre Behav Res 2019;25(1):84-111. [doi: 10.1108/ijebr-08-2017-0299]
65. Bundesamt für Statistik. Ständige Wohnbevölkerung ab 15 Jahren nach Hauptsprachen, nach Kanton und Stadt [Permanent resident population aged 15 and over according to main languages, canton and city] URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/sprachen-religionen/sprachen.assetdetail.4242797.html [accessed 2019-11-06]
66. Cameron AC, Trivedi PK. Regression-based tests for overdispersion in the Poisson model. J Econom 1990;46(3):347-364. [doi: 10.1016/0304-4076(90)90014-k]
67. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013;36(1):27-46. [doi: 10.1111/j.1600-0587.2012.07348.x]
68. Corporate Xing. Facts and Figures URL: https://corporate.xing.com/en/about-new-work-se/facts-and-figures/ [accessed 2019-11-06]
69. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: Motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
70. Prang K, Canaway R, Bismark M, Dunt D, Miller JA, Kelaher M. Public performance reporting and hospital choice: A cross-sectional study of patients undergoing cancer surgery in the Australian private healthcare sector. BMJ Open 2018 Apr 27;8(4):e020644 [FREE Full text] [doi: 10.1136/bmjopen-2017-020644] [Medline: 29703855]
71. Wiedenhöfer D, Keppler S. Free hospital choice in Switzerland - patients’ decision criteria and sources of information. Interdiscip Manag Res 2014;10:763-773 [FREE Full text]
72. XING Coaches + Trainer. URL: https://coaches.xing.com/ [accessed 2019-11-06]
73. Brennan R, Croft R. The use of social media in B2B marketing and branding: An exploratory study. J Cust Behav 2012;11(2):101-115. [doi: 10.1362/147539212x13420906144552]
74. Beck N, Walgenbach P. Technical efficiency or adaptation to institutionalized expectations? The adoption of ISO 9000 standards in the German mechanical engineering industry. Organ Stud 2005;26(6):841-866. [doi: 10.1177/0170840605054599]
75. Nah S, Saxton GD. Modeling the adoption and use of social media by nonprofit organizations. New Media Soc 2013;15(2):294-313. [doi: 10.1177/1461444812452411]
76. IGEM. 2017. IGEM-digiMONITOR 2017 URL: https://www.igem.ch/download/IGEM-digiMONITOR_Factsheet-2017.pdf [accessed 2019-11-06]
77. Relling M, Schnittka O, Sattler H, Johnen M. Each can help or hurt: Negative and positive word of mouth in social network brand communities. Int J Res Mark 2016 Mar;33(1):42-58. [doi: 10.1016/j.ijresmar.2015.11.001]

78. Frenkel S. The New York Times. 2018 Jul 25. Facebook Starts Paying a Price for Scandals URL: https://www.nytimes.com/2018/07/25/technology/facebook-revenue-scandals.html [accessed 2019-11-06]
79. Tuttle H. Risk Management. 2018 May 1. Facebook Scandal Raises Data Privacy Concerns URL: http://www.rmmagazine.com/2018/05/01/facebook-scandal-raises-data-privacy-concerns/ [accessed 2020-02-21]
80. Mergel I. Social media adoption and resulting tactics in the US federal government. Gov Inf Q 2013;30(2):123-130. [doi: 10.1016/j.giq.2012.12.004]
81. Pittman J. Facebook, speech, and the national labor relations act. South J Bus Ethics 2015;7:150-158.

Abbreviations

DRG: diagnosis-related group
FOPH: Federal Office of Public Health
FSO: Federal Statistical Office
IRR: incidence rate ratio
OR: odds ratio
TOE: technology-organization-environment
VHI: voluntary health insurance

Edited by G Eysenbach; submitted 10.11.19; peer-reviewed by S Speedie, P Santana-Mancilla; comments to author 12.12.19; revised version received 25.12.19; accepted 26.01.20; published 09.03.20.

Please cite as:
Beier M, Früh S
Technological, Organizational, and Environmental Factors Influencing Social Media Adoption by Hospitals in Switzerland: Cross-Sectional Study
J Med Internet Res 2020;22(3):e16995
URL: https://www.jmir.org/2020/3/e16995
doi:10.2196/16995
PMID:32149718

Original Paper

Successful Moderation in Online Patient Communities: Inductive Case Study

Tanner Skousen[1], MS; Hani Safadi[1], PhD; Colleen Young[2], BA; Elena Karahanna[1], PhD; Sami Safadi[3], MD; Fouad Chebib[2], MD

[1]Management Information Systems, Terry College of Business, University of Georgia, Athens, GA, United States
[2]Mayo Clinic, Rochester, MN, United States
[3]Critical Care Medicine Fellowship, University of Maryland Medical Center, Baltimore, MD, United States

Corresponding Author:
Tanner Skousen, MS
Management Information Systems
Terry College of Business
University of Georgia
620 S Lumpkin St
Athens, GA, 30602
United States
Phone: 1 3852018669
Email: tanner.skousen@uga.edu

Abstract

Background: Online patient communities are becoming more prevalent as a resource to help patients take control of their health. However, online patient communities experience challenges that require active moderation.

Objective: This study aimed to identify the challenges of sustaining a thriving online patient community and the moderation practices employed to address the challenges and manage the online patient community successfully.

Methods: An inductive case study of Mayo Clinic Connect was analyzed using the grounded theory methodology. Insights for the analysis were obtained from semistructured interviews with community managers and community members. Secondary data sources, such as community management documents, observational meeting notes, and community postings, were used to validate and triangulate the findings.

Results: We identified four challenges unique to online patient communities. These challenges include passion, nonmedical advice, personal information, and community participation. We identified five categories of practices that community members used to address these challenges and moderate the community successfully. These practices include instructive, semantic, connective, administrative, and policing practices.

Conclusions: Successful moderation in online patient communities requires a multitude of practices to manage the challenges that arise in these communities. Some practices are implemented as preventive measures while other practices are more interventive. Additionally, practices can come from both authority figures and exemplary members.

(J Med Internet Res 2020;22(3):e15983) doi:10.2196/15983

KEYWORDS

online patient communities; online social support; online community moderation; community management

Introduction

Background

Patients increasingly rely on the internet to look for medical information; ask questions; find peers with similar health concerns; read commentaries and experiences about health issues; and consult reviews and rankings of treatments, doctors, and hospitals [1-4]. According to a 2013 study, 72% of internet users looked at health information online [5]. Online patient communities are becoming forums for patients to share their stories, gain peer support, and search for medical information [6,7]. Studies have shown that membership in patient communities is associated with better health, behavior, and medical knowledge [7-10].

We define successful online patient communities as communities that are vibrant, supportive, and active. They welcome members, encourage active participation and interactions, foster relationships, and provide accurate medical information. Among the many potential concerns and challenges for online patient communities, the quality and accuracy of medical knowledge provided in these communities are questioned [11-14]. This is especially important given that members vary considerably in their medical expertise and may have difficulty discerning information quality. To address these issues and assure their success, some online patient communities use moderators [15]. Little academic research exists on effective moderation practices in online patient communities. Our objective was to systematically study the challenges of online patient communities and identify moderation practices that are employed to address these challenges and maintain a successful community.

To gain a deep understanding of this phenomenon, we conducted an inductive case study of Mayo Clinic Connect, a leading online patient community. Using the grounded theory methodology, we collected observational and interview data from key members over 3 years. Our findings highlighted five moderation practices in the community: instructive, connective, semantic, administrative, and policing. Furthermore, these moderation practices were not only enacted by the community's management but also by the enlisted volunteer community members who assist in achieving community success. Together, the practices used by community management and community members successfully address several challenges to create a thriving community.

Prior Work

Online communities bring together members who share common interests and contribute valuable knowledge and expertise and are recognized as generators of significant knowledge [16]. In health care, online patient communities have become a medium for patients to share their stories, gain peer support, and search for medical information [6]. A healthy community retains its existing members, attracts new ones, and elicits their contribution [17]. Research has consistently shown the importance of feedback in driving members' contributions to online communities [18-20]. One challenge facing online communities is managing the sheer number of contributors and contributions in the absence of monetary incentives and formal managerial structures [21].

Compared with other online communities, online patient communities face additional unique challenges. First and foremost, the quality, accuracy, and trustworthiness of medical knowledge provided over these platforms are questioned [11-14]. Second, there are ethical and privacy issues related to sharing and disseminating patients' information online [22]. Third, there is a mismatch of motives and expectations among members who use these platforms as communication and marketing tools and members who use them for social support and exchanging advice [23].

Community leadership plays an essential role in dealing with online community challenges [20,24,25]. Effective leaders and moderators are characterized by inclusivity, helpfulness, and sociability [20,26,27]. Community management can be assisted by elevating the status of exemplary members to become trusted volunteer peer leaders [19]. In online patient communities, the involvement of volunteer peer leaders and echoing the voice of patients are essential for community success [28]. However, this inclusivity should not ignore the quality of information and the trustworthiness of conversation. Such tension is not very prevalent in other grassroots communities because the conversation reflects the shared interests of members without much further consideration. However, understanding how online communities can be other accurate sources of information and remain inclusive to members' opinions remains an important research area [16,21].

Methods

Research Context

We conducted a study of Mayo Clinic Connect (referred to as Connect hereafter), a leading online patient community sponsored by Mayo Clinic. Connect is designed to "connect patients and family caregivers with each other" [29]. This community enables patients and their families with different medical backgrounds to share experiences, find support, and exchange information with others who have faced similar experiences. The community has open boundaries, which means it is open to the public and not restricted to Mayo Clinic patients only. At the time of publication, the community has 93,000 registered members and more than 10,000 active members who post at least once a month.

Data Collection

Our primary source of data came from semistructured interviews that were collected over five phases. Levitt et al [30] claim that fidelity to data "may be procured by inviting participants to interview or describe their experiences." First, data collection was initiated when 1 author became acquainted with the director of Connect and participated in an early interview with her. This interview was instrumental in establishing our initial understanding of the community and in guiding our sampling of subsequent interviewees through a snowballing approach [31]. Second, we conducted four interviews with managers of the community over the phone. Third, we attended a community meeting of managers and members who discussed the concerns, successes, and goals of the community. In this meeting, we participated as silent observers to understand the community in general and collected four onsite interviews from community members who were participating.

After the initial analysis of these interviews, we focused on theoretical sampling for subsequent interviews. Therefore, our fourth phase included the collection of additional 13 phone interviews from members of the community. Through probing questions, we asked about their participation in Connect and what they felt the community did to achieve success. We asked them to expound on their experiences with the community, including challenging ones. Overall, we let all interviewees talk about being a member of Connect and their interactions with others. Each interview lasted approximately 30 min and was recorded and transcribed verbatim. We continued to interview individuals until we reached theoretical saturation and gained no additional insights from additional interviews [32]. Fifth, we conducted a final interview with the director of Connect to discuss and validate our findings.

After the completion of the interviews, we supplemented this primary data source with additional observational data sources to triangulate the findings from the interviews. First, all authors joined the community as passive members and observed the community members' contributions and interactions over the last 3 years. Second, we obtained archival community documents, which included standardized response templates, a reference guide for exemplary posts, orientation packages, and community management guidelines. The observations gathered from these data sources were in line with the findings that emerged in the interviews, strengthening the fidelity of the findings.

Data Analysis

We adopted a grounded theory coding perspective on the semistructured interviews to let theory emerge from the data. Our initial unfamiliarity with the community was instrumental in allowing the data to inform our insights, rather than preconceived theory. However, we were aware of how our prior experiences as researchers might shape the interpretation of the data [33]. Therefore, as an added measure to ensure the coding was grounded in the data, the author who coded the majority of the data was the least experienced in online community literature. This perspective management in data analysis enhanced the fidelity of our data [30]. We followed a constant comparative process throughout our analysis, that is, we repeatedly compared codes and findings in frequent discussions among the authors [34].

Furthermore, 2 of the authors conducted line-by-line open coding techniques [35] in the first five interviews. After both the authors completed the coding of these first interviews, they discussed their findings to agree on emergent themes. One author coded the remaining interviews, and all authors met frequently during the coding process to discuss the inductive insights from the codes. After creating an extensive list of open codes, we conducted a second round of coding to ensure that minor differences in codes from open coding were combined to form unique ideas that did not overlap in the data. During open coding, memos were kept on each interview to provide extended notes and insights.

After completing open coding and discussing emergent themes, we followed an axial coding process to combine and relate codes together to form more abstract themes and relationships [36]. Axial coding helps to identify the who, what, when, where, why, and how of the emerging theory. In our case, it illustrated a clear picture of the moderation practices in the community.

We felt that the results from interview data were sufficient to generate insights, yet we used the observational data of archival documents and community postings to validate those results. Although this observational data were not formally coded in the same manner as the interviews, it was used as a source to validate the findings and provide examples of our presented results.

Results

Community Challenges

Our interviews highlighted certain challenges for online patient communities, specifically that require active moderation to address. These challenges include passion, giving nonmedical advice, providing personal information, and community participation.

First, many members are passionate about Connect and their contributions to it. These members experienced or overcame medical conditions and are passionate to share their experiences and help others. Thus, passion can yield positive consequences, such as increased engagement and participation. Passion, however, leads many individuals to argue about different beliefs and opinions. Ugly and argumentative posts are at times a challenge for Connect:

> *We might have a member that comes in really opinionated.* [Moderator 3]

Second, sharing personal experiences is perhaps the overarching activity of members in Connect. Members with similar problems share experiences to encourage, inform, and support one another. Overall, this activity creates positive outcomes, and many members are very grateful for the feeling that they are "not alone." However, challenges arise when experiences are viewed as or claimed to be "medical advice." Members in the community are not medical practitioners but solely patients who have experienced a similar disease state. Therefore, nonmedical advice creates the potential for disappointment if the advice does not yield positive outcomes, or worse, legal repercussions if that advice causes further physical or emotional harm:

> *When it comes to alternative medicine, complementary medicine, that's where we run into problems. Because members are free to express their views, but when they advise their discussion group and they instruct them to stop taking traditional medications and go for homeopathy or herbal things, that's when we have to step in.* [Moderator 4]

Third, soliciting personal information is helpful for community members to know how to respond to other members. More information regarding a member's condition and situation helps others relate to the member's issues and provide appropriate support. Members often ask questions such as "how was this diagnosed," "how long have you had the problem," and "what kinds of treatment have you had." However, sometimes this is taken too far. Violations to privacy may occur when a member shares excessive personal health information:

> *In a HIPAA sense, this individual is asking this person very specific questions on the day, the time and the results of a medical test, and I'm not sure that that's an appropriate place for a public discussion board.* [Mentor 6]

Fourth, community participation is necessary by definition for the community to thrive. Therefore, a lack of participation is a challenge to the community. Yet, perhaps, the more nuanced challenge is that the motivation or form of the participation can vary. For example, many members will come to the community only for answers to their questions or to find a way to schedule an appointment with Mayo Clinic, in contrast to joining to establish relationships and gain social support. When members do not participate in the community to establish relationships, the turnover of community members is high, which will hurt the community.

In addition to the four challenges mentioned above—passion, nonmedical advice, personal information, and community participation—we also noticed secondary challenges that either were not mentioned sufficiently in the interviews or were related to the main challenges addressed. These challenges included ambiguous identities, lack of physician interest, generational gaps, and technology challenges. However, these were secondary to the main story we saw in the data, and therefore, we will focus on the four main challenges for the duration of the paper.

Moderation Structure in the Community

To respond successfully to these challenges, the community must maximize the potential positive benefits and minimize the potential negative outcomes of addressing them. In the case of Connect, these challenges are addressed by community management. The management team of Connect consists of a director and 5 staffers. These managers are referred to as moderators in the nomenclature of Connect. Moderators have set formal guidelines such as being careful when giving out medical advice, being respectful, and no commercial advertising. At the outset of the community, the moderators ran the community, moderated others' posting, and provided the support to members who started to join the community.

As the community grew, moderators recognized the need to scale community moderation to maintain members' contributions. At the same time, moderators noted that some members were taking increasingly active leadership roles. They identified these emerging peer leaders and invited some of them to help monitor the community. In the nomenclature of Connect, the emerging peer leaders selected by the moderators are called mentors. Once selected, mentors, who volunteer their time, are trained and guided by moderators to act in accordance with the guidelines of the community. Multiple training practices are implemented to onboard and train mentors (Table 1).

Table 1. Training mechanisms for mentors.

Training mechanism	Example from interviews
Standard documentation	"I'll send you three or four documents that you can look through and understand a little more what we're looking for in a mentor."
Setting expectations	"We're hopeful that you'll agree to spend at least X number of hours a month visiting the board and watching things."
Coaching and individual training	"When you are first invited, you spend time with the director, a one-to-one, doing screen shares and discussing different scenarios."
Mentor meetups	"The mentors and moderators, we get together once a year. It is a little bit more training and sort of continuing education of being a mentor."
Remote training	"Every quarter we will have a phone call where we can share what concerns us and build up a discussion."
Behavior emulation	"We [Moderators] ask them to emulate us as far as the way we would respond."

These training practices communicate and transfer the goals of moderators to mentors. This process brings three advantages. First, it reduces the workload of moderators, which is becoming infeasible with the continuous growth of the community, receiving around 10,000 posts per month. Second, mentors are uniquely qualified to assist members because of their shared experience—an experience that the moderators often do not have. Third, it enables moderators and mentors to specialize in different tasks focusing on desired outcomes. Moderators focus on administrative responsibilities, policing the community, and resolving tensions. Mentors can focus on patient success through the daily postings in community threads.

Moderation Practices in the Community

Through the analysis of the interviews, we found that moderators and mentors enacted different practices to address the challenges of the community. Groups of similar practices were abstracted into five high-level categories to add parsimony to our understanding of the moderation. Although some practices are employed by *both* moderators and mentors, in general, practices were not shared equally between moderators and mentors. The three practice groups associated more often with mentors were instructive, semantic, and connective practices. The two practice groups associated more often with moderators were administrative and policing practices.

Instructive practices are actions that provide medical information relevant to member experiences. Although the community acts as emotional support for individuals, the community also strives to empower members by educating them on how to make better decisions about their health. Specific examples of instructive practices include pointing a member to a doctor, stating that the community is not a replacement for professional help. Mentors not only point people to doctors but also assist them by sharing appropriate questions that could be asked or tips of ways to talk to a doctor. In addition, mentors in the community have become proficient in using tools such as Google Scholar to search and provide relevant links to patients who have medical-related questions, inspiring them to learn for themselves through peer-reviewed literature. At times, moderators invite doctors to participate in Connect through live webinars that answer common questions that patients have asked recently.

Semantic practices are actions that focus on the content of messages and responses, that is, what is being communicated in a post. Mentors are instructed to share beneficial content in addition to sharing it in a caring and empathetic manner. Some of the beneficial content that mentors can post are their own experiences with past illnesses that relate to patients and show them that they are not alone in their journey. Mentors can also post words of welcome to new members to establish a sense of security, love, and care. In addition, mentors ask questions to spur further discussion and better understand the individuals they are communicating with. The content of a message, however, is only half of the meaning of semantic practices—how the message is delivered is also important. Mentors must, when appropriate, validate the concerns of the posts through their words to make patients feel heard and important. This is enhanced when mentors use encouraging language and positive reinforcement, rather than critical and negative comments that are so often seen online.

Connective practices are actions that create relationships inside the community and bring individual members into a community setting. This connection is achieved through the platform design of *tagging* members as seen in many social media sites. Mentors go beyond tagging, however, and take a proactive approach to create relationships. Two examples include proactively following up with individuals who have not posted recently and attempting to connect them with similar members. In addition, mentors who receive private messages (a feature on Connect) will often ask to bring them public so that the entire community can see the questions and participate in supporting that individual. So, although the feature of tagging is often how connections are made, the proactive mindset to foster relationships is a significant element for connective practices.

Administrative practices are performed by moderators outside of discussion threads to keep the community growing with a *positive community feel*. What is done behind the scenes plays an important role in creating an environment for the community discussion to stay vibrant and productive. Administrative practices include increasing the number of volunteer mentors to help scale the moderation of the community. Once mentors are selected, moderators coach them on the proper conduct in the community. In addition, moderators will, at times, coach regular members who act inappropriately. Finally, moderators of the community attempt to build community unity by spotlighting select members frequently.

Policing practices are interventions against antisocial behaviors and departures from community guidelines. Ideally, the community would like to not have to implement these practices, but because of the sensitivity of medical information, they are implemented when needed. Moderators act as the authoritative hand more than mentors. Mentors may indirectly police the actions of members by asking for evidence of claims or diffusing negative and criticizing remarks, but typically, they are used as eyes on the community to report posts. Moderators alone implement policing interventions through muting members, taking down posts, and acting as an intervention when posts get out of hand.

Addressing Challenges Through Moderation Practices

After we independently identified specific challenges and moderation practices of Connect, we used axial coding to relate the challenges and practices together. By doing so, we uncovered processes by which mentors and moderators worked together to maximize the benefits that come from challenges and minimize their negative impacts. Below, we identify each challenge that was introduced above and discuss the practices implemented by moderators and mentors to address the challenge. We highlight the insights with relevant excerpts that came from interviews of community members.

Passion

Passion was fostered by mentors showing interest in members' posts. Particularly, semantic practices, such as asking questions and validating concerns, were used to leverage the positive benefit of passion and expound on the member's interest. When a member's passion turned negative and argumentative, however, mentors addressed this challenge through indirect policing practices such as diffusing negative comments and observing the community and reporting to moderators:

> *We've had our share of those people who come in and are negative, and as a mentor, I honor that because we're not all going to be positive every day. I will recognize that and say, "Hey, it sounds like you're having a really bad day. Is there anything that caused that?"* [Mentor 8]
>
> *We had a member who was really opinionated, and couldn't be convinced he was giving bad advice, so in the end, what I had to do was call on the director to take a hand in the matter.* [Mentor 4]

However, when the discussion became too ugly, moderators would intervene and address the situation. Moderators typically did not remove posts or mute members who shared passionate opinions because they recognize that passion is what strengthens the community. Rather, moderators implemented administrative practices such as coaching to work with members and correct members' contribution to the community:

> *And there was one particular gentleman who was very combative, talked down to members a lot, wrote in all caps, which in an online community, is considered yelling. And with a lot of coaching, that particular person, over the course of many months, ended up becoming a really valuable member.* [Moderator 2]

If members did not respond to coaching or if the negative behavior continued, then direct policing practices such as muting members or removing posts were implemented as a final option:

> *In one case, the moderators actually did ban the person from the board and say, "We don't allow people in here who are going to call people names and take things to that extreme."* [Mentor 6]

Nonmedical Advice

Giving *advice in a proper way* is a manner in which the community achieves its goals of educating and supporting patients. In lieu of advice, mentors share their experiences, a semantic practice, which follows the guidelines of Connect. Their post with their experience not only benefits the patient in need but also, hopefully, acts as a model for others to follow. They attempt to include the phrase, “this is not medical advice” before each of their experiences:

> *You definitely want people to realize that I’m not a doctor. But you want to be able to offer help to people. So, we have to work in certain ways. And I usually just say, “based on my experience, this is what happened.”* [Mentor 14]

In addition, mentors have been instructed on how to question nonmedical advice by asking, “Where did you find that information?” They use instructive practices that educate members on how to do their research to find valid medical information and sometimes share those sources with patients:

> *When I see something that I am not familiar with I will definitely say, “I am not familiar with that. Have you checked that research? And we do have guidelines to help refer people to approved sites.”* [Mentor 5]

Moderators invite doctors to participate in live webinars to provide a professional answer regarding questions that have spurred nonmedical advice. Hopefully, these webinars provide the needed answers that the community can refer to when they encounter an instance of nonmedical advice. However, moderators will use policing practices as well to intervene in discussion threads if there is any information shared that could be harmful to patients:

> *We will do a live Q&A session between a physician and the surgeon. Members are able to ask questions prior. They can watch the video live, and then we link that to Mayo Connect so that the information is housed on Connect too.* [Moderator 3]

> *There are times where it is bad advice. That’s where we step in, in those situations. And we’ll, I wouldn’t say squash the conversation, but relay how important it is to seek an appointment with a medical provider.* [Moderator 2]

Personal Information

To gather information to help patients, mentors will use semantic practices to validate concerns and ask questions to members. These questions prompt further discussion and deeper insights that allow the community to help a member’s concerns. Many members, however, do not understand the proper information to divulge and to retain when they join the community. Mentors will help new members with welcome messages that establish expectations, often teaching them what is appropriate, what are the goals of Connect, and who are some of the other people in the community:

> *I try to relate empathically to what they’re going through. That makes it easier to discuss how we feel, or have felt in a similar situation.* [Mentor 12]

> *Asking questions to the person is a good thing. I know that people don’t always like to post on a public forum, so I always start a sentence with, “if you are comfortable sharing more information about yourself, would you please tell us...etc.”* [Mentor 7]

> *If they’re brand new, I welcome them to the site and explain a little bit how it works.* [Mentor 9]

Moderators will promote the sharing of personal information by spotlighting members. These spotlights are a good example of personal information that is appropriate to share to help others know how to communicate with a patient. Moderators also use the administrative practice of coaching to help instruct members when they are sharing too much. In addition, moderators use policing practices of deleting posts to remove information that is potentially Health Insurance Portability and Accountability Act of 1996 sensitive:

> *We do a spotlight to get to know people personally. Their favorite things, where they live, what they like to do, and beyond having a condition, that we are people too.* [Mentor 8]

> *You might not want to give all that information and I’ve had to refer them to the director and say, “There is a little too much information shared here. Let’s talk to this person, edit their post, and take out the exact department building where they are living.”* [Mentor 8]

Community Participation

Mentors enact multiple practices to maintain member participation. Participation includes the amount of content that a user posts, but it can also signify if a member remains involved in the community. Mentors help to increase the amount of participation by semantic practices, namely, asking questions and welcoming members. Mentors additionally focus on retaining members through connective practices, particularly tagging members to foster relationships and proactively following up with patients who have not been heard from recently:

> *I think you just ask a lot of open-ended questions. Usually, I say, “I’m fascinated by what you’ve told us, please tell us something else.”* [Mentor 10]

> *If a member has been posting and then suddenly stops, we tag members to connect them to our members.* [Moderator 4]

Moderators also implemented connective practices in the same manner as mentors, without a major distinction between roles. Our interpretation is that moderators of the community were the initial *mentors* of the community when the community was in its early stages, and those moderators have continued to maintain the relationships that were created then. In addition, moderators implement connective practices to model how to properly tag members:

> *When I first started, it was really more about fostering relationships and, no pun intended with the word connect, but in the beginning really our job was to connect members with other members who had talked about the same thing.* [Moderator 2]

We provide 2 figures to help readers understand the context and the findings. First, in Figure 1, we include a redacted screenshot of a post from Connect to illustrate the practices implemented by moderators and mentors. We highlight some specific semantic, connective, and instructive practices to the right. However, our main purpose for including the screenshot was for readers to form their own opinions of the community and familiarize themselves with the context. Second, in Figure 2, we provide a graphical summary of the roles that implement the practices, the five main practices identified, and the four challenges that those practices address. We provide this graphic to help quickly and visually outline the dynamics of the community.

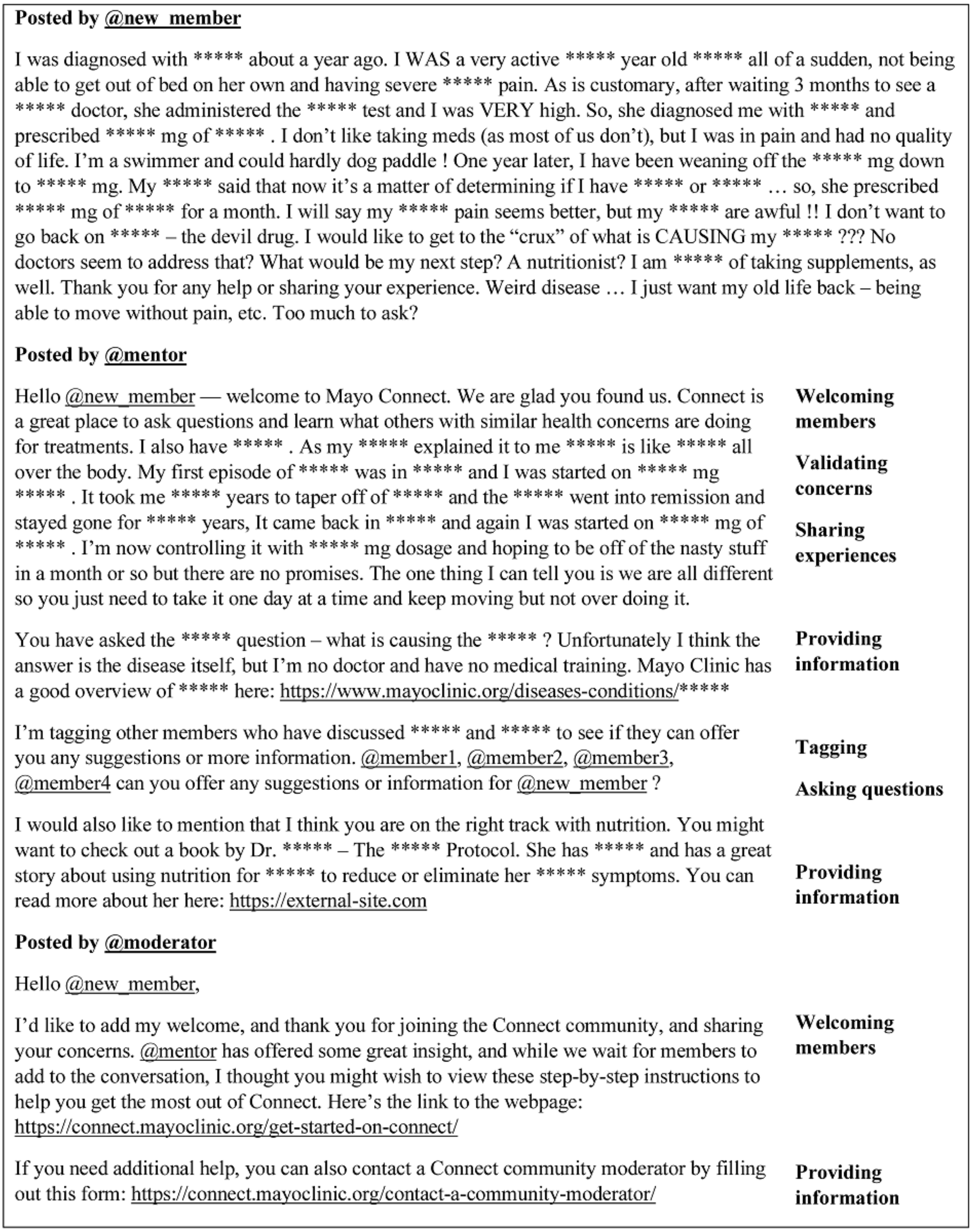


Figure 1. Moderation practices employed in one example post.

Posted by @new_member

I was diagnosed with ***** about a year ago. I WAS a very active ***** year old ***** all of a sudden, not being able to get out of bed on her own and having severe ***** pain. As is customary, after waiting 3 months to see a ***** doctor, she administered the ***** test and I was VERY high. So, she diagnosed me with ***** and prescribed ***** mg of ***** . I don't like taking meds (as most of us don't), but I was in pain and had no quality of life. I'm a swimmer and could hardly dog paddle ! One year later, I have been weaning off the ***** mg down to ***** mg. My ***** said that now it's a matter of determining if I have ***** or ***** … so, she prescribed ***** mg of ***** for a month. I will say my ***** pain seems better, but my ***** are awful !! I don't want to go back on ***** – the devil drug. I would like to get to the "crux" of what is CAUSING my ***** ??? No doctors seem to address that? What would be my next step? A nutritionist? I am ***** of taking supplements, as well. Thank you for any help or sharing your experience. Weird disease … I just want my old life back – being able to move without pain, etc. Too much to ask?

Posted by @mentor

Hello @new_member — welcome to Mayo Connect. We are glad you found us. Connect is a great place to ask questions and learn what others with similar health concerns are doing for treatments. I also have ***** . As my ***** explained it to me ***** is like ***** all over the body. My first episode of ***** was in ***** and I was started on ***** mg ***** . It took me ***** years to taper off of ***** and the ***** went into remission and stayed gone for ***** years, It came back in ***** and again I was started on ***** mg of ***** . I'm now controlling it with ***** mg dosage and hoping to be off of the nasty stuff in a month or so but there are no promises. The one thing I can tell you is we are all different so you just need to take it one day at a time and keep moving but not over doing it.

You have asked the ***** question – what is causing the ***** ? Unfortunately I think the answer is the disease itself, but I'm no doctor and have no medical training. Mayo Clinic has a good overview of ***** here: https://www.mayoclinic.org/diseases-conditions/*****

I'm tagging other members who have discussed ***** and ***** to see if they can offer you any suggestions or more information. @member1, @member2, @member3, @member4 can you offer any suggestions or information for @new_member ?

I would also like to mention that I think you are on the right track with nutrition. You might want to check out a book by Dr. ***** – The ***** Protocol. She has ***** and has a great story about using nutrition for ***** to reduce or eliminate her ***** symptoms. You can read more about her here: https://external-site.com

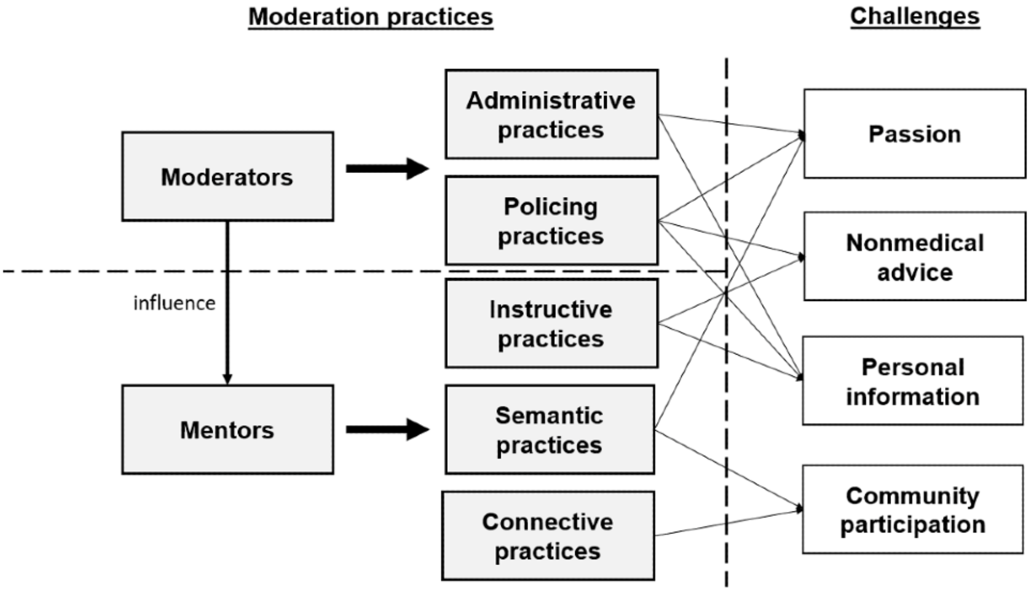
Posted by @moderator

Hello @new_member,

I'd like to add my welcome, and thank you for joining the Connect community, and sharing your concerns. @mentor has offered some great insight, and while we wait for members to add to the conversation, I thought you might wish to view these step-by-step instructions to help you get the most out of Connect. Here's the link to the webpage: https://connect.mayoclinic.org/get-started-on-connect/

If you need additional help, you can also contact a Connect community moderator by filling out this form: https://connect.mayoclinic.org/contact-a-community-moderator/

Figure 2. Graphical summary of our findings.



Discussion

Like most online communities, Connect experiences multiple challenges. We have identified and explicated many practices jointly used by Connect moderators and mentors to address these challenges and maintain a thriving community. We have provided a graphical summary of our findings in Figure 2. These practices are not solely emergent or wholly prescribed. The nature of the practice matches the needs of the community and its members. Some of these practices are *interventive*. They are mostly used to provide a quick and direct resolution to time-sensitive challenges. Other practices are *preventive*. They promote members' behavior that leads to desired outcomes by avoiding tensions. These practices can be a guide and reference for an online patient community to promote the success and well-being of its members.

Beyond the context of health care, these findings illustrate how knowledge-creating online communities [16] can balance the need to promote the contribution of accurate and trustworthy knowledge with the need to remain inclusive to members' opinions and interests. Such an endeavor requires the investment of time and effort to promote a culture of trust and the joint stewardship of the community by its management and volunteer members. Finally, this research presented the findings of one case study. Future work can improve the generalizability of the findings to other online communities in other contexts.

Acknowledgments

The authors would like to thank the moderators and mentors in Mayo Clinic Connect for their generous collaboration. The authors benefited from the feedback by participants at the 2019 International Conference on Information Systems, the 2019 Terry College of Business management information system (MIS) Seminar, and Prof Amrit Tiwana. HS acknowledges financial support from the 2017 Faculty Research Grant sponsored by the University of Georgia Research Foundation (grant number 10-21-RX064-679), the 2019 Terry-Sanford Summer Research Award, and the MIS department in the Terry College of Business.

Conflicts of Interest

None declared.

References

1. Fox S. Pew Research Center. 2011 May 12. The Social Life of Health Information, 2011 URL: https://www.pewresearch.org/internet/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2020-01-21]
2. 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 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
3. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]

4. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
5. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health Online 2013 URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2020-01-21]
6. Yan L, Peng J, Tan Y. Network dynamics: how can we find patients like us? Inf Syst Res 2015 Sep;26(3):473-636. [doi: 10.1287/isre.2015.0585]
7. Goh JM, Gao G, Agarwal R. The creation of social value: can an online health community reduce rural-urban health disparities? Manag Inf Syst Q 2016;40(1):247-263. [doi: 10.25300/MISQ/2016/40.1.11]
8. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
9. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
10. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289 [FREE Full text] [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]
11. Ogah I, Wassersug RJ. How reliable are 'reputable sources' for medical information on the internet? The case of hormonal therapy to treat prostate cancer. Urol Oncol 2013 Nov;31(8):1546-1552. [doi: 10.1016/j.urolonc.2012.08.003] [Medline: 23141782]
12. Benigeri M, Pluye P. Shortcomings of health information on the internet. Health Promot Int 2003 Dec;18(4):381-386. [doi: 10.1093/heapro/dag409] [Medline: 14695369]
13. Marrie RA, Salter AR, Tyry T, Fox RJ, Cutter GR. Preferred sources of health information in persons with multiple sclerosis: degree of trust and information sought. J Med Internet Res 2013 Mar 17;15(4):e67 [FREE Full text] [doi: 10.2196/jmir.2466] [Medline: 23635393]
14. Rössler B, Lahner D, Schebesta K, Chiari A, Plöchl W. Medical information on the internet: quality assessment of lumbar puncture and neuroaxial block techniques on YouTube. Clin Neurol Neurosurg 2012 Jul;114(6):655-658. [doi: 10.1016/j.clineuro.2011.12.048] [Medline: 22310998]
15. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013 Jun 11;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
16. Faraj S, von Krogh G, Monteiro E, Lakhani KR. Special section introduction—online community as space for knowledge flows. Inf Syst Res 2016;27(4):665-991. [doi: 10.1287/isre.2016.0682]
17. Wasko MM, Faraj S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. Manag Inf Syst Q 2005;29(1):35-57. [doi: 10.2307/25148667]
18. Moon JY, Sproull LS. The role of feedback in managing the Internet-based volunteer work force. Inf Syst Res 2008;19(4):494-515. [doi: 10.1287/isre.1080.0208]
19. Zhang C, Hahn J, De P. Research note—Continued participation in online innovation communities: does community response matter equally for everyone? Inf Syst Res 2013;24(4):883-1167. [doi: 10.1287/isre.2013.0485]
20. Johnson SL, Safadi H, Faraj S. The emergence of online community leadership. Inf Syst Res 2015;26(1):165-187. [doi: 10.1287/isre.2014.0562]
21. Faraj S, Jarvenpaa SL, Majchrzak A. Knowledge collaboration in online communities. Organ Sci 2011;22(5):1224-1239. [doi: 10.1287/orsc.1100.0614]
22. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
23. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
24. Kane GC, Johnson J, Majchrzak A. Emergent life cycle: the tension between knowledge change and knowledge retention in open online coproduction communities. Manag Sci 2014;60(12):3026-3048. [doi: 10.1287/mnsc.2013.1855]
25. Joyce E, Pike JC, Butler BS. Rules and roles vs consensus: self-governed deliberative mass collaboration bureaucracies. Am Behav Sci 2013;57(5):576-594. [doi: 10.1177/0002764212469366]
26. Collier B, Kraut R. arXiv e-Print archive. 2012. Leading the Collective: Social Capital and the Development of Leaders in Core-Periphery Organizations URL: https://arxiv.org/abs/1204.3682 [accessed 2020-01-21]
27. Faraj S, Kudaravalli S, Wasko M. Leading collaboration in online communities. Manag Inf Syst Q 2015;39(2):393-412. [doi: 10.25300/misq/2015/39.2.06]
28. Brubaker JR, Lustig C, Hayes GR. Semantic Scholar. 2010. PatientsLikeMe: Empowerment and Representation in a Patient-Centered Social Network URL: https://pdfs.semanticscholar.org/1917/dcade334dba91f15f9fdb259e7c128374151.pdf?_ga=2.223172955.1954147832.1579589925-1679671381.1567599385 [accessed 2020-01-21]
29. Mayo Clinic Connect. 2019. About Connect: Who, What & Why URL: https://connect.mayoclinic.org/page/about-connect/tab/aboutconnect/ [accessed 2019-08-19]

30. Levitt HM, Motulsky SL, Wertz FJ, Morrow SL, Ponterotto JG. Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. Qual Psychol 2017;4(1):2-22. [doi: 10.1037/qup0000082]
31. Morse JM. Sampling in grounded theory. In: Bryant A, Charmaz K, editors. The SAGE Handbook of Grounded Theory. London: SAGE Publications Ltd; 2007:656.
32. Glaser BG. Conceptualization: on theory and theorizing using grounded theory. Int J Qual Methods 2002;1(2):23-38. [doi: 10.1177/160940690200100203]
33. Finlay L. 'Outing' the researcher: the provenance, process, and practice of reflexivity. Qual Health Res 2002 Apr;12(4):531-545. [doi: 10.1177/104973202129120052] [Medline: 11939252]
34. Charmaz K. Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. London: SAGE Publications Ltd; 2006.
35. Strauss A, Corbin J. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. London: SAGE Publications Inc; 1990.
36. Corbin JM, Strauss A. Grounded theory research: procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3-21 [FREE Full text]

Abbreviations

MIS: management information system

Edited by G Eysenbach; submitted 23.08.19; peer-reviewed by M Holter, T Aslanidis; comments to author 28.09.19; revised version received 23.11.19; accepted 19.12.19; published 17.03.20.

Please cite as:
Skousen T, Safadi H, Young C, Karahanna E, Safadi S, Chebib F
Successful Moderation in Online Patient Communities: Inductive Case Study
J Med Internet Res 2020;22(3):e15983
URL: http://www.jmir.org/2020/3/e15983/
doi:10.2196/15983
PMID:32181743

Original Paper

Examination of Gender Stereotypes and Norms in Health-Related Content Posted to Snapchat Discover Channels: Qualitative Content Analysis

Kelsea LeBeau[1], MPH; Cary Carr[1], MPH; Mark Hart[1], MALS, EdD

[1]College of Public Health and Health Professions, University of Florida, Gainesville, FL, United States

Corresponding Author:
Kelsea LeBeau, MPH
College of Public Health and Health Professions
University of Florida
1225 Center Drive
Health Professions, Nursing, and Pharmacy Room 4176
Gainesville, FL, 32610
United States
Phone: 1 8504991952
Fax: 1 3522736048
Email: klebeau@ufl.edu

Abstract

Background: Snapchat has seen one of the most rapid, and unprecedented, growths in the history of social networking sites and social media with 3 billion Snapchats sent daily. In 2015, Snapchat introduced a new feature, Snapchat Discover, providing a unique way for publishers, such as magazines, to connect their content to Snapchat users.

Objective: This study aimed to evaluate qualitatively the health-related content distributed among male-focused and female-focused Discover channels and to determine whether differences exist between the content posted to these channels.

Methods: Magazine Discover channels with male and female target audiences were identified based on the magazine's claimed audience and a search of Snapchat Discover's magazine publishers, resulting in the selection of two male-focused and two female-focused channels. Stories were collected daily from each of the selected channels during a 4-week period. Using the constant comparative method, 406 Discover stories were collected and analyzed.

Results: Differences in health content coverage existed between male- and female-focused channels. General health stories from male channels comprised 7.5% (10/134) of total stories compared with 22.8% (62/272) for female channels. Sexual health stories from male channels comprised 3.0% (4/134) of total stories compared with 18.8% (51/272) for female channels. Moreover, female-focused channels' content was more comprehensive. Female audiences were portrayed as being health information seekers, concerned with sexual health and male satisfaction, primarily responsible for contraception and pregnancy prevention, and less informed about sex. Male audiences were portrayed as being less likely to seek health information, obsessed with and driven by sex, and less concerned with sexual health.

Conclusions: Understanding the content shared to social media is important, especially when considering the implications content may have for behavior. In terms of content, these findings suggest Discover channels appear to promote gender stereotypes and norms for health and sexual health through the information posted.

(J Med Internet Res 2020;22(3):e15330) doi:10.2196/15330

KEYWORDS

social media; online social networking; health behavior; sexual health; social norms; gender; gender role; mobile applications

Introduction

Snapchat Overview

Technology has become a major part of everyday life for both men and women, providing almost immediate and unlimited access to mass media, social networking sites (SNS), and social media platforms [1]. Although several popular SNS and social media platforms have been around for a while (eg, Facebook and Instagram), newer platforms have been gaining in popularity. One such platform is Snapchat (Snap, Inc), a social media app, which was founded in September 2011 and has become increasingly popular among smartphone users. Among those aged 18 to 34 years, Snapchat is now the third most popular app after Facebook and Instagram [2].

The growth of Snapchat has been one of the most rapid, and unprecedented, in the history of SNS and social media [3]. Snapchat had an estimated growth of approximately 90 to 100 million users from 2012 to 2015 [4,5]. As of 2018, there were roughly 3 billion Snapchats sent daily, 80 million daily active users in the United States, and 190 million daily active users worldwide [4]. Approximately 18% of US social media users use Snapchat, and the daily average time spent per user is 30 or more minutes [6]. Around 75% of Snapchat users are younger than 34 years old, and 90% of Snapchat users are between the ages of 13 and 24 years [6,7]. In addition, Snapchat has a disproportionate number of female users compared with male users, with roughly 70% of Snapchat users being female [6].

Snapchat is a mobile photo messaging and multimedia sharing app allowing users to send *snaps* consisting of photos or videos to people of their choosing. Ephemerality, the concept of something lasting temporarily, is a key feature differentiating Snapchat from other social media and makes the platform appealing to users [8]. Snapchat users can choose to send a snap with a specified time limit (between one and 10 seconds) or with no time limit (the snap does not disappear until you touch the screen again), offering users a more private and intimate form of communication compared with other social media [2].

Snapchat Discover

In 2015, Snapchat introduced a new feature, Snapchat Discover, which is used by many Snapchatters [9]. In 2016, Snapchat reported more than 60 million people viewing its Discover platform each month [10]. The Snapchat Discover feature allows users to explore stories from various editorial teams on a daily basis. Essentially, Snapchat Discover is a network of media partners—ranging from National Geographic to Cosmopolitan—in which each partner has its own media channel where they can post articles, videos, and stories for their mobile audiences and subscribers [11]. Many channels are hosted by publishers that also have outside magazines, websites, or newspapers (eg, Cosmopolitan, National Geographic, BuzzFeed, and the Wall Street Journal).

Discover stories posted by publishers have similar transient features as personal snaps. Depending on the channel's frequency of posting, stories can be available for as little as 24 hours to as long as a few days. Users are able to share stories with friends, but beyond that there is no way for a user to interact directly with a Discover story (ie, no commenting, liking, or direct saving).

Gender Stereotypes, Gender Norms, and Social Media

Gender stereotypes and norms are common in our society [12,13] and are often endorsed across multiple media platforms, including SNS and social media [14-16]. These stereotypes and norms are usually constructed according to cultural values and practices, as well as cultural definitions of masculinity and femininity [12,14,17,18]. They tend to shape the perceptions of acceptable behaviors for men and women [18-21]. Masculinity is commonly associated with being strong, tough, aggressive, independent, and self-reliant. Masculine attributes include having and exercising power and control; the denial of weakness or vulnerability; participating in risky activities or behaviors; sexual infidelity or being less interested in committed, monogamous relationships; and avoiding open expression of emotions [12-14,17,19-21]. In contrast, femininity is commonly associated with being nurturing, affectionate, emotional, and sentimental; avoiding conflict; sexual fidelity; being more acquiescent, submissive, and dependent than males; and investing in physical appearance [12-15,20,21]. The stereotypes and norms endorsed by media and society also shape beliefs with regard to appropriate and expected health behaviors for males and females, including general health and sexual health [18,19,22-24]. In line with stereotypes for masculinity and femininity, being concerned with health, or seeking medical advice about general or sexual health, is seen as a feminine quality and something men do not need to engage in [12,17,25].

Pressure to conform to gender stereotypes and norms can have implications for health. In fact, certain behaviors considered normative for a particular gender have been shown to be associated with health consequences [12,18,26]. The impact gender stereotypes and norms have on health behaviors differ for men and women, with men often engaging in risky health behaviors and women often experiencing limitations in their ability to engage in behaviors that allow them to take control of their health [18]. Furthermore, the perpetuation of gender stereotypes and norms has been shown to influence sexual behaviors and beliefs and reinforce a sexual double standard [19,27].

In general, media sources are saturated with gendered messages that reinforce stereotypes and norms. Exposure to media and internet representations of gender have been shown to influence the endorsement of stereotypical expectations of men and women [16,27,28]. With the growing popularity and evolution of social media, however, it is unclear whether newer media platforms (eg, Snapchat) are reproducing gendered messages or are moving away from stereotypical representations of gender. As there is a limited amount of research in this area, it is important to explore whether gender stereotypes and norms are being reproduced on newer social media platforms.

Review of Snapchat Research

Although its use has become increasingly prevalent [3], research has not kept pace with Snapchat's rapid emergence. Research interests generally center on the use of Snapchat, often focusing on businesses and using Snapchat for brand building. Some research has begun to investigate the effects Snapchat has on interpersonal relationships [1,2]; the uses sought and gratifications obtained through Snapchat [5]; the influence Snapchat has on the social, emotional, and psychological experiences of its users [8]; the impact of social media use on psychosocial functioning during early adolescence [29]; the deterrent mechanisms associated with Snapchat use while driving [30]; the patterns of use on Snapchat [3]; and users' perceptions of different aspects of Snapchat [31]. The focus of the majority of research studies is on the user's interaction with Snapchat and not Snapchat's interactions with its users.

Objectives

A research area receiving little, if any, attention is an examination of newer Snapchat features that allow Snapchat to interact with its users, such as its Discover channels. Snapchat is able to disseminate a variety of information (such as news, current events, and health information) through these channels to an audience who often receives this type of information from SNS and social media platforms [32,33]. In a day and age in which health, sexual health, and gender norms are consistently being questioned and transformed, Snapchat has the opportunity to convey content in a new way that is both relevant and transformative to its audience.

To the researchers' knowledge, no current research has examined the health content shared to Snapchat Discover channels. More specifically, no research has investigated the health content being shared to Discover channels as it relates to gender stereotypes and norms. Thus, an investigation of Discover channels is warranted to provide a better understanding of the content being posted with regard to health and gender. The purpose of this research study, then, was to evaluate qualitatively the general health and sexual health content posted to Discover channels. Further, this study intended to determine whether differences exist between the content posted to male-focused and female-focused channels. The study was guided by the following two research questions:

1. What health content, related to both general health and sexual health, is being shared to Snapchat Discover channels?
2. How does the content vary from male-focused channels to female-focused channels?

Methods

Selection of Discover Channels

Women's and men's magazines, in partnership with Snapchat, have created their own Discover channels and have become publishers on this platform. Magazines often have a stated gendered audience on their respective websites and cater their content to attract this audience [34]. Furthermore, magazines are often specifically labeled as either women's magazines or men's magazines, separating the type of magazine by gender [34,35]. It is reasonable, then, to assume that Discover channels affiliated with women's and men's magazines would have a similar target audience as the magazine's specified audience outside of its Snapchat domain (posting content catered to either males or females). For this reason, magazine Snapchat publishers were chosen because of the ability to determine their gendered audience.

Data Collection

Discover channels were selected from the Discover section on the Snapchat mobile app. To help sift through Discover channels on Snapchat and ensure a gendered audience could be identified, channels were selected according to a set of predetermined inclusion and exclusion criteria. Inclusion criteria were as follows: The channel was a designated publisher with its own publisher stories (content publishers created in partnership with Snapchat); the channel was affiliated with an existing magazine, and the affiliated magazine had information about the gender of their target audience; and an equal number of channels with a primary focus on a male audience and with a primary focus on a female audience had to be selected. Exclusion criteria were as follows: The channel could not be an influencer, and if information about the affiliated magazine's audience was not available then the channel would not be used. The Snapchat Discover page on the mobile app was searched according to these criteria, and four publisher channels were selected: *GQ*, *Esquire*, *Cosmopolitan*, and *SELF*. These four channels were selected because they were the best available options meeting the inclusion and exclusion criteria at the time of data collection.

The target audiences for the four channels were determined according to their associated magazine website. *GQ* (formerly *Gentlemen's Quarterly*) states it is the "premier men's magazine" with the latest tips and advice for men [36]. *Esquire*'s website explains, "it's a magazine for men…it is, and has been for nearly seventy years, a magazine about the interests, the curiosity, the passions, of men" [37]. *Cosmopolitan* is described as the "best-selling young women's magazine in the U.S." [38]. The *SELF* magazine website reads, "SELF Magazine: Women's Workouts, Health Advice & Beauty Tips," demonstrating a predominantly female audience.

Owing to the transient nature of Snapchat, data collection occurred daily. Each Discover channel updated their stories on a different schedule; therefore, data had to be collected and posts had to be saved every day. *Cosmopolitan* was the only channel that updated its stories every 24 hours; it was imperative data collection occurred the day it was posted, or the content would disappear without an alternative way to retrieve it. *GQ*, *Esquire*, and *SELF* all posted on different schedules: *GQ* posted stories two times a week (every Sunday and Wednesday), *Esquire* posted two times a week (every Tuesday and Friday), and *SELF* posted three times a week (every Wednesday, Saturday, and Monday).

It is important to note that saving content from Snapchat was challenging for this study. At the time of data collection, there was no direct way to download or save stories from the app, and stories were available only for a limited amount of time. Several months after the data collection for this study was completed, Snapchat released an update which made stories available for a longer period of time. This update made it possible for Discover stories to be revisited, creating more opportunities for data collection. The update has implications for future research on Discover channels because it would make the data collection process easier.

Data were collected from each of the four Snapchat channels during a specified 4-week period from January 21, 2018, to February 17, 2018. All titles and headlines for stories, with descriptions of what each story entailed, were documented. Screenshots of all stories specifically relevant to general health and sexual health were captured to assist with analyses. For this study, stories were considered related to general health if they mentioned physical, mental, spiritual, or emotional health and well-being. Stories were considered related to sexual health if they mentioned sexual activity, contraception or birth control, reproductive health or hygiene, sexual harassment, or sex-related items (ie, sex positions or sex toys). A total of 406 Discover stories (N=406) were collected and subsequently analyzed using the constant comparative method.

Data Analysis

The constant comparative method was used for the qualitative analysis of the collected Snapchat data [39-42]. The constant comparative method is an iterative process of analyzing qualitative data, involving recoding and recategorizing data as they present themselves and creating new categories when codes do not fit into existing ones. It ensures the systematic comparison of all data with all other data in the data set [42] and is used to reduce data into manageable units and coded information [39-41]. An inductive approach to data analysis was used in which the coding of data was performed without using a pre-existing coding frame. This allowed for an organic exploration of themes as they emerged from the data. In this study, the codes, categories, and themes were directly drawn from the Snapchat data set; no preconceived codes or categories were used to inform the analysis process.

The constant comparative method has been increasingly utilized in SNS and social media research. Hart et al [43] used the method to categorize public health professionals' tweets to evaluate how public health professionals are using Twitter as a platform to further the mission of public health. Fox and Moreland [44] used constant comparison to explore users' negative emotional experiences within Facebook. In another study, Fox et al [45] examined the effects of Facebook on romantic relationships using the constant comparative method for analysis. In their study about social media use by physicians, Campbell et al [46] used the constant comparative method to understand the perceived risks and benefits of using social media as a health professional and the perceptions on social media use in health care. Finally, in their mixed method study, Bayer et al [8] used the constant comparative method to better understand college students' social and emotional experiences on Snapchat.

The data analysis process began with an initial read through to gain an overall understanding of the data. Next, two researchers trained in qualitative research methods (KL and CC) independently open-coded all the Snapchat stories, sorting similar stories into categories. Open coding has been defined as "the process of breaking down, examining, comparing, conceptualizing, and categorizing data" [39-41]. Once initial codes were identified through open coding, comparative analyses were independently conducted within and across Snapchat stories using axial coding [41]. This allowed the researchers to refine the coding structure and group codes into higher level categories [41]. Then, the two researchers collectively identified major themes of the grouped data. Following the development of major themes, the researchers utilized selective coding to compare all stories with their major theme, using constant comparison to refine and sort stories to the most fitting theme. In addition, results were quantified by counting the number of units of data (stories) in each theme and subtheme, considering this information an objective measure to help summarize the results and guide interpretation of the findings.

All stories collected during the data collection period (N=406) were sorted into subthemes and themes through the constant comparative process; there were no stories considered *miscellaneous* or *other* that were not placed under one of the major themes. The researchers coded each story to only one theme or subtheme. Although a story may have related to more than one theme or subtheme, the theme or subtheme with which the story most related was the theme or subtheme that the story was coded under. If there was any disagreement between the two trained researchers, they discussed the topic of the story until an agreement was made. In addition, a third-party individual was available to help with disagreements.

Results

Discover Channels

The data collection process yielded 406 Snapchat Discover stories (N=406), all of which were included in the final analysis. Of the 406 stories, 61 (15.0%) came from *GQ*, 73 (18.0%) came from *Esquire*, 191 (47.0%) came from *Cosmopolitan*, and 81 (20.0%) came from *SELF*. *Cosmopolitan* stories made up almost half of the total stories collected. This is likely because it was the only channel that posted new content every day, whereas the other three channels posted stories two or three times per week.

GQ was one of the male-focused Discover channels used for comparison in this study. From the analysis, 16 subthemes were coded into five main themes: entertainment, health, fashion, dating and relationships, and grooming (see Multimedia Appendix 1). The theme with the most stories was fashion (27/61, 44%), with entertainment coming in second (13/61, 21%). This is not surprising given *GQ* magazine emphasizes fashion and style for men. The theme with the least number of stories was health (5/61, 8%). *GQ* was also the channel with the least number of stories overall compared with the other channels.

Esquire was the other male-focused Discover channel used for comparison in this study. There were 13 subthemes coded into five main themes: entertainment, health, sexual orientation, dating and relationships, and grooming (see Multimedia Appendix 1). The theme with the most stories was entertainment, comprising more than half (42/73, 58%) of all the stories posted by this publisher. The theme with the least number of stories was dating and relationships (2/73, 3%). The health theme was second to last, accounting for 8% (6/73) of stories.

Cosmopolitan was one of the female-focused Discover channels used for comparison in this study. *Cosmopolitan* had the most stories of all the channels analyzed, as well as the most themes and subthemes. There were 28 subthemes coded into seven main themes: entertainment, health, sex and relationships, appearances, politics, food and beverages, and home and home goods (see Multimedia Appendix 1). Themes with the most stories were entertainment (56/191, 29.3%), sex and relationships (52/191, 27.2%), and health (35/191, 18.3%). The themes with the least stories were home and home goods (9/191, 4.7%) and politics (5/191, 2.6%).

SELF was the second female-focused Discover channel used for comparison in this study. *SELF* had the second largest number of stories out of the four channels. As seen in Multimedia Appendix 1, *SELF* had 17 subthemes coded into six main themes: entertainment, health, fitness, sexual and reproductive health, dating and relationships, and appearance. Of the six major themes, the theme with the most stories was health (28/81, 35%), followed by fitness (15/81, 19%). This was the only channel that had a health theme with the most stories which is likely because *SELF* is billed as a health magazine. The theme with the least stories was dating and relationships (3/81, 4%).

Health Content Comparison

The male-focused channels had few stories with general health content. When combined together, the male-focused channels had a total of 10 health stories, accounting for only 7.5% (10/134) of the total stories among the two channels (Table 1). The female-focused channels had a total of 62 stories with health content, accounting for 22.8% (62/272; approximately three times as many health stories as the male-focused channels) of the total stories among the two channels (Table 1).

Table 1. Number and percentages of general health stories from male- and female-focused Discover channels.

Discover channels	General health stories, n (%[a])
Male-focused channels (n=134)	
Combined	10 (7.5)
GQ[b]	4 (3.0)
Esquire	6 (4.5)
Female-focused channels (n=272)	
Combined	62 (22.8)
Cosmopolitan	19 (7.0)
SELF	43 (15.8)

[a] Percentage of health stories from the combined total stories.

[b] GQ: Gentlemen Quarterly.

Differences in subthemes and general health content coverage existed between the male- and female-focused channels. The scope of general health content was greater for female-focused channels compared with male-focused channels, with female-focused channels covering a much wider range of topics (Table 2). Moreover, the health content addressed by male-focused channels covered most of their health topics only once. For example, the male-focused channels posted only one story about video games and health titled *One very, very good reason to start gaming*, which discussed the link found by researchers between video games and a reduced risk of Alzheimer disease (Figure 1). The male-focused channels also only posted one story about mental health titled *How to: Not be sad this winter*, which discussed the signs, symptoms, and treatment of seasonal affective disorder (Figure 1). The female-focused channels, however, often posted multiple separate stories about the same general health topic.

Table 2. Health content subthemes of male- and female-focused general health stories.

Gendered audience	Subthemes
Male-focused channels	
GQ[a]	Exercise, mental health, and food
Esquire	Exercise, video games and health, and health podcast recommendations
Female-focused channels	
Cosmopolitan	Exercise, health information and advice, skin care, and emotional health
SELF	Health information and advice, physical health, diet, exercise, exercise equipment, and recovery

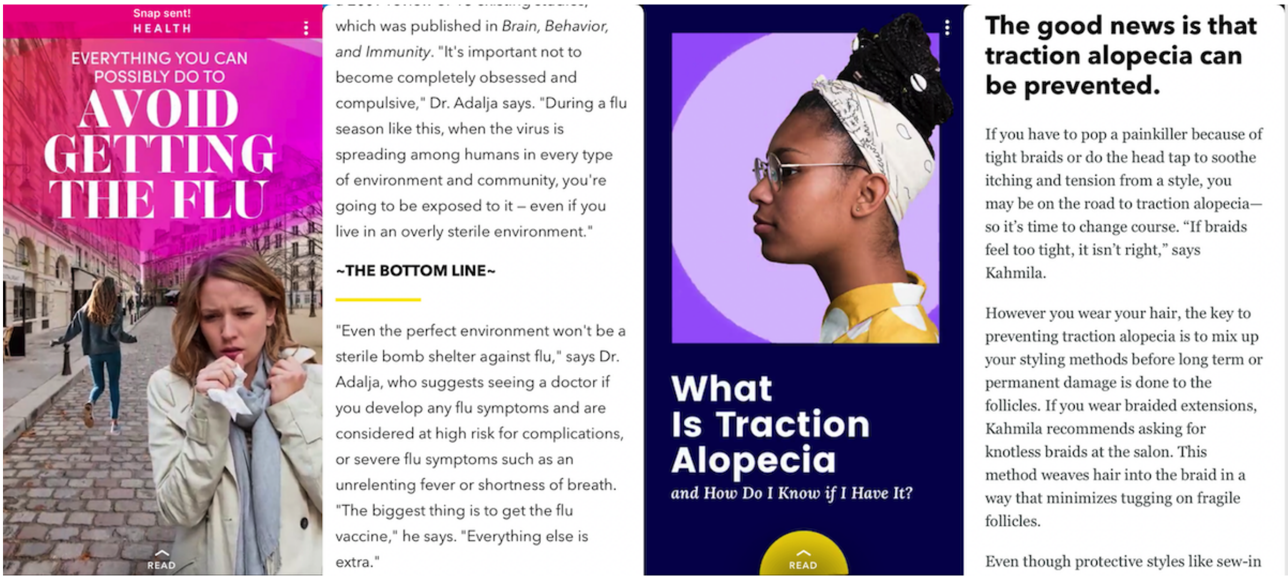
[a] GQ: Gentlemen Quarterly.

Figure 1. Examples of general health topics posted to male-focused Discover channels.

Start with dynamic, movement-based warm-ups, which prepare the body for controlled, explosive, and forceful movement through the full range of motion.👉

The male-focused channels were more general and less extensive in their approach to providing health content. A majority of the posted stories included only one or two brief paragraphs or a short list describing the health topic, including *How to Properly Warm Up for Your Next Gym Day* or *The Breakup Workout...Because Revenge is Best Served as a Six Pack of Abs* (Figure 1). In contrast, the general health content delivered by the female-focused channels posted more specific and descriptive health content. The channels' stories offered tangible advice and in-depth explanations about the posted health topic, such as *Everything You Can Possibly Do To Avoid Getting the Flu*, *7 Weird Signs You're Not Drinking Enough Water*, and *What Is Traction Alopecia and How Do I Know if I Have It?* (Figure 2).

Figure 2. Examples of general health topics posted to female-focused Discover channels.



which was published in *Brain, Behavior, and Immunity*. "It's important not to become completely obsessed and compulsive," Dr. Adalja says. "During a flu season like this, when the virus is spreading among humans in every type of environment and community, you're going to be exposed to it – even if you live in an overly sterile environment."

~THE BOTTOM LINE~

"Even the perfect environment won't be a sterile bomb shelter against flu," says Dr. Adalja, who suggests seeing a doctor if you develop any flu symptoms and are considered at high risk for complications, or severe flu symptoms such as an unrelenting fever or shortness of breath. "The biggest thing is to get the flu vaccine," he says. "Everything else is extra."

The good news is that traction alopecia can be prevented.

If you have to pop a painkiller because of tight braids or do the head tap to soothe itching and tension from a style, you may be on the road to traction alopecia—so it's time to change course. "If braids feel too tight, it isn't right," says Kahmila.

However you wear your hair, the key to preventing traction alopecia is to mix up your styling methods before long term or permanent damage is done to the follicles. If you wear braided extensions, Kahmila recommends asking for knotless braids at the salon. This method weaves hair into the braid in a way that minimizes tugging on fragile follicles.

Even though protective styles like sew-in

Sexual Health Content Comparison

The male-focused channels posted less stories covering sexual health content than the female-focused channels. Male-focused channels posted a total of four sexual health stories, comprising 3.0% (4/134) of the total stories among the two male-focused channels. The four sexual health stories only came from the *GQ*; *Esquire* did not post any stories with sexual health content (Table 3). The female-focused channels posted a total of 51 sexual health stories, comprising 18.8% (51/272) of the total stories among the two channels (Table 3). More than half of the sexual health posts came from *Cosmopolitan*. The female-focused channels also covered a larger variety of topics. The male-focused channels had two main subthemes addressing sexual health, whereas the female-focused channels had a number of subthemes addressing sexual health (Table 4).

Table 3. Number and percentages of sexual health stories from male- and female-focused Discover channels.

Discover channels	Sexual health stories, n (%[a])
Male-focused channels (n=134)	
Combined	4 (3.0)
GQ[b]	4 (3.0)
Esquire	0 (0.0)
Female-focused channels (n=272)	
Combined	51 (18.8)
Cosmopolitan	35 (12.9)
SELF	16 (5.9)

[a] Percentage of health stories from the combined total stories.

[b] GQ: Gentlemen Quarterly.

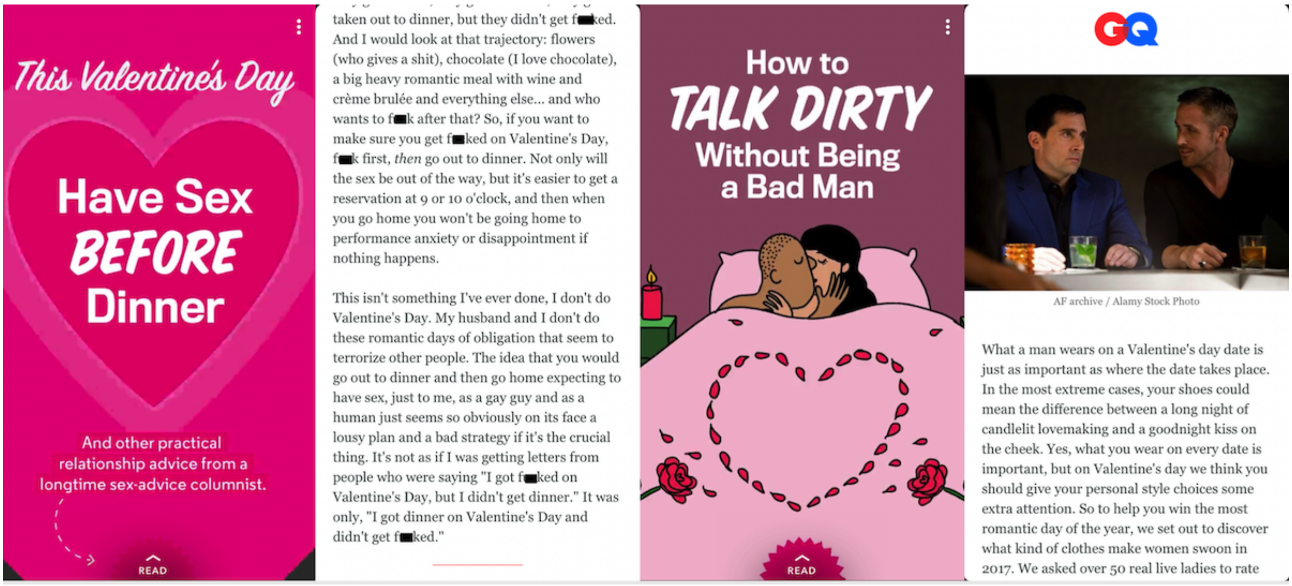
Table 4. Sexual health content subthemes of male- and female-focused sexual health stories.

Gendered audience	Subthemes
Male-focused channels	
GQ[a]	Dating and sex
Esquire	None
Female-focused channels	
Cosmopolitan	Pregnancy, sexual and reproductive health, sexual assault and harassment, dating, sex toys, and sex secrets
SELF	Birth control, feminine hygiene, reproductive health, women's rights, and dating

[a] GQ: Gentlemen Quarterly.

Although there were only four stories for the male-focused channels, apparent differences in content emerged. The male-focused channels did not post any sex-related stories that specifically included sexual health information; rather, they posted about sexual activity and sexual success. For example, one of the stories titled *This Valentine's Day Have Sex Before Dinner* discussed engaging in sexual activity before going out to dinner on Valentine's Day to guarantee you successfully engage in sexual activity. The man giving the advice wrote, "if you want to make sure you get f****ed on Valentine's day, f*** first, *then* go out to dinner" (Figure 3). Another story titled *Never Play Acoustic Guitar for a Woman* told men to "never play guitar for someone you're trying to sleep with" because it is not seductive and might be a turn off. The third story titled *What Should You Wear on Valentine's Day?* gave advice about what a man should wear to avoid a "goodnight kiss on the cheek" and successfully ensure a "long night of candlelit lovemaking" (Figure 3).

Figure 3. Examples of sexual health topics posted to male-focused Discover channels.



taken out to dinner, but they didn't get f[redacted]ked. And I would look at that trajectory: flowers (who gives a shit), chocolate (I love chocolate), a big heavy romantic meal with wine and crème brulée and everything else... and who wants to f[redacted]k after that? So, if you want to make sure you get f[redacted]ked on Valentine's Day, f[redacted]k first, *then* go out to dinner. Not only will the sex be out of the way, but it's easier to get a reservation at 9 or 10 o'clock, and then when you go home you won't be going home to performance anxiety or disappointment if nothing happens.

This isn't something I've ever done, I don't do Valentine's Day. My husband and I don't do these romantic days of obligation that seem to terrorize other people. The idea that you would go out to dinner and then go home expecting to have sex, just to me, as a gay guy and as a human just seems so obviously on its face a lousy plan and a bad strategy if it's the crucial thing. It's not as if I was getting letters from people who were saying "I got f[redacted]ked on Valentine's Day, but I didn't get dinner." It was only, "I got dinner on Valentine's Day and didn't get f[redacted]ked."

AF archive / Alamy Stock Photo

What a man wears on a Valentine's day date is just as important as where the date takes place. In the most extreme cases, your shoes could mean the difference between a long night of candlelit lovemaking and a goodnight kiss on the cheek. Yes, what you wear on every date is important, but on Valentine's day we think you should give your personal style choices some extra attention. So to help you win the most romantic day of the year, we set out to discover what kind of clothes make women swoon in 2017. We asked over 50 real live ladies to rate

The female-focused channels posted stories with a variety of sex-related topics, many of which included sexual health-related information, and were the only channels to post stories about contraception. Stories about contraception included *7 Best Places to Get Free Condoms*, *How to Find the Best Condom for You*, *5 Signs an IUD is the Best Birth Control for You*, and *Here are the Signs an IUD Isn't Right for You* (Figure 4). Interestingly, the female-focused channels did mention males providing condoms or preparing for sexual intercourse by bringing condoms in a couple of their stories. The male-focused channels made no mention of contraceptive use in any of their stories about sex.

Figure 4. Examples of sexual health topics posted to female-focused Discover channels.

The female-focused channels also posted sexual health information about reproductive health (such as female internal and external genitalia) and feminine hygiene. This information was often presented in the form of advice or recommendations from reputable sources (eg, doctors, specialists, and experts). Examples include stories such as *7 Things Your OB/GYN Wants You to Know About Your Vagina*, *11 Things You Need to Know About Your Vulva*, *Is It Normal to Have a Really Long Menstrual Cycle?*, and *Keeping Your Vagina Clean* (Figure 4). The male-focused channels did not offer similar information about male reproductive health or hygiene in their posted stories.

Another difference in the sexual health content across female- and male-focused channels was the way in which they were framed. Stories posted to female-focused channels often made it seem as if men possess secrets about sexual intercourse and sexual activities that they do not want women to know. Examples of these stories include *9 Sixty-Nine Secrets Guys Won't Tell You*, *10 Anal Sex Secrets Guys Won't Tell You*, and *10 Sex Toy Secrets Guys Won't Tell You* (Figure 4). The focus of these stories was often about what men think or how men feel without mentioning the same for women. In contrast, as stated previously, the main focus of the stories from the male-focused channels was on setting up men for sexual success and sexual pleasure.

Discussion

Principal Findings

This study sought to examine the general health and sexual health content being posted to Snapchat Discover channels, as well as to determine whether differences exist between the content posted based on the gender-specific audiences. In examining the content of Discover channels, this preliminary study moves beyond the traditional focus on the effect of users' interactions with main Snapchat features to explore a different feature that offers a variety of health information to users through some of its publisher channels. Compared with the male-focused channels, female-focused channels had more stories offering general health–related and sexual health–related content. The scope of their general health and sexual health content was greater than that of the male-focused channels, and the number of stories addressing each general health and sexual health topic was greater for female-focused than male-focused channels. The findings from this study demonstrate that the analyzed Snapchat Discover channels reflect and reiterate stereotypical conceptions of gender through the content posted. If users are adopting behaviors learned from content posted to these channels, then it is possible that users not only are being exposed to this content but also might be engaging in behaviors that reinforce gender stereotypes and norms for general health and sexual health.

General Health Content and Promotion of Gender Stereotypes

A focus of this study was to investigate the general health content being shared to Discover channels and the differences in content posted to male- and female-focused channels. It is evident from the analysis that target audiences are perceived differently by their respective publishers with regard to their overall health behaviors and health information–seeking behaviors. Stories posted to the male-focused channels were more superficial in their general health content. These channels mostly posted health information with the intent of providing ways to make you appear more attractive, not necessarily with the intent of providing helpful information to enhance or maintain your overall health. This reflects the gender stereotype that participating in health-promoting behaviors is perceived as less masculine [12,17,18]. In contrast, the female-focused channels established a sense of agency and control for their audience over their health and well-being. These channels addressed more aspects of overall health with an emphasis on prevention, self-care, healthy habits, and taking ownership of your health.

These findings are in line with research that has documented significant gender differences in health behaviors [12,17,18]. In fact, gender has been found to be one of the most important factors that influences health-related behavior, with women generally adopting healthier behavior practices than men [12]. Traditional masculine ideals endorse the gender stereotype that men should be strong, independent, and self-reliant, and therefore do not need to practice health-seeking behaviors [12,17,18]. Whether intentionally or not, by reiterating this gender stereotype through content posted (or, in some cases, not posted), it is possible that these male-focused channels are influencing social norms among their target audience. The content shared to these channels is promoting the idea that health behaviors should be more important to one gender than the other, and that health-related behaviors are defined by one's gender. There is a need to undo the stereotypes and norms that make health behaviors typical of a single gender, and instead inclusive of all genders. Social media could take steps toward this by representing health behaviors in new ways and posting about health behaviors to both male- and female-focused Snapchat channels.

Sexual Health Content and Promotion of Gender Stereotypes

The findings suggest that Discover channels also promote sexual health gender stereotypes and norms. There was a difference in the number of posts with sexual health–related content, with female-focused channels posting more content. This difference suggests that sexual relationships and sexual health are primarily topics for women to concern themselves with and topics men do not have to worry about. Thus, women are seen as wanting and needing sexual health information, whereas similar information is seen as unimportant or unnecessary for men, or knowledge men already possess.

In line with findings from previous research [19,35], when sex-related topics were posted to the male-focused channels, the primary focus was on men's sexual freedom and sexual success in heterosexual relationships. All of the stories with sexual content focused on how to successfully ensure you have sex with a woman. Men were told how to dress attractively to have sex with their date, how to plan their evening around having sex, how to appropriately *talk dirty*, and what not to do to have sex with a woman. These stories endorse the stereotypes that link being obsessed with and driven by sex to masculinity [12,22,35], undermining the idea that men should be interested and involved in their sexual health. They also promote the stereotype that women serve to provide sex to men on men's terms [14,35].

Furthermore, the lack of sexual health–related information among male-focused channels suggests that casual sexual intercourse and activities are normal for men without any sexual health concerns or responsibilities attached to them. The male-focused channels posted four stories specifically about having sex without any mention of contraception or sexual health. This contributes to the perception of sex as a carefree activity for men and reinforces the stereotype that safe sex practices (in heterosexual relationships) are primarily the responsibility of women [20,23,35]. Female-focused channels posted some content about condoms and birth control. Condoms were described to females as cost-effective forms of birth control that can protect against sexually transmitted infections (STIs) when used correctly. All condoms discussed were male condoms, with no mention of female condoms, which is interesting considering these channels have mainly female audiences. The presence of sexual health content on female-focused channels, and not on male-focused channels, reinforces the stereotype that women bear the responsibility for contraception and pregnancy prevention [22,35]. It is important to note that the responsibility for safe sex in heterosexual relationships often hinges on the woman's ability to negotiate safe sex practices. Negotiations such as these require some degree of assertiveness, an attribute that is societally viewed as contradictory to femininity. This conflict between assertiveness and passivity complicates conversations about sexual safety for women [35]. Research has found that power imbalances in relationships make it difficult to plan to have safe sex, such as taking precautions to protect against STIs and pregnancy [24]. Thus, the findings also demonstrate that the stigma attached to women carrying condoms, in combination with the expectations for women to be responsible for pregnancy prevention, might put women in a difficult situation that could lead to negative consequences, such as STIs or unplanned pregnancies [19,24].

Moreover, the stories posted to female-focused channels often instructed women about how to understand men's sexual interests and behaviors by disclosing the *secrets* men are keeping from them. Men are portrayed as possessing secrets about their own sexual behavior women want to know. Essentially, Snapchat stories are acting as portals that unlock access to secrets, so women can learn how men think and feel about sexual intercourse and activities, with the goal of enabling women to serve men's desires more successfully. The same types of secrets are not shared through stories on the male-focused channels. This difference in content suggests an interesting power differential, aligning with similar findings from a study about sexual double standards [19,20]. Being portrayed as possessing all of the knowledge they need about sex puts men in a stereotypical masculine position of power and control [12,19,35]. It is possible that the female-focused channels are attempting to address this power differential by revealing the secrets men are keeping; however, the way the content is framed promotes the idea that men do, in fact, have more knowledge about sex than women, and women should be most concerned with providing male pleasure.

The amount of information about reproductive health and hygiene posted to the Discover channels also reinforced gender stereotypes and norms regarding sexual health. The female-focused channels posted several stories about female internal and external genitalia and appropriate feminine hygiene, whereas the male-focused channels posted no sexual health content about reproductive health and hygiene. This agrees with research that have found norms surrounding femininity encourage women to understand and examine their own bodies and recognize subtle signs of change [25]. Norms surrounding masculinity discourage men from understanding or examining their bodies [25]. Men tend to avoid discussing sexual health issues, which could be why male-focused channels also avoid posting stories containing this type of content. These channels may be providing a disservice to their male audiences by failing to post meaningful discussions that consider the role of masculinity in relation to sexual health.

Limitations and Recommendations

This study is not without limitations. First, the data are subjective based on the researchers' interpretations, as is the nature of qualitative research. The information gained from a qualitative process, however, provides a nuanced understanding into the content posted by Snapchat publishers for their gender-specific audiences. Another limitation of this study is that the data collection period was only 4 weeks long. It is possible that different themes would have emerged had the data collection period been longer. The limited number of stories and short data collection period reduce the explanatory power and generalizability of the findings. Future research examining Snapchat Discover stories should collect data for a longer period of time to enhance explanatory power and generalizability. In addition, there are limitations to the selected channels themselves. As the channels selected for this study had to be linked to magazines (based on the inclusion and exclusion criteria), the choices of channels were limited. The nature of the channels selected might not be the most representative or comparable with respect to audience and content. For example, *GQ* is often described as a style magazine and is likely to have a different male audience than other magazines. The same could be said for the audiences of *SELF* and *Cosmopolitan*. The content posted to these channels is likely to align with the purpose and target audience of the magazine. Thus, we cannot generalize these findings to all Discover channels on Snapchat. Future research could benefit by identifying Discover channels that have more comparable target audiences and focus mostly on health and sexual health content. It might also be important to consider the number of subscribers a channel has. This study did not select channels based on the number of subscribers; however, it is possible that the popularity of a channel might impact the content being posted and how much influence a channel could have on its users.

As social media continues to evolve, opportunities are presented for these platforms to conform with, challenge, or defy societal expectations and normative ideas [14]. Social media apps have the ability to break away from gender stereotypes and norms by changing the type of content they deliver, and the way relevant topics are discussed. Snapchat Discover channels, in particular, are capable of doing this by posting general health and sexual health content in a way that does not conform to gender stereotypes and norms. In doing so, these platforms can transform alongside their audiences who are slowly changing and building new meanings of health, gender, sexuality, and identity [34].

The authors have a few practical recommendations to aid these processes. First, social media publishers should acknowledge the pervasiveness of gender stereotypes and accept that these stereotypes might be biasing the content being posted. This could help to identify and correct such biases [28]. Second, social media publishers could work to educate themselves and their content curators about gender stereotypes [28]. This includes learning about the implications that reinforcing and endorsing stereotypes might have for consumers of the content. It is important for social media publishers to be more aware of the potential implications gendered messages can have for users, especially with regard to health and health behaviors. It is also important for social media platforms to bear responsibility for their content decisions. Third, social media publishers could seek feedback from members of a different gender before posting potential content. This could be in the form of encouraging equal representation of genders on content creation teams or hiring a gender expert to consult on content creation. Finally, social media publishers should work to create new understandings of how they view their gendered audiences. Publishers could then use their new understandings to generate content that incorporates more effective and inclusive gender portrayals.

Conclusions

This study has provided a first look at the general health and sexual health content posted to Snapchat Discover channels. The findings suggest that publishers posting content to the male-focused channels view their audiences as less likely to seek health information than their female counterparts, as obsessed with and driven by sex, and as less concerned with sexual health. Publishers creating content for the female-focused channels view their audiences as health information seekers, as concerned with sexual health, as responsible for contraception and pregnancy prevention, and as less informed about sex than their male counterparts. The content posted to the male- and female-focused channels seems to be promoting gender stereotypes and responsibilities for health information seeking and for sexual health and sexual relationships. By posting relevant content in new ways that neither represents traditional femininity nor hegemonic masculinity, social media can help to slowly transform restrictive gender norms and socially constructed stereotypes.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Themes and underlying subthemes of Snapchat Discover channel stories.

[DOCX File , 16 KB - jmir_v22i3e15330_app1.docx]

References

1. Vaterlaus JM, Barnett K, Roche C, Young JA. 'Snapchat is more personal': An exploratory study on Snapchat behaviors and young adult interpersonal relationships. Comput Hum Behav 2016;62:594-601. [doi: 10.1016/j.chb.2016.04.029]
2. Utz S, Muscanell N, Khalid C. Snapchat elicits more jealousy than Facebook: a comparison of Snapchat and Facebook use. Cyberpsychol Behav Soc Netw 2015 Mar;18(3):141-146. [doi: 10.1089/cyber.2014.0479] [Medline: 25667961]
3. Piwek L, Joinson A. 'What do they snapchat about?' Patterns of use in time-limited instant messaging service. Comput Hum Behav 2016;54:358-367. [doi: 10.1016/j.chb.2015.08.026]
4. Clement J. Statista. 2019 Feb 11. Snapchat - Statistics & Facts URL: https://www.statista.com/topics/2882/snapchat/ [accessed 2020-01-24]
5. Punyanunt-Carter NM, De La Cruz JJ, Wrench JS. Investigating the relationships among college students' satisfaction, addiction, needs, communication apprehension, motives, and uses & gratifications with Snapchat. Comput Hum Behav 2017;75:870-875. [doi: 10.1016/j.chb.2017.06.034]
6. Aslam S. Omnicore Agency. 2020 Jan 13. Snapchat by the Numbers: Stats, Demographics & Fun Facts URL: https://www.omnicoreagency.com/snapchat-statistics/ [accessed 2019-06-11]
7. Smith A, Anderson M. Pew Research Center. 2018 Mar 1. Social Media Use in 2018 URL: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ [accessed 2019-06-11]
8. Bayer JB, Ellison NB, Schoenebeck SY, Falk EB. Sharing the small moments: ephemeral social interaction on Snapchat. Inf Commun Soc 2015 Sep 18;19(7):956-977. [doi: 10.1080/1369118X.2015.1084349]
9. Wallenstein A, Ault S. Variety. 2016 Mar 24. Snapchat Content Survey: How Much Millennials Actually Use Live Stories, Discover and More URL: https://variety.com/2016/data/news/snapchat-content-survey-how-much-millennials-actually-use-live-stories-discover-and-more-1201736616/ [accessed 2019-06-11]
10. Hutchinson A. Social Media Today. 2016 Jun 7. Snapchat Revamps Discover to Boost Readership URL: https://www.socialmediatoday.com/social-networks/snapchat-revamps-discover-boost-readership [accessed 2019-06-11]
11. Flynn K. International Business Times. 2016 Jan 27. Snapchat Discover One Year Later: How 23 Media Companies Are Building Stories For Evan Spiegel URL: https://www.ibtimes.com/snapchat-discover-one-year-later-how-23-media-companies-are-building-stories-evan-2281851 [accessed 2019-06-11]
12. 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]
13. Eisenchlas SA. Gender roles and expectations: any changes online? SAGE Open 2013;3(4):1-11. [doi: 10.1177/2158244013506446]
14. Rose J, Mackey-Kallis S, Shyles L, Barry K, Biagini D, Hart C, et al. Face it: The impact of gender on social media images. Commun Q 2012;60(5):588-607. [doi: 10.1080/01463373.2012.725005]
15. Bailey J, Steeves V, Burkell J, Regan P. Negotiating with gender stereotypes on social networking sites: from 'bicycle face' to Facebook. J Commun Inquiry 2013;37(2):91-112. [doi: 10.1177/0196859912473777]
16. Wood JT. Gendered media: the influence of media on views of gender. In: Gendered Lives: Communication, Gender and Culture. Belmont, CA: Wadsworth; 1994:231-244.

17. Mahalik JR, Burns SM, Syzdek M. Masculinity and perceived normative health behaviors as predictors of men's health behaviors. Soc Sci Med 2007 Jun;64(11):2201-2209. [doi: 10.1016/j.socscimed.2007.02.035] [Medline: 17383784]
18. Fleming PJ, Agnew-Brune C. Current Trends in the study of Gender Norms and Health Behaviors. Curr Opin Psychol 2015 Oct 1;5:72-77 [FREE Full text] [doi: 10.1016/j.copsyc.2015.05.001] [Medline: 26075291]
19. Lefkowitz ES, Shearer CL, Gillen MM, Espinosa-Hernandez G. How gendered attitudes relate to women's and men's sexual behaviors and beliefs. Sex Cult 2014 Mar;18(4):833-846 [FREE Full text] [doi: 10.1007/s12119-014-9225-6] [Medline: 25435760]
20. Petersen JL, Hyde JS. Gender differences in sexual attitudes and behaviors: a review of meta-analytic results and large datasets. J Sex Res 2011 Mar;48(2-3):149-165. [doi: 10.1080/00224499.2011.551851] [Medline: 21409712]
21. Prentice DA, Carranza E. What women and men should be, shouldn't be, are allowed to be, and don't have to be: the contents of prescriptive gender stereotypes. Psychol Women Q 2002;26(4):269-281. [doi: 10.1111/1471-6402.t01-1-00066]
22. Hust SJ, Brown JD, L'Engle KL. Boys will be boys and girls better be prepared: An analysis of the rare sexual health messages in young adolescents' media. Mass Commun Soc 2008;11(1):3-23. [doi: 10.1080/15205430701668139]
23. Marston C, King E. Factors that shape young people's sexual behaviour: a systematic review. Lancet 2006 Nov 4;368(9547):1581-1586. [doi: 10.1016/S0140-6736(06)69662-1] [Medline: 17084758]
24. Brown S. 'They think it's all up to the girls': gender, risk and responsibility for contraception. Cult Health Sex 2015;17(3):312-325. [doi: 10.1080/13691058.2014.950983] [Medline: 25270238]
25. Owen C, Campbell C. How do men's magazines talk about penises? J Health Psychol 2018 Feb;23(2):332-344. [doi: 10.1177/1359105317745333] [Medline: 29198153]
26. Weber AM, Cislaghi B, Meausoone V, Abdalla S, Mejía-Guevara I, Loftus P, Gender Equality, Norms and Health Steering Committee. Gender norms and health: insights from global survey data. Lancet 2019 Jun 15;393(10189):2455-2468. [doi: 10.1016/S0140-6736(19)30765-2] [Medline: 31155273]
27. Ward LM, Harrison K. The impact of media use on girls' beliefs about gender roles, their bodies, and sexual relationships: a research synthesis. In: Cole E, Daniel JH, editors. Featuring Females: Feminist Analyses of Media. Washington, DC, US: American Psychological Association; 2005:3-23.
28. Ellemers N. Gender stereotypes. Annu Rev Psychol 2018 Jan 4;69:275-298. [doi: 10.1146/annurev-psych-122216-011719] [Medline: 28961059]
29. Vannucci A, McCauley Ohannessian C. Social media use subgroups differentially predict psychosocial well-being during early adolescence. J Youth Adolesc 2019 Aug;48(8):1469-1493. [doi: 10.1007/s10964-019-01060-9] [Medline: 31256313]
30. Truelove V, Freeman J, Davey J. 'I Snapchat and Drive!' A mixed methods approach examining snapchat use while driving and deterrent perceptions among young adults. Accid Anal Prev 2019 Oct;131:146-156. [doi: 10.1016/j.aap.2019.06.008] [Medline: 31255800]
31. Rauzzino A, Correa JC. Millennials sex differences on Snapchat perceived privacy. Suma Psicológica 2017;24(2):129-134. [doi: 10.1016/j.sumpsi.2017.08.002]
32. Shearer E, Matsa KE. Pew Research Center. 2018 Sep 10. News Use Across Social Media Platforms 2018 URL: https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/ [accessed 2020-01-24]
33. Basch CH, MacLean SA, Romero R, Ethan D. Health information seeking behavior among college students. J Community Health 2018 Dec;43(6):1094-1099. [doi: 10.1007/s10900-018-0526-9] [Medline: 29779074]
34. Gauntlett D. Media, Gender and Identity. Second Edition. New York, NY: Routledge; 2008.
35. Moran C, Lee C. On his terms: representations of sexuality in women's magazines and the implications for negotiating safe sex. Psychol Sex 2011;2(2):159-180. [doi: 10.1080/19419899.2010.534489]
36. Condé Nast. GQ URL: https://www.condenast.com/brands/gq [accessed 2019-06-11]
37. Esquire. 2007 Feb 15. What is Esquire? URL: https://www.esquire.com/about/a2198/what-is-esquire/ [accessed 2019-06-11]
38. Hearst. Cosmopolitan URL: https://www.hearst.com/magazines/cosmopolitan [accessed 2019-06-11]
39. Glaser BG, Strauss AL. Discovery of Grounded Theory: Strategies for Qualitative Research. New York, NY: Transaction Publishers; 1967.
40. Patton MQ. Qualitative Evaluation and Research Methods. Thousand Oaks, CA: Sage Publications, Inc; 2002.
41. Strauss A, Corbin JM. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage Publications, Inc; 1990.
42. Fram SM. The constant comparative analysis method outside of grounded theory. Qual Rep 2013;18(1):1-25 [FREE Full text]
43. Hart M, Stetten N, Islam S, Pizarro K. Twitter and Public Health (Part 2): qualitative analysis of how individual health professionals outside organizations use microblogging to promote and disseminate health-related information. JMIR Public Health Surveill 2017 Oct 4;3(4):e54 [FREE Full text] [doi: 10.2196/publichealth.6796] [Medline: 28978500]
44. Fox J, Moreland JJ. The dark side of social networking sites: an exploration of the relational and psychological stressors associated with Facebook use and affordances. Comput Hum Behav 2015;45:168-176 [FREE Full text] [doi: 10.1016/j.chb.2014.11.083]

45. Fox J, Osborn JL, Warber KM. Relational dialectics and social networking sites: the role of Facebook in romantic relationship escalation, maintenance, conflict, and dissolution. Comput Hum Behav 2014;35:527-534 [FREE Full text] [doi: 10.1016/j.chb.2014.02.031]
46. Campbell L, Evans Y, Pumper M, Moreno MA. Social media use by physicians: a qualitative study of the new frontier of medicine. BMC Med Inform Decis Mak 2016 Jul 15;16:91 [FREE Full text] [doi: 10.1186/s12911-016-0327-y] [Medline: 27418201]

Abbreviations

SNS: social networking sites
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 06.07.19; peer-reviewed by L Shen, N Kordzadeh; comments to author 11.08.19; revised version received 04.11.19; accepted 16.12.19; published 20.03.20.

Please cite as:
LeBeau K, Carr C, Hart M
Examination of Gender Stereotypes and Norms in Health-Related Content Posted to Snapchat Discover Channels: Qualitative Content Analysis
J Med Internet Res 2020;22(3):e15330
URL: http://www.jmir.org/2020/3/e15330/
doi:10.2196/15330
PMID:32196461

Original Paper

Impact of the Use of Electronic Health Tools on the Psychological and Emotional Well-Being of Electronic Health Service Users (The Seventh Tromsø Study - Part 3): Population-Based Questionnaire Study

Andrius Budrionis[1], MSc, PhD; Rolf Wynn[2,3], MD, PhD; Luis Marco-Ruiz[1], MSc, PhD; Kassaye Yitbarek Yigzaw[1], MSc, PhD; Svein Bergvik[4], PhD; Sunday Oluwafemi Oyeyemi[5], MSc, MD, MPH; Johan Gustav Bellika[1,2], MSc, PhD

[1]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway
[2]Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway
[3]Division of Mental Health and Addictions, University Hospital of North Norway, Tromsø, Norway
[4]Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway
[5]Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Johan Gustav Bellika, MSc, PhD
Department of Clinical Medicine
Faculty of Health Sciences
UiT The Arctic University of Norway
Postboks 6050 Langnes
Tromsø, 9037
Norway
Phone: 47 95748049
Email: johan.gustav.bellika@ehealthresearch.no

Abstract

Background: Electronic health (eHealth) has been described as a silver bullet for addressing how challenges of the current health care system may be solved by technological solutions in future strategies and visions for modern health care. However, the evidence of its effects on service quality and cost effectiveness remains unclear. In addition, patients' psychological and emotional reactions to using eHealth tools are rarely addressed by the scientific literature.

Objective: This study aimed to assess how the psychological and emotional well-being of eHealth service users is affected by the use of eHealth tools.

Methods: We analyzed data from a population-based survey in Norway, conducted in the years 2015-2016 and representing 10,604 eHealth users aged over 40 years, to identify how the use of eHealth tools was associated with feeling anxious, confused, knowledgeable, or reassured. Associations between these four emotional outcomes and the use of four types of eHealth services (Web search engines, video search engines, health apps, and social media) were analyzed using logistic regression models.

Results: The use of eHealth tools made 72.41% (6740/9308) of the participants feel more knowledgeable and 47.49% (4421/9308) of the participants feel more reassured about their health status. However, 25.69% (2392/9308) reported feeling more anxious and 27.88% (2595/9308) reported feeling more confused using eHealth tools. A high level of education and not having a full-time job were associated with positive reactions and emotions (feeling more knowledgeable and reassured), whereas low self-reported health status and not having enough friends who could provide help and support predicted negative reactions and emotions (ie, feeling anxious and confused). Overall, the positive emotional effects of eHealth use (feeling knowledgeable and reassured) were relatively more prevalent among users aged over 40 years than the negative emotional effects (ie, feeling anxious and confused). About one-fourth of eHealth users reported being more confused and anxious after using eHealth services.

Conclusions: The search for health information on the internet can be motivated by a range of factors and needs (not studied in this study), and people may experience a range of reactions and feelings following health information searching on the Web. Drawing on prior studies, we categorized reactions as positive and negative reactions. Some participants had negative reactions, which is challenging to resolve and should be taken into consideration by eHealth service providers when designing services (ie, including concrete information about how users can get more help and support). There is a need for more studies examining a greater range of reactions to online health information and factors that might predict negative reactions to health information on the Web.

(J Med Internet Res 2020;22(3):e13118) doi:10.2196/13118

KEYWORDS

eHealth; telemedicine; health information; Tromsø study; health-related internet use; emotions; effect; anxious; confused; knowledgeable; reassured

Introduction

In Norway, the demand for health care services is expected to increase by 40% by 2040 [1]. Comparable estimates may be expected for other developed countries, predicting a major increase in health care budgets. Electronic health (eHealth) is often presented as a solution to this situation [2-4].

Although the pursuit of eHealth is global, evidence of its effect and impact remains conflicting and limited. It is even more difficult to find support for cost-effectiveness of eHealth tools, regardless of the claims policy makers are using to attract interest and funding for large-scale deployments [5,6]. Even though the policies may be driven by expectations rather than scientific evidence, it is easy to recognize the potential of eHealth to shape the services for the future [3]. Much can be done to enhance the experience of care, improve health of populations, and reduce per capita costs of services [7] by not only improving available technologies but also adopting and implementing novelties in the field. The use of eHealth services may increase access in remote areas and underserved populations [8-10]. However, not all population groups utilize the technology to the same degree; this has been referred to as a *digital divide* [11-15]. Recently, some studies have discussed how the consequences of internet use may differ between different groups, referred to as the *third-level digital divide* [12]. However, with some important exceptions, few have empirically studied determinants of this third-level digital divide [12,16-19]. In our study, we look at the positive and negative psychological consequences of health information seeking, which can be understood as an application of the third-level digital divide concept.

Understanding people's reactions to eHealth services is an imperative part of evidence-based eHealth. However, the eHealth literature seems to focus more on potentially positive aspects of eHealth from the patient's point of view, such as perceived support and increased health literacy, as major effects of eHealth tools [20-22], rather than raising critical questions. Social support is a well-established protective factor against mental health problems including anxiety, explained as a buffer against stress [23]. Many have emphasized the potentially positive effects of eHealth on shared decision making and patient empowerment [24-26]. The impact of eHealth service use on health care participation, patient involvement, and health status has been reported, whereas findings on issues such as psychological well-being, anxiety, and depression are inconsistent [20,21]. Although some of these findings were observed among patients with cancer, we may expect similar trends in other patient groups and in the general population. Overall, eHealth seems to have a positive impact on patient-related outcomes [27,28]. However, inconsistencies and variations in the literature may relate to insufficient consideration of human factors in measures and study designs, resulting in biased findings [29]. Moreover, there has been a wide range of different methodologies used, and there is still much uncertainty with regard to what constitutes eHealth, contributing to the inconsistent findings [2].

People may search for health information online for various reasons, typically to obtain more knowledge on a specific matter [24]. However, searching and reacting to online health information cognitive processes that are also driven by emotions [30-32]. Emotions are central to the human experience and infiltrate every aspect of our existence [33]. The Information Search Process Model [34] suggests that there are six steps in information seeking and understands this process as complex and multifactorial, and it also considers emotional factors. These factors vary depending on the stage of the model from optimism to confusion and from frustration and doubt to satisfaction or disappointment [30,34,35]. This study draws on this insight into the importance of emotional factors and especially the finding that the process of information searching may increase feelings such as uncertainty [30,35]. However, we focus only on the final stage of the information searching process—that is, how people feel after having found online health information. In this study, we focus particularly on the emotional aspects and how information seekers, in retrospect, appraise the outcome of their information seeking activity. Although some may feel reassured and more knowledgeable by the information they found, others might become worried and concerned, which may lead to frustration and, perhaps, symptoms of anxiety. This may relate to a range of factors such as low income, low education, and an avoidant coping style. These attributes have been associated with vulnerability to misinformation and unfounded claims [36]. Health-related misinformation is commonly encountered on the internet [37,38]. Searching for health information may also relate to a tendency for health anxiety, defined as a concern about health in the absence of a pathology or excessive concern when there is only some degree of pathology [39]. The preoccupation with thoughts about illness in health-anxious individuals is associated with a need to search for health information [40-42]. In a study by Baumgartner and Hartmann [32], the results indicated that online health information from trustworthy websites leads to increased worries among health-anxious individuals but not among non–health-anxious individuals. Although frequent online searches for health information among health-anxious individuals were associated with an increased number of doctor appointments relating to the information found online, a negative association between online searching and doctor appointments has been found among individuals with low health anxiety, that is, frequent online health information searching among individuals with low health anxiety was found to decrease their number of doctor visits [43]. Thus, the motivation for online searching for health information and the outcome of the searching activity may be associated with feeling more knowledgeable and reassured but also with negative feelings of worries, anxiety, and frustration. Although this has been studied in limited samples [27], there is a need for large-scale studies of the relative prevalence and implications of online health-searching behavior within the general population.

In a series of four papers, we explored data on the use of eHealth related to a range of other variables that were measured in a population survey in Norway (the Tromsø Study). In part 1 [44], we presented the main findings regarding the characteristics of participants and their use of eHealth. In part 2 [45], we presented and discussed how having different illnesses influences the use of eHealth. In paper 3 (this paper), we have examined some outcomes of the use of eHealth. In part 4 [46], we studied how eHealth consumption influences actual doctor visits (KY Yigsaw, PhD, unpublished data, 2018).

Aiming to increase the understanding of the psychological effects of eHealth tools, we examined the emotions of respondents who had used eHealth tools to obtain health information. Specifically, we examined the positive emotions of feeling more knowledgeable and reassured and the negative emotions of feeling anxious and confused. On the basis of existing literature, we hypothesized that the positive reactions, in general, would be more prevalent than the negative reactions.

Methods

Overview

To research the influence of eHealth use on the respondents' emotional state, data from the seventh survey of the Tromsø Study (Tromsø 7) population-based study were analyzed. The Tromsø Study is a representative survey collecting a wide range of data from the population in the municipality of Tromsø in Northern Norway [47]. In the seventh version of the study, people aged over 40 years were included, and data on the use of eHealth services were collected for the first time.

The survey was conducted in the years 2015-2016 and included 21,083 participants in the overall study. A total of 10,604 respondents reported at least one single use of the internet service (Web search engines, video search engines, health apps, and social media) for searching health information during the last year [44]. Only these internet service users were included in this study.

Dependent Variables

The dependent variables studied in this paper were participants' responses to questions regarding whether they felt either anxious, confused, more knowledgeable, or more reassured after using eHealth tools (Web search engines, video search engines, health apps, and social media). Responses were provided on a Likert scale format with the values *never*, *once*, *a few times*, and *often*. For the analysis, these were recoded into binary variables indicating *never* or *once or more* (collapsing the *once*, *few times*, and *often* categories).

Independent Variables

Demographic variables such as age, sex, self-reported health status, education, household income, and occupation were included in the analyses. Age was recoded into an ordinal variable of four age groups (40-49, 50-59, 60-69, and ≥70 years). The occupation groups *Disability benefits receivers* and *Family income supplement receivers* were joined into a single group titled *Social benefits receivers*. It resulted in a total of seven occupation groups used in the analysis (*Works full-time*, *Works part-time*, *Unemployed*, *Housekeeping*, *Retired*, *Student or military service*, and *Social benefits receivers*). Initial household income groups (<150,000 kr / <15,963 US, 150,000-250,000 kr / 15,963-26,605 US, 251,000-350,000 kr / 26,712-37,247 US, 351,000-550,000 kr / 37,354-58,532 US, 551,000-750,000 kr / 58,638-79,816 US, 751,000-1,000,000 kr / 79,923-106,422 US, and >1,000,000 kr / >106,422 US) were reorganized to obtain a better balance in the number of participants per group (0-250,000 kr / 0-26,605 US, 251,000-450,000 kr / 26,712-47,890 US, 451,000-750,000 kr / 47,996-79,816 US, 751,000-1,000,000 kr / 79,923-106,422 US, and >1,000,000 kr / >106,422 US). No preprocessing was applied to sex, self-reported health status, and education variables.

Additional variables were included, such as respondents living with a spouse, having enough friends who could provide help and support, and having enough friends to talk confidentially with. These variables were coded in a binary form (*Yes* or *No*).

Current or past medical conditions (high blood pressure, heart attack, heart failure, atrial fibrillation, angina pectoris, stroke, diabetes, kidney diseases, bronchitis, asthma, cancer, rheumatoid arthritis, arthrosis, migraine, psychological problems, and chronic pain) were also included in the analyses. The medical condition variables were converted into a single binary variable, indicating the existence of at least one condition currently or in the past. A detailed analysis of Tromsø 7 data with regard to medical conditions and eHealth use is presented in part 2 of this paper series [45].

All independent variables are summarized in Table 1.

Table 1. Characteristics of the study sample (N=9308).

Variables	Values, n (%)
Age group (years)	
40-49	3601 (38.68)
50-59	2595 (27.88)
60-69	2427 (26.07)
≥70	685 (7.36)
Gender	
Women	5213 (56.01)
Men	4095 (43.99)
Self-reported health status	
Very bad	28 (0.30)
Bad	504 (5.41)
Neither good nor bad	2124 (22.82)
Good	5164 (55.48)
Excellent	1488 (15.99)
Education	
Primary/partly secondary	1056 (11.35)
Upper secondary	2361 (25.37)
Tertiary short (<4 years of college)	2104 (22.60)
Tertiary long (≥4 years of college)	3787 (40.69)
Occupation	
Full-time worker	6495 (69.78)
Part-time worker	819 (8.80)
Unemployed	73 (0.78)
Housekeeping	33 (0.35)
Retired	1049 (11.27)
Student or military service	37 (0.40)
Social benefits receiver	802 (8.62)
Household income (kr/US)[a]	
0-250,000 kr / 0-26,605 US	204 (2.19)
251,000-450,000 kr / 26,712-47,890 US	1039 (11.17)
451,000-750,000 kr / 47,996-79,816 US	2463 (26.46)
751,000-1,000,000 kr / 79,923-106,422 US	2559 (27.49)
>1,000,000 kr / >106,422 US	3043 (32.69)
Lives with a spouse	
No	1989 (21.37)
Yes	7319 (78.63)
Has enough friends to talk confidentially with	
No	1298 (13.94)
Yes	8010 (86.14)
Has enough friends who could give help and support	
No	1028 (11.04)
Yes	8280 (88.96)

Variables	Values, n (%)
Medical condition currently or in the past	
No	2442 (26.24)
Yes	6866 (73.76)

[a] Norwegian kroner (kr) / American dollar (US).

Data Analysis

The removal of cases containing missing values in the dependent or independent variables resulted in a complete dataset of N=9308 included in the further analyses. Data were analyzed with the IBM SPSS Statistics for Macintosh, Version 25.0 (IBM Corp, Armonk, NY). Descriptive statistics were used for data exploration, whereas associations between dependent and independent variables were analyzed using a logistic regression model. Goodness of fit was assessed using a Holsmer-Lemeshow test, and only statistically significant models are reported.

The following interactions between variables were tested in all four logistic regression models: age group and self-reported health status; occupation and household income; education and household income; age group and household income; age group and any medical condition; and having enough friends who could provide help and support and having enough friends to talk confidentially with.

Analyses were run including all independent variables into the regression models. Interaction terms that were considered important on the basis of the domain knowledge or that could act as confounders were also included. The interactions were tested before the regression analyses. Only statistically significant variables are reported.

Ethics

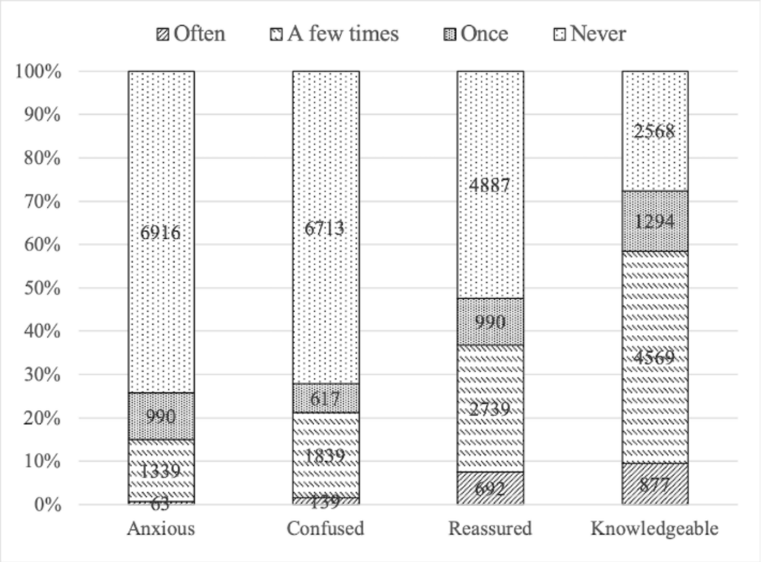
The Regional Ethical Committee for Medical and Health Research Ethics approved Tromsø 7 (REK Nord, reference 2014/940). All participants provided written consent.

Results

Descriptive Analysis

Looking at the distribution of the dependent variables (Figure 1), it is evident that the use of eHealth tools is more often associated with positive than negative feelings. Almost half of the users (4421/9308, 47.49%) reported feeling reassured and close to three-quarters (6740/9308, 72.41%) reported feeling more knowledgeable, compared with less than one-third of the respondents who reported becoming anxious (2392/9308, 25.69%) or confused (2595/9308, 27.88%) after using eHealth tools. At the same time, more than two-thirds of the participants reported no anxiety (6916/9308, 74.31%) or confusion (6713/9308, 72.12%) associated with their use of eHealth tools, whereas the ratio of respondents who had never felt reassured or more knowledgeable was 52.51% (4887/9308) and 27.59% (2468/9308), respectively.

Figure 1. Distribution of dependent variables (N=9308).



Relation Between Dependent and Independent Variables

Feeling Anxious

We conducted a logistic regression analysis with the dependent variable *feeling anxious* following the use of eHealth tools. The following predictors were included: age group, sex, self-reported health status, education, occupation, household income, lives with a spouse, has enough friends to talk confidentially with, has enough friends who could give help and support, and medical condition currently or in the past. The logistic regression model fitted well with data (Hosmer-Lemeshow goodness-of-fit chi-square=6.949, *P*=.54). The full model was statistically significant (χ^2_5=2676.6; *P*<.001). None of the interactions were significant.

Seven variables made unique independent contributions to the model. One variable had a significant positive contribution to feeling anxious after using eHealth tools. Participants having a medical condition currently or in the past had relatively higher odds (odds ratio [OR] 1.239) to feel anxiety than the participants without any disease (Table 2).

Table 2. Significant contributions to feeling anxious after using electronic health tools.

Significant variables[a]	Odds ratio (95% CI)	*P* value
Age group (years)		
40-49	1.00	Ref[b]
50-59	0.818 (0.728-0.920)	.001
60-69	0.679 (0.595-0.776)	<.001
≥70	0.699 (0.610-0.746)	.02
Sex		
Women	1.00	Ref
Men	0.674 (0.610-0.746)	<.001
Self-reported health status		
Very bad	1.00	Ref
Excellent	0.606 (0.417-0.879)	.008
Occupation		
Full-time worker	1.00	Ref
Retired	0.760 (0.586-0.985)	.04
Household income (1000 kr)		
0-250	1.00	Ref
251-450	0.719 (0.524-0.986)	.04
751-1000	0.666 (0.476-0.932)	.02
>1000	0.633 (0.447-0.895)	.01
Enough friends who could provide help and support		
No	1.00	Ref
Yes	0.662 (0.554-0.791)	<.001
Medical condition currently or in the past		
No	1.00	Ref
Yes	1.239 (1.099-1.396)	<.001

[a] Complete list of variables included in the model: age group (40-49, 50-59, 60-69, and ≥70), sex (women and men), self-reported health status (very bad, bad, neither good nor bad, good, and excellent), education (primary/partly secondary, upper secondary, tertiary short, and tertiary long), occupation (works full-time, works part-time, unemployed, housekeeping, retired, student or military service, and social benefits receiver), household income (0-250,000 kr / 0-26,605 US, 251,000-450,000 kr / 26,712-47,890 US, 451,000-750,000 kr / 47,996-79,816 US, 751,000-1,000,000 kr / 79,923-106,422 US, and >1,000,000 kr / >106,422 US), lives with a spouse (no and yes), has enough friends to talk confidentially with (no and yes), has enough friends who could give help and support (no and yes), and medical condition currently or in the past (no and yes).

[b] Ref: reference group.

In total, six independent variables contributed negatively to feeling anxious after using eHealth tools. The OR of feeling anxious decreased with age; a few cases of anxiety were observed among participants older than 60 years. Men were less likely (OR 0.674) to feel anxious than women. Participants who rated their own health as excellent had lower odds (OR 0.606) of feeling anxious than those rating their health as very bad. Retired participants were less likely (OR 0.760) to feel anxious than full-time employees. Higher income was associated with a lower chance of anxiety. The tendency to become anxious was the least prevalent in the tow highest income groups (OR 0.666 and OR 0.633). Participants who had enough friends that could provide help and support had lower odds (OR 0.662) of feeling anxious after using eHealth tools. Significant predictors of feeling anxious among eHealth users are summarized in Table 2.

To sum up, those who were more likely to become anxious after using eHealth tools tended to be younger women rating their own health lower than *excellent*, currently or previously suffering from a medical condition, and not having enough friends who could provide help and support. They were likely in a lower household income group and were not retired.

Feeling Confused

We performed a logistic regression analysis with the dependent variable *feeling confused* following the use of eHealth tools. The following predictors were included: age group, sex, self-reported health status, education, occupation, household income, lives with a spouse, has enough friends to talk confidentially with, has enough friends who could give help and support, and medical condition currently or in the past. The logistic regression model fitted well with data (Hosmer-Lemeshow goodness-of-fit chi-square=6.376, *P*=.61). The full model was statistically significant (χ^2_6=2220.5; *P*<.001). The influence of the independent variables on feeling confused after using eHealth tools is summarized in Table 3.

A total of five variables made unique independent contributions to the model. Having a current or previous medical condition had a positive contribution to feeling confused. Participants who had one or more medical conditions had higher odds (OR 1.16) of feeling confused after using eHealth tools in comparison with the healthy individuals.

In total, four variables contributed negatively to confusion after using eHealth tools. Increase in age contributed negatively to feeling confused, and the level of confusion was relatively lower in age groups 3 and 4 (≥60 years). Women were more likely to feel confused than men (OR 0.694). Participants who rated their health as *good* or *excellent* were less likely (OR 0.696 and OR 0.541, respectively) to feel confused than with the ones rating their health as *very bad*. The retired had lower odds (OR 0.769) of feeling confused than full-time workers.

In summary, the use of eHealth tools was more likely to cause confusion in younger women who currently or previously were had any medical condition. They were more likely to have relatively poor self-rated health status and were less likely to be retired.

A significant interaction was observed between the variables having enough friends who could provide help and support and having enough friends to talk confidentially with (OR 0.690, 95% CI 0.488-0.976; *P*=.04). This shows that the odds for becoming confused were lower for those reporting both enough friends to talk confidentially with and who could provide help and support (OR 0.690) compared with those who only responded yes on one of these questions.

Table 3. Significant contributions to feeling confused after using electronic health tools.

Significant variables[a]	Odds ratio (95% CI)	*P* value
Age group (years)		
40-49	1.00	Ref[b]
50-59	0.782 (0.698-0.875)	<.001
60-69	0.612 (0.537-0.697)	<.001
≥70	0.703 (0.520-0.951)	.02
Sex		
Women	1.00	Ref
Men	0.694 (0.629-0.766)	<.001
Self-reported health status		
Very bad	1.00	Ref
Good	0.696 (0.489-0.989)	.04
Excellent	0.541 (0.374-0.782)	.001
Occupation		
Full-time worker	1.00	Ref
Retired	0.769 (0.596-0.992)	.04
Medical condition currently or in the past		
No	1.00	Ref
Yes	1.16 (1.035-1.302)	.01
Significant interaction observed between variables having enough friends who could give help and support and enough friends to talk confidentially with	0.690 (0.488-0.976)	.04

[a] The complete list of variables included in the model is as follows: age group (40-49, 50-59, 60-69, and ≥70), sex (women and men), self-reported health status (very bad, bad, neither good nor bad, good, and excellent), education (primary/partly secondary, upper secondary, tertiary short, and tertiary long), occupation (works full-time, works part-time, unemployed, housekeeping, retired, student or military service, and social benefits receiver), household income (0-250,000 kr / 0-26,605 US, 251,000-450,000 kr / 26,712-47,890 US, 451,000-750,000 kr / 47,996-79,816 US, 751,000-1,000,000 kr / 79,923-106,422 US, and >1,000,000 kr / >106,422 US), lives with a spouse (no and yes), has enough friends to talk confidentially with (no and yes), has enough friends who could give help and support (no and yes), and medical condition currently or in the past (no and yes).

[b] Ref: reference group.

Feeling Knowledgeable

We conducted a logistic regression analysis with the dependent variable *feeling knowledgeable* following the use of eHealth tools. The following predictors were included: age group, sex, self-reported health status, education, occupation, household income, lives with a spouse, has enough friends to talk confidentially with, has enough friends who could give help and support, and medical condition currently or in the past. The logistic regression model fitted well with data (the Hosmer-Lemeshow goodness-of-fit chi-square=2.661, *P*=.95). The full model was statistically significant (χ^2_6=2254.7; *P*<.001). The influence of the independent variables on feeling more knowledgeable after eHealth use is summarized in Table 4.

A total of four variables made unique independent contributions to the model. The level of education contributed positively to feeling more knowledgeable. A higher level of education was associated with higher odds for feeling more knowledgeable after using eHealth tools: Participants who had upper secondary education (OR 1.601), short university education (OR 2.526), and long university education (OR 2.858) had higher odds of feeling more knowledgeable in comparison with individuals who had primary or partly secondary education (Table 4). Having a medical condition currently or in the past had a significant positive contribution to feeling more knowledgeable. Participants who had any of the earlier mentioned conditions were more likely (OR 1.119) to feel more knowledgeable after using eHealth tools. The occupation variable contributed positively to feeling knowledgeable. Part-time workers (OR 1.211), unemployed (OR 2.429), and social benefits receivers (OR 1.448) had higher odds of feeling more knowledgeable than full-time workers.

Age group 3 (participants aged 60-69 years) had significantly lower odds (OR 0.787) of feeling knowledgeable than age group 1 (individuals aged 40-49 years).

In summary, a typical eHealth user who felt more knowledgeable after using eHealth tools was a younger person (no significant gender differences) with higher education, who currently had a medical condition or had a history of a medical condition. He/she worked part-time, was unemployed, or received social benefits.

A significant interaction was observed between the variables having enough friends who could give help and support and having enough friends to talk confidentially with (OR 0.631, 95% CI 0.427-0.933; *P*=.021). This shows that the odds of having enough friends to talk confidentially with were lower (OR 0.631) in participants who had enough friends who could provide help and support in comparison with the ones who did not.

Table 4. Significant contributions to feeling more knowledgeable after using electronic health tools.

Significant variables[a]	Odds ratio (95% CI)	*P* value
Age group (years)		
40-49	1.00	Ref[b]
60-69	0.787 (0.693-0.893)	<.001
Education		
Primary	1.00	Ref
Upper secondary	1.601 (1.375-1.865)	<.001
Tertiary short[c]	2.526 (2.142-2.980)	<.001
Tertiary long[d]	2.858 (2.437-3.351)	<.001
Occupation		
Full-time worker	1.00	Ref
Part-time worker	1.211 (1.015-1.444)	.03
Unemployed	2.429 (1.263-4.672)	.008
Social benefits receiver	1.448 (1.193-1.759)	<.001
Medical condition currently or in the past		
No	1.00	Ref
Yes	1.119 (1.001-1.250)	.047
Significant interaction observed between variables has enough friends who could give help and support and has enough friends to talk confidentially with	0.631 (0.427-0.933)	.02

[a] The complete list of variables included in the model is as follows: age group (40-49, 50-59, 60-69, and ≥70), sex (women and men), self-reported health status (very bad, bad, neither good nor bad, good, and excellent), education (primary/partly secondary, upper secondary, tertiary short, and tertiary long), occupation (works full-time, works part-time, unemployed, housekeeping, retired, student or military service, and social benefits receiver), household income (0-250,000 kr / 0-26,605 US, 251,000-450,000 kr / 26,712-47,890 US, 451,000-750,000 kr / 47,996-79,816 US, 751,000-1,000,000 kr / 79,923-106,422 US, and >1,000,000 kr / >106,422 US), lives with a spouse (no and yes), has enough friends to talk confidentially with (no and yes), has enough friends who could give help and support (no and yes), and medical condition currently or in the past (no and yes).

[b] Ref: reference group.

[c] College/university, <4 years.

[d] College/university, ≥4 years.

Feeling Reassured

We performed a logistic regression analysis with the dependent variable *feeling reassured* following the use of eHealth tools. The following predictors were included: age group, sex, self-reported health status, education, occupation, household income, lives with a spouse, has enough friends to talk confidentially with, has enough friends who could give help and support, and medical condition currently or in the past. The logistic regression model fitted well with data (Hosmer-Lemeshow goodness-of-fit chi-square=4.953, *P*=.76). The full model was statistically significant (χ^2_5=191.7; *P*<.001). None of the interactions were significant. The influence of the independent variables on feeling more reassured after using eHealth tools use is summarized in Table 5.

A total of four variables made unique independent contributions to the model. Education contributed positively to feeling reassured after using eHealth tools. The largest effect size was observed in the short university/collage education group (OR 1.339). The occupation variable had a positive contribution to feeling reassured. Unemployed participants and social benefits receivers were groups that differed the most from other participants. The OR of an unemployed person feeling reassured after using eHealth tools was 1.67, whereas for receivers of disability benefits, it was 1.22 in comparison with full-time employees (Table 5).

The age group and gender variables contributed negatively to feeling reassured after using eHealth tools. Men were less likely (OR 0.657) to feel reassured after using eHealth tools in comparison with women. The biggest difference with regard to age was between groups 1 (40-49 years) and 3 (60-69 years; OR 0.779).

In short, those who were more likely to feel reassured after using eHealth tools were the young, female, highly educated, unemployed, or receiving social benefits.

Table 5. Significant contributions to feeling reassured after using electronic health tools.

Significant variables[a]	Odds ratio (95% CI)	*P* value
Age group (years)		
40-49	1.00	Ref[b]
60-69	0.779 (0.695-0.873)	<.001
Sex		
Women	1.00	Ref
Men	0.657 (0.603-0.717)	<.001
Education		
Primary	1.00	Ref
Upper secondary	1.179 (1.016-1.369)	.03
Tertiary short[c]	1.339 (1.147-1.565)	<.001
Tertiary long[d]	1.192 (1.026-1.385)	.02
Occupation		
Full-time worker	1.00	Ref
Unemployed	1.673 (1.040-2.690)	.03
Social benefits receiver	1.220 (1.029-1.446)	.02

[a] The complete list of variables included in the model is as follows: age group (40-49, 50-59, 60-69, and ≥70), sex (women and men), self-reported health status (very bad, bad, neither good nor bad, good, and excellent), education (primary/partly secondary, upper secondary, tertiary short, and tertiary long), occupation (works full-time, works part-time, unemployed, housekeeping, retired, student or military service, and social benefits receiver), household income (0-250,000 kr / 0-26,605 US, 251,000-450,000 kr / 26,712-47,890 US, 451,000-750,000 kr / 47,996-79,816 US, 751,000-1,000,000 kr / 79,923-106,422 US, and >1,000,000 kr / >106,422 US), lives with a spouse (no and yes), has enough friends to talk confidentially with (no and yes), has enough friends who could give help and support (no and yes), and medical condition currently or in the past (no and yes).

[b] Ref: reference group.

[c] College/university, <4 years.

[d] College/university, ≥4 years.

Summary of the Findings

Our findings on the psychological effects of eHealth tools are consistent with previous studies performed in several European countries [27,28] and are summarized in Table 6.

Table 6. Summary of findings (significant predictors).

Significant variables	Feeling reassured and feeling more knowledgeable	Feeling anxious and feeling confused
Gender (men and women)	Women (reassured), both genders (more knowledgeable)	Women
Age group (years; 40-49, 50-59, 60-69, and ≥70 years)	Younger	Younger
Self-rated health status (very bad, bad, neither good nor bad, good, and excellent)	Not significant	Below average (neutral, bad, and very bad)
Education (primary/partly secondary, upper secondary, tertiary short, and tertiary long)	High level of education	Not significant
Employment (works full-time, works part-time, unemployed, housekeeping, retired, student or military service, and social benefits receiver)	Part-time, unemployed, student or military service, or receives social benefits	Not retired
Medical condition currently or in the past (yes and no)	Yes	Yes
Enough friends, who could provide help and support (yes and no)	Not significant	No

Discussion

Principal Findings

Despite the global trends of accelerating technological developments and a rapid implementation of digitalized care processes in health systems, few studies have investigated how users feel and react after health information searching. The results of this study shed some light on an area of knowledge that has remained unclear owing to contradictory results of previous research [20,21]. Illness may have a huge impact on people's lives, including on well-being and other psychological outcomes. It is imperative that patients are informed in a clear and balanced way, so that they can play an active and constructive role in the management of their illness. Inaccurate, imbalanced, or misleading information can generate confusion and lead patients to wrong choices [16,17].

Kuhlthau's theory related to information searching points out that information searching is informed by emotions and cognitive capabilities [30,35,36]. In our study, we have drawn on Kuhlthau's insight and focused on the final part of the information search process and the emotional reactions to online health information. We found that while most had positive reactions following health information searching, about a quarter had negative reactions. How people react to health information will be determined by a range of factors, including cognitive and emotional factors, and people with health anxiety are in general likely to respond more negatively than others [32]. However, online health information may be designed in such a way that it may reduce some of the stress and negative emotions related to the information search process itself [31]. Considering our finding that approximately a quarter of participants experienced negative feelings after health information searching, it might, for instance, be helpful and reduce stress and negative feelings if provider/institution names and contact information were systematically provided together with the online health information.

The psychological effects of health information will vary and some information is likely to produce stronger emotional reactions than other types, such as information about life-threatening diseases [48]. Information provision is the main purpose of many eHealth services, based on the assumption that successful eHealth services may increase knowledge on health issues among their users. We examined some reactions to the use of eHealth services, and our results confirmed that the majority of users felt more knowledgeable and reassured after using eHealth services. In this section, findings from this study are discussed in the light of previous studies reporting relevant results.

The eHealth Trends Study [27], a telephone survey of 7903 respondents from Norway, Denmark, Germany, Greece, Poland, Portugal, and Latvia in 2005, found that 30% of internet users felt reassured or relieved after searching for information about health or illness online, whereas only 15% reported feeling anxious. In a subset of the general Norwegian population data [49], almost a quarter of the users (23%) reported feeling reassured by online health information, whereas 10% reported increased anxiety from the same type of information. The relative numbers of the participants who felt reassured and anxious doubled (42.6% and 17.3%, respectively), if only the eHealth user population was considered [50]. In our study, performed 10 years later, we found a slightly higher percentage of those feeling reassured (47.5%) and those feeling anxious (25.7%) among the eHealth users. Thus, the findings in our study are consistent with previous studies.

Medlock et al [28] reported similar findings in their small sample study from 2011 of Dutch seniors (aged 49-94 years; N=100; 85% older than 65 years). Feelings of anxiety (38%) and confusion (39%) were relatively higher than the figures in our study (25.7% and 27.9%, respectively), whereas more comparable figures were reported on feeling reassured in 56% (47.5% in Tromsø 7) and feeling more knowledgeable in 69% (72.4% in Tromsø 7). We may only speculate on the reasons for these differences, as the timing, the population, and the design of the studies vary. A stereotypical assumption that users of higher age, as in the Dutch study, would be more easily confused or worried by eHealth services is not confirmed by our data. On the contrary, tendencies of confusion and anxiety diminished by increasing age. However, the older age groups felt that they tended to be less reassured and less knowledgeable.

Several publications emanating from the US Health Information National Trends Survey have examined associations between online health information and participants' knowledge and health behavior [16-19]. Although these have not centered primarily on emotional reactions, they show that there are important differences in how participants respond to online health information, supporting the concept of a *third-level digital divide*. In our study, we draw on the concept of the *third-level digital divide* and demonstrate empirically that the concept is relevant also for psychological reactions and emotions.

In our study, seven variables made significant contributions in predicting individual feelings after using eHealth tools (Table 6). Female gender (except feeling more knowledgeable) and younger age were associated with both positive and negative feelings. This is supported by earlier studies, showing that women are generally more engaged in eHealth activities [27,51], thus larger effects, both positive and negative, can be expected among women. Gender differences are also well documented in mental health issues, with reports of more worries and concerns and higher rates of anxiety and depression in women, compared with men [52]. The reduced anxiety and confusion among older users may reflect general age differences in perceptions and worries of health issues. Epidemiological studies report a general tendency of reduced anxiety with increasing age [53]. In our study, increasing age contributed negatively to both positive and negative feelings. On the basis of our data, we cannot conclude, but may speculate, if this may be related to decreasing levels of computer literacy in older participant groups, which may lower the expectations for eHealth tools and, consequently, the probability of both positive and negative effects. It may also be that older participants in general have more experience and thereby have developed coping strategies that make them less prone to be influenced emotionally by online health information.

Chronic illness is common among adults in the Western countries, and 45% of US adults have at least one chronic illness [54]. Having a current or past medical condition predicted both increased positive and negative emotions following eHealth use. One way to understand this finding is that the stakes can be higher for people who have (or have had) a medical condition. They have a personal experience of being ill and might be searching for information pertaining directly to themselves, which might make the health information feel more important or relevant to them than people who are completely healthy and well. Consequently, for people with a medical condition, information that is perceived as negative might result in stronger negative emotions and information that is perceived as positive might result in stronger positive reactions. Some of those who state that they have (or have had) a medical condition suffer from psychological problems, including different forms of anxiety. Prior research has suggested that people with health anxiety react stronger and more negatively to health information than others [32,43].

Socioeconomic status (SES) is a well-known and central indicator of health and use of health services, and low scores on SES indicators such as educational level, household income, and work status have been associated with relatively poorer health and higher use of certain health services [55,56]. However, in Norway, the use of specialist services has been higher among the more educated [57]. A detailed analysis of Tromsø 7 data with regard to health care services consumption and eHealth use is presented in part 4 of this paper series.

Our results confirm that feelings and reactions after using eHealth services are associated with SES and also show some interesting patterns that can be considered to reflect the *third-level digital divide*, that is, different groups of users have different emotional reactions to the use of the services [11,12]. Although the negative reaction of feeling anxious is associated with lower income, the positive reactions of feeling reassured and more knowledgeable are both associated with higher education. In general, anxiety and depression are more prevalent in lower-income populations [58], and this was also reflected in our results. We may speculate whether this indicates that higher education may provide individuals with cognitive tools to process health information, and perhaps critically judge the credibility of the information and the source of information. As an indicator of SES, education may reflect being more resourceful. Thus, the effect of education may relate to the group of higher educated participants tending to be more resourceful and may more easily navigate in the heath care system and get access to specialized health services, compared with the group of lower educated participants. Indeed, previous research indicates that the educational level is a significant predictor of health disparities and that health literacy partly mediates this relation [59,60].

Another interesting finding is the importance of social support on feeling anxious following the use of eHealth services. Notably, there was a significant effect of the interaction between having enough friends, expected to provide help and support, and having enough friends to talk confidentially with on feeling confused after using eHealth services. Our data suggest that talking to friends who they perceive as supportive may protect them from being confused by their use of eHealth services. This supports what has been established in the research on social support, that having a social network or talking to friends is not in itself sufficient to gain an effect of social support. The effect is seen when social networks are perceived as supportive [61].

Limitations

The population of the Tromsø 7 participants originates from the city of Tromsø in the North of Norway. Although the Tromsø 7 population represents Tromsø well, it may not be completely representative to other geographical locations.

The limitations of establishing any causality between the use of eHealth resources and feeling anxious, confused, reassured, or more knowledgeable is a general weakness of cross-sectional design studies. Owing to the design of the study and the questionnaire, we cannot draw any conclusions on the effect of eHealth tools on the positive and negative feelings in users. The findings are merely associations, and we cannot tell whether the use of eHealth tools caused any changes in the positive and negative health-related feelings users may have had beforehand. Results discussed in this paper present the status of feelings in eHealth users; however, this status may not be directly attributed to the use of eHealth. Additional measures, such as health anxiety, personality profiles, or coping styles or strategies, may have contributed to the interpretation of the findings, but were not available for this study.

Although we have included a range of highly relevant variables, we lack some variables that might be relevant, such as previous experience with health services, technological literacy, and psychological factors such as personality, cognitive capacity, coping strategies, self-efficacy, and tendencies to worrying, anxiousness, and depression. We also lack a broader range of variables on social support and variables that could give even more nuanced information about the importance of partners to the reactions to the use of eHealth tools. Future studies should consider including some of these variables. Our study underlines the importance of including appraisals and emotional reactions to the information seeking process. However, as we only have retrospective data on emotional reactions to the health information–seeking process, we are unable to link our results to the information seeking process itself.

Conclusions

This study demonstrated that positive (reassured and more knowledgeable) feelings after using eHealth services are approximately 2.2 times more common than negative (anxious and confused) feelings in eHealth users. Women have stronger effects than men, and younger users are more likely to have both positive and negative effects than the older ones.

A person, who is likely to have positive outcomes of eHealth, is highly educated, but not in full-time employment. He/she is likely experiencing a medical condition currently or has experienced one in the past. Negative outcomes are more common in people with a poorer self-reported health condition; these people are also likely experiencing a medical condition currently or have experienced one in the past. They are likely not retired and do not have enough friends who could provide help and support.

Online health searching may be motivated by many different factors and needs that we have not studied in this work. Furthermore, people may experience different reactions and emotions when they have searched for health information online. In our study, drawing on prior literature, we have classified reactions into *positive* and *negative*. It is important that some participants reacted negatively, and this issue should be addressed by eHealth service providers when designing services—for instance, by including concrete information about how users can get more help and support. More studies that examine a greater range of reactions to online health information are needed, as are studies that examine more detailed factors that might predict negative reactions to online health information.

Acknowledgments

We thank the organizers, management, and the technical staff of the seventh survey of the Tromsø Study (Tromsø 7) for their valuable work in preparing and collecting the data. Above all, we thank the residents of Tromsø. Their willingness to participate is fundamental to this research. This research was made possible by a grant from the Research Council of Norway to the Norwegian Centre for eHealth Research, University Hospital of North Norway, Grant No 248150/O70, and by UiT The Arctic University of Norway. The publication charges for this paper have been funded by a grant from the Publication Fund of UiT The Arctic University of Norway.

Conflicts of Interest

None declared.

References

1. Bråthen R, Hjemås G, Holmøy E, Ottersen I. SSB: Forside. 2015 Jun 22. Staffing needs in the specialist health service towards 2040 URL: http://www.ssb.no/helse/artikler-og-publikasjoner/bemanningsbehov-i-spesialisthelsetjenesten-mot-2040 [accessed 2019-10-29]
2. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
3. Hill JW, Powell P. The national healthcare crisis: Is eHealth a key solution? Bus Horiz 2009 May;52(3):265-277. [doi: 10.1016/j.bushor.2009.01.006]
4. Scott RE, Mars M. Principles and framework for eHealth strategy development. J Med Internet Res 2013 Jul 30;15(7):e155 [FREE Full text] [doi: 10.2196/jmir.2250] [Medline: 23900066]
5. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
6. 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]
7. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769. [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]

8. Bennett GG, Glasgow RE. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
9. Woldaregay AZ, Walderhaug S, Hartvigsen G. Telemedicine services for the arctic: a systematic review. JMIR Med Inform 2017 Jul 28;5(2):e16 [FREE Full text] [doi: 10.2196/medinform.6323] [Medline: 28659257]
10. Acharibasam JW, Wynn R. Telemental health in low- and middle-income countries: a systematic review. Int J Telemed Appl 2018;2018:9602821 [FREE Full text] [doi: 10.1155/2018/9602821] [Medline: 30519259]
11. Wei K, Teo H, Chan HC, Tan BC. Conceptualizing and testing a social cognitive model of the digital divide. Inf Syst Res 2010 Mar;22(1):170-187. [doi: 10.1287/isre.1090.0273]
12. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
13. 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]
14. Nguyen GT, Bellamy SL. Cancer information seeking preferences and experiences: disparities between Asian Americans and Whites in the Health Information National Trends Survey (HINTS). J Health Commun 2006;11(Suppl 1):173-180. [doi: 10.1080/10810730600639620] [Medline: 16641082]
15. Rutten LJ, Squiers L, Hesse B. Cancer-related information seeking: hints from the 2003 Health Information National Trends Survey (HINTS). J Health Commun 2006;11(Suppl 1):147-156. [doi: 10.1080/10810730600637574] [Medline: 16641080]
16. Langford A, Resnicow K, An L. Clinical trial awareness among racial/ethnic minorities in HINTS 2007: sociodemographic, attitudinal, and knowledge correlates. J Health Commun 2010;15(Suppl 3):92-101 [FREE Full text] [doi: 10.1080/10810730.2010.525296] [Medline: 21154086]
17. Hawkins NA, Berkowitz Z, Peipins LA. What does the public know about preventing cancer? Results from the Health Information National Trends Survey (HINTS). Health Educ Behav 2010 Aug;37(4):490-503. [doi: 10.1177/1090198106296770] [Medline: 17478600]
18. Oh A, Shaikh A, Waters E, Atienza A, Moser RP, Perna F. Health disparities in awareness of physical activity and cancer prevention: findings from the National Cancer Institute's 2007 Health Information National Trends Survey (HINTS). J Health Commun 2010;15(Suppl 3):60-77 [FREE Full text] [doi: 10.1080/10810730.2010.522694] [Medline: 21154084]
19. Nelson W, Moser RP, Gaffey A, Waldron W. Adherence to cervical cancer screening guidelines for US women aged 25-64: data from the 2005 Health Information National Trends Survey (HINTS). J Womens Health (Larchmt) 2009 Dec;18(11):1759-1768 [FREE Full text] [doi: 10.1089/jwh.2009.1430] [Medline: 19951209]
20. Slev VN, Mistiaen P, Pasman HR, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
21. Rising CJ, Bol N, Kreps GL. Age-related use and perceptions of eHealth in men with prostate cancer: a web-based survey. JMIR Cancer 2015 Jul 25;1(1):e6 [FREE Full text] [doi: 10.2196/cancer.4178] [Medline: 28410165]
22. Quaglio G, Sørensen K, Rübig P, Bertinato L, Brand H, Karapiperis T, et al. Accelerating the health literacy agenda in Europe. Health Promot Int 2017 Dec 1;32(6):1074-1080. [doi: 10.1093/heapro/daw028] [Medline: 27099240]
23. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Oct;98(2):310-357. [doi: 10.1037/0033-2909.98.2.310] [Medline: 3901065]
24. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki CE, Prokosch H, Wynn R. Informed citizen and empowered citizen in health: results from an European survey. BMC Fam Pract 2011 May 16;12:20 [FREE Full text] [doi: 10.1186/1471-2296-12-20] [Medline: 21496309]
25. 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]
26. Chewning B, Bylund CL, Shah B, Arora NK, Gueguen JA, Makoul G. Patient preferences for shared decisions: a systematic review. Patient Educ Couns 2012 Jan;86(1):9-18 [FREE Full text] [doi: 10.1016/j.pec.2011.02.004] [Medline: 21474265]
27. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007 May 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
28. Medlock S, Eslami S, Askari M, Sent D, de Rooij SE, Abu-Hanna A. The consequences of seniors seeking health information using the internet and other sources. Stud Health Technol Inform 2013;192:457-460. [Medline: 23920596]
29. Wildenbos GA, Peute LW, Jaspers MW. Impact of patient-centered eHealth applications on patient outcomes: a review on the mediating influence of human factor issues. Yearb Med Inform 2016 Dec 10(1):113-119 [FREE Full text] [doi: 10.15265/IY-2016-031] [Medline: 27830238]
30. Kuhlthau CC. Seeking Meaning: A Process Approach to Library and Information Services. Second Edition. Santa Barbara, CA: Libraries Unlimited; 2004.
31. Kalbach J. 'I'm feeling lucky': The role of emotions in seeking information on the Web. J Am Soc Inf Sci Technol 2006;57(6):813-818. [doi: 10.1002/asi.20299]
32. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):613-618. [doi: 10.1089/cyber.2010.0425] [Medline: 21548797]

33. Cornelius RR. The Science of Emotion: Research and Tradition in the Psychology of Emotions. London, England: Pearson; 1996.
34. Kuhlthau CC, Heinström J, Todd RJ. InformationR. 2008. The 'information search process' revisited: is the model still useful? URL: http://informationr.net/ir/13-4/paper355.html [accessed 2019-06-03]
35. Kuhlthau CC. The role of experience in the information search process of an early career information worker: Perceptions of uncertainty, complexity, construction, and sources. J Am Soc Inf Sci 1999;50(5):399-412. [doi: 10.1002/(sici)1097-4571(1999)50:5<399::aid-asi3>3.0.co;2-l]
36. Kalichman SC, Cherry C, Cain D, Weinhardt LS, Benotsch E, Pope H, et al. Health information on the Internet and people living with HIV/AIDS: information evaluation and coping styles. Health Psychol 2006 Mar;25(2):205-210. [doi: 10.1037/0278-6133.25.2.205] [Medline: 16569112]
37. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? Br Med J 2014 Oct 14;349:g6178. [doi: 10.1136/bmj.g6178] [Medline: 25315514]
38. Gabarron E, Serrano JA, Wynn R, Lau AY. Tweet content related to sexually transmitted diseases: no joking matter. J Med Internet Res 2014 Oct 6;16(10):e228 [FREE Full text] [doi: 10.2196/jmir.3259] [Medline: 25289463]
39. Lucock M, Morley S. The Health Anxiety questionnaire. Br J Health Psychol 1996 May;1(2):137-150. [doi: 10.1111/j.2044-8287.1996.tb00498.x]
40. Muse K, McManus F, Leung C, Meghreblian B, Williams JM. Cyberchondriasis: fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the Internet. J Anxiety Disord 2012 Jan;26(1):189-196. [doi: 10.1016/j.janxdis.2011.11.005] [Medline: 22137465]
41. Marcus DK, Gurley JR, Marchi MM, Bauer C. Cognitive and perceptual variables in hypochondriasis and health anxiety: a systematic review. Clin Psychol Rev 2007 Mar;27(2):127-139. [doi: 10.1016/j.cpr.2006.09.003] [Medline: 17084495]
42. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
43. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]
44. Wynn R, Oyeyemi SO, Budrionis A, Marco-Ruiz L, Yigzaw KY, Bellika JG. Electronic Health Use in a Representative Sample of 18,497 Respondents in Norway (The Seventh Tromsø Study - Part 1): Population-Based Questionnaire Study. JMIR Med Inform 2020 Mar 05;22(3):e13106 [FREE Full text] [doi: 10.2196/13106]
45. Marco-Ruiz L, Wynn R, Oyeyemi SO, Budrionis A, Yigzaw KY, Bellika JG. Impact of Illness on Electronic Health Use (The Seventh Tromsø Study - Part 2): Population-Based Questionnaire Study. J Med Internet Res 2020 Mar 05;22(3):e13116 [FREE Full text] [doi: 10.2196/13116]
46. Yigzaw KY, Ruiz L, Budrionis A, Oyeyemi S, Fagerlund A, Wynn R, et al. The Association Between Health Information Seeking on the Internet and Doctor Visits (The Seventh Tromsø Study - Part 4): Population-Based Questionnaire Study. J Med Internet Res 2020 Mar 05;22(3):e13120 [FREE Full text] [doi: 10.2196/13120]
47. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njølstad I. Cohort profile: the Tromso study. Int J Epidemiol 2012 Aug;41(4):961-967 [FREE Full text] [doi: 10.1093/ije/dyr049] [Medline: 21422063]
48. Nolan T, Dack C, Pal K, Ross J, Stevenson FA, Peacock R, et al. Patient reactions to a web-based cardiovascular risk calculator in type 2 diabetes: a qualitative study in primary care. Br J Gen Pract 2015 Mar;65(632):e152-e160 [FREE Full text] [doi: 10.3399/bjgp15X683953] [Medline: 25733436]
49. Andreassen HK, Wangberg SC, Wynn R, Sørensen T, Hjortdahl P. [Health-related use of the Internet in the Norwegian population]. Tidsskr Nor Laegeforen 2006 Dec 16;126(22):2950-2952 [FREE Full text] [Medline: 17117194]
50. Wangberg S, Andreassen H, Kummervold P, Wynn R, Sørensen T. Use of the internet for health purposes: trends in Norway 2000-2010. Scand J Caring Sci 2009 Dec;23(4):691-696. [doi: 10.1111/j.1471-6712.2008.00662.x] [Medline: 19473317]
51. Escoffery C. Gender similarities and differences for e-health behaviors among US adults. Telemed J E Health 2018 May;24(5):335-343. [doi: 10.1089/tmj.2017.0136] [Medline: 28813630]
52. McLean CP, Asnaani A, Litz BT, Hofmann SG. Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. J Psychiatr Res 2011 Aug;45(8):1027-1035 [FREE Full text] [doi: 10.1016/j.jpsychires.2011.03.006] [Medline: 21439576]
53. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2015 Oct;17(3):327-335 [FREE Full text] [Medline: 26487813]
54. Fox S, Duggan M. Pew Research Center. The Diagnosis Difference URL: https://www.pewresearch.org/science/2013/11/26/the-diagnosis-difference/ [accessed 2019-10-29] [WebCite Cache ID 73J3bPp5s]
55. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129(Suppl 2):19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
56. Filc D, Davidovich N, Novack L, Balicer RD. Is socioeconomic status associated with utilization of health care services in a single-payer universal health care system? Int J Equity Health 2014 Dec 28;13:115 [FREE Full text] [doi: 10.1186/s12939-014-0115-1] [Medline: 25431139]

57. Vikum E, Bjørngaard JH, Westin S, Krokstad S. Socio-economic inequalities in Norwegian health care utilization over 3 decades: the HUNT Study. Eur J Public Health 2013 Dec;23(6):1003-1010. [doi: 10.1093/eurpub/ckt053] [Medline: 23729479]
58. Freeman A, Tyrovolas S, Koyanagi A, Chatterji S, Leonardi M, Ayuso-Mateos JL, et al. The role of socio-economic status in depression: results from the COURAGE (aging survey in Europe). BMC Public Health 2016 Oct 19;16(1):1098 [FREE Full text] [doi: 10.1186/s12889-016-3638-0] [Medline: 27760538]
59. van der Heide I, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health, education, and health literacy: results from the Dutch Adult Literacy and Life Skills Survey. J Health Commun 2013;18(Suppl 1):172-184 [FREE Full text] [doi: 10.1080/10810730.2013.825668] [Medline: 24093354]
60. Ginossar T. Predictors of online cancer prevention information seeking among patients and caregivers across the digital divide: a cross-sectional, correlational study. JMIR Cancer 2016 Mar 9;2(1):e2 [FREE Full text] [doi: 10.2196/cancer.5108] [Medline: 28410177]
61. Cohen S. Social relationships and health. Am Psychol 2004;59(8):676-684. [doi: 10.1037/0003-066x.59.8.676]

Abbreviations

eHealth: electronic health
OR: odds ratio
SES: socioeconomic status
Tromsø 7: seventh survey of the Tromsø Study

Edited by G Eysenbach; submitted 12.12.18; peer-reviewed by M Earl, E Da Silva, H Mehdizadeh, C Escoffery, T Ginossar; comments to author 28.02.19; revised version received 24.04.19; accepted 31.08.19; published 05.03.20.

Please cite as:
Budrionis A, Wynn R, Marco-Ruiz L, Yigzaw KY, Bergvik S, Oyeyemi SO, Bellika JG
Impact of the Use of Electronic Health Tools on the Psychological and Emotional Well-Being of Electronic Health Service Users (The Seventh Tromsø Study - Part 3): Population-Based Questionnaire Study
J Med Internet Res 2020;22(3):e13118
URL: https://www.jmir.org/2020/3/e13118
doi:10.2196/13118
PMID:32134396

Original Paper

Affect-Focused Psychodynamic Internet-Based Therapy for Adolescent Depression: Randomized Controlled Trial

Karin Lindqvist[1], MSc; Jakob Mechler[1], MSc; Per Carlbring[1], PhD; Peter Lilliengren[2], PhD; Fredrik Falkenström[3], PhD; Gerhard Andersson[3,4], PhD; Robert Johansson[1], PhD; Julian Edbrooke-Childs[5,6], PhD; Hanne-Sofie J Dahl[7,8], PhD; Katja Lindert Bergsten[9], PhD; Nick Midgley[6,10], PhD; Rolf Sandell[11], PhD; Agneta Thorén[12], PhD; Naira Topooco[3,13], PhD; Randi Ulberg[8,14], PhD; Björn Philips[1], PhD

[1]Department of Psychology, Stockholm University, Stockholm, Sweden
[2]Ersta Sköndal Bräcke University College, Stockholm, Sweden
[3]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden
[4]Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden
[5]Evidence Based Practice Unit, Anna Freud National Centre for Children and Families, London, United Kingdom
[6]Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom
[7]Vestfold Hospital Trust, Oslo, Norway
[8]Division of Mental Health and Addiction, University of Oslo, Oslo, Norway
[9]Department of Psychology, Uppsala University, Uppsala, Sweden
[10]Child Attachment and Psychological Therapies Research Unit, Anna Freud National Centre for Children and Families, London, United Kingdom
[11]Department of Psychology, Lund University, Lund, Sweden
[12]The Erica Foundation, Stockholm, Sweden
[13]Center for m2Health, Palo Alto, CA, United States
[14]Department of Psychiatry, Diakonhjemmet Hospital, Oslo, Norway

Corresponding Author:
Björn Philips, PhD
Department of Psychology
Stockholm University
SE-106 91
Stockholm
Sweden
Phone: 46 (0)8 16 20 10
Email: bjorn.philips@psychology.su.se

Abstract

Background: Adolescent depression is one of the largest health issues in the world and there is a pressing need for effective and accessible treatments.

Objective: This trial examines whether affect-focused internet-based psychodynamic therapy (IPDT) with therapist support is more effective than an internet-based supportive control condition on reducing depression in adolescents.

Methods: The trial included 76 adolescents (61/76, 80% female; mean age 16.6 years), self-referred via an open access website and fulfilling criteria for major depressive disorder. Adolescents were randomized to 8 weeks of IPDT (38/76, 50%) or supportive control (38/76, 50%). The primary outcome was self-reported depressive symptoms, measured with the Quick Inventory of Depressive Symptomatology for Adolescents (QIDS-A17-SR). Secondary outcomes were anxiety severity, emotion regulation, self-compassion, and an additional depression measure. Assessments were made at baseline, postassessment, and at 6 months follow-up, in addition to weekly assessments of the primary outcome measure as well as emotion regulation during treatment.

Results: IPDT was significantly more effective than the control condition in reducing depression (*d*=0.82, *P*=.01), the result of which was corroborated by the second depression measure (*d*=0.80, *P*<.001). IPDT was also significantly more effective in reducing anxiety (*d*=0.78, *P*<.001) and increasing emotion regulation (*d*=0.97, *P*<.001) and self-compassion (*d*=0.65, *P*=.003). Significantly more patients in the IPDT group compared to the control group met criteria for response (56% vs 21%, respectively) and remission (35% vs 8%, respectively). Results on depression and anxiety symptoms were stable at 6 months follow-up. On average, participants completed 5.8 (SD 2.4) of the 8 modules.

Conclusions: IPDT may be an effective intervention to reduce adolescent depression. Further research is needed, including comparisons with other treatments.

Trial Registration: International Standard Randomised Controlled Trial Number (ISRCTN) 16206254; http://www.isrctn.com/ISRCTN16206254

(J Med Internet Res 2020;22(3):e18047) doi:10.2196/18047

KEYWORDS

depressive disorder; adolescents; psychodynamic; internet-based treatment; treatment outcome; mobile phone

Introduction

According to the World Health Organization [1], adolescent depression is the fourth-leading cause of illness and disability among young people 15-19 years of age. Adolescent depression is associated with higher risk of adversity throughout life, including recurrent depressive episodes, other mental health issues, relational problems, and lower educational attainment [2]. Still, only a minority of young people suffering from depression seek and/or receive professional help [3,4]. Barriers to seeking help include practical issues of accessing services, perceived stigma and feelings of shame, desire to be self-reliant, and difficulties recognizing symptoms of mental health disorders [5].

Internet-based interventions may address some of the barriers to seeking and receiving treatment. For example, people living in rural areas may access care in their home rather than travelling potentially long distances to services [6]. Internet-based interventions may potentially also reach patients who otherwise would avoid seeking treatment due to social stigma [7]. Moreover, it has been suggested that internet-based treatments might engage adolescents with more severe symptoms who are reluctant to seek care [8]. Thus, making treatment more accessible for adolescents via the internet may lead to future health benefits in the population as it enables people to receive treatment at an earlier stage of their psychiatric illness.

To date, the literature on internet-based treatments for adolescent depression is lagging behind the research on adult populations [9]. Given the affinity young people have with communication via the internet, it is rather surprising that research on this group is scarce. The research that exists suggests that internet-based cognitive behavioral therapy (ICBT) is an effective treatment for adolescent and childhood psychopathology [10]. However, few studies have specifically targeted adolescent and/or childhood depression, and the ones that exist have shown a mixed pattern of findings [11-13]. Two recent randomized controlled trials (RCTs) indicated positive results for ICBT with enhanced synchronous chat support in the treatment of adolescent depression. ICBT led to a 50% or higher reduction of depressive symptoms in 42% of cases [14] and 46% achieved clinically significant improvement [15]. Although these results are highly promising, it is possible that some patients might be better helped by another type of treatment, meaning that treatment alternatives should be developed and tested.

One such treatment alternative is psychodynamic psychotherapy (PDT). Albeit not as extensively researched as cognitive behavioral therapy (CBT), several meta-analyses support the efficacy of PDT as a treatment for adults suffering from psychopathology [16,17]. The available research also supports PDT as an effective treatment for children and adolescents, but the research is much less robust due to the limited number of controlled studies [18,19]. Results from the largest RCT on psychotherapy for adolescents with depression suggested that PDT can be as clinically effective and cost-effective as CBT [20]. Recently, internet-based interventions based on psychodynamic principles, such as internet-based psychodynamic therapy (IPDT), have been developed and tested with promising results in adult populations, both as a transdiagnostic approach (eg, Johansson et al [21] and Zwerenz et al [22]) and an approach targeting specific mood and anxiety disorders [23-25]. However, to the authors' knowledge, no study has yet tested the effects of an IPDT adaptation in an adolescent population suffering from depression.

Thus, this RCT aimed to evaluate a newly developed, affect-focused IPDT program for adolescent depression. The treatment was given as a guided self-help program with therapist support and weekly chat sessions and was compared to a control condition in which participants were given online therapist support. It was hypothesized that the treatment would be significantly more effective than the control condition for depression as well as for secondary outcomes, such as anxiety, emotion regulation, and self-compassion. Furthermore, treatment effects on depression and anxiety were investigated at a 6-month follow-up after treatment completion. As this was a trial of a newly developed treatment, acceptability and attitudes toward the treatment were also investigated.

Methods

Overview

This study was carried out in accordance with the Consolidated Standards of Reporting Trials (CONSORT) statement for clinical trials [26]. The ISRCTN (International Standard Randomised Controlled Trial Number) registration ID is 16206254. The trial was approved by the Regional Ethics Board of Stockholm, Sweden (number: 2018/2268-31/5). Participants received the treatment at no cost. Written informed consent was obtained from all participants via the online treatment platform. During the diagnostic interview, participants were also given the same information and had the opportunity to ask questions.

Participants were not paid in any way for their participation or completion of measures. The project was undertaken by Stockholm University in collaboration with Linköping University.

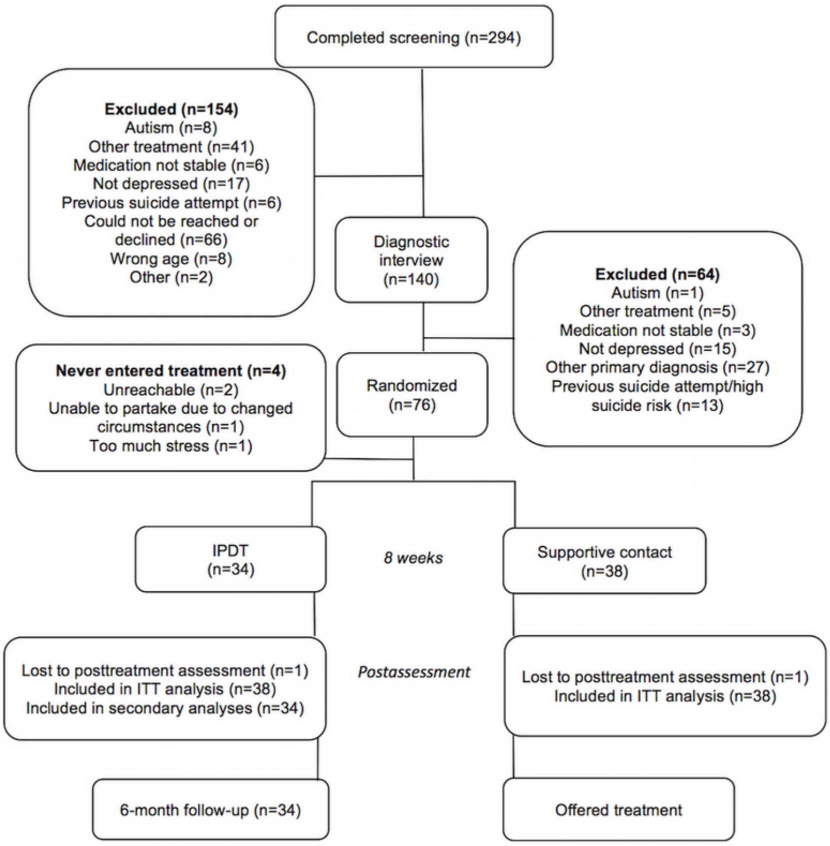
Recruitment and Participants

Adolescents were recruited via social media as well as through information via schools, youth centers, youth mental health care providers, and other similar locations during January and February 2019. Eligibility criteria were being aged 15-18 years; fulfilling a diagnosis of unipolar major depressive disorder according to DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) criteria, as established by scoring at least 10 points on the Quick Inventory of Depressive Symptomatology for Adolescents (QIDS-A17-SR) [27]; and fulfilling criteria according to the Mini International Neuropsychiatric Interview 7.0 (MINI 7.0) [28]. Furthermore, participants had to have access to a computer, smartphone, or tablet with internet connection and had to be able to read, write, and speak Swedish without the aid of an interpreter. Exclusion criteria included a substantial risk of suicidality, as indicated by clearly stated intent and plans and/or earlier suicide attempts; partaking in other concurrent psychological treatment; psychotropic medication dosage not stable for at least 3 months; other primary diagnoses in need of other treatment; and current fulfilment of any of the following diagnoses: any psychotic disorder, bipolar I/II disorder, antisocial personality disorder, autism-spectrum disorder, or any substance use disorder.

Procedure

Potential participants were directed to a study website with information about the project and online registration. Informed consent was given upon registration. After applying for the study, participants were given access to an online screening survey consisting of demographic questions as well as online versions of the screening and outcome measures. If initial inclusion criteria were met (ie, scoring ≥10 on the QIDS-A17-SR and not meeting any of the exclusion criteria), participants were contacted within the following few days for a diagnostic interview (ie, MINI 7.0) over the phone with study therapists or clinical psychologists in the research group. During this interview, each participant's identity was confirmed through full name, social security number, and address. If the interview confirmed a current major depressive episode and no exclusion criteria were fulfilled, patients were included. All cases were discussed with the principal investigator and coordinators of the study as well as with a senior psychiatrist to determine inclusion or exclusion. Included participants were randomized to either the treatment or control condition in a 1:1 ratio, the process of which was conducted by an independent researcher using an online randomization tool [29]. See the CONSORT flowchart in Figure 1 for an illustration of the procedure. All participants who were excluded were given this information personally and offered advice or a referral to appropriate care. Neither participants nor therapist, per se, could be blinded to treatment condition. Since only self-report measures were employed after baseline assessment, any further blinding was redundant.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart. IPDT: internet-based psychodynamic therapy; ITT: intention-to-treat.

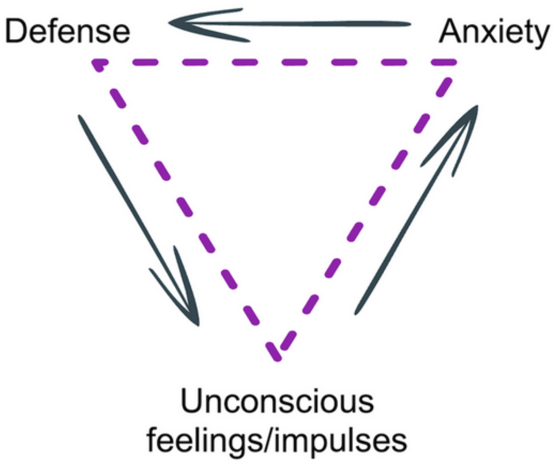


Interventions

The IPDT intervention consisted of eight therapist-supported self-help modules delivered over 8 weeks on a secure online platform [30]. Modules consisted of texts and videos followed by exercises that participants completed and sent to their therapist upon which they received feedback, typically within 24 hours on working days. In addition, the treatment included one 30-minute chat session between participants and their therapist each week. Chat sessions were unstructured, but therapists were instructed to focus on emotional conflicts underlying depressive symptomatology and relationship events affecting depression, as well as lending support and guidance throughout the treatment. Participants could also contact their therapist at any time via the integrated message system on the platform. Participants that had been inactive for a week received an encouraging message on the platform. Textbox 1 details the content of the eight modules.

The IPDT program was developed specifically for this project (by the first and second author) and was based on similar principles as a treatment program with efficacy for adults in several RCTs (eg, Johansson et al [21] and Zwerenz et al [22]). The aim of the intervention was to decrease emotional avoidance and to increase awareness, experience, and expression of emotions. Following the “triangle of conflict” [31], participants were first introduced to the idea that emotional conflicts may underlie depressive symptoms. According to this theory, feelings that are perceived as threatening to key relationships (ie, when someone important to us reacts to our emotions by expressing discomfort, withdrawal, or by expressing anger) will evoke anxiety and, hence, be suppressed and/or distorted in order to maintain the relationship [32]. The triangle of conflict depicts how these unconscious feelings and/or impulses generate anxiety, which leads to the use of defenses (ie, experiential avoidance) to ward off the anxiety-laden affects and, thus, keep them out of awareness (see Figure 2). However, rigid use of defenses will lead to negative consequences in the long run, such as depression and anxiety as well as hampering our ability to grow and mature. Throughout the treatment participants were encouraged to become aware of their own defenses, notice and regulate anxiety, and gradually approach previously warded-off feelings related to situations that could trigger depressive symptoms. The final part of the program contained material on how to express previously avoided affects in close relationships, with specific attention to psychological developmental issues relevant to adolescence (ie, Blatt [33]). Compared to existing internet-based psychodynamic treatments for adults, the modules were substantially shorter and easier to read, with vignettes that were more recognizable for the age group.

Figure 2. Malan's (1979) triangle of conflict.



Textbox 1. Treatment modules.

Module 1: Introducing theory on the interplay between basic emotions and attachment. Emphasis lies on how attachment needs are given priority above our emotions, thus leading to affect phobias (illustrated using the “triangle of conflict”).

Module 2: Superego, shame, and self-compassion. Focus on building the capacity for self-observation and differentiating between old habits of self-neglect and self-criticizing versus more healthy parts of the ego.

Module 3: Differentiation between optimal and too-high levels of anxiety. Anxiety regulation through an increased capacity for self-observation and breathing exercises.

Module 4: Affect theory and the visceral experience of affect. Exposure to warded-off feelings through an expressive writing exercise.

Module 5: Identifying and understanding defensive patterns. Identifying different defensive maneuvers and the long-term negative consequences connected to them.

Module 6: Especially problematic feelings: anger, sadness, and guilt. Mixed and complex emotions. How to notice, accept, and experience them viscerally.

Module 7: Interpersonal patterns of relatedness and self-definition. Identifying one’s predominant relationship patterns according to Sidney Blatt’s theory on anaclitic and introjective polarities of personality [33]. Participants are also taught how this is connected to avoidance of our emotions and how to break these patterns by going against them.

Module 8: Communicating and expressing affects appropriately and identifying and repairing relationship ruptures. Moving forward and maintaining gains.

The control condition consisted of supportive contact over the internet with weekly monitoring of symptoms and well-being. Each week, participants were contacted with a text message by their personally assigned therapist—both licensed clinical psychologists (JM and KL)—typically containing questions regarding their well-being during the previous week. All messages from participants were responded to, typically within 24 hours. Participants could also contact their therapist at any time via the integrated message system on the platform. Therapists provided basic support, expressed empathy and validated emotions and experiences, and encouraged participants to describe and discuss distressing life events but used no psychotherapeutic techniques or interventions. Participants who expressed suicidality, either in weekly ratings or verbally through text, were immediately contacted via phone or text message for an in-depth assessment and were offered appropriate support.

The control condition did not include any treatment material or chat sessions. After 10 weeks, participants in the control condition were offered the IPDT program without chat sessions (not presented in this study).

Therapists

A total of 11 therapists were included in the project and were either licensed psychologists (n=2) or students of clinical psychology in the last semesters of their psychologist training (n=9). All therapists had chosen to specialize in PDT for their clinical training and, thus, had taken courses in PDT theory and practice. Student therapists received a 1-day training by the treatment developers (JM and KL) and treated the majority of participants in the treatment (n=32). All therapists were supervised in group (n = 5-6) during 90-minute weekly sessions by an experienced psychotherapist (PL) specialized in experiential psychodynamic psychotherapies.

Instruments

Psychiatric diagnoses at inclusion were assessed during telephone interviews using the MINI 7.0 [28]. The MINI 7.0 interview was modified for adolescents by adding the irritability criteria to the depression module as well as the separation anxiety module from the MINI for Children and Adolescents (MINI KID). The MINI 7.0 is a well-validated diagnostic measure. It has not been validated in a Swedish adolescent population, but a study of an earlier MINI version exists [34]. Furthermore, the suicidality module was replaced by the Columbia-Suicide Severity Rating Scale (C-SSRS) [35], baseline/screening version 090114. The C-SSRS was chosen due to it being more easily administered and also recommended by the United States Food and Drug Administration [36] as the instrument of choice in clinical trials. The C-SSRS has been used in samples of Swedish young adults [37] and validated in a population consisting of Danish adolescents [38].

The diagnostic interviews were conducted by licensed psychologists (n=3) from the research team as well as psychology students (n=3) who received training in both the MINI 7.0 and the C-SSRS.

The primary outcome was self-reported depressive symptoms, measured with the QIDS-A17-SR, a reliable self-rated measure of depressive symptoms validated for both adults and adolescents [27,39]. Using all available data from all time points, an average Cronbach alpha of .76 (range .71-.85) was found, suggesting an acceptable internal consistency. Assessments were made via internet-delivered self-rating forms pretreatment, weekly during treatment, and posttreatment. Furthermore, the treatment group was assessed 6 months following treatment termination.

Secondary outcomes were measured using the Generalized Anxiety Disorder 7-item scale (GAD-7) [40], the Montgomery Åsberg Depression Rating Scale–self-rated (MADRS-S) [41,42], and the Self-Compassion Scale short-form [43], which were administered pre- and posttreatment, as well as the Difficulties in Emotion Regulation Scale [44], which was administered weekly. Cronbach alpha values suggested that internal consistency on secondary outcome measures ranged from acceptable to good (alpha=.76, .82, .71, and .89, respectively). All secondary outcome instruments have been validated in adolescent populations [42,45-47] and official translations were used.

Statistical Analysis

Sample size was determined a priori based on an expected between-group effect size of *d*=0.70 previously observed for the comparison of ICBT participants and wait-list controls for anxiety and depression in adolescents [11]. At a 5% significance level, 80% power, and *d*=0.70, a total sample size of 72 was needed.

Primary Outcome

Since the primary outcome was measured weekly within subjects, we applied multilevel modeling (MLM) [48,49] to account for individual changes over time and to test between-group differences in change rates. MLM adequately handles nested data structures and uses all available data from randomized participants, constituting a full intention-to-treat (ITT) analysis. Further, MLM provides unbiased estimates with a relatively unrestrictive assumption about missing data (ie, missing at random [50]). Model building started with estimating a basic time model that included random intercepts and fixed and random slopes for *time*. *Time* was coded as 0 for pretreatment, 1-9 for the weekly assessments during treatment, and 10 for the follow-up assessment point. To account for possible nonlinearity in the data, a quadratic term for *time* (ie, *time* × *time*) was also tested. It was found significant as a main effect as well as improving model fit for both measures (ie, reduction in Akaike's Information Criterion >2) and was, thus, retained; a random quadratic effect was discarded since the model did not converge. Lastly, *group*—coded as 0 for control and 1 for IPDT—was entered both as a main effect, to test for possible differences between IPDT and control groups at pretreatment assessment, as well as in interaction with *time*, to test for group differences in change rates over time. All models were fitted with maximum-likelihood estimation and an unstructured covariance structure for the random effects. The MLM analyses were done using SPSS, version 25.0.0.1. (IBM Corp).

Secondary Outcomes

For all secondary outcome measures, missing data at posttreatment (n=3) were addressed using multiple imputation in R, version 1.7 (The R Foundation), packages Mice [51] and Miceadds [52]. Between-group differences in the secondary measures were determined by an analysis of covariance (ANCOVA), controlling for individual differences on the respective measures at baseline.

Between-group effect size for the primary outcome was calculated using model-estimated means at posttreatment and the observed pretreatment sample SD, as recommended by Feingold [53]. For secondary measures, between-group effect sizes were transformed from eta2 from the ANCOVA, according to the formula described by Cohen [54].

Response and Remission Rates

Response to treatment was calculated using the Reliable Change Index (RCI) [55] and was defined as fulfilling the RCI criteria while scoring 2 SD below the pretreatment mean. Baseline Cronbach alpha from this study was used. Remission was defined as a QIDS-A17-SR score of 6 or below [39].

Results

Participants

A total of 76 participants were included in the study (IPDT 38/76, 50%; control 38/76, 50%). Out of 76 participants, 61 (80%) were female and the sample had a mean age of 16.6 years (SD 1.1). A total of 4 participants randomized to IPDT never entered treatment, meaning that they never participated in any of the exercises or chat sessions and had no contact with the therapist, except in 2 cases to say that they wanted to drop out. A total of 3 of these participants never opened the initial study message, meaning that they dropped out without knowing their allocation (ie, treatment or control). These 4 nonstarters were still included in the ITT analysis for the primary outcome measure, but not for the secondary outcome measures.

Patients' characteristics for the full ITT sample are presented in Table 1. All reported psychotropic medications were selective serotonin reuptake inhibitors (SSRIs) with dosages stable for at least 3 months. A total of 2 of the 4 patients on SSRIs at baseline also had prescriptions for anxiety medication—1 prometazin and 1 hydrocizine—and a sleeping aid (ie, melatonin) when needed. A total of 30 participants out of 76 (39%) reported living in a large city, 28 (37%) reported living in a smaller town, and 18 (24%) reported living in a rural area. A total of 73 participants out of 76 (96%) reported going to school. Out of 76 participants, 47 (62%) reported living with both parents; 23 (30%) reported living with either one parent, moving between two parents, or living with a remarried parent; and 6 (8%) reported living alone or with a friend, sibling, or partner.

Table 1. Demographic data at baseline[a].

Characteristic	IPDT[b] group (n=38)	Control group (n=38)
Gender, n (%)		
Female	31 (82)	30 (79)
Uncertain or other	4 (11)	0 (0)
Age (years), mean (SD)	16.6 (1.11)	16.5 (1.13)
Disorder, n (%)		
Major depressive disorder[c]	38 (100)	38 (100)
Any anxiety disorder[c]	22 (58)	23 (62)
Posttraumatic stress disorder[c]	4 (11)	1 (3)
Eating disorder[c,d]	2 (5)	1 (3)
Self-harm behavior, n (%)		
Ever	20 (53)	10 (26)
Current	9 (24)	4 (11)
Currently on psychotropic medication, n (%)	2 (5)	2 (5)

[a] There were no significant between-group differences on any of the data at baseline.

[b] IPDT: internet-based psychodynamic therapy.

[c] Confirmed by the Mini International Neuropsychiatric Interview 7.0 (MINI 7.0).

[d] Bulimia nervosa and binge-eating disorder.

Primary Outcome

Fixed-effect estimates from our MLM analysis for QIDS-A17-SR are displayed in Table 2. The model intercept (15.62) represents the average baseline score in the control group while the IPDT estimate (–0.32) represents the group difference at baseline between IPDT and control groups, which was nonsignificant (*P*=.67). The slope estimate (–0.74) represents the initial symptom reduction rate in the control group and the *time* × *time* (estimate=0.05) means that for each session, symptoms were reduced at a slightly slower rate. The IPDT versus control estimate (–0.29) represents the additional weekly decrease in symptoms for patients in the IPDT group. This term proved significant (*P*=.01), indicating that patients in the IPDT group had a significantly steeper decline in symptoms compared to those in the control group. The between-group effect size at the posttreatment assessment point was large (*d*=0.82, 95% CI 0.35-1.29) and in favor of IPDT. The observed and estimated change on the QIDS-A17-SR is illustrated in Figure 3.

Secondary Outcomes

The ANCOVAs revealed that improvements on all secondary measures were significantly greater in the treatment group. Detailed results and effect sizes are presented in Table 3.

Table 2. Multilevel models estimating changes over time in the primary outcome measure, QIDS-A17-SR[a].

Model estimates	Estimate	95% CI	*P* value
Baseline score			
Intercept	15.62	14.57 to 16.68	<.001
IPDT[b]	–0.32	–1.76 to 1.13	.67
Rate of change			
Slope	–0.73	–1.01 to –0.46	<.001
Time × time	0.05	0.02 to 0.07	<.001
IPDT versus control	–0.29	–0.51 to –0.07	.01
Variance components			
Residual variance	6.10	5.40 to 6.89	<.001
Intercept	7.61	5.04 to 11.50	<.001
Slope	0.13	0.08 to 0.23	<.001
Correlation	–0.05	–0.29 to 0.38	.78
Between-group effect size (Cohen *d*)	0.82	0.35 to 1.29	

[a] QIDS-A17-SR: Quick Inventory of Depressive Symptomatology for Adolescents.

[b] IPDT: internet-based psychodynamic therapy.

Figure 3. Weekly mean scores on the QIDS-A17-SR (Quick Inventory of Depressive Symptomatology for Adolescents). IPDT: internet-based psychodynamic therapy.

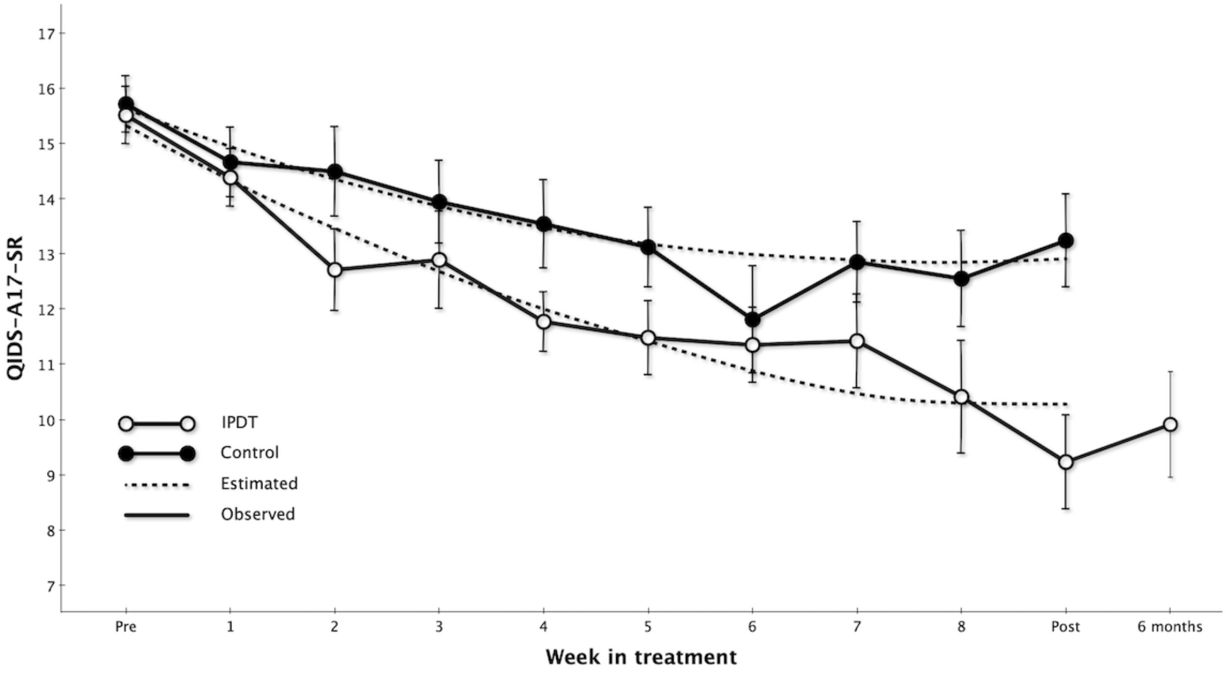


Table 3. Secondary outcome results.

Measure	Pretreatment, mean (SD)	Posttreatment, mean (SD)	Follow-up, mean (SD)	Posttreatment, between-group Cohen *d* (95% CI)	*P* value
MADRS-S[a]					
IPDT[b]	24.97 (7.06)	18.97 (7.53)	N/A[c]	0.80 (0.32-1.28)	<.001
Control[d]	26.21 (6.93)	25.84 (8.51)			
GAD-7[e]					
IPDT	12.35 (4.11)	8.18 (4.62)	8.41 (5.54)	0.78 (0.30-1.26)	<.001
Control	10.45 (3.88)	10.42 (4.65)			
DERS[f]					
IPDT	55.44 (12.12)	41.53 (14.47)	N/A	0.97 (0.48-1.46)	<.001
Control	56.42 (10.64)	53.0 (14.41)			
SCS-SF[g]					
IPDT	25.91 (5.68)	31.06 (7.31)	N/A	0.65 (0.18-1.12)	.003
Control	26.87 (5.65)	27.08 (7.69)			

[a] MADRS-S: Montgomery Åsberg Depression Rating Scale–self-rated.

[b] IPDT: internet-based psychodynamic therapy, n=34.

[c] Not applicable.

[d] Control, n=38.

[e] GAD-7: Generalized Anxiety Disorder 7-item scale.

[f] DERS: Difficulties in Emotion Regulation Scale.

[g] SCS-SF: Self-Compassion Scale short-form.

Response and Remission Rates

In all analyses of response and remission rates, missing cases (n=2, 1 in each group) were categorized as not improved. Using fulfillment of RCI and scoring 2 SDs below the pretreatment mean as criteria for response, 56% (19/34) of the treatment group compared to 21% (8/38) of the control group were categorized as responders. There was a significant difference between the groups in favor of the treatment group (χ^2_4=10.9, *P*=.03). Using a QIDS-A17-SR score of 6 or under as a cutoff for remission, 35% (12/34) of the treatment group compared to 8% (3/38) of the control group were categorized as remitted. A higher proportion of participants in the IPDT group fulfilled the criteria for remission (χ^2_1=8.2, *P*=.004).

Follow-Up Assessment

Follow-up assessments at 6 months posttreatment were conducted using the QIDS-A17-SR and GAD-7. Data were obtained from all participants who entered the treatment (n=34). Pairwise *t* tests between post- and follow-up assessments indicated no significant differences for either measure (QIDS-A17-SR mean difference −0.61, t_{32}=−0.66, *P*=.51; GAD-7 mean difference -0.09, t_{32}=−0.12, *P*=.91), suggesting results for depression and anxiety were maintained during the follow-up period.

Negative Effects and Adverse Events

Posttreatment QIDS-A17-SR scores and an open-ended question concerning negative effects [56] were used to assess potential negative effects during the trial. No participant in the treatment condition deteriorated reliably on the QIDS-A17-SR, while this was true for 3 participants in the control condition.

When responding to the open-ended question regarding negative effects, the vast majority did not report any (28/34, 82%). Out of 34 participants, 1 (3%) participant described that online text-based interactions were not optimal due to feelings of loneliness, and another participant (3%) described increased awareness of feelings of anger and that this was painful and distressing in the short term; however, the same participant then described the mastery of these angry feelings gained through the treatment as positive in the long term. Out of 34 participants, 2 (6%) described feelings of distress in connection with facing previously avoided thoughts and feelings, and 2 (6%) participants found the treatment format stressful; 1 (3%) of these participants also described feelings of shame in connection with not completing exercises on time. Thus, 6 patients in total out of 34 (18%) reported negative effects of the treatment. No serious adverse events were reported during the trial.

Program Use and Treatment Acceptability

Of the 34 participants that entered treatment, 4 (12%) dropped out of treatment, meaning that they stopped opening modules, attending chat sessions, or responding to messages before week 7. The completion of modules was defined as completing at least one exercise in the module. The mean number of completed modules was 5.8 (SD 2.4) of the 8 that were available. Excluding the 4 dropouts, the mean number was 6.2 (SD 1.9) and the median was 7. The mean number of chat sessions attended was

6.6 (SD 2.1) of the 8 available. Excluding the 4 dropouts, the mean number of chat sessions attended was 7.1 (SD 1.4). Of the 34 participants, 1 (3%) who did not drop out completed zero modules but continued attending chat sessions throughout the treatment period. Spearman correlations showed no dose-response relationship.

Discussion

Principal Findings

This RCT aimed to evaluate an affect-focused psychodynamic internet-based treatment for depression in a sample of adolescents (15-18 years of age). The results indicated that the IPDT treatment was effective in reducing depression and anxiety, as well as in enhancing emotion regulation and self-compassion, compared to a supportive contact control condition. Results indicated that the treatment facilitated clinically significant changes [55] as 56% of the treatment group recovered according to RCI and 35% scored below clinical cutoff on the primary outcome measure after completing the treatment. Treatment gains for depression and anxiety symptoms were maintained at the 6-month follow-up. In addition, the IPDT treatment seemed acceptable and tolerable overall, although some distress was reported related to taking part in the treatment. None of the participants deteriorated reliably during the intervention. The vast majority of patients (28/34, 82%) who entered treatment did not report any negative effects, and those that were reported were relatively mild. The occurrence and nature of negative experiences were similar to what has previously been reported for ICBT in treatment with adults and adolescents [14,56]. The absence of deterioration and negative effects is especially important considering that deterioration seems common among young people in psychotherapy [57].

Between-group effect sizes were similar to those found in previous trials on IPDT targeting adult psychopathology [21] as well as ICBT targeting adolescent depression [11,14,15]. The significant result for anxiety has not been found in ICBT treatments in previous trials for the same target group [14,15] and this is a promising finding, albeit in need of replication. Although the target groups appear similar, there might be population differences affecting these results. Previous trials have used another anxiety measure, rendering comparisons of initial anxiety levels difficult. However, the results are in line with those shown in a meta-analysis by Driessen et al [16] who found that short-term PDT targeting depression was significantly more efficacious than other treatments on anxiety measures. One reason for this might be the more transdiagnostic nature of PDT, which, although focused on depression symptoms to a certain extent, might target a wider scope of symptoms. The model for depression in affect-focused psychotherapy is based on the triangle of conflict, which is the same across different psychiatric disorders [31].

Although some of the more classical ingredients of psychodynamic therapy are lost in the translation to a guided self-help program, being significantly more directive and not including transference work to the same extent, the treatment is clearly based on a psychodynamic model. This is an important point since this distinguishes the treatment from ones that have been previously developed and assessed for adolescent depression. The model for understanding depression, which was also presented to the participants, is the triangle of conflict [31]. Participants are taught about attachment theory and are encouraged to reflect on their own patterns of relating based on this and whether there are any patterns they would like to change. Furthermore, participants are encouraged to reflect on and experience affects that have been perceived as threatening to key attachment relationships. The treatment relies heavily on working with defenses against these anxiety-laden affects, leading to exposure to painful emotions [32]. Furthermore, the therapists were instructed not to give advice, but rather to invite exploration and to interpret problems using the triangle of conflict. In ICBT for adolescents, chat sessions are more structured, while the chat sessions in this RCT's treatment were unstructured and the participants chose the subject; however, therapists were instructed to be symptom-focused, as the primary aim of the treatment was to uncover inner emotional conflicts creating and perpetuating symptoms of depression. While including exercises in psychodynamic treatments is not common practice, this is included in some psychodynamic modalities when deemed appropriate [58]. A future study will aim to examine the nature of therapist interventions in chat sessions during the treatment in detail. Future studies should also address whether these different treatments may be suitable for different target groups of depressed adolescents.

This study did not comprise an evaluation of adding synchronous chat sessions to the standard guided self-help format, so it is not possible to separate out the impact of each element to the overall effectiveness of the intervention. However, some aspect of the positive results may be related to the text-chat component, as it might have increased motivation and hence continued participation in the treatment. In order to minimize the workload, therapists were permitted to have two parallel chat sessions at the same time. The added value of synchronous chat sessions is an area for further research in the future.

While the chat sessions make the treatment format more time-consuming compared to internet-based interventions targeting adult depression, it is comparable to some other internet-based treatment protocols targeting adolescents [14,15,59] that enhance treatment protocols with blended real-time elements (ie, chat for 30-45 minutes or 20 minutes of weekly phone calls for 14 weeks, respectively). Also, 8 weeks of treatment is substantially shorter than other existing psychodynamic face-to-face protocols in the treatment of adolescent depression [20]. The rates of participation and completion of modules were lower than studies on adults typically show [60]. However, studies of face-to-face treatments for adolescents generally report high levels of attrition (eg, Abbass et al [19]).

The sample of the study is predominantly female. This is in line with other psychotherapy trials on face-to-face treatment (eg, Abbass et al [19]) as well as internet interventions [61]. There is also some evidence that the female predominance is even larger for internet-based interventions than in those conducted face to face [61]. One suggested reason for that is that males tend to need more influence from family and/or peers in order to seek support, and that seeking internet-based interventions relies more on individual motivation and choice. The dominance of females in the study sample could be seen as a problem, although common in the field of psychotherapy research, with generalizability of the results. It is clear that an important challenge for future treatments as well as trials is to find ways to reach and motivate males to apply as well.

Strengths and Limitations

Apart from the randomized design, a considerable strength is that a "pure" wait list was not used [62]. Instead, IPDT was compared to weekly support. In addition, a structured diagnostic interview was carried out before inclusion to confirm the depression diagnosis. The study approach was to use weekly measures of the primary outcome to detect group differences, which allowed for more sophisticated analyses. Furthermore, the trial had a low rate of attrition both at the postassessment and follow-up measurements, increasing the reliability of the results. The sample in this trial is representative of the depressed adolescent population and is a clearly psychiatric one. Participants had relatively high comorbidity and a substantial amount had current and/or past self-harm behaviors. The exclusion of severely suicidal teenagers is also in line with other research in the field (eg, Sander et al [63]).

Limitations of the study include the fact that no diagnostic interviews were conducted after the treatment and that 6-month follow-up data were only collected for two of the outcome measures. The reason for this was to avoid burdening the young people with too many measures and interviews, as it might have heightened the risk of missing data. Therefore, the most relevant instruments were prioritized for both postassessment and follow-up measurements. Also, although all interviewers were trained in the application of the MINI 7.0 and the C-SSRS, no measures of interrater reliability were conducted as interviews were not recorded. All measures in the study have been validated in adolescent populations and official translations were used. However, except for the MADRS-S [42], they have not been validated in Swedish adolescent populations. The lack of validation studies for teenage populations is a common problem in the field and one future aim is to publish psychometric data for the Swedish translations of the questionnaires. As shown, they have acceptable-to-good internal consistency in this study's sample.

Conclusions

This RCT's results provide preliminary support for affect-focused IPDT for adolescent depression. This study furthers the evidence that psychodynamic models can be translated into internet-based treatments and strengthens the evidence for the effectiveness of psychodynamic treatments in general.

Future research could aim to explore whether this treatment, in an adapted format, would be suitable for a transdiagnostic sample or to target adolescent anxiety. This study is preceding a noninferiority trial comparing IPDT to the already-proven effective ICBT program for depression in adolescents [14,15]. Furthermore, future research should address the efficacy of the treatment in other settings (ie, other countries).

Acknowledgments

This research is supported by the Kavli Trust. We thank the participating adolescents, and the study therapists: Anna Carrascosa Molin, Henrik Hallberg, Isak Jansmyr, Pontus Kinnmark, Therese Lindgren, Charlotta Olerud, Per-Åke Olsson, Katarina Streberg Carstorp, and Fredrik Strid. We also thank George Vlaescu for platform development and support and Anna Steen, MD, for psychiatric consultations.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1184 KB - jmir_v22i3e18047_app1.pdf]

References

1. World Health Organization. 2019 Oct 23. Adolescent mental health URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2019-11-30]
2. Alaie I, Philipson A, Ssegonja R, Hagberg L, Feldman I, Sampaio F, et al. Uppsala Longitudinal Adolescent Depression Study (ULADS). BMJ Open 2019 Mar 01;9(3):e024939 [FREE Full text] [doi: 10.1136/bmjopen-2018-024939] [Medline: 30826765]
3. Essau CA. Frequency and patterns of mental health services utilization among adolescents with anxiety and depressive disorders. Depress Anxiety 2005;22(3):130-137. [doi: 10.1002/da.20115] [Medline: 16175563]
4. Rocha TB, Graeff-Martins AS, Kieling C, Rohde LA. Provision of mental healthcare for children and adolescents: A worldwide view. Curr Opin Psychiatry 2015 Jul;28(4):330-335. [doi: 10.1097/YCO.0000000000000169] [Medline: 26001925]

5. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
6. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: From innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
7. Foroushani PS, Schneider J, Assareh N. Meta-review of the effectiveness of computerised CBT in treating depression. BMC Psychiatry 2011 Aug 12;11:131 [FREE Full text] [doi: 10.1186/1471-244X-11-131] [Medline: 21838902]
8. Salzer S, Stefini A, Kronmüller KT, Leibing E, Leichsenring F, Henningsen P, et al. Cognitive-behavioral and psychodynamic therapy in adolescents with social anxiety disorder: A multicenter randomized controlled trial. Psychother Psychosom 2018;87(4):223-233. [doi: 10.1159/000488990] [Medline: 29895001]
9. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: A narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
10. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: A systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
11. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: A meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
12. Grist R, Croker A, Denne M, Stallard P. Technology delivered interventions for depression and anxiety in children and adolescents: A systematic review and meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
13. Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: A systematic review and meta-analysis. BMC Health Serv Res 2014 Jul 18;14:313 [FREE Full text] [doi: 10.1186/1472-6963-14-313] [Medline: 25037951]
14. Topooco N, Berg M, Johansson S, Liljethörn L, Radvogin E, Vlaescu G, et al. Chat- and internet-based cognitive-behavioural therapy in treatment of adolescent depression: Randomised controlled trial. BJPsych Open 2018 Jul;4(4):199-207 [FREE Full text] [doi: 10.1192/bjo.2018.18] [Medline: 29988969]
15. Topooco N, Byléhn S, Dahlström Nysäter E, Holmlund J, Lindegaard J, Johansson S, et al. Evaluating the efficacy of internet-delivered cognitive behavioral therapy blended with synchronous chat sessions to treat adolescent depression: Randomized controlled trial. J Med Internet Res 2019 Nov 01;21(11):e13393 [FREE Full text] [doi: 10.2196/13393] [Medline: 31682572]
16. Driessen E, Hegelmaier LM, Abbass AA, Barber JP, Dekker JJ, Van HL, et al. The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis update. Clin Psychol Rev 2015 Dec;42:1-15. [doi: 10.1016/j.cpr.2015.07.004] [Medline: 26281018]
17. Keefe JR, McCarthy KS, Dinger U, Zilcha-Mano S, Barber JP. A meta-analytic review of psychodynamic therapies for anxiety disorders. Clin Psychol Rev 2014 Jun;34(4):309-323. [doi: 10.1016/j.cpr.2014.03.004] [Medline: 24786897]
18. Midgley N, O'Keeffe S, French L, Kennedy E. Psychodynamic psychotherapy for children and adolescents: An updated narrative review of the evidence base. J Child Psychother 2017 Jun 26;43(3):307-329. [doi: 10.1080/0075417x.2017.1323945]
19. Abbass AA, Rabung S, Leichsenring F, Refseth JS, Midgley N. Psychodynamic psychotherapy for children and adolescents: A meta-analysis of short-term psychodynamic models. J Am Acad Child Adolesc Psychiatry 2013 Aug;52(8):863-875. [doi: 10.1016/j.jaac.2013.05.014] [Medline: 23880496]
20. Goodyer IM, Reynolds S, Barrett B, Byford S, Dubicka B, Hill J, et al. Cognitive behavioural therapy and short-term psychoanalytical psychotherapy versus a brief psychosocial intervention in adolescents with unipolar major depressive disorder (IMPACT): A multicentre, pragmatic, observer-blind, randomised controlled superiority trial. Lancet Psychiatry 2017 Feb;4(2):109-119 [FREE Full text] [doi: 10.1016/S2215-0366(16)30378-9] [Medline: 27914903]
21. Johansson R, Björklund M, Hornborg C, Karlsson S, Hesser H, Ljótsson B, et al. Affect-focused psychodynamic psychotherapy for depression and anxiety through the internet: A randomized controlled trial. PeerJ 2013;1:e102 [FREE Full text] [doi: 10.7717/peerj.102] [Medline: 23862104]
22. Zwerenz R, Becker J, Johansson R, Frederick RJ, Andersson G, Beutel ME. Transdiagnostic, psychodynamic Web-based self-help intervention following inpatient psychotherapy: Results of a feasibility study and randomized controlled trial. JMIR Ment Health 2017 Oct 16;4(4):e41 [FREE Full text] [doi: 10.2196/mental.7889] [Medline: 29038094]
23. Johansson R, Hesslow T, Ljótsson B, Jansson A, Jonsson L, Färdig S, et al. Internet-based affect-focused psychodynamic therapy for social anxiety disorder: A randomized controlled trial with 2-year follow-up. Psychotherapy (Chic) 2017 Dec;54(4):351-360. [doi: 10.1037/pst0000147] [Medline: 29251954]
24. Johansson R, Ekbladh S, Hebert A, Lindström M, Möller S, Petitt E, et al. Psychodynamic guided self-help for adult depression through the internet: A randomised controlled trial. PLoS One 2012;7(5):e38021 [FREE Full text] [doi: 10.1371/journal.pone.0038021] [Medline: 22741027]

25. Andersson G, Paxling B, Roch-Norlund P, Östman G, Norgren A, Almlöv J, et al. Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: A randomized controlled trial. Psychother Psychosom 2012;81(6):344-355 [FREE Full text] [doi: 10.1159/000339371] [Medline: 22964540]
26. Schulz K, Altman D, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
27. Bernstein IH, Rush AJ, Trivedi MH, Hughes CW, Macleod L, Witte BP, et al. Psychometric properties of the Quick Inventory of Depressive Symptomatology in adolescents. Int J Methods Psychiatr Res 2010 Dec;19(4):185-194 [FREE Full text] [doi: 10.1002/mpr.321] [Medline: 20683845]
28. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34-57 [FREE Full text] [Medline: 9881538]
29. RANDOM.ORG. URL: https://www.random.org/ [accessed 2020-03-13]
30. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Interv 2016 Nov;6:107-114 [FREE Full text] [doi: 10.1016/j.invent.2016.09.006] [Medline: 30135819]
31. Malan DH. Individual Psychotherapy and the Science of Psychodynamics. London, UK: Butterworths; 1979.
32. Fosha D. The dyadic regulation of affect. J Clin Psychol 2001 Feb;57(2):227-242. [doi: 10.1002/1097-4679(200102)57:2<227::aid-jclp8>3.0.co;2-1]
33. Blatt SJ. Polarities of Experience: Relatedness and Self-Definition in Personality Development, Psychopathology, and the Therapeutic Process. Washington, DC: American Psychological Association; 2008.
34. Högberg C, Billstedt E, Björck C, Björck PO, Ehlers S, Gustle L, et al. Diagnostic validity of the MINI-KID disorder classifications in specialized child and adolescent psychiatric outpatient clinics in Sweden. BMC Psychiatry 2019 May 09;19(1):142 [FREE Full text] [doi: 10.1186/s12888-019-2121-8] [Medline: 31072319]
35. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
36. Guidance for Industry. Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials.: US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER); 2012 Aug. URL: https://www.fda.gov/media/79482/download
37. Beckman K, Lindh AU, Waern M, Stromsten L, Renberg ES, Runeson B, et al. Impulsive suicide attempts among young people: A prospective multicentre cohort study in Sweden. J Affect Disord 2019 Jan 15;243:421-426. [doi: 10.1016/j.jad.2018.09.070] [Medline: 30268958]
38. Conway PM, Erlangsen A, Teasdale TW, Jakobsen IS, Larsen KJ. Predictive validity of the Columbia-Suicide Severity Rating Scale for short-term suicidal behavior: A Danish study of adolescents at a high risk of suicide. Arch Suicide Res 2017 Jul 03;21(3):455-469. [doi: 10.1080/13811118.2016.1222318] [Medline: 27602917]
39. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 01;54(5):573-583. [doi: 10.1016/s0006-3223(02)01866-8] [Medline: 12946886]
40. Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
41. Svanborg P, Asberg M. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr Scand 1994 Jan;89(1):21-28. [doi: 10.1111/j.1600-0447.1994.tb01480.x] [Medline: 8140903]
42. Ntini I, Vadlin S, Olofsdotter S, Ramklint M, Nilsson KW, Engström I, et al. The Montgomery and Åsberg Depression Rating Scale: Self-assessment for use in adolescents: An evaluation of psychometric and diagnostic accuracy. Nord J Psychiatry 2020 Mar 03:1-8. [doi: 10.1080/08039488.2020.1733077] [Medline: 32125211]
43. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the Self-Compassion Scale. Clin Psychol Psychother 2011;18(3):250-255. [doi: 10.1002/cpp.702] [Medline: 21584907]
44. Bjureberg J, Ljótsson B, Tull MT, Hedman E, Sahlin H, Lundh L, et al. Development and validation of a brief version of the Difficulties in Emotion Regulation Scale: The DERS-16. J Psychopathol Behav Assess 2016 Jun;38(2):284-296 [FREE Full text] [doi: 10.1007/s10862-015-9514-x] [Medline: 27239096]
45. Tiirikainen K, Haravuori H, Ranta K, Kaltiala-Heino R, Marttunen M. Psychometric properties of the 7-item Generalized Anxiety Disorder Scale (GAD-7) in a large representative sample of Finnish adolescents. Psychiatry Res 2019 Feb;272:30-35. [doi: 10.1016/j.psychres.2018.12.004] [Medline: 30579178]
46. Cunha M, Xavier A, Castilho P. Understanding self-compassion in adolescents: Validation study of the Self-Compassion Scale. Pers Individ Dif 2016 Apr;93:56-62. [doi: 10.1016/j.paid.2015.09.023]

47. Neumann A, van Lier PA, Gratz KL, Koot HM. Multidimensional assessment of emotion regulation difficulties in adolescents using the Difficulties in Emotion Regulation Scale. Assessment 2010 Mar;17(1):138-149. [doi: 10.1177/1073191109349579] [Medline: 19915198]
48. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York, NY: Oxford University Press; 2003.
49. Gueorguieva R, Krystal JH. Move over ANOVA: Progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
50. Enders CK. Analyzing longitudinal data with missing values. Rehabil Psychol 2011 Nov;56(4):267-288. [doi: 10.1037/a0025579] [Medline: 21967118]
51. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1-67 [FREE Full text] [doi: 10.18637/jss.v045.i03]
52. Robitzsch A, Grund S. miceadds: Some Additional Multiple Imputation Functions, Especially for 'mice'. 2020 Feb 17. URL: https://cran.r-project.org/web/packages/miceadds/miceadds.pdf [accessed 2019-11-29]
53. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009 Mar;14(1):43-53 [FREE Full text] [doi: 10.1037/a0014699] [Medline: 19271847]
54. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
55. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
56. Rozental A, Castonguay L, Dimidjian S, Lambert M, Shafran R, Andersson G, et al. Negative effects in psychotherapy: Commentary and recommendations for future research and clinical practice. BJPsych Open 2018 Jul;4(4):307-312 [FREE Full text] [doi: 10.1192/bjo.2018.42] [Medline: 30083384]
57. Warren JS, Nelson PL, Mondragon SA, Baldwin SA, Burlingame GM. Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. J Consult Clin Psychol 2010 Apr;78(2):144-155. [doi: 10.1037/a0018544] [Medline: 20350026]
58. Levenson H. Brief Dynamic Therapy. 2nd edition. Washington, DC: American Psychological Association; 2017.
59. Stjerneklar S, Hougaard E, McLellan LF, Thastum M. A randomized controlled trial examining the efficacy of an internet-based cognitive behavioral therapy program for adolescents with anxiety disorders. PLoS One 2019;14(9):e0222485 [FREE Full text] [doi: 10.1371/journal.pone.0222485] [Medline: 31532802]
60. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided Web-based interventions for depression: An 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
61. Rickwood D, Webb M, Kennedy V, Telford N. Who are the young people choosing Web-based mental health support? Findings from the implementation of Australia's national Web-based youth mental health service, eheadspace. JMIR Ment Health 2016 Aug 25;3(3):e40 [FREE Full text] [doi: 10.2196/mental.5988] [Medline: 27562729]
62. Cuijpers P, Cristea IA. How to prove that your therapy is effective, even when it is not: A guideline. Epidemiol Psychiatr Sci 2016 Oct;25(5):428-435. [doi: 10.1017/S2045796015000864] [Medline: 26411384]
63. Sander L, Gerhardinger K, Bailey E, Robinson J, Lin J, Cuijpers P, et al. Suicide risk management in research on internet-based interventions for depression: A synthesis of the current state and recommendations for future research. J Affect Disord 2020 Feb 15;263:676-683. [doi: 10.1016/j.jad.2019.11.045] [Medline: 31757622]

Abbreviations

ANCOVA: analysis of covariance
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
C-SSRS: Columbia-Suicide Severity Rating Scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
GAD-7: Generalized Anxiety Disorder 7-item scale
ICBT: internet-based cognitive behavioral therapy
IPDT: internet-based psychodynamic therapy
ISRCTN: International Standard Randomised Controlled Trial Number
ITT: intention-to-treat
MADRS-S: Montgomery Åsberg Depression Rating Scale–self-rated
MINI 7.0: Mini International Neuropsychiatric Interview 7.0
MINI KID: Mini International Neuropsychiatric Interview for Children and Adolescents
MLM: multilevel modelling
PDT: psychodynamic psychotherapy

QIDS-A17-SR: Quick Inventory of Depressive Symptomatology for Adolescents
RCI: Reliable Change Index
RCT: randomized controlled trial
SSRI: selective serotonin reuptake inhibitor

Edited by G Eysenbach; submitted 30.01.20; peer-reviewed by A Abbass, S Erlandsson; comments to author 18.02.20; revised version received 23.02.20; accepted 26.02.20; published 30.03.20.

Please cite as:
Lindqvist K, Mechler J, Carlbring P, Lilliengren P, Falkenström F, Andersson G, Johansson R, Edbrooke-Childs J, Dahl HSJ, Lindert Bergsten K, Midgley N, Sandell R, Thorén A, Topooco N, Ulberg R, Philips B
Affect-Focused Psychodynamic Internet-Based Therapy for Adolescent Depression: Randomized Controlled Trial
J Med Internet Res 2020;22(3):e18047
URL: http://www.jmir.org/2020/3/e18047/
doi:10.2196/18047
PMID:32224489

Original Paper

Detecting Lung Cancer Trends by Leveraging Real-World and Internet-Based Data: Infodemiology Study

Chenjie Xu[1*], MSc; Hongxi Yang[1,2*], MSc; Li Sun[3*], PhD; Xinxi Cao[1], MSc; Yabing Hou[1], MSc; Qiliang Cai[1,4], PhD; Peng Jia[5,6,7], PhD; Yaogang Wang[1], MD, PhD

[1]School of Public Health, Tianjin Medical University, Tianjin, China
[2]School of Public Health, Yale University, New Haven, CT, United States
[3]School of Nursing, Tianjin Medical University, Tianjin, China
[4]The Second Hospital of Tianjin Medical University, Tianjin, China
[5]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[6]International Initiative on Spatial Lifecourse Epidemiology, Hong Kong, China
[7]Faculty of Geo-information Science and Earth Observation, University of Twente, Enschede, Netherlands
[*]these authors contributed equally

Corresponding Author:
Yaogang Wang, MD, PhD
School of Public Health
Tianjin Medical University
No. 22 Qixiangtai Road
Heping District
Tianjin, 300070
China
Phone: 86 13820046130
Email: wyg@tmu.edu.cn

Abstract

Background: Internet search data on health-related terms can reflect people’s concerns about their health status in near real time, and hence serve as a supplementary metric of disease characteristics. However, studies using internet search data to monitor and predict chronic diseases at a geographically finer state-level scale are sparse.

Objective: The aim of this study was to explore the associations of internet search volumes for lung cancer with published cancer incidence and mortality data in the United States.

Methods: We used Google relative search volumes, which represent the search frequency of specific search terms in Google. We performed cross-sectional analyses of the original and disease metrics at both national and state levels. A smoothed time series of relative search volumes was created to eliminate the effects of irregular changes on the search frequencies and obtain the long-term trends of search volumes for lung cancer at both the national and state levels. We also performed analyses of decomposed Google relative search volume data and disease metrics at the national and state levels.

Results: The monthly trends of lung cancer-related internet hits were consistent with the trends of reported lung cancer rates at the national level. Ohio had the highest frequency for lung cancer-related search terms. At the state level, the relative search volume was significantly correlated with lung cancer incidence rates in 42 states, with correlation coefficients ranging from 0.58 in Virginia to 0.94 in Oregon. Relative search volume was also significantly correlated with mortality in 47 states, with correlation coefficients ranging from 0.58 in Oklahoma to 0.94 in North Carolina. Both the incidence and mortality rates of lung cancer were correlated with decomposed relative search volumes in all states excluding Vermont.

Conclusions: Internet search behaviors could reflect public awareness of lung cancer. Research on internet search behaviors could be a novel and timely approach to monitor and estimate the prevalence, incidence, and mortality rates of a broader range of cancers and even more health issues.

(J Med Internet Res 2020;22(3):e16184) doi:10.2196/16184

KEYWORDS

lung cancer; incidence; mortality; internet searches; infodemiology

Introduction

Cancer affects people at all socioeconomic levels and has become a worldwide public health problem. In 2018, the International Agency for Research on Cancer reported the substantial global burden of cancer [1]. Lung cancer is the most commonly diagnosed cancer type (11.6% of all cases) and is the leading cause of cancer-related deaths (18.4% of all cases). In the United States, cancer is the second leading cause of death, resulting in approximately 150,000 deaths per year [2,3]. Moreover, the mortality rate of lung cancer is the highest in the United States. Current data on cancer incidence, mortality, and survival have been mainly collected by the US Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI). However, the year of disclosing such data usually lags about 3 years behind the year in which the data are relevant owing to the time required for data collection, compilation, quality control, and dissemination [4]. Despite an increasing demand for up-to-date knowledge regarding cancers, the lack of real-time data continues to be a major impediment to timely and effective cancer surveillance [5]. Hence, new methods in the era of big data are needed to help supplement current strategies and improve the monitoring of lung cancer.

Along with the rapid spread of social media and medical forums, people frequently search internet resources for symptom-related information, basic medical advice, and to exchange information [6-10]. As more and more people utilize the internet, data from internet searches are increasingly able to better reflect real-world data. With the advent of big data, information and communication technologies have made it possible to reflect trends of real-world diseases based on search data [11-13].

The past decade has witnessed an exponential increase in internet penetration. As of April 2019, 56% of the world's population was reported to have internet access, with rates in the developed world now exceeding 80% [14,15]. According to the World Bank, individuals who use the internet accounted for 76% of the total population in the United States in 2017 [16]. Additionally, 95% of Americans now own a cell phone of some kind, and approximately 77% of the US population now owns a smartphone [17]. Digital data passively generated from user online search behaviors may be utilized to estimate disease metrics in different states until verification data from a traditional source becomes available [18,19]. Internet technologies have the potential to mitigate some of the shortcomings of current health monitoring systems and might be used to supplement existing disease surveillance methods.

Previous studies have attempted to utilize search engines (eg, Google and Baidu) to improve the surveillance of some epidemics. Some scholars have made use of online social networks (eg, Facebook and Twitter) to mine public interest in medical-related issues such as diagnosis and treatment knowledge, patient experience sentiments, and quality of medical services. Meanwhile, more and more studies have elucidated trends in disease communications through internet communities (eg, Reddit) [20-22]. Since 2004, Google Trends [23] has provided national- and state-level search data for free via entering related search query terms [23]. Previous studies have revealed that it is possible to improve the surveillance and prediction of infectious diseases and examine public interest in multiple health topics by monitoring the search behaviors of millions of users and conducting data mining through such search engines [24-29]. However, few studies have focused on leveraging internet search data to monitor and predict cancers at a geographically finer state-level scale [30,31].

In the present study, we tested the hypothesis that the volumes of internet search queries related to lung cancer reflect real-world spatiotemporal variation in the incidence and mortality rates of lung cancer in the United States. Our findings suggest that internet search volumes may reflect disease characteristics of lung cancer (such as incidence and mortality) and provide an additional means of national- and state-level cancer surveillance in the United States.

Methods

Internet Search Data

We collected monthly search data for lung cancer through Google Trends [23] from 2004 to 2018, at both the state and national levels. The search data were downloaded from Google Trends in January 2019. A research firm reported that Google accounted for up to 92% of the market share of search engines in 2018, representing an increase of 57% and 89% in 2004 and 2015, respectively; thus, Google is currently the most widely used search engine in the world [32].

When downloading search data, we could freely choose our own time range. The earliest search data available for Google Trends are from 2004. Therefore, to illustrate and analyze more complete changes in search trends, we first chose to download data from 2004. Although the search data are updated in real time, as mentioned above, the publication of cancer registration data usually lag behind data collection for several years; thus, the latest incidence and mortality rates of lung cancer available for the present analysis were those published in 2015 by the CDC. To explore the association of internet search volumes for lung cancer with published cancer incidence and mortality rates in the United States, we downloaded the relevant data for the same period of time and set the year 2015 as our first deadline. We downloaded data in January 2019 so that we could obtain complete search data from 2004 to 2018 before conducting the subsequent analysis.

Search Query Terms

Lung cancer awareness was examined on the basis of the general population's ability to seek such information or pay attention to it [33]. We selected 12 different query terms among the most commonly used terms for lung cancer [34]. The selected terms were not searched in quotes. Search query volumes were filtered by the term "health" using the Google query category feature to discard non health-related queries that may have confounded the results. Each data point represented the relative search volume (RSV) of specific query terms on a normalized scale of 0-100. The data were divided by total searches of a particular geographic location and the particular time range they represent to compare the relative popularity of the query terms. For example, compared with the total search volumes, if a particular region had a higher number of specific query terms, the RSV would be closer to 100. Data of internet searches used in this study are publicly available, anonymous, and cannot be traced back to identifiable individuals.

Cancer Data

The age-adjusted incidence and mortality rates per 100,000 individuals with lung cancer in both sexes were obtained at the state and national levels from the CDC for the period of 2004-2015 [2]. Incidence and mortality rates were then merged by state and integrated with the monthly RSV data; the processed dataset included 144 data points for each state for a total of 7344 data points.

Statistical Analyses

RSVs at national and state levels are represented as time-series data. Each time series was divided into four parts: long-term trend, seasonal change, cycling, and random fluctuation. We used a time-series decomposition method to eliminate the effects of irregular changes and obtained the long-term trends of the internet search data from January 2004 to December 2018. Each long-term trend is a continuously increasing or decreasing trend that an objective phenomenon exhibits over a long period of time, which may be due to a fundamental cause. The purposes of studying long-term trends are to understand the regularity of the development of internet search data, provide the necessary conditions for statistical prediction, and remove them from the time series to ultimately analyze the influence of other factors on the time series. The steps of this procedure were as follows. First, we chose a multiplication model based on the time-series graph of the RSV of each state. Second, we used a moving-average method to smooth the time series, and used a monthly average method to calculate the seasonal index. Third, we drew a scatter plot and selected the long-term trend of the appropriate curve model to fit the sequence and obtain the long-term trend. Finally, after decomposing the seasonal index and the periodic variation factors, the remaining factors represented the long-term trend.

We performed Spearman correlation analysis to evaluate the relationship between the known lung cancer incidence and mortality rates and Google RSVs over the period of 2004-2015. As Spearman correlation analysis is a nonparametric method, no assumptions of distributions were required on the original variables. We also performed Spearman correlation analysis to evaluate the relationship among the incidence and mortality rates and search data after time-series decomposition.

We assumed that trends in lung cancer-related internet hits would be consistent with those reported for national lung cancer incidence. As the search data could be downloaded in real time, we downloaded complete search data from 2004 to 2018. However, publication of incidence and mortality rates usually lags 3 years behind the actual data. Hence, after the correlation analysis, we conducted a simple linear prediction of the incidence and mortality rates of lung cancer from 2016 to 2018 based on previous epidemiological data, and explored whether trends in lung cancer internet hits would be consistent with the future incidence and mortality rates of lung cancer in the same manner as described above. Therefore, we set the year 2018 as our second deadline for the present analysis.

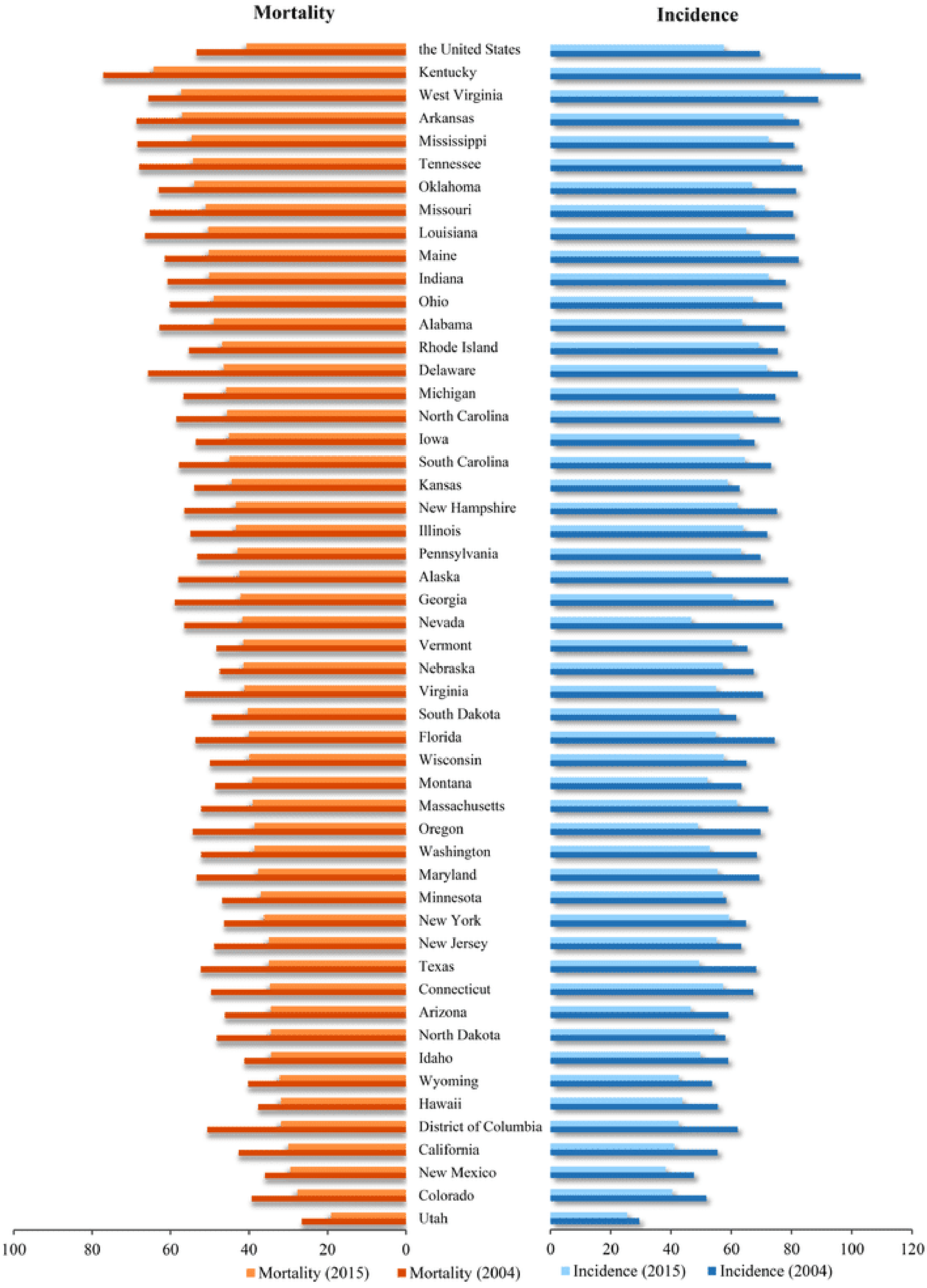
Statistical analysis was conducted using IBM SPSS (version 22.0, IBM Corporation, Armonk, NY, USA), EViews (version 8, IHS Global Inc, London, UK), and R project (version 3.4, R Development Core Team, Vienna, Austria). The statistical significance of the correlation for each dataset was computed using a two-tailed Student *t* test. The significance level was set at alpha=.05 for all tests.

Results

Burden of Lung Cancer

Figure 1 shows the ranking of lung cancer rates in each state in 2004 and 2015. Compared with those in 2004, both rates declined to varying degrees in all states by 2015. Kentucky showed the highest incidence and mortality rates of lung cancer, whereas Utah showed the lowest incidence and mortality rates of lung cancer.

Figure 1. Mortality and incidence rates of lung cancer in the United States broken down by state in 2004 and 2015.



Inclusion of Search Query Terms

We separately extracted the Google RSVs for each of the 12 keywords selected in the above method. However, except for “lung cancer,” the other query terms lacked Google RSVs for all states. Thus, we only selected the common lay term “lung cancer” in this study in consideration of the integrity of the data (see Multimedia Appendix 1).

Regional Distribution of Relative Search Volumes

Figure 2 shows the regional distributions of the Google RSVs, mortality, and incidence data for lung cancer in 2004, 2015, and 2018. The change in color from green to red in the heatmap indicates low to high values of the datasets, respectively. From 2004 to 2015, the following states had the highest incidences of lung cancer (ranked in the top 5 at any point): Kentucky, West Virginia, Tennessee, Delaware, Arkansas, Mississippi, Maine, Oklahoma, Missouri, and Indiana. The following states had the highest mortality rates (ranked in the top 5 at any point): Kentucky, Arkansas, West Virginia, Tennessee, Mississippi, Oklahoma, and Louisiana. The five states with higher average RSVs (>50 from 2004 to 2015) were as follows: Ohio, Maryland, Connecticut, Indiana, and Pennsylvania. The three states with lower average RSVs (<20) were as follows: Mississippi, Wyoming, and North Dakota. Thus, the states with higher incidence and mortality rates usually had higher RSVs (Figure 2).

Figure 2. Regional distribution of mortality, incidence rates, and relative search volumes (RSVs) for lung cancer in the United States in 2004, 2015, and 2018. The color of the map changes from blue to red; the closer the color is to red, the higher the value.

Trends in Internet Searches Related to the Incidence and Mortality Rates of Lung Cancer

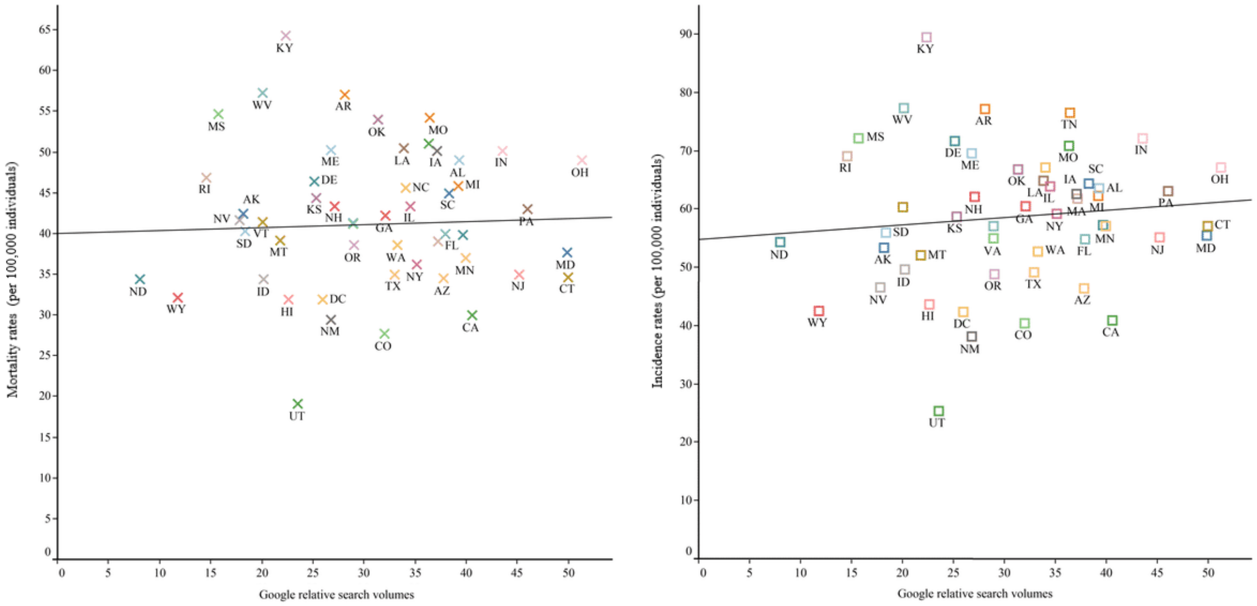
Figure 3 shows a time series of the Google RSVs, along with the incidence (blue lines) and mortality (red lines) rates of lung cancer of all states from 2004 to 2018. The green lines represent the original RSVs, the gray lines represent the new RSVs after seasonal decomposition, and the dotted lines represent the predicted value (from 2016 to 2018). The trends of RSVs for these states initially fluctuated but eventually showed a steady trend over time. The incidence and mortality rates of lung cancer in each state showed a downward trend with time.

Figure 3. Time series of internet search data, incidence, and mortality rates from 2004 to 2018 in the United States. The vertical axis represents the rate and Google relative search volumes, and the horizontal axis represents time. The unit of rate is the number of patients or deaths per 100,000 people.

Correlational Analyses

Multimedia Appendix 2 shows the Spearman correlation coefficients among incidence rates of lung cancer, mortality rates of lung cancer, and Google RSVs by state in the United States. We found statistically significant correlations between incidence rates and original Google RSVs for lung cancer in 42 of 51 states, with the only exceptions being Arkansas, California, Connecticut, District of Columbia, North Dakota, Rhode Island, Utah, Vermont, and West Virginia. We also found statistically significant correlations between mortality rates and Google RSVs for lung cancer for all but four states. For California, Connecticut, and Vermont, there were no statistically significant correlations between incidence rates and relative Google search volumes, and there were also no significant correlations between mortality rates and Google RSVs. Following time-series decomposition of the RSVs, both the lung cancer incidence and mortality rates were correlated with Google RSVs in 50 states, except for Vermont. Figure 4 shows a representative scatter plot of incidence rates, mortality rates, and relative Google RSVs of lung cancer by state in 2015.

Figure 4. Scatter plots with fitted linear regression lines of mortality and incidence rates vs relative search volumes for lung cancer by state. Note that linear regression lines are not equivalent to correlation coefficients and are only included for visual purposes.



Discussion

Principal Findings

Previous studies have mainly focused on the epidemiology of infectious diseases by analyzing internet searches or using social media data sources (eg, Twitter) to conduct health research. The present study examined the association between internet searches and the incidence and mortality rates of lung cancer at national and state levels in the United States over a 12-year period, and predicted the incidence and mortality rates of lung cancer based on their correlations with internet searches. We found an association between RSVs and actual incidence rates of lung cancer for 42 states. Upon examining lung cancer mortality rates, we found statistically significant correlations between mortality rates and RSVs for all but four states.

Our results support the concept of using internet search data and broader public access in health topics for estimating disease characteristics such as incidence and mortality rates. For Kentucky, which had the highest lung cancer incidence and mortality rates in the United States, both incidence and mortality rates were strongly correlated with Google RSVs. In Utah, which had the lowest incidence and mortality rates of lung cancer, RSVs were not correlated with either of these two rates. One possible explanation for this pattern could be that prior online search activities are amplified by individuals at risk for lung cancer, their caregivers, or those who want to acquire knowledge on lung cancer [35-37]. Our results further suggest that search behaviors may reflect, at least partially, the actual prevalence of lung cancer in each state. These types of data sources can be particularly useful when real-time information is needed, because the publication of cancer registration data usually lag several years from data collection. Previous studies have also confirmed that many data elements discernible from a user's social media, especially demographics, can provide new opportunities to characterize the users whose data are analyzed in health research. Google and Twitter-based health research is a growing field that can be utilized to conduct content analysis, surveillance, engagement, intervention, and network analysis in health fields [38,39]. In 37 states, the correlation coefficients between mortality rates and RSVs were higher than those between incidence rates and RSVs. This situation may indicate that people pay more attention to the death caused by lung cancer, implying a general lack awareness of prevention, and that relevant departments are not paying enough attention to the prevention and intervention of lung cancer.

Many other factors may contribute to this association. Although most of the states surveyed in the present study showed statistically significant correlations of RSVs with incidence and mortality rates, some states did not show such correlations (eg, Vermont). However, our study contained only one query term, which was not able to reflect people's search needs and changes in search patterns. Additionally, data downloaded from Google Trends are not raw and unprocessed data, and Google's search algorithm is not static. The algorithm itself is constantly being tested and improved. The instability of Google searching caused by algorithmic dynamics may induce Google Trends to also offer an unstable depiction of the occurrence of lung cancer. Algorithmic dynamics represent changes made by engineers to improve business services and consumer use of their search services. Google's search algorithms and user behavior changes can affect the tracking of Google Trends. The multitude of algorithmic changes is the result of Google's numerous programmers working on small units combined with the use of search engines by millions of consumers around the world.

Hence, it is difficult to reproduce the original algorithm of Google Trends. When users search for disease information, Google may automatically provide more entries that are helpful for diagnosis, such that changes in each user's search result in changes in the Google Trends due to the recommendation of the search bar. Google does offer a function called "related queries," which allows users to identify the retrieved data after a given time series, but this only applies to data at the national level. The key concepts are the lack of transparency regarding absolute search volumes, the exact query text and search Boolean logic used to calculate search volumes, and preprocessing in the temporal domain. After processing by Google, the filtered data may have lost much of the information contained in the original raw data [40]. Several studies have attempted to ameliorate this shortcoming by supplementing search data from other search engines such as Bing and Baidu. These search engines are capable of providing more granular search information (eg, gender, age, and regional information of the search population) than is available from Google Trends [41].

Twitter, Facebook, Google, and the entire internet are changing day by day. Studying these changes, trends, and algorithms may help us to develop more efficient data analysis models. This situation may also be influenced by various public health activities related to lung cancer. The launch of these activities may increase the public's online searches for lung cancer, regardless of disease indicators. One study found that RSVs increased in lung cancer screening after trials reported their potential to reduce the risk of death in heavy smokers [42]. Previous studies have found that after the public health campaign for a disease has been launched, information search behavior related to the disease increases. For example, every October, the annual breast cancer campaign in the United States stimulates online search activities, and the number of Google searches for "breast cancer" has since increased significantly [43]. Moreover, RSVs for lung cancer increased after trials reported the potential of screening to reduce the risk of death among heavy smokers [44]. Reports of cancer in famous individuals may also lead to an increase in internet searches, which has been called "the celebrity effect" [30].

There are also regional differences in search volumes. In recent years, the emergence of corresponding therapy and new early screening methods for lung cancer have helped to prevent and treat lung cancer, but the level of development of lung cancer prevention and treatment, along with the diagnostic rate and survival time of lung cancer, vary from state to state. It is well known that indoor and outdoor air pollution increases the incidence of lung cancer, including smoking, air pollution, secondhand smoke, and radon [45,46]. For example, Kentucky has a high smoking rate and a relatively high incidence of lung cancer, whereas Utah has a low smoking rate and a relatively low incidence of lung cancer. Local authorities should strive to ensure that online health information is available to the public, especially in areas with high smoking rates and lung cancer incidence. Researchers and public health practitioners can further explore this situation by accessing and analyzing recent Google Trends data, which may assist with predicting regions of concern for lung cancer. Additional research on this topic may help to determine how organizations might use Google Trends data as a tool for prediction and targeted interventions. Google Trends can also be used to measure the development of knowledge and interest in new cancer screening tests or specific screening tests. Some scholars have found that people are considerably interested in lung cancer screenings and virtual colonoscopies, but are not very interested in prostate cancer screening [44]. Researchers could explore the psychology behind why people search for lung cancer, such as whether they search primarily for information related to experiencing potential symptoms, have heard of public figures suffering from lung cancer, or have read other information on the news to prompt such searches [47-49]. Additionally, as keyword predictors vary over time, additional research could investigate how searches may relate to current behavioral trends. Our study suggests that Google Trends data could be a new data source for researchers and organizations focused on addressing lung cancer. In the United States, where lung cancer is reported by the CDC and the NCI Surveillance, Epidemiology, and End Results Program through traditional epidemiological methods, supplementing case report data with extant data sources like Google Trends data may help enhance current surveillance by forecasting changes in informing targeted awareness campaigns.

We also forecasted the incidence and mortality rates of lung cancer in all states from 2016 to 2018 based on correlations between these rates and Google RSVs. Our direction for future work may focus on predicting the incidence of lung cancer by combining RSV data and real-world medical data. This approach may not only benefit the development of early warning, intervention, prevention, and control measures for lung cancer but may also extend the research findings of lung cancer to other chronic diseases.

Limitations

Studying search engine data is inevitably restricted by some random factors. As such, the present study had some limitations. First, since the internet search data of other query terms in some states were not included in Google Trends, the present study only contained one search term. Second, Google Trends is not a real epidemiological instrument. The use of Google Trends to estimate lung cancer is not fully representative of the general population, since only individuals with access to the internet can be accounted for via this approach. Third, we were unable to determine which types of internet users conducted search activities. Fourth, we only utilized search data from one search engine, which may not represent search preferences of the whole population, although Google leads the explicit core search market in the United States [32,50]. We hope to continue to find novel ways to identify and reduce biases in search engine data for ultimately utilizing internet search data to provide useful information for cancer surveillance, evaluation of public cancer awareness, and education programs.

Conclusions

The widespread proliferation and increasing utility of the internet have fundamental impacts on the ways in which people seek and acquire medical information. Studying internet search data can augment traditional methods for cancer surveillance when registry data are lagging, and help to achieve improved prevention and control of diseases. In addition, underscoring the potential utility of internet searches may be beneficial to identify people or regions at risk for cancers or other chronic noncommunicable diseases.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (No. 91746205). HY was supported in part by the National Construction of High-level University Public Graduate Project sponsored by the China Scholarship Council.

Authors' Contributions

YW directed the study. HY downloaded all of the original Google Trends data in the United States. CX processed and analyzed the data, and then wrote the first draft of the manuscript. All authors critically revised and reviewed the manuscript, and contributed to the final approved version. YW had the final responsibility of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Selected search query terms and data available in each state.
[DOC File , 166 KB - jmir_v22i3e16184_app1.doc]

Multimedia Appendix 2
Correlation coefficients between lung cancer rates and relative search volumes (RSVs).
[DOCX File , 34 KB - jmir_v22i3e16184_app2.docx]

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. [doi: 10.3322/caac.21492] [Medline: 30207593]
2. United States Cancer Statistics: Data Visualizations.: Centers for Disease Control and Prevention; 2017. URL: https://gis.cdc.gov/Cancer/USCS/DataViz.html [accessed 2020-02-28]
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018 Jan;68(1):7-30. [doi: 10.3322/caac.21442] [Medline: 29313949]
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019 Jan;69(1):7-34. [doi: 10.3322/caac.21551] [Medline: 30620402]
5. McCabe MS, Bhatia S, Oeffinger KC, Reaman GH, Tyne C, Wollins DS, et al. American Society of Clinical Oncology statement: achieving high-quality cancer survivorship care. J Clin Oncol 2013 Feb 10;31(5):631-640 [FREE Full text] [doi: 10.1200/JCO.2012.46.6854] [Medline: 23295805]
6. Bove R. Social Media in the Age of the. N Engl J Med 2019 Mar 28;380(13):1195-1197. [doi: 10.1056/NEJMp1813390] [Medline: 30917256]
7. Merchant RM, Elmer S, Lurie N. Integrating social media into emergency-preparedness efforts. N Engl J Med 2011 Jul 28;365(4):289-291. [doi: 10.1056/NEJMp1103591] [Medline: 21793742]
8. Hartzband P, Groopman J. Untangling the Web--patients, doctors, and the Internet. N Engl J Med 2010 Mar 25;362(12):1063-1066. [doi: 10.1056/NEJMp0911938] [Medline: 20335581]
9. Lemire M, Paré G, Sicotte C, Harvey C. Determinants of Internet use as a preferred source of information on personal health. Int J Med Inform 2008 Nov;77(11):723-734. [doi: 10.1016/j.ijmedinf.2008.03.002] [Medline: 18434246]
10. Chen Y, Li C, Liang J, Tsai C. Health Information Obtained From the Internet and Changes in Medical Decision Making: Questionnaire Development and Cross-Sectional Survey. J Med Internet Res 2018 Feb 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]
11. Leonelli S. The challenges of big data biology. Elife 2019 Apr 05;8:e47381. [doi: 10.7554/eLife.47381] [Medline: 30950793]
12. Munevar S. Unlocking Big Data for better health. Nat Biotechnol 2017 Jul 12;35(7):684-686. [doi: 10.1038/nbt.3918] [Medline: 28700551]
13. Simonsen L, Gog JR, Olson D, Viboud C. Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. J Infect Dis 2016 Dec 01;214(suppl_4):S380-S385 [FREE Full text] [doi: 10.1093/infdis/jiw376] [Medline: 28830112]
14. Miniwatts MG. World internet users statistics and 2019 world population stats. URL: https://www.internetworldstats.com/stats.htm [accessed 2020-02-28]

15. Kepios P. Digital 2019: Global Digital Overview. 2019. URL: https://datareportal.com/reports/digital-2019-global-digital-overview [accessed 2020-02-28]
16. The World Bank. World Development Indicators: The Information Society, 2017 URL: http://wdi.worldbank.org/table/5.12 [accessed 2020-02-28]
17. Pew Research Center. Mobile Fact Sheet, February 2018. URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2020-02-28]
18. Salathé M. Digital Pharmacovigilance and Disease Surveillance: Combining Traditional and Big-Data Systems for Better Public Health. J Infect Dis 2016 Dec 01;214(suppl_4):S399-S403 [FREE Full text] [doi: 10.1093/infdis/jiw281] [Medline: 28830106]
19. Bidargaddi N, Musiat P, Makinen V, Ermes M, Schrader G, Licinio J. Digital footprints: facilitating large-scale environmental psychiatric research in naturalistic settings through data from everyday technologies. Mol Psychiatry 2017 Feb;22(2):164-169 [FREE Full text] [doi: 10.1038/mp.2016.224] [Medline: 27922603]
20. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
21. Okon E, Rachakonda V, Hong HJ, Callison-Burch C, Lipoff JB. Natural language processing of Reddit data to evaluate dermatology patient experiences and therapeutics. J Am Acad Dermatol 2019 Jul 12:pii: S0190-9622(19)32371-0. [doi: 10.1016/j.jaad.2019.07.014] [Medline: 31306722]
22. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying Topics for E-Cigarette User-Generated Contents: A Case Study From Multiple Social Media Platforms. J Med Internet Res 2017 Jan 20;19(1):e24 [FREE Full text] [doi: 10.2196/jmir.5780] [Medline: 28108428]
23. Google. Google Trends. URL: https://trends.google.com/trends/ [accessed 2020-02-28]
24. Young SD, Torrone EA, Urata J, Aral SO. Using Search Engine Data as a Tool to Predict Syphilis. Epidemiology 2018 Jul;29(4):574-578 [FREE Full text] [doi: 10.1097/EDE.0000000000000836] [Medline: 29864105]
25. Deiner MS, Lietman TM, McLeod SD, Chodosh J, Porco TC. Surveillance Tools Emerging From Search Engines and Social Media Data for Determining Eye Disease Patterns. JAMA Ophthalmol 2016 Sep 01;134(9):1024-1030 [FREE Full text] [doi: 10.1001/jamaophthalmol.2016.2267] [Medline: 27416554]
26. Santillana M. Editorial Commentary: Perspectives on the Future of Internet Search Engines and Biosurveillance Systems. Clin Infect Dis 2017 Jan 01;64(1):42-43. [doi: 10.1093/cid/ciw660] [Medline: 27678086]
27. 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]
28. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci U S A 2015 Nov 24;112(47):14473-14478 [FREE Full text] [doi: 10.1073/pnas.1515373112] [Medline: 26553980]
29. Lu FS, Hattab MW, Clemente CL, Biggerstaff M, Santillana M. Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches. Nat Commun 2019 Jan 11;10(1):147 [FREE Full text] [doi: 10.1038/s41467-018-08082-0] [Medline: 30635558]
30. Wehner MR, Nead KT, Linos E. Correlation Among Cancer Incidence and Mortality Rates and Internet Searches in the United States. JAMA Dermatol 2017 Sep 01;153(9):911-914 [FREE Full text] [doi: 10.1001/jamadermatol.2017.1870] [Medline: 28658470]
31. Phillips CA, Barz Leahy A, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship Between State-Level Google Online Search Volume and Cancer Incidence in the United States: Retrospective Study. J Med Internet Res 2018 Jan 08;20(1):e6 [FREE Full text] [doi: 10.2196/jmir.8870] [Medline: 29311051]
32. Statcounter. Search engine market share worldwide. 2019 Mar. URL: http://gs.statcounter.com/search-engine-market-share [accessed 2020-02-28]
33. Ji Z, Zhang Y, Xu J, Chen X, Wu Y, Xu H. Comparing Cancer Information Needs for Consumers in the US and China. Stud Health Technol Inform 2017;245:126-130 [FREE Full text] [Medline: 29295066]
34. Merck Manual Professional Edition, Online edition. Lung Carcinoma: Tumors of the Lungs. URL: https://www.merckmanuals.com/professional/pulmonary-disorders/tumors-of-the-lungs/lung-carcinoma#sec05-ch062-ch062b-1405 [accessed 2020-02-28]
35. Arora NK, Hesse BW, Rimer BK, Viswanath K, Clayman ML, Croyle RT. Frustrated and confused: the American public rates its cancer-related information-seeking experiences. J Gen Intern Med 2008 Mar;23(3):223-228 [FREE Full text] [doi: 10.1007/s11606-007-0406-y] [Medline: 17922166]
36. Kowalski C, Kahana E, Kuhr K, Ansmann L, Pfaff H. Changes over time in the utilization of disease-related Internet information in newly diagnosed breast cancer patients 2007 to 2013. J Med Internet Res 2014 Aug 26;16(8):e195 [FREE Full text] [doi: 10.2196/jmir.3289] [Medline: 25158744]
37. Castleton K, Fong T, Wang-Gillam A, Waqar MA, Jeffe DB, Kehlenbrink L, et al. A survey of Internet utilization among patients with cancer. Support Care Cancer 2011 Aug;19(8):1183-1190. [doi: 10.1007/s00520-010-0935-5] [Medline: 20556435]
38. Huang X, Baade P, Youlden DR, Youl PH, Hu W, Kimlin MG. Google as a cancer control tool in Queensland. BMC Cancer 2017 Dec 04;17(1):816 [FREE Full text] [doi: 10.1186/s12885-017-3828-x] [Medline: 29202718]

39. Pershad Y, Hangge P, Albadawi H, Oklu R. Social Medicine: Twitter in Healthcare. J Clin Med 2018 May 28;7(6):E121 [FREE Full text] [doi: 10.3390/jcm7060121] [Medline: 29843360]
40. Google. How Trends data is adjusted. URL: https://support.google.com/trends/answer/4365533?hl=zh-Hans&ref_topic=6248052 [accessed 2020-02-28]
41. Xu C, Wang Y, Yang H, Hou J, Sun L, Zhang X, et al. Association Between Cancer Incidence and Mortality in Web-Based Data in China: Infodemiology Study. J Med Internet Res 2019 Jan 29;21(1):e10677 [FREE Full text] [doi: 10.2196/10677] [Medline: 30694203]
42. McCook A. More signs lung cancer screening could save lives, 2010. URL: https://uk.reuters.com/article/health-us-lung-cancer-screening-idUKTRE6BR3E720101228 [accessed 2020-02-28]
43. Glynn RW, Kelly JC, Coffey N, Sweeney KJ, Kerin MJ. The effect of breast cancer awareness month on internet search activity--a comparison with awareness campaigns for lung and prostate cancer. BMC Cancer 2011 Oct 12;11:442 [FREE Full text] [doi: 10.1186/1471-2407-11-442] [Medline: 21993136]
44. Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, et al. The utility of Google Trends data to examine interest in cancer screening. BMJ Open 2015 Jun 08;5(6):e006678 [FREE Full text] [doi: 10.1136/bmjopen-2014-006678] [Medline: 26056120]
45. Simon S. American Cancer Society. Lung cancer risks for non-smokers URL: https://www.cancer.org/latest-news/why-lung-cancer-strikes-nonsmokers.html [accessed 2020-02-28]
46. Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013 May;143(5 Suppl):e1S-e29S [FREE Full text] [doi: 10.1378/chest.12-2345] [Medline: 23649439]
47. Ellery PJ, Vaughn W, Ellery J, Bott J, Ritchey K, Byers L. Understanding internet health search patterns: An early exploration into the usefulness of Google Trends. J Commun Healthc 2013 Nov 29;1(4):441-456. [doi: 10.1179/cih.2008.1.4.441]
48. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
49. Steehler KR, Steehler MK, Pierce ML, Harley EH. Social media's role in otolaryngology-head and neck surgery: informing clinicians, empowering patients. Otolaryngol Head Neck Surg 2013 Oct;149(4):521-524. [doi: 10.1177/0194599813501463] [Medline: 23959818]
50. Comscore Releases February 2016. U.S. Desktop Search Engine Rankings URL: https://www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings [accessed 2020-02-28]

Abbreviations

CDC: US Centers for Disease Control and Prevention
NCI: National Cancer Institute
RSV: relative search volume

Edited by G Eysenbach; submitted 08.09.19; peer-reviewed by T Freeman, M Lozano-Lozano, C Ng; comments to author 02.10.19; revised version received 28.10.19; accepted 22.02.20; published 12.03.20.

Please cite as:
Xu C, Yang H, Sun L, Cao X, Hou Y, Cai Q, Jia P, Wang Y
Detecting Lung Cancer Trends by Leveraging Real-World and Internet-Based Data: Infodemiology Study
J Med Internet Res 2020;22(3):e16184
URL: http://www.jmir.org/2020/3/e16184/
doi:10.2196/16184
PMID:32163035

Original Paper

Characteristics of Gun Advertisements on Social Media: Systematic Search and Content Analysis of Twitter and YouTube Posts

Lisa Jordan[1*], PhD; James Kalin[1*]; Colleen Dabrowski[1*], BA

[1] Drew University, Madison, NJ, United States

[*] all authors contributed equally

Corresponding Author:
Lisa Jordan, PhD
Drew University
36 Madison Ave
Madison, NJ, 07940 1434
United States
Phone: 1 9734083740
Email: ljordan@drew.edu

Abstract

Background: Although gun violence has been identified as a major public health concern, the scope and significance of internet gun advertising is not known.

Objective: This study aimed to quantify the characteristics of gun advertising on social media and to compare the reach of posts by manufacturers with those of influencers.

Methods: Using a systematic search, we created a database of recent and popular Twitter and YouTube posts made public by major firearm manufacturers and influencers. From our sample of social media posts, we reviewed the content of the posts on the basis of 19 different characteristics, such as type of gun, presence of women, and military or police references. Our content analysis summarized statistical differences in the information conveyed in posts to compare advertising approaches across social media platforms.

Results: Sample posts revealed that firearm manufacturers use social media to attract audiences to websites that sell firearms: 14.1% (131/928; ±2.9) of Twitter posts, 53.6% (228/425; ±6.2) of YouTube videos, and 89.5% (214/239; ±5.1) of YouTube influencer videos link to websites that facilitate sales. Advertisements included women in efforts to market handguns and pistols for the purpose of protection: videos with women included protection themes 2.5 times more often than videos without women. Top manufacturers of domestic firearms received 98 million channel views, compared with 6.1 billion channel views received by the top 12 YouTube influencers.

Conclusions: Firearm companies use social media as an advertising platform to connect viewers to websites that sell guns. Gun manufacturers appropriate YouTube servers, video streaming services, and the work of YouTube influencers to reach large audiences to promote the widespread sale of consumer firearms. YouTube and Twitter subsidize gun advertising by offering server and streaming services at no cost to gun manufacturers, to the commercial benefit of Google and Twitter's corporate ownership.

(J Med Internet Res 2020;22(3):e15736) doi:10.2196/15736

KEYWORDS

firearms; advertising; social media; internet; gender identity

Introduction

Background

Gun production and imports of guns in the United States have risen significantly over the past 30 years. According to the Bureau of Alcohol, Tobacco, Firearms, and Explosives, there has been a three-fold increase in the total number of guns made in the United States from 3 million in 1986 to over 11 million in 2016. Imports grew from 0.7 million in 1986 to 4.49 million in 2017 [1]. As the international regulation of firearms in most industrialized countries tends to be more restrictive than the

United States, some foreign manufacturers, such as those in Japan, export over 90% of their annual production to the United States [2,3]. Proliferation of firearms presents significant challenges to domestic and global public health: weapons from US markets have been linked to elevated crime, violence, and homicide in other countries [3-8]. Public health responses to firearms require a closer examination of the company practices that facilitate the widespread distribution and increased lethality of small arms in the United States.

US gun control research has much to learn from successful public health programs that sought to reduce widespread injuries from motor vehicle accidents and harm from tobacco use. For example, in outlining his recommendations for a public health approach to guns, Hemenway [3] borrowed the Haddon matrix [9] from injury prevention studies to illustrate opportunities for health interventions before, during, and after injury events involving firearms. Hemenway also recommended that public health scholarship examine the approaches that firearm manufacturers use to promote gun sales, drawing comparisons to very successful health interventions in tobacco advertising and sales [3]. Interventions aimed at restricting the distribution of firearms, limiting gun advertising, and challenging the normalization of gun use are examples of prevention programs that reduce firearm-related injuries.

Although several past studies describe the characteristics of firearm advertising in print magazines and catalogs [2,10,11], the landscape for advertising, sales, and communications has changed radically with the advent of Web-based marketing. New public health scholarship in this area is needed [12,13]. However, despite diversification in the media to include Web-based videos, websites, blogs, podcasts, and social media, many of the advertising messages developed over the past few decades are similar to current approaches, making previous research still relevant for informing contemporary studies of social media and Web-based gun sales.

Building on work to characterize print advertising [2,10,14], this paper examined major themes in firearm advertising used in social media. First, we developed a system for sampling and comparing Twitter use by firearm manufacturers to distribute advertising, then we expanded our work to classify YouTube posts made by gun manufacturers and gun influencers to study firearm advertising. We have begun with a descriptive review of US civilian firearm ownership, the health consequences of firearm advertising, and the role of social media in advertising. Our methods for studying the advertising characteristics used by firearm manufacturers have been described, and the descriptive results of our analysis have been provided. Finally, we concluded with a list of possible interventions to curtail small arms proliferation in US civilian markets and a list of recommendations for future work.

Characteristics of Civilian Gun Ownership in the United States

Recent estimates of US civilian firearms place the total number of private guns at 393 million: higher than the US population [15]. Despite the high overall volume of weapons, the General Social Survey and Pew Research Surveys found a trend toward declining US gun ownership, from over half of adults in 1980 to less than one-third of adults in 2015 [16]. Concurrent with the decline in ownership is a rise in the number of guns per owner: the National Firearms Survey estimated that half the US civilian gun stock was owned by 14% of the gun owners, comprising roughly 3% of the US population [17].

Trends in US gun ownership are stratified by time, place, and demography. At no time in US history has ownership of this lethal weapon been equal. The General Social Survey results reveal significant regional and temporal variations in ownership over the past 50 years. The highest ownership rates were found in 1976, where over 80% of households located in East South Central states (Alabama, Kentucky, Mississippi, and Tennessee) reported owning a gun; conversely, in 2018, sample data for Middle Atlantic states (New York, New Jersey, and Pennsylvania) estimate ownership rates at 19% [18]. Contemporary studies found that 67% of gun owners respond that protection is a major reason for owning a gun [19].

The Pew Research Center found substantive demographic differences in US gun ownership: 39% of men and 22% of women recently surveyed own a gun [20]. White gun ownership is 50% greater than black gun ownership (36% compared with 24%), whereas Hispanic gun ownership is less than half the rate of white gun ownership (Hispanic gun ownership is estimated at 15%) [20]. One study found that public perception of gun ownership significantly overestimates actual ownership, which may contribute to more moderate views of gun control [21]. To reiterate, the majority of people in the United States do not own a gun, but male and white populations own guns at significantly higher rates.

Health Outcomes and Gun Advertising

The health impact of firearms in the United States is widespread. In total, 44% of US residents know someone who has been shot, and a higher proportion, 51% of US gun owners surveyed, know someone who was shot [22]. When combined across causes of death, including firearm deaths from homicides, suicides, and accidental shootings, the annual age-adjusted death rate declined from a peak in 1993, but the total number of firearm-related deaths in the United States has increased in recent years and in 2017 it reached 39,773, exceeding total deaths from motor vehicle accidents [22,23]. According to the Centers for Disease Control, from 1999 to 2017, 612,000 deaths by firearms occurred in the United States [23].

Many scholars have observed the significance of firearms on children’s health: gun-related injuries are the second leading cause of death for children and adolescents in the United States [24]. Over 70% of gun owners have small children, and firearm-related deaths are more frequent when handgun ownership is higher [25]. A study of school-associated homicides in the United States found that although rates of single-victim homicides remained unchanged over the past 20 years, multiple-victim incidence rates increased significantly from 2009 to 2018 [26]. In total, 95% of multiple-victim school-associated homicides in the United States were from firearm-related injuries, compared with the rate of 62.8% for single-victim homicides [26].

The presence of firearms is a known hazard. Occupations that require firearm use have been connected to elevated risk of fatality by suicide. Data from the northeastern United States found that 13% of suicide attempts resulted in fatality; however, 91% of suicide attempts by firearms were lethal [27]. Suicides among US police officers, veterans, and members of the armed forces have come under particular scrutiny. Recent studies have shown that the rate of suicide among veterans and service members is twice the rate of suicide in the civilian population and that firearms were used in 70% of service member suicide deaths [28]. New York Police Department suicides have been declared a mental health emergency [29]. Evidence-based interventions in the Israeli military that restricted officer access to firearms over the weekend reduced suicide deaths by 40% [30].

The risk of firearm injury found in occupational health extends to the general population. Research in preventive medicine found US gun ownership and youth suicide rates to be closely correlated [31] and that state legal restrictions on firearms reduced intimate partner homicide [32]. Permissive state gun laws are also significantly associated with greater numbers of mass shootings [33].

Perhaps in response to the high burden of deaths from firearms, 67% of Americans surveyed by the Associated Press in partnership with the University of Chicago in 2017 and 2018 favor stricter gun laws [34]. Community-based interventions that provide gun storage for families with suicidal individuals may also save lives, suggesting that where fewer state laws exist, law enforcement and gun dealers may step in to promote safe storage [35]. A recent survey of gun owners identified law enforcement and active military as the most effective groups to educate about safe gun storage [36].

Some scholars equate advertising with disease promotion. Freudenberg [37] argued:

> *Advertising seeks to create new customers and encourage existing ones to purchase more. When the product being advertised is lethal (as in the case of tobacco or guns) or can easily be used in ways that harm health (e.g. alcohol, SUVs, and some pharmaceuticals), advertising falls squarely within the rubric of disease promotion.*

Both the expansion of US civilian gun markets and the intensification of gun ownership to include more weapons are troubling trends, given recent evidence that individuals in the United States who self-report *impulsive angry behavior* and gun ownership currently comprise an estimated 8.9% of the population [38]. Reducing the appeal of assault weapons, and limiting advertisements of guns to police, veterans, and armed service members, by better understanding and intervening in gun advertising could translate into fewer gun-related deaths.

Gun Advertising on Social Media

Decisions by major print, television, and Web-based media sources to restrict paid firearm advertising is remarkable: gun advertisements used to circulate among major US newspapers and magazines commonly found in homes [11]. Notable advocacy campaigns, such as *Close the Loophole on Gun Advertising*, went from newspaper to newspaper to negotiate policies that would limit gun solicitations in classified advertisements [37]. Comcast followed NBC, Time Warner Cable, Fox, and ESPN to ban firearm and ammunition advertising in most channels in 2013 [39].

Despite encouraging statements by private media companies, public health responses to predatory advertising have been limited by the 2005 passage of the federal Protection of Lawful Commerce in Arms Act (PLCAA), which protects gun makers from lawsuits related to weapons misuse [14,40]. Although firearm advertisers have acknowledged constraints, such as exclusions from Super Bowl advertising, they have welcomed looser, more diversified Web-based options through both mainstream and alternative internet and social media [41].

The shift to internet advertising has been rapid. Since 2005, the number of US adults who used a social media site grew from 5% to 69% in 2018 [42]. The number is highest among adults aged between 18 and 29 years: 88% [42]. In 2019, the most commonly used social media platform by US adults was YouTube (73%) followed by Facebook (69%) and Twitter (22%) [43]. Among adults aged between 18 and 29 years, over 90% said they have used YouTube, and 44% of adults aged 18 to 24 years said they have used Twitter [43]. According to YouTube, 1.9 billion users log on each month, from over 90 countries [44].

Most US adult YouTube users agree that the site is very important for “figuring out how to do things they haven’t done before” [45]. However, users often identify problematic content. Among US adults viewing YouTube, 61% say they frequently or sometimes have observed videos with “people engaging in dangerous or troubling behavior” [45]. Moreover, 81% of US parents let their children watch videos on YouTube, and 61% of those parents have felt that their child regularly or occasionally “encountered content that they felt was unsuitable for children” [45].

The relative efficacy of social media advertising leading to firearm sales is unknown: most manufacturers are privately owned companies, so documents on advertising expenditures and company profitability are not public [2,3]. However, there is an emerging, but rich, body of research on the use of social media, particularly YouTube, for spreading tobacco and e-cigarette advertising [46-51]. Given there are restrictions on legal channels for advertising, tobacco companies have increasingly turned to internet promotion [52]. Platforms such as YouTube present challenges for consumer information, because it is difficult to differentiate paid advertising from purely creative content, and the authenticity of YouTube videos creates relationships between the video personalities and the viewers, which are particularly influential with young people [53]. In the case of tobacco, exposure to Web-based marketing was found to be a risk factor for use [54-56].

Celebrity endorsements have been found to positively influence sales, and previous research identifies Twitter and other social media platforms as important mechanisms to share endorsements [57]. In addition to celebrities, an increasingly scrutinized source of behavioral change is the role of the internet or social media influencers. For example, recent studies connected tobacco use behaviors to influencers on Twitter [58], and research on eating behaviors in youths found that YouTube influencers significantly affected unhealthy food consumption [59]. Studies now connect *engagement* strategies with increased advertising effectiveness [55,60]. Research also found that tobacco engagement marketing led to an increase in "the risk of initiation and progression and decrease in likelihood of cessation" [55].

Pew Research Center surveys reveal some of the patterns in media consumption among gun owners: 43% of male gun owners (33% of female gun owners) watch television programs or videos about guns, and 39% of male gun owners (28% of female gun owners) visit websites about guns [20]. Researchers also found that gun owners with more guns were more likely to watch gun videos or visit gun websites: 53% of gun owners with 5 or more firearms watch videos about guns, compared with 32% of gun owners with only 1 gun [20]. In total, 51% of gun owners with five or more guns visited gun websites, compared with 22% who owned only 1 gun [20].

Protection is a dominant concern for gun owners and a common marketing message used by manufacturers of guns. According to Pew Research Center findings, two-thirds of gun owners say that protection is the major reason that they own a gun [22]. Handgun advertising has been shown to exploit this rationale, despite evidence that households with firearms are at elevated risks of gun violence from homicide, suicide, and accidental injury [61].

Social networks also appear to influence a variety of health behaviors [62]. *Screenagers*, a documentary by Delaney Ruston, a primary care physician, describes the wide range of negative health and behavioral outcomes emerging from social media use by the youth [63]. Adoption of social media to create discord and harm poses a threat to public health: one study has explored violence and crime-related Twitter use by gangs in Detroit, Michigan [64]. Another recent article found troubling evidence for the use of Russian bots to distribute misinformation about vaccines on Twitter [65].

In some ways, internet and social media companies have been responsive to gun violence and other issues. Social media companies agreed to a variety of restrictions after the Parkland school shooting [66]. YouTube expanded its previous ban on videos demonstrating the use or construction of bump stock modifications following the Las Vegas shooting in 2016 to include more rigorous restrictions [67]. According to stated policy, videos intending to sell firearms, providing instruction on the construction of firearms, ammunition, or accessories, or providing instruction on the installation of accessories are not allowed and will result in the removal of the video [68]. However, the details of the policy description focus on person-to-person sales and do-it-yourself fabrication as the emphasis of content moderation. As of June 2019, Twitter's policy is as follows: "Twitter prohibits the promotion of weapons and weapon accessories globally" [69]. Nevertheless, many are concerned that protections for viewers, particularly for children, are too relaxed and that implementation of company policies has been less effective than socially desired [70].

Recent research on commercial content moderation challenges the face value of social media policies, demonstrating the intentionality behind vague user guidelines and the purposeful cover-up of company instruments of control [71,72]. For example, most commercial content moderation requires human intervention that is not automated [71,72]. The employment and working conditions, particularly with respect to occupational health and mental health, are shocking, unethical, and hidden behind restrictive nondisclosure agreements or by recruiting work through contract labor, piecemeal labor, and offshore arrangements [72].

This study aimed to characterize the contemporary use of social media for the purposes of gun advertising by gun companies and YouTube influencers. We quantified the frequency of common themes found in gun advertisements, as first designated by research pertaining to print advertisements [10]. This study builds on past research of print advertisements by conducting a systematic search of Twitter and YouTube use by major US gun manufacturers, identifying links to internet gun sales made in Twitter and YouTube posts, and conducting a systematic search of gun promotion found in YouTube influencer posts.

Methods

To study the ways that firearm manufacturers use social media for advertising, we performed a systematic search and content analysis. We began by identifying top gun manufacturers. Next, we located the publicly accessible Twitter profiles and YouTube channels connected to the manufacturing companies. We sampled the manufacturer posts by examining the most recent and most popular posts. From our sample, we systematically reviewed the text, images, and video information delivered on each post for the presence of 19 advertising themes, originally explored in a previous study by Saylor et al [10]. These variables were coded into two databases: Twitter posts and YouTube posts.

In addition to surveying information distributed by firearm manufacturers on social media, we chose to explore the recent and most popular posts made by YouTube influencers. On the basis of marketing reports from the firearm advertising industry, we searched for the most influential YouTube channels that focus on sharing information about firearms. This sample of social media posts was also reviewed for the presence of gun advertising themes. After the observations of social media posts were complete, we conducted a content analysis by calculating summary statistics to describe the characteristics of posts made by manufacturers on Twitter and YouTube and by influencers on YouTube. We compared the recent posts with the most popular posts, across platforms and owners.

The specific process for identifying social media advertisements is outlined below. We began with a systematic search of top domestic gun producers and identified top foreign imports to the United States. We explored the relative and cumulative impact of their production of guns, and their relative and cumulative impact on social media advertising. We also specified the database design and contents.

Producers of firearms and ammunition in the United States and the quantities of guns made annually are listed in the *Annual Firearms Manufacturing and Export Report*, published by the

Bureau of Alcohol, Firearms, Tobacco, and Explosives (ATF; see Table 1 for a summary) [73]. Information aggregated by the ATF demonstrates that production of firearms is concentrated in a few companies. The top domestic manufacturer, Sturm, Ruger & Company, produced 1.6 million firearms in 2017, which comprised 19.50% (1,631,554/8,366,943) of all new domestic firearms. The top 4 companies manufacture the majority of all firearms (4,521,925/8,366,942, 54.04%), and the top manufacturers producing over 50,000 firearms in 2017, included 23 companies, accounting for 87.78% (7,345,049/8,366,942) of US production. Of the 2111 firearm manufacturers identified by the ATF, over half (n=1120) manufactured fewer than 10 firearms.

Table 1. Firearm manufacturing characteristics of top domestic producers (over 50,000 firearms produced in 2017).

Federal firearms licensed manufacturers	Types of firearms produced		Number of firearms manufactured (includes rifles, pistols, shotguns, revolvers, and miscellaneous), n (%)
	Rifles, n (%)	Pistols, n (%)	
Sturm, Ruger & Company	661,155 (40.52)	781,623 (47.91)	1,631,554 (19.50)
Smith & Wesson Corp	265,356 (17.62)	1,032,450 (68.54)	1,506,256 (18.00)
Remington Arms	448,513 (55.28)	59,581 (7.34)	811,421 (9.70)
Sig Sauer Inc	35,920 (6.27)	536,774 (93.73)	572,694 (6.84)
Maverick Arms Inc	80,275 (16.08)	0 (0.00)	499,100 (5.97)
Henry Rac Holding Corp	235,037 (100.00)	0 (0.00)	235,037 (2.81)
Heritage Manufacturing	0 (0.00)	0 (0.00)	226,065 (2.70)
Kimber Mfg Inc	11,378 (5.25)	183,858 (84.89)	216,585 (2.59)
WM C Anderson Inc	2295 (1.07)	1448 (0.67)	215,125 (2.57)
Glock Inc	0 (0.00)	175,696 (100.00)	175,696 (2.10)
Palmetto State Armory	28,562 (17.80)	3326 (2.07)	160,417 (1.92)
Springfield Inc	69,352 (46.01)	81,377 (53.99)	150,729 (1.80)
SCCY Industries LLC	0 (0.00)	150,647 (100.00)	150,647 (1.80)
Kel Tec CNC Industries	66,235 (43.97)	58,982 (39.16)	150,630 (1.80)
Radical Firearms LLC	88,430 (96.96)	2775 (3.04)	91,205 (1.09)
Strassells Machine Inc	40,511 (46.82)	46,015 (53.18)	86,526 (1.03)
Aero Precision LLC	1490 (1.87)	2 (0.00)	79,525 (0.95)
Beretta USA Corp	2778 (3.60)	57,411 (74.35)	77,214 (0.92)
FN America LLC	15,614 (20.25)	61,510 (79.75)	77,124 (0.92)
Taurus International	103 (0.15)	69,123 (100.00)	69,226 (0.83)
Colt's Manufacturing	13,942 (23.40)	31,987 (53.68)	59,591 (0.71)
Browning Arms	668 (1.30)	50,331 (97.82)	51,452 (0.61)
Diamondback Firearms	26,960 (52.63)	24,270 (47.37)	51,230 (0.61)
Total among top manufacturers (listed)	2094,574 (28.52)	3,409,186 (46.41)	7,345,049 (87.79)

For this paper, we chose to explore the top manufacturers that produced over 50,000 firearms. Given that a significant portion of firearms enter the United States as imports [2,74], and are not included in the domestic manufacturing list, we identified top manufacturers in other countries with sizable exports of firearms to the United States: Croatia (HS Produkt), Turkey (MKE, also known as Zenith Firearms), the Czech Republic (CZ firearms), and the Philippines (Armscor). Within the list of major producers, a few did not have Twitter handles (Sturm, Ruger & Company; Colt's Manufacturing Company; and HS Produkt), and some did not have YouTube channels (Heritage Manufacturing; HS Produkt; and Radical Firearms). The SCCY Network YouTube channel was excluded because it only included two videos.

By searching Twitter, we were able to identify Twitter handles for 24 manufacturers, representing 68% of the domestic firearm production (Table 2). From these, we archived the 44,571 most recent tweets, ending May 15, 2019, representing 65% of all tweets from these manufacturers, using Twitonomy, a social media analytics service [75]. From the recent tweets obtained, we explored the 20 most recent tweets, ending May 15, 2019, and the 20 most retweeted tweets, for each company: generating a sample size of 928 tweets. The choice to explore 20 was arbitrarily made by the authors to create a sufficiently large sample to code within reasonable time constraints.

By searching YouTube, we found 24 channels hosted by the major firearm manufacturers, which account for 85% of the domestic firearm production (Table 3). Of the over 3600 videos posted by these manufacturers, we chose to classify the 10 most recent videos and the 10 most viewed videos. The number 10 was chosen arbitrarily by the authors. Our sample totaled to 425 videos, ranging in dates from May 2008 to May 2019. Overall, these channels covered over 0.5 million subscribers and 98 million views. The videos we reviewed account for 11.6% of all videos from these channels, but a sizable proportion of all views, at just over 44 million views.

Table 2. Twitter archive summary for top domestic firearm manufacturers by followers.

Manufacturer	Twitter handle	Total number of tweets, n	Followers, n
Glock Inc	@GLOCKInc	4954	290,240
Smith & Wesson Corp	@SmithWessonCorp	2844	265,906
Remington Arms Company LLC	@RemingtonArms	4924	240,378
Beretta USA Corp	@Beretta_USA	14,831	189,945
Springfield Inc	@Springfield_Inc	1634	144,283
Sig Sauer Inc	@sigsauerinc	2356	108,388
Kimber Mfg Inc	@kimberamerica	7646	75,738
Fn America, LLC	@FN_America	1695	68,417
CZ	@czusafirearms	3493	50,328
Aero Precision LLC	@aero_precision	691	43,939
Maverick Arms, INC (subsidiary of Mossberg & Sons)	@MossbergCorp	3480	40,109
Taurus International Manufacturing Inc	@TarususUSA	1785	23,941
Henry RAC Holding Corp	@HenryRifles	1251	17,227
Browning Arms Company	@BrowningArms	766	16,297
WM C Anderson Inc	@andersonrifles	6290	11,434
Sccy Industries LLC	@SCCYguns	1244	11,175
Kel Tec CNC Industries Inc	@keltecweapons	1330	9645
Palmetto State Armory, LLC	@PalmettoArmory	3891	9318
Diamondback Firearms LLC (owned by Taurus)	@DBFirearms	1048	5118
Armscor	@ArmscorRIA	571	3134
Strassells Machine Inc (also known as Hi-Point Firearms)	@HiPointFirearms	489	1559
Radical Firearms LLC	@RadicalFirearms	97	1112
MKE (also known as Zenith Firearms)	@ZenithFirearms	922	560
Heritage Manufacturing Inc	@heritagemfginc	473	131
Total	N/A[a]	68,705	1,628,322

[a] Not applicable.

Table 3. Summary of top firearm manufacturers on YouTube by total views.

Manufacturer	Subscribers, n	Total number of videos, n	Total number of views. n
SCCY Firearms, Sccy Industries LLC	Private	2	Private
Springfield Armory, Springfield Inc	Private	349	Private
Beretta USA Corp, Browning Arms Company	65,407	372	17,553,843
Sturm, Ruger and Co	74,282	364	15,133,053
Browning	18,532	617	13,191,067
Sig Sauer, Sig Sauer Inc	95,927	256	9,894,623
Smith and Wesson, Smith & Wesson Corp	43,865	233	8,696,129
Remington Arms, Remington Arms Company LLC	33,756	242	6,858,982
Mossberg, Maverick Arms, Inc (subsidiary of Mossberg & Sons)	27,593	115	5,631,875
Glock, Glock Inc	56,865	78	4,863,570
CZUSA, CZ	16,781	106	4,263,317
Taurus USA, Taurus International Manufacturing Inc	12,628	34	3,159,038
ArmscorRIA, Armscor	19,039	366	2,988,741
Henry Repeating Arms, Henry RAC Holding Corp	33,783	54	2,289,310
Kimber Firearms, Kimber Mfg Inc	12,418	188	1,489,968
Palmetto State Armory, Palmetto State Armory, LLC	17,190	57	870,149
Colt Manufacturing Co	7031	42	440,015
FN, Fn America, LLC	10,386	32	208,947
Aero Precision, Aero Precision LLC	9617	18	175,656
Zenith Firearms, MKE (also known as Zenith Firearms)	1528	50	146,328
Kel-Tec, Kel Tec CNC Industries Inc	441	41	61,618
Hi-Point, Strassells Machine Inc (also known as Hi-Point Firearms)	1,433	8	54,250
Anderson Manufacturing, WM C Anderson Inc	2,025	20	41,862
Diamondback Firearms, Diamondback Firearms LLC (owned by Taurus)	627	13	32,603
Total	561,154	3657	98,044,944

Firearm advertisers and advertisements refer to the important role of *influencers* in communicating information about new firearms and ammunition and promoting gun ownership and use for recreation and home protection. According to the Danger Close Media (DCM) Group, a firearm advertising organization, influencers are essential to firearm advertisers as social media outlets increase restrictions [76]. From the DCM Group, a list of top influencers was identified. We explored influencer channels on YouTube and Social Blade, a YouTube analytics website, to identify influencers that were not recognized by the DCM Group. Among over 4 dozen influencers found, we inspected the top 12 influencers, with over 150 million channel views. The list of influencers, along with channel names, views, subscribers, and short summaries can be found in Table 4. The top 10 and most recent 10 videos were reviewed for each channel based on the same attributes explored for Twitter and YouTube posts shared by manufacturers. The sample of influencers included a total of 239 videos.

Drawing from the methodology and categories originally identified by Saylor et al [10], which explored themes in firearm advertisements commonly appearing in print media, we reviewed our samples of 928 tweets, 425 videos by manufacturers, and 239 videos by influencers, across 19 characteristics, described in Table 5. We expanded our classification system for influencers to identify various forms of paid promotions. All characteristics listed in Table 5 were coded as binary variables.

Tweets by manufacturers typically included photos. Tweets often include video content, so Twitter and YouTube samples overlap to some extent. In describing the content characteristics of the tweet or video post, we considered the descriptive text, photo, and video included in the message. Manufacturers and influencers post videos in a variety of YouTube categories, including education, sports, entertainment, people and blogs, science and technology, and more.

Table 4. Characteristics of YouTube firearm influencers by total channel views.

Influencer channel	Subscribers	Total channel views	Description of influencer	Most viewed video
hickok45	4,201,360	1,143,690,107	Retired middle school teacher Greg Kinman collaborates with his son to review historic and modern firearms	460 Magnum versus Watermelons
DemolitionRanch	6,832,879	983,053,254	Popular YouTube personality and veterinarian Matt Carriker produces vlog-style gun reviews	How Deep into Dirt Will It Go?
FPSRussia	6,569,240	851,074,661	Now inactive; was one of the first and most popular gun channels on YouTube	AA-12 Fully Automatic Shotgun!!!
Iraqveteran8888	2,168,324	561,054,559	Iraq War veteran Eric Blandford makes gun and podcast videos with strong second amendment themes	RANGE TEST: THE ULTIMATE AR-15 MALL NINJA TACTICAL ZOMBIE DESTROYER!
Active Self Protection	1,385,888	557,194,826	Navy Veteran John Correia posts videos of him analyzing various self-defense encounters taken from security camera videos	More Proof that Evil Exists in Our World | Active Self Protection
Edwin Sarkissian	1,874,572	428,975,314	Entrepreneur makes range shooting videos with his friend	How many PUBG Cast Iron skillets does it take to stop a bullet?
Forgotten Weapons	1,211,469	359,042,434	Ian McCollum showcases historical guns and modern guns with unique histories	World's Smallest Pistol—2.7mm Kolibri
FullMag	2,406,155	352,093,297	Richard Ryan and a team of others make gun videos featuring a related brand, Black Rifle Coffee, and military themes	Will Bulletproof Glass Stop A .50 Cal? slow motion Richardson Ryan
nutnfancy	796,686	298,717,269	A retired Air Force pilot reviews guns, knives and weapon accessories. His videos occasionally feature his son, photos of military service, and his wife	Weasel vs Ground Squirrel: Nature's Combat
sootch00	837,830	231,285,288	Sootch makes gun review and podcast style videos with his sons and daughter: “God bless America, long live the republic.”	Classic Firearms Tour! Surplus Gun Heaven!
TFB TV	652,752	175,649,925	A team of people produce high quantity of gun review videos especially at events like the SHOT Show	Top 5 Hilariously Bad Carry Guns | TFBTV
Military Arms Channel	946,843	158,139,701	Tim Harmsen, Marine Corps veteran, reviews modern and historical military guns; he is known for his anti-NRA[a] stance	5.7×28mm versus 22 Magnum
Total	29,883,998	6,099,970,635	N/A[b]	N/A

[a] NRA: National Rifle Association.

[b] Not applicable.

Table 5. Characteristics assessed in Twitter and YouTube activity of major firearm manufacturers and firearm influencers.

Characteristics	Description
Handgun	Handgun or pistol present in photo, text, or video
Shotgun	Shotgun present in photo, text, or video
Rifle	Rifle present in photo, text, or video
Attributes	Characteristics of firearm described in photo, text, or video
Protection	Post indicates or exhibits firearm use for protection
Hunting	Post indicates or exhibits hunting themes
Recreation	Post indicates or exhibits firearm use for recreation
2A	Post indicates or references the US Constitution’s Second Amendment
NRA	Post references the National Rifle Association
Conceal Carry	Post indicates or references firearm for concealed carry
Family	Post indicates or exhibits family themes
Kids	Post indicates or exhibits child use of firearms
Female	Post includes a woman or quotes a woman
Patriotism	Post indicates or exhibits patriotic themes (flags, leaders, and American pride)
Veterans	Post references or exhibits veterans
Military	Post references or exhibits military themes (soldiers, military use of weapons, and endorsement)
Police	Post references or exhibits law enforcement themes (police, thin blue line, and endorsement)
Western	Post references or exhibits western or cowboy themes
Weblink	Post links to website with sales, visually displays the link, or presents the link verbally
Gun Brand is Mentioned[a]	Video or description identifies specific gun brand(s)
Gun-Related Paid Promotion[a]	Video or description features paid promotion from the firearm or ammunition company
Nongun-Related Paid Promotion[a]	Video or description features paid promotion from the nongun company

[a] Additional characteristics identified for influencers.

Results

Quantitative Aggregation

A quantitative aggregation of the themes found in social media advertisements made by gun manufacturers and influencers helps characterize the prevailing strategies used by companies to increase purchases of consumer guns. To summarize our findings in social media advertising, we compared the use of themes across social media platforms, and we explored the types of guns advertised, advertised uses of guns, and how women, military, and police themes enter into advertising.

Content Density and Gun Types in Firearm Advertising

The content and themes presented in YouTube posts exceeded the information conveyed by Twitter posts. An average of 2.2 themes was found for our sample of Twitter posts (n=928); whereas, twice as many themes, 4.4, were found in YouTube posts by manufacturers (n=425), and 5.3 themes were found on average among YouTube influencer videos (n=239). In summary, we found that twice as many themes are communicated in YouTube posts, compared with Twitter posts, indicating a higher density of content per YouTube post.

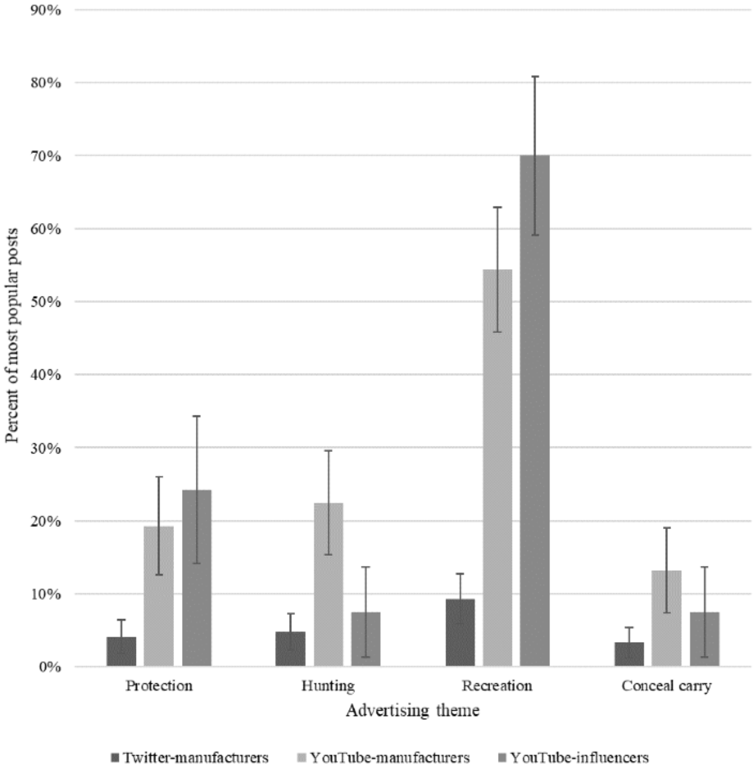
We found that handguns were advertised more frequently by manufacturers, and rifles were advertised more frequently by YouTube influencers. In comparison to shotguns or rifles, manufacturers on Twitter and YouTube more frequently display or refer to handguns or pistols (311/928, 30.2% of Twitter posts, and 243/425, 56.1% of YouTube posts). YouTube influencers, however, displayed rifles more often than other gun types (134/239, 58.3% of posts displayed rifles). This means that over half of the YouTube posts made by manufacturers intend to promote handguns but over half of the YouTube influencers promote rifles.

Gun Use in Advertisements

Of the major themes we examined, recreation was among the most common themes in YouTube video posts. We identified the recreational use for posts that displayed or discussed gun range shooting or shooting targets. The hunting use was classified in its own category (see Figure 1 for comparison). When we split the sample to explore only the most viewed YouTube posts, we found that 54.4% (124/228) of posts by manufacturers and 70.0% (84/120) of influencer posts displayed recreational gun use.

Military, patriotic, and law enforcement themes were also commonplace: 46.7% (56/120) of the top viewed influencer posts depicted military themes. Retweets from all branches of the armed forces were identified among Twitter posts made by gun manufacturers. Patriotic theme prevalence varied depending on the source: 1 in 5 of the most retweeted Twitter posts (108 of 484 posts), 1 in 4 of the most viewed YouTube posts by manufacturers (63 of 228 posts), and 1 in 3 of the most viewed YouTube posts by influencers (40 of 120 posts) conveyed US patriotism.

Figure 1. Suggested use for guns by most popular posts (1% CIs displayed).



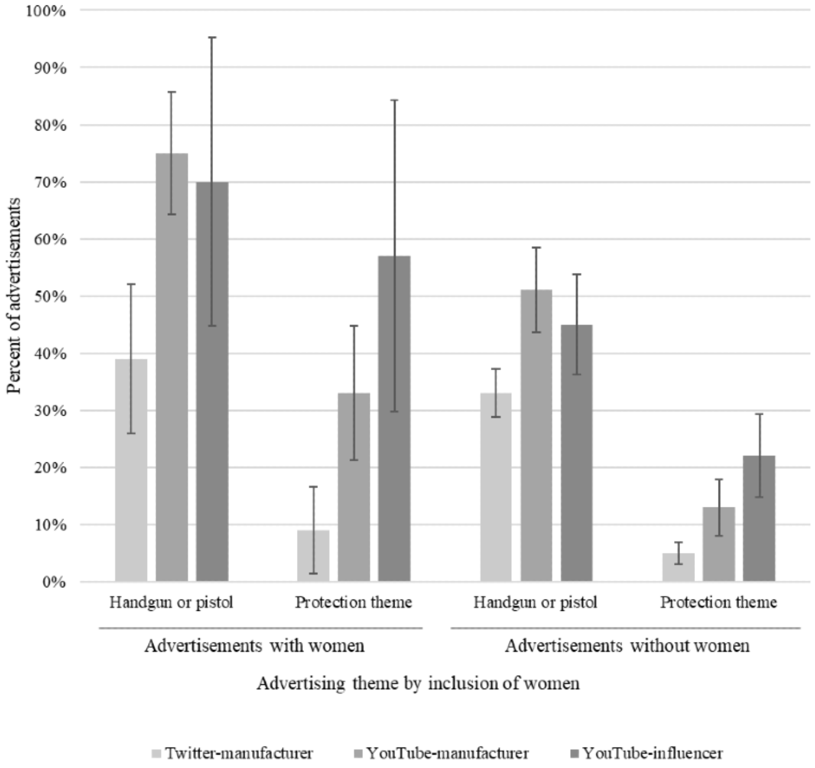
Women, Children, and Other Themes in Gun Advertising

The appearance of women in firearm advertising also varied significantly by source: the presence of a female in an advertisement was found in more than 1 out of 5 YouTube posts by manufacturers (108 of 425 posts), but in less than 1 out of 10 posts on Twitter (93 of 928 posts) or in those made by YouTube influencers (23 of 239 posts). Of the posts that do include women, handguns and gun protection themes were more prevalent (Figure 2).

Family themes were found in fewer than 10% of Twitter or YouTube posts. In all, 4.0% (17/426) of YouTube posts by manufacturers included children. Posts with children showed young children observing or participating in gun fire at shooting ranges, displaying hunting weapons, or receiving firearms as gifts. Although posts with children are less common than the other themes explored, children appear twice as often in video posts with women. The once common western and cowboy themes were found less than 5% of the time across sources.

Second amendment themes were also found in less than 5% of YouTube posts and 6.8% (64/928) of Twitter posts. The National Rifle Association (NRA) mentions or promotions were identified in 9.1% (85/928) of Twitter posts, 2.1% (9/425) of YouTube manufacturing posts, and 14.6% (35/239) of YouTube influencer posts. Although we also coded a theme labeled *attribute*, we found that the vast majority of gun advertisements describe the attributes of the guns. We chose to exclude this category from substantive comparison.

Figure 2. Prevalence of handgun- and protection-themed advertising in advertisements with and without women (1% CIs displayed).



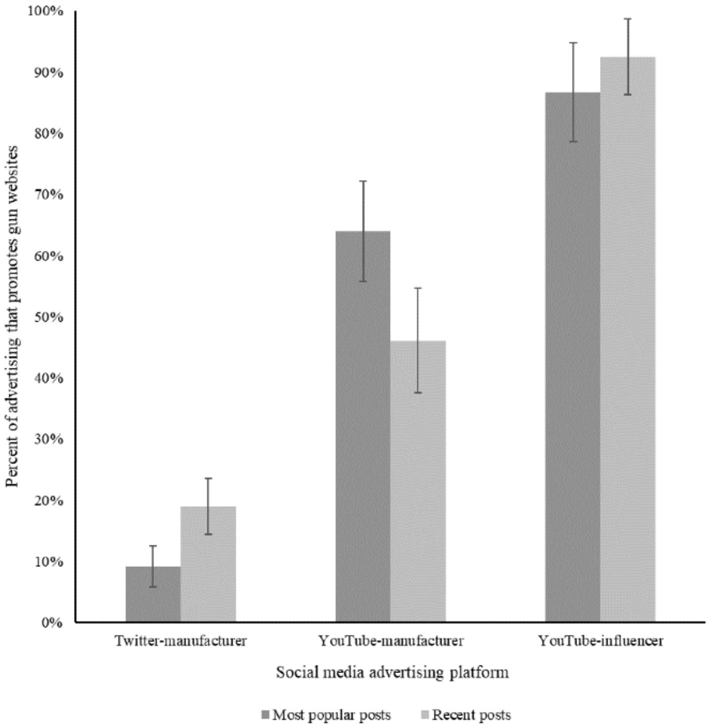
Promotion of Internet Sales

In contrast to generally stated social media policies, social media posts connect viewers to websites with gun sales (Figure 3). YouTube posts more frequently provided links than Twitter posts. YouTube influencer posts were the most likely to link to gun sales: 8 out of 10 of the most popular YouTube influencer videos connected viewers to sales (104 of 120 posts), and 9 out of 10 of the most recent videos connected viewers to sales (111 of 120 posts).

In total, 64.4% (154/239) of YouTube influencer posts mention specific gun brand names, and 19% (45/239) include gun-related paid promotions. In total, 6.6% (16/239) of YouTube influencer posts contain nongun paid promotions; 18.9% (91/480) of recent Twitter posts made by gun manufacturers link to websites, but only 9.0% (44/484) of the most popular Twitter posts link to websites. Half of the YouTube videos by manufacturers provided links to internet sales (228 of 425 posts).

Among Twitter posts, firearm giveaways were not uncommon, and some manufacturers advertise and giveaway *builds* for custom rifles. At the time of analysis, firearm sponsorship in social media depicted gun use by the Major League Baseball pitcher, Andrew Cashner; company logos on a National Hockey League zamboni for the Nashville Predators; and as a featured sponsor of the National Association for Stock Car Auto Racing. Actress Halle Berry and actor Keanu Reeves were featured in video and photo Twitter posts for Sig Sauer MPX. One of the most popular retweeted posts across all manufacturers was a post by Anderson Rifles, who retweeted a post originally made by Texas governor, Greg Abbott: “I’m EMBARRASSED: Texas #2 in nation for new gun purchases, behind CALIFORNIA. Let’s pick up the pace Texans. @NRA.” The post also linked to the *Houston Chronicle* article that reported the state ranks for new gun purchases [77].

Figure 3. Percentage of recent and most popular posts that promote gun sales websites by social media type (1% CIs displayed).



Examples of Gun Video Content

Exploitation of stranger rape and glorification of military gun use were easily found in contemporary gun advertising. The following descriptions of a few important case studies exhibit gun advertising themes of protection for women, sexualization of gun use, promotion of assault weapons use for nonmilitary buyers, and arms production for the purpose of undermining gun control. The guns featured in these examples include Glock handguns, assault rifles and pistols from Sig Sauer, and AK-47s from Palmetto State Armory.

The *GLOCK and Gunny—Wrong Girl* advertisement was first posted by Glock on YouTube in 2013. This 2-min video is the most viewed manufacturer video with over 1.3 million views. In the video, a young woman at home, watching television in her pajama shorts and top, is being stalked by a bearded man in a van. She hears a knock at the door, does not see anyone, and goes to her room to remove a gun, presumably loaded, from a safe under her bed. Banging at the front door continues. She stands pointing her handgun at the door, and when it bursts open, the attacker faints. The police arrest him. No words are spoken until R Lee Ermey, playing the medic who constrains the attacker, concludes: “Somebody picked the wrong girl.” A link to the Glock website is shown across the screen.

The Glock video appears to be well received, with over 7000 likes, and a like to dislike ratio of 9:1. A user with the pseudonym *UziNineMillomeetah* commented, “Smokin’ hot red head who likes Glocks? My dream girl.” The video conveys handgun use for home protection and protection for women against stranger rape, while normalizing gun use and supporting defensive gun use as the first and only self-defense tactic. Gun use is sexualized. R Lee Ermey posted an *Extended Version* on his website, which received over 3 million views, but is identical to the post by Glock. Other videos in this series of advertisements include *GLOCK and Gunny—Wrong Convenience Store* and *Gunny & Glock—Wrong Guy*. All of the advertisements in this series convey the message that Glock guns are for protection: a message that also seeks to undercut public health evidence that the possession of firearms is a major risk factor for violent death from homicide and suicide and accidental death by a firearm.

Another category of advertisement worth exploring includes messages that glorify combat weaponry for private consumption. The Sig Sauer Sig MCX VIRTUS Mission series, including 4 advertising videos with a combined 0.75 million views, labeled under the YouTube category *Sports*, offers a case study in promotional militarism. The advertisements include the following: Mission 1: Overwatch, Mission 2: Target Identified, Mission 3: Vehicle Assault, and Mission 4: Tango Down. These videos feature gun use in cinematic combat settings, specifying the location as Ramadi, in Iraq. In *Mission 1*, the Sig MCX VIRTUS is shown in use by a sniper. The text on the Mission 1 post declares: “The world’s most innovative battle rifle, the SIG MCX VIRTUS, is ready for any mission.” The text on YouTube accompanying the Mission 2 video includes links to the sales website and explains, “Built from the ground up for suppressed operation, the MCX VIRTUS can go places - and do things - that no other rifle can.” Two messages are evident from these advertisements: first, the Sig MCX VIRTUS is intended for combat and lethality and second, Sig Sauer seeks to provide products with highly customized combat performance features to gun buyers.

Although Glock and Sig Sauer chose highly tailored video advertisements, Palmetto State Armory, a US gunmaker of AK-47s, posted a candid description of the company’s history and goals. Injured veteran and CEO Jamin McCallum stated the following:

> *I hope in thirty years people look back and say…we tried to pass gun control…but it wasn't very effective…because this pesky company made 20 million ARs…they got them into circulation…and now the regulations we put into effect have little effect because there's so much of it out there already.*

Normalization of gun use with the intent to *reach the masses*, was an object of discussion by YouTube influencer, sootch00, who stated the following:

> *We have got to stay on YouTube because this is where the masses are coming. I'm telling you guys all over the place, people tell me all the time: "I got on YouTube, I saw a gun video, I was like, oh, I hadn't even thought about guns in a long time." Before long they were watching different gun videos. They went out and bought a gun. They got to concealed carry. They joined the NRA, or GOA, or whatever. They became second amendment activists. If I go to Full30, I'm just preaching to the choir.*

Full30 is a reference to a firearm video website that is sponsored by firearm manufacturers and contains almost exclusively videos on guns. Palmetto State Armory's video collection includes a video of *Sarah*, a gunsmith, demonstrating how to assemble an AK-47 from parts that can be purchased from the manufacturer [78].

YouTube influencers occasionally give weapons poor reviews. In these moments, influencers may convey certain health warnings on a variety of factors. Influencers observed *gas in your face* from weapons use, commented on offending smells, eye irritation, potential risks of hearing loss, concerns about extremely high temperatures of the firearms after use, and observations about weight and recoil of weapons. Complaints like these by military personnel have been important in directing health research to explore the respiratory effects of occupational gunfire, including two recent Norwegian studies that found significant declines in forced expiratory volume even 24 hours after shooting practice, and reports of respiratory symptoms similar to metal fume fever in almost all participants [79,80]. The authors explain that "soldiers are exposed to emissions of CO, particulate matter (dust), combustion products, copper, zinc, bismuth, lead and tin," and "bullets without potentially harmful emissions are not available" [80]. Unfortunately, these important health messages are lost to the more common tropes from influencers describing the *fun* of recreational shooting.

When Saylor et al [10] explored firearm advertising in gun magazines from 2001 to 2002, they found a circulation of 4.2 million, with hunting and outdoors (20.4%) and patriotism (15.0%) as the most common themes identified in their sample. These themes continue to be prevalent. However, the reach of videos, based on measurements of views, appears to be significantly wider than print materials.

Discussion

Principal Findings

We demonstrate that firearm manufacturers use Twitter and YouTube to promote the sale of guns to millions of viewers. The renewed interest in and attention to firearm advertising permits an important deepening and sophistication to public debate in the United States on the scope of possible responses to gun violence. Like all consumer products, firearms have a life cycle that requires scrutiny and is subject to intervention through multiple phases: design, manufacturing, retail, distribution, and disposal.

A focus on firearm advertising demonstrates that marketing to promote retail and distribution is one area, among many, that may be considered for both research and revised rule making. Better understanding the life cycle of guns raises both opportunities and significant challenges to where and how we work as communities to reduce lethal violence. This section first explores opportunities for changes in the US legislation that can help reduce the widespread proliferation of small arms in the United States by limiting advertising and suggests avenues for future research. After identifying opportunities, major challenges are listed to propose subsequent research and dialog on these more embedded obstacles to achieving peaceful communities.

Opportunities

Specific changes in the legislation are possible. The American Public Health Association collaborated with 14 partners to write to the US Congress in favor of federal research funding on gun violence, currently limited by the 1996 Dickey amendment [81]. Several public health scholars have argued in favor of repealing the PLCAA [14,40]. Repealing the PLCAA would remove protections for manufacturers and dealers from lawsuits related to harmful gun use and false advertising pertaining to gun use as a form of protection.

Organizations such as Moms Demand Action have made progress in lobbying retail stores, such as Starbucks and Target, to prohibit guns from their premises. Furthermore, private sector policies may shape retail behavior. The *Washington Post* reported an example where the software marketing agency, Salesforce, threatened to withhold services for distributors of military-style rifles [82]. Both private and public actions, from grassroots and from the national stage, can help build a culture of nonviolence; and, should the PLCAA and Tiahrt Amendment be repealed, 1 advertisement agency has already proposed graphic warning labels for ammunition [83].

The work presented here studies how firearm manufacturers use social media for advertising. However, present advocacy efforts to reduce gun violence have made significant progress in connecting gun control advocates by using social media to promote community engagement and campaigns to incentivize corporate changes. For example, the Moms Demand Action campaign explicitly discusses the strategy to use social media to unite gun control advocates [84]. Every town for Gun Safety [85] has similarly adopted a social media approach. Web-based media coverage of gun violence from *The Trace* provides an example of generating a counternarrative to the NRA's Web-based media [86]. Academic consortiums, such as those emerging from work by the Coalition to Stop Gun Violence, can assist particularly during strategic opportunities when lawmakers are focused on gun violence [87]. All this is to say gun control advocacy and its relative success in relation to gun promotion deserves deeper consideration and can contribute to best practice guidelines or *lessons learned* materials for local health departments and advocacy groups.

Additional research can contribute to our understanding of firearm advertising. For instance, future work in this area could explore the local use of social media advertising by small business producers of firearms and the interaction of these uses with larger companies. The state of New Jersey offers a simple example. Among the states with relatively few firearm manufacturers, New Jersey ranked 46th in the number of firearm manufacturers, with 28 federally licensed firearm manufacturers and one major producer: Henry Repeating Arms. However, 79% (22/28) of New Jersey manufacturers maintain websites, most of which connect users to sales and social media. Although the sales may be relatively small for these producers, their community impact on firearm social media may be significant.

As suggested in the literature on health communications generally [88], understanding the media commonly viewed by gun owners can help in the design of brochures, posters, webpages, and informational materials that may reinforce recommendations made by physicians on the topic of gun ownership (eg, facts on gun ownership and health risks introduced into homes with guns and evidence of decades of predatory advertising for the purpose of increasing sales). Twitter is also used by local health departments for health communications, and scholars have suggested that these programs might engage community members, to a greater extent, in a dialog about local health [89]. Social media also offers the possibility for public health engagement in the chat features, as were identified in some protobacco YouTube posts [52]. Some research has gone as far to suggest that policy might help ensure a prohealth balance to available media [90].

Obstacles

The globalization of media presents an obstacle for regulation, although some countries do legislate restrictions [91]. An emerging concern from the public health literature on social media and tobacco use is that social media are not neutral platforms. Social media companies are commercial entities, not public goods, despite efforts by these companies to appear free and open [71,72]. Significant progress made toward restrictions on tobacco, eg, are facing reversal owing to the combination of e-cigarettes, internet advertising, and internet sales. Youths' use of e-cigarettes has been declared an epidemic [92].

The power of industry, where full-time employees are paid to produce advertising content—photographs, videos, entertainment, and compelling messaging—for social media platforms, is not counteracted by advocacy messaging or ordinary people who hold the majority opinion [93]: firearm ownership and tobacco use are not health behaviors that should be encouraged or modeled. Firearm manufacturers are making efforts to fully exploit social media channels for commercial purposes.

The incentive to profit from consumer products exceeds the current social capacity to counterpose these messages, suggesting that health interventions are necessary. However, guidelines are available for communicating public health objectives effectively [94,95]; and, several organizational structures serve to advocate for changes in corporate behavior: national organizations, coalitions, health professionals and researchers, legal groups, local organizations, and other participants in campaigns [37]. These groups can, then, engage in information gathering, legislative action, electoral activities, litigation, actions aimed at corporations, and education, information, and mobilization campaigns [37]. Roberts suggests that democratic policy interventions and partnerships with public libraries may be more suitable approaches to combating misinformation spread by social media than working within social media settings [72]. This paper contributes to the larger discussion about opportunities for regulating the media that promote harmful health behaviors in the United States and internationally.

Unfortunately, rather than entering at the peak of adoption, public health work is now addressing firearms, tobacco, and marijuana in a phase of escalation. From this point of view, gun violence and health effects from tobacco, e-cigarettes, and marijuana are anticipated to increase and not decline in the foreseeable future. As some firearm manufacturers and dealers have increased sales of gun parts to avoid regulation, concerns about circulation and social media distribution of 3D printed guns remain on the horizon. A future legislation and policy intervention to limit the advertising and content viewed by children is likely [70]. Such work might consider the characteristics and strategies used by influencers.

The domestic production and distribution of small arms in the United States contributes to regional and global health challenges. While reporting on the trial of Joaquin Guzman, the Mexican drug lord known as *El Chapo*, the PBS Newshour stated that "one of El Chapo's deputies testified that 99 percent of the guns he purchased came from the United States" [96]. Actions taken to understand and intervene in the life cycle of US firearms can reduce gun violence not only in the United States, but regionally and around the world. In this way, by focusing upstream on gun production, US public health actors can make substantive contributions to the UN Sustainable Development Goal 16: "Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels." Specifically, Target 16.1 seeks to "significantly reduce all forms of violence and related death rates everywhere" [97].

In building consensus and shifting cultural norms on gun ownership and use, US military and law enforcement present the largest obstacles, but also the most important opportunities for reducing gun violence. Public health collaboration with the military and law enforcement is absolutely necessary for building effective actions. For example, public health review, health impact assessments, and environmental impact assessments of military and law enforcement small arms purchasing and communications could provide agencies with vital information about the civilian and regional health consequences of their choices.

Any restrictions on military and law enforcement agencies from training soldiers and police officers on gun safety in the home should be lifted. Section 1062 of Public Law 111-383, which “forbids the Department of Defense to ‘collect or record’ any information about private firearms of members of the military or its civilian employees,…” [14] should be repealed. Social media produced by the military and law enforcement currently works to the benefit of firearm producers: future social media produced by these government agencies should consider opportunities to promote peace and stability, in our homes, communities, and around the world. Perhaps the Department of Defense (DoD) might reconsider social media use altogether given the commercial, nonpublic, and obscurely regulated nature of these platforms.

In some cases, helpful changes to firearm advertising could emerge from enforcement of existing policy. According to DoD instructions on visual information:

> *Service members must comply with DoD 5500.07-R, DoDI 8550.01, and DoDI 1334.01 prior to permitting NFEs (Non-Federal Entities) to use their image in uniform. Both active duty Military Service members and former members are prohibited from wearing their uniform in connection with commercial interests when an inference of official sponsorship for the activity or interest could be drawn. [98]*

Several influencers, such as nutnfancy, may be in violation of this policy, and the DoD may be able to take steps to reduce the reach of influencers who gain credibility in promoting firearm sales by visually displaying their current or past military experience. The DoD disclaimer, “The appearance of U.S. Department of Defense (DoD) visual information does not imply or constitute DoD endorsement,” was not identified in firearm advertising that used armed service symbols and images. Sig Sauer, eg, explicitly uses DoD visual material as an endorsement, without disclaimers.

Limitations and Future Research

By focusing primarily on social media advertising, we were not able to demonstrate why, despite tighter restrictions on cable advertising, firearm production rose and stayed high, during both Obama and Trump administrations. A market analysis that includes careful consideration of the firearm economy over the past decade may help reveal the causes in the overall purchasing trends for guns. Twitter and YouTube may be used for a time-series analysis and possibly for uncovering the relative market power of social media in influencing global firearm sales.

The methodology that this paper pursued was based on previous research on print advertising. By selecting the most recent and most popular Twitter and YouTube posts, we generated a systematic search and analysis of major gun manufacturer and popular influencer use of these services. Our search was not a simple random sample. We cannot make inferences about the less-viewed influencers that we did not assess. We also cannot infer social media themes used by very small gun manufacturers. We were also not able to examine gun use on Instagram or Facebook, which are also commonly used social media services. The exploration of these forums is an opportunity for future research.

By choosing to evaluate the 20 most recent tweets and 10 most recent videos, the periods for the evaluated posts varied depending on the frequency and temporal distribution of posts made by manufacturers and influencers. We sought to explore variations in posts based on popularity and recency to identify differences between the two. Our intention was to characterize the advertising themes and presence of links to sales found in social media by top gun producers and influencers, and not to examine causal relationships. Future research could benefit from better understanding contemporary firearm consumers and the influence of advertising on their decision making.

Continued work in the domain of firearm advertising can help deepen the understanding of other themes and strategies used by manufacturers. For example, safety can be a specific theme explored in detail. Connections to religion were observed and worth exploring further. Race, ethnicity, and sexual orientation were not explored in detail here, but would be interesting to study further. Our work was limited to English-only social media: advertising in other languages would make for an interesting comparison. Podcasts are another new source of internet content distribution that would be helpful to analyze.

Social media and print media across audiences, specifically the police and military, need more attention. Gun advertisements appear to be printed by *staff writers* across magazines and Web-based sources. While researching for this paper, we found a Glock advertisement surface word-for-word in articles, not as advertisements or sponsored content, but as reader content for *Police* magazine [99] and *Military Times* [100]. The copied text includes links to the Glock website. How frequently are gun advertisements appearing as news articles? What are the perceptions about the legitimacy of this media? Who are the audiences for these magazines? Future research could help shed light on these questions.

The informational material on the US Army Picatinny Arsenal website, which engages in small arms acquisition for the armed forces, offers an interesting contrast to the way firearm advertisers characterize guns. The website describes the purpose of the arsenal as providing nearly 90% of the Army’s *lethality* [101]. Lethality seems a more suitable term for describing the purpose of firearms. In analyzing firearm advertising, we observed numerous instances of sharing of information from the armed services and promotional videos highlighting the bravery of specific servicemen. Future work could explore social media use by the US armed services and law enforcement to understand how these agencies respond to or incorporate advertising from firearm manufacturers into their information sharing with the communities that follow them.

Acknowledgments

LJ contributed to the text, data collection, classification, and visualization. CD designed and classified most of the YouTube manufacturer database. JK designed and classified most of the YouTube influencer database. All authors contributed to data collection of all datasets, text, and measures used in visualization, providing verification and validation. The authors would like to acknowledge the Drew University Digital Humanities initiative, supported by a grant from Andrew W Mellon Foundation. Codirectors Wendy Kolmar and Lee Arnold supported our research financially and intellectually through the Digital Humanities Summer Institute (DHSI). We also received helpful feedback from the student and faculty research teams in DHSI.

Conflicts of Interest

None declared.

References

1. Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF). Washington, DC: US Department of Justice; 2018. Firearms Commerce in the United States: Annual Statistical Update 2018 URL: https://www.atf.gov/file/130436/download [accessed 2020-01-27]
2. Diaz T. Making a Killing: The Business of Guns in America. New York: New Press; 1999.
3. Hemenway D. Private Guns, Public Health. Ann Arbor: University of Michigan Press; 2004.
4. The Economist. 2019 May 23. Law and ordnance: Guns from the United States are flooding Latin America URL: https://www.economist.com/the-americas/2019/05/23/guns-from-the-united-states-are-flooding-latin-america [accessed 2020-01-27]
5. McDougal T, Shirk DA, Muggah R, Patterson JH. Justice in Mexico. San Diego, CA: University of San Diego Trans-Border Institute; 2013 Mar. The Way of the Gun: Estimating Firearms Traffic across the US-Mexico Border URL: https://justiceinmexico.org/wp-content/uploads/2014/09/2013_WayoftheGun.pdf [accessed 2020-01-27]
6. McDougal T, Shirk D, Muggah R, Patterson J. The way of the gun: estimating firearms trafficking across the US–Mexico border. J Econ Geogr 2015;15(2):297-327. [doi: 10.1093/jeg/lbu021]
7. Zamudio NP, Vargas LG, Lindsay-Poland J, Téllez MM. Stop US Arms to Mexico. San Francisco, CA: Mexican Commission for the Defense and Promotion of Human Rights; 2018 Aug. Gross Human Rights Abuses: The Legal and Illegal Gun Trade to Mexico URL: https://www.stopusarmstomexico.org/wp-content/uploads/2018/08/THE-LEGAL-AND-ILLEGAL-GUN-TRADE-TO-MEXICO_August2018.pdf [accessed 2020-01-27]
8. Agrawal D. Stop US Arms to Mexico. Berkeley, CA: The University of California; 2019. Combating US Gun Trafficking to Mexico URL: https://stopusarmstomexico.org/wp-content/uploads/2019/10/U.S.-Guns-to-Mexico-Final.pdf [accessed 2020-01-29]
9. Runyan CW. Using the Haddon matrix: introducing the third dimension. Inj Prev 1998 Dec;4(4):302-307 [FREE Full text] [doi: 10.1136/ip.4.4.302] [Medline: 9887425]
10. Saylor EA, Vittes KA, Sorenson SB. Firearm advertising: product depiction in consumer gun magazines. Eval Rev 2004 Oct;28(5):420-433. [doi: 10.1177/0193841X04267389] [Medline: 15358905]
11. Vernick JS, Teret SP, Webster DW. Regulating firearm advertisements that promise home protection. A public health intervention. J Am Med Assoc 1997 May 7;277(17):1391-1397. [Medline: 9134946]
12. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar 14;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
13. Kostygina G, Tran H, Shi Y, Kim Y, Emery S. 'Sweeter Than a Swisher': amount and themes of little cigar and cigarillo content on Twitter. Tob Control 2016 Oct;25(Suppl 1):i75-i82 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053094] [Medline: 27697951]
14. Diaz T. The Last Gun: How Changes in the Gun Industry Are Killing Americans and What It Will Take to Stop It. New York: The New Press; 2013.
15. Karp A. Small Arms Survey. Geneva: nbmedia; 2018 Jun. Estimating Global Civilian-Held Firearms Numbers URL: http://www.smallarmssurvey.org/fileadmin/docs/T-Briefing-Papers/SAS-BP-Civilian-Firearms-Numbers.pdf [accessed 2020-01-27]
16. Smith TW, Son J. NORC at the University of Chicago. Chicago, IL: NORC at the University of Chicago; 2015 Mar. Trends in Gun Ownership in the United States, 1972-2014 URL: http://www.norc.org/PDFs/GSS%20Reports/GSS_Trends%20in%20Gun%20Ownership_US_1972-2014.pdf [accessed 2020-01-27]
17. Azrael D, Hepburn L, Hemenway D, Miller M. The stock and flow of US Firearms: results from The 2015 National Firearms Survey. RSF 2017;3(5):38-57. [doi: 10.7758/rsf.2017.3.5.02]
18. Smith TW, Marsden PV, Hout M, Kim J. icpsr - University of Michigan. 2016. General Social Survey, 1972-2014 URL: https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36319 [accessed 2020-01-27]
19. Igielnik R, Brown A. Pew Research Center. 2017 Jun 22. Key Takeaways on Americans' Views of Guns and Gun Ownership URL: https://www.pewresearch.org/fact-tank/2017/06/22/key-takeaways-on-americans-views-of-guns-and-gun-ownership/ [accessed 2020-01-27]

20. Parker K, Horowitz JM, Igielnik R, Oliphant JB, Brown A. Pew Research Center.: Pew Research Center; 2017 Jun 22. America’s Complex Relationship With Guns: An in-depth Look at the Attitudes and Experiences of US Adults URL: https://www.pewsocialtrends.org/2017/06/22/americas-complex-relationship-with-guns/ [accessed 2020-01-27]
21. Joslyn M, Haider-Markel DP. The Washington Post. 2018 May 7. Americans Vastly Overestimate the Number of Gun Owners URL: https://www.washingtonpost.com/news/monkey-cage/wp/2018/05/07/americans-vastly-overestimate-the-number-of-gun-owners-thats-a-problem/ [accessed 2020-01-27]
22. Gramlich J, Schaeffer K. Pew Research Center. 2019 Oct 2. 7 Facts About Guns in the US URL: https://www.pewresearch.org/fact-tank/2018/12/27/facts-about-guns-in-united-states/ [accessed 2020-01-27]
23. Centers for Disease Control and Prevention WONDER. 2018. Underlying Cause of Death, 1999-2017 URL: http://wonder.cdc.gov/ucd-icd10.html [accessed 2020-01-27]
24. Cunningham RM, Walton MA, Carter PM. The major causes of death in children and adolescents in the United States. N Engl J Med 2018 Dec 20;379(25):2468-2475 [FREE Full text] [doi: 10.1056/NEJMsr1804754] [Medline: 30575483]
25. Prickett KC, Gutierrez C, Deb S. Family firearm ownership and firearm-related mortality among young children: 1976-2016. Pediatrics 2019 Feb;143(2):pii: e20181171. [doi: 10.1542/peds.2018-1171] [Medline: 30835244]
26. Holland KM, Hall JE, Wang J, Gaylor EM, Johnson LL, Shelby D, School-Associated Violent Deaths Study Group. Characteristics of school-associated youth homicides - United States, 1994-2018. MMWR Morb Mortal Wkly Rep 2019 Jan 25;68(3):53-60 [FREE Full text] [doi: 10.15585/mmwr.mm6803a1] [Medline: 30677007]
27. Miller M, Azrael D, Hemenway D. The epidemiology of case fatality rates for suicide in the northeast. Ann Emerg Med 2004 Jun;43(6):723-730. [doi: 10.1016/S0196064404000691] [Medline: 15159703]
28. National Violent Death Reporting System (NVDRS). Atlanta, GA: Safe States Alliance; 2018. Deaths From Suicide Among US Veterans & Armed Forces in 16 States. A Special Report With Data From the National Violent Death Reporting System, 2010-2014 URL: https://cdn.ymaws.com/www.safestates.org/resource/resmgr/nvdrs/Suicide_Report_2018/FINAL_REPORT_-_NVDRS_Special.pdf [accessed 2020-01-29]
29. Sisak M, Mustian J. PBS News Hour. 2019 Aug 15. Police Departments Confront 'epidemic' in Officer Suicides URL: https://www.pbs.org/newshour/nation/police-departments-confront-epidemic-in-officer-suicides [accessed 2020-01-27]
30. Lubin G, Werbeloff N, Halperin D, Shmushkevitch M, Weiser M, Knobler HY. Decrease in suicide rates after a change of policy reducing access to firearms in adolescents: a naturalistic epidemiological study. Suicide Life Threat Behav 2010 Oct;40(5):421-424. [doi: 10.1521/suli.2010.40.5.421] [Medline: 21034205]
31. Knopov A, Sherman RJ, Raifman JR, Larson E, Siegel MB. Household gun ownership and youth suicide rates at the state level, 2005-2015. Am J Prev Med 2019 Mar;56(3):335-342. [doi: 10.1016/j.amepre.2018.10.027] [Medline: 30661885]
32. Sivaraman JJ, Ranapurwala SI, Moracco KE, Marshall SW. Association of state firearm legislation with female intimate partner homicide. Am J Prev Med 2019 Jan;56(1):125-133. [doi: 10.1016/j.amepre.2018.09.007] [Medline: 30573141]
33. Reeping PM, Cerdá M, Kalesan B, Wiebe DJ, Galea S, Branas CC. State gun laws, gun ownership, and mass shootings in the US: cross sectional time series. Br Med J 2019 Mar 6;364:l542 [FREE Full text] [doi: 10.1136/bmj.l542] [Medline: 30842105]
34. AP-NORC. 2019. Gun Laws and Public Safety URL: http://www.apnorc.org/projects/Pages/Gun-Laws-and-Public-Safety-.aspx [accessed 2020-01-27]
35. Runyan CW, Brooks-Russell A, Brandspigel S, Betz M, Tung G, Novins D, et al. Law enforcement and gun retailers as partners for safely storing guns to prevent suicide: A study in 8 Mountain West States. Am J Public Health 2017 Nov;107(11):1789-1794. [doi: 10.2105/AJPH.2017.304013] [Medline: 28933926]
36. Crifasi CK, Doucette ML, McGinty EE, Webster DW, Barry CL. Storage practices of US gun owners in 2016. Am J Public Health 2018 Apr;108(4):532-537. [doi: 10.2105/AJPH.2017.304262] [Medline: 29470124]
37. Freudenberg N. Public health advocacy to change corporate practices: implications for health education practice and research. Health Educ Behav 2005 Jun;32(3):298-319; discussion 355. [doi: 10.1177/1090198105275044] [Medline: 15851541]
38. Swanson JW, Sampson NA, Petukhova MV, Zaslavsky AM, Appelbaum PS, Swartz MS, et al. Guns, impulsive angry behavior, and mental disorders: Results from the National Comorbidity Survey Replication (NCS-R). Behav Sci Law 2015 Jun;33(2-3):199-212 [FREE Full text] [doi: 10.1002/bsl.2172] [Medline: 25850688]
39. Contrera J. USA Today. 2013 Mar 23. Comcast Bans Gun, Ammo Ads URL: https://www.usatoday.com/story/money/business/2013/03/22/comcast-gun-ads/2009717/ [accessed 2020-01-27]
40. Webster DW, Vernick JS. Reducing Gun Violence in America: Informing Policy with Evidence and Analysis. Baltimore, MD: Johns Hopkins University Press; 2013.
41. Claflin J. Garrison Everest. 2019 Apr 18. 5 Firearm Marketing Alternatives You Haven’t Thought Of Yet URL: https://www.garrisoneverest.com/firearms-and-hunting/5-firearm-marketing-alternatives-you-havent-thought-of-yet/ [accessed 2020-01-27]
42. Pew Research Center. 2019. Social Media Fact Sheet URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2020-01-27]
43. Perrin A, Anderson M. Pew Research Center. 2019 Apr 10. Share of US Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018 URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2020-01-27]

44. YouTube. YouTube for Press URL: https://www.youtube.com/intl/en-GB/yt/about/press/ [accessed 2020-01-27]
45. Smith A, Toor S, van Kessel P. Pew Research Center. 2018 Nov 7. Many Turn to YouTube for Children's Content, News, How-To Lessons URL: https://www.pewinternet.org/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/ [accessed 2020-01-27]
46. Bromberg JE, Augustson EM, Backinger CL. Portrayal of smokeless tobacco in YouTube videos. Nicotine Tob Res 2012 Apr;14(4):455-462 [FREE Full text] [doi: 10.1093/ntr/ntr235] [Medline: 22080585]
47. Hébert ET, Case KR, Kelder SH, Delk J, Perry CL, Harrell MB. Exposure and engagement with tobacco- and e-cigarette-related social media. J Adolesc Health 2017 Sep;61(3):371-377 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.04.003] [Medline: 28669801]
48. Huang J, Kornfield R, Emery SL. 100 million views of electronic cigarette YouTube videos and counting: quantification, content evaluation, and engagement levels of videos. J Med Internet Res 2016 Mar 18;18(3):e67 [FREE Full text] [doi: 10.2196/jmir.4265] [Medline: 26993213]
49. Kim K, Paek HJ, Lynn J. A content analysis of smoking fetish videos on YouTube: regulatory implications for tobacco control. Health Commun 2010 Mar;25(2):97-106. [doi: 10.1080/10410230903544415] [Medline: 20390676]
50. Liang Y, Zheng X, Zeng DD, Zhou X, Leischow SJ, Chung W. Exploring how the tobacco industry presents and promotes itself in social media. J Med Internet Res 2015 Jan 21;17(1):e24 [FREE Full text] [doi: 10.2196/jmir.3665] [Medline: 25608524]
51. Seidenberg AB, Rodgers EJ, Rees VW, Connolly GN. Youth access, creation, and content of smokeless tobacco ('dip') videos in social media. J Adolesc Health 2012 Apr;50(4):334-338. [doi: 10.1016/j.jadohealth.2011.09.003] [Medline: 22443835]
52. Freeman B, Chapman S. Is 'YouTube' telling or selling you something? Tobacco content on the YouTube video-sharing website. Tob Control 2007 Jun;16(3):207-210 [FREE Full text] [doi: 10.1136/tc.2007.020024] [Medline: 17565142]
53. Schwemmer C, Ziewiecki S. Social media sellout: the increasing role of product promotion on YouTube. Soc Media Soc 2018;4(3):205630511878672. [doi: 10.1177/2056305118786720]
54. Depue JB, Southwell BG, Betzner AE, Walsh BM. Encoded exposure to tobacco use in social media predicts subsequent smoking behavior. Am J Health Promot 2015;29(4):259-261. [doi: 10.4278/ajhp.130214-ARB-69] [Medline: 24670071]
55. Soneji S, Yang J, Knutzen KE, Moran MB, Tan AS, Sargent J, et al. Online Tobacco Marketing and Subsequent Tobacco Use. Pediatrics 2018 Feb;141(2):pii: e20172927 [FREE Full text] [doi: 10.1542/peds.2017-2927] [Medline: 29295893]
56. Wellman RJ, Sugarman DB, DiFranza JR, Winickoff JP. The extent to which tobacco marketing and tobacco use in films contribute to children's use of tobacco: a meta-analysis. Arch Pediatr Adolesc Med 2006 Dec;160(12):1285-1296. [doi: 10.1001/archpedi.160.12.1285] [Medline: 17146027]
57. Bergkvist L, Zhou KQ. Celebrity endorsements: a literature review and research agenda. Int J Advert 2016;35(4):642-663. [doi: 10.1080/02650487.2015.1137537]
58. Chu KH, Majmundar A, Allem JP, Soto DW, Cruz TB, Unger JB. Tobacco use behaviors, attitudes, and demographic characteristics of tobacco opinion leaders and their followers: Twitter analysis. J Med Internet Res 2019 Jun 4;21(6):e12676 [FREE Full text] [doi: 10.2196/12676] [Medline: 31165716]
59. Coates AE, Hardman CA, Halford JC, Christiansen P, Boyland EJ. Social media influencer marketing and children's food intake: a randomized trial. Pediatrics 2019 Apr;143(4):pii: e20182554. [doi: 10.1542/peds.2018-2554] [Medline: 30833297]
60. Calder BJ, Malthouse EC, Schaedel U. An experimental study of the relationship between online engagement and advertising effectiveness. J Interact Mark 2009;23(4):321-331. [doi: 10.1016/j.intmar.2009.07.002]
61. Kellermann AL. Comment: Gunsmoke--changing public attitudes toward smoking and firearms. Am J Public Health 1997 Jun;87(6):910-913. [doi: 10.2105/ajph.87.6.910] [Medline: 9224166]
62. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. Stat Med 2013 Feb 20;32(4):556-577 [FREE Full text] [doi: 10.1002/sim.5408] [Medline: 22711416]
63. Screenagers Movie. URL: https://www.screenagersmovie.com/ [accessed 2020-01-27]
64. Patton DU, Patel S, Hong JS, Ranney ML, Crandall M, Dungy L. Tweets, gangs, and guns: a snapshot of gang communications in Detroit. Violence Vict 2017 Oct 1;32(5):919-934 [FREE Full text] [doi: 10.1891/0886-6708.VV-D-16-00040] [Medline: 28810937]
65. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
66. Chokshi N. New York Times. 2018 Mar 22. YouTube to Ban Videos Promoting Gun Sales URL: https://www.nytimes.com/2018/03/22/business/youtube-gun-ban.html [accessed 2020-01-27]
67. Heater B. TechCrunch. 2017 Oct 10. YouTube Bans Bump Stock Instructional Videos Following Las Vegas Shooting URL: http://social.techcrunch.com/2017/10/09/youtube-bans-bump-stock-instructional-videos-following-las-vegas-shooting/ [accessed 2020-01-27]
68. YouTube. Policies on Content Featuring Firearms URL: https://support.google.com/youtube/answer/7667605?hl=en [accessed 2020-01-27]

69. Twitter. Weapons and Weapon Accessories URL: https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/weapons-and-weapon-accessories.html [accessed 2020-01-27]
70. The Economist. 2019 May 25. Kids Aren't Alright: How to Make Social Media Safe for Children URL: https://www.economist.com/business/2019/05/25/how-to-make-social-media-safe-for-children [accessed 2020-01-27]
71. Gillespie T. Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. New Haven: Yale UP; 2018.
72. Roberts ST. Behind the Screen: Content Moderation in the Shadows of Social Media. New Haven: Yale University Press; 2019.
73. Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF). Washington, DC: Bureau of Alcohol, Tobacco, Firearms and Explosives; 2017. Annual Firearms Manufacturing and Export Report URL: https://www.atf.gov/file/133476/download [accessed 2020-01-29]
74. Mosendz P, Barrett P, Rojanasakul M. Bloomberg. 2018 Nov 27. How Foreign Guns Invaded the US URL: https://www.bloomberg.com/graphics/2018-us-gun-imports/ [accessed 2020-01-27]
75. Twitonomy. 2019. URL: http://twitonomy.com/ [accessed 2020-01-27]
76. Danger Close Media Group. 2019 Jan 12. Most Influential Gun Industry Influencers 2019 URL: http://dangerclosegroup.com/influential-gun-industry-influencers-2019/ [accessed 2020-01-27]
77. Schiller D. Houston Chronicle. 2015 Oct 28. Texas Request to Buy Guns Tops 1 Million for Year URL: https://blog.chron.com/narcoconfidential/2015/10/texas-request-to-buy-guns-tops-1-million-for-year/ [accessed 2020-01-27]
78. Palmetto State Armory. YouTube. 2017 Jan 20. How to Build your PSAK-47 URL: https://www.youtube.com/watch?v=livBrRdpgdk [accessed 2020-01-27]
79. Voie Ø, Borander A, Sikkeland LI, Grahnstedt S, Johnsen A, Danielsen TE, et al. Health effects after firing small arms comparing leaded and unleaded ammunition. Inhal Toxicol 2014 Dec;26(14):873-879. [doi: 10.3109/08958378.2014.970783] [Medline: 25373830]
80. Borander AK, Voie Ø, Longva K, Danielsen TE, Grahnstedt S, Sandvik L, et al. Military small arms fire in association with acute decrements in lung function. Occup Environ Med 2017 Sep;74(9):639-644. [doi: 10.1136/oemed-2016-104207] [Medline: 28408655]
81. Barna M. The Nation's Health. 2019 Apr. APHA Advocates: Gun violence research funding needed URL: http://thenationshealth.aphapublications.org/content/49/2/2.2 [accessed 2020-01-27]
82. Greene J. The Washington Post. 2019 May 30. Tech Giant Brings Software to a Gun Fight URL: https://www.washingtonpost.com/technology/2019/05/30/tech-giant-brings-software-gun-fight/ [accessed 2020-01-27]
83. Mascia J. The Trace. 2019 Jun 06. Meet the Ad Execs Who Want to Put Graphic Warnings on Ammunition URL: https://www.thetrace.org/2019/06/meet-the-ad-execs-who-want-to-put-graphic-warnings-on-ammunition/ [accessed 2020-01-27]
84. Moms Demand Action for Gun Sense in America. YouTube. 2015 Nov 3. Moms Demand Action 101 URL: https://www.youtube.com/watch?v=dgF8Nlj4maM&feature=youtu.be [accessed 2020-01-27]
85. We are Everytown for Gun Safety. 2020. Who we are URL: https://everytown.org/who-we-are/ [accessed 2020-01-27]
86. The Trace. 2020. Reporting on Guns and Gun Violence in America URL: https://www.thetrace.org/ [accessed 2020-01-27]
87. The Coalition to Stop Gun Violence (CSGV). URL: https://www.csgv.org/ [accessed 2020-01-27]
88. Schiavo R. Health Communication: From Theory to Practice. New York: John Wiley & Sons; 2013.
89. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013 Aug 19;15(8):e177 [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]
90. Forsyth SR, Malone RE. 'I'll be your cigarette--light me up and get on with it': Examining Smoking Imagery on YouTube. Nicotine Tob Res 2010 Aug;12(8):810-816 [FREE Full text] [doi: 10.1093/ntr/ntq101] [Medline: 20634267]
91. EUR-Lex. 2016. Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016PC0593 [accessed 2020-01-27]
92. Office of the US Surgeon General. 2018. Surgeon General's Advisory on E-cigarette Use Among Youth URL: https://e-cigarettes.surgeongeneral.gov/documents/surgeon-generals-advisory-on-e-cigarette-use-among-youth-2018.pdf [accessed 2020-01-27]
93. Bohme SR, Zorabedian J, Egilman DS. Maximizing profit and endangering health: corporate strategies to avoid litigation and regulation. Int J Occup Environ Health 2005;11(4):338-348. [doi: 10.1179/oeh.2005.11.4.338] [Medline: 16350467]
94. Dorfman L, Wallack L, Woodruff K. More than a message: framing public health advocacy to change corporate practices. Health Educ Behav 2005 Jun;32(3):320-36; discussion 355. [doi: 10.1177/1090198105275046] [Medline: 15851542]
95. Nathanson CA. Social movements as catalysts for policy change: the case of smoking and guns. J Health Polit Policy Law 1999 Jun;24(3):421-488. [doi: 10.1215/03616878-24-3-421] [Medline: 10386324]
96. PBS News Hour. 2019 Feb 12. The Shocking Violence That Characterized the Reign of 'El Chapo' URL: https://www.pbs.org/newshour/show/the-shocking-violence-that-characterized-the-reign-of-el-chapo [accessed 2020-01-27]
97. Sustainable Development Knowledge Platform. Sustainable Development Goal 16 URL: https://sustainabledevelopment.un.org/sdg16 [accessed 2020-01-27]

98. Executive Services Directorate. Washington, DC: Office of the Assistant to the Secretary of Defense for Public Affairs; 2016. DoD Instruction 5410.20. Public Affairs Relations with For-Profit Businesses and Business Industry Organizations URL: https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/541020p.pdf
99. Police: The Law Enforcement Magazine. 2019 Apr 8. Glock Signs Chuck Norris as Spokesperson URL: http://www.policemag.com/509573/glock-signs-chuck-norris-as-spokesperson [accessed 2020-01-27]
100. Military Times. Chuck Norris loves Glock — so you better love Glock URL: https://www.militarytimes.com/off-duty/gearscout/2019/04/09/chuck-norris-loves-glock-so-you-better-love-glock/ [accessed 2020-01-27]
101. U.S. Army. 2019. Picatinny Arsenal - The Joint Center of Excellence for Guns and Ammunition URL: https://web.archive.org/web/20200319201030/https://www.pica.army.mil/Picatinny/about/?AspxAutoDetectCookieSupport=1 [accessed 2020-01-27]

Abbreviations

ATF: Bureau of Alcohol, Firearms, Tobacco, and Explosives
DCM: Danger Close Media
DHSI: Digital Humanities Summer Institute
NRA: National Rifle Association
PLCAA: Protection of Lawful Commerce in Arms Act

Edited by G Eysenbach; submitted 01.08.19; peer-reviewed by D Capurro, E Da Silva; comments to author 20.09.19; revised version received 15.11.19; accepted 13.01.20; published 27.03.20.

Please cite as:
Jordan L, Kalin J, Dabrowski C
Characteristics of Gun Advertisements on Social Media: Systematic Search and Content Analysis of Twitter and YouTube Posts
J Med Internet Res 2020;22(3):e15736
URL: http://www.jmir.org/2020/3/e15736/
doi:10.2196/15736
PMID:32217496

Original Paper

Medical Help-Seeking Strategies for Perinatal Women With Obstetric and Mental Health Problems and Changes in Medical Decision Making Based on Online Health Information: Path Analysis

Kyungmi Chung[1,2], PhD; Hee Young Cho[3], MD; Young Ran Kim[3], MD, PhD; Kyungun Jhung[4], MD, PhD; Hwa Seon Koo[5], MD, PhD; Jin Young Park[1,2], MD, PhD

[1]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
[2]Department of Psychiatry, Yonsei University College of Medicine, Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea
[3]Department of Obstetrics and Gynecology, CHA Bundang Medical Center, CHA University School of Medicine, Seongnam, Republic of Korea
[4]Department of Psychiatry and Behavioral Neuroscience, International St. Mary's Hospital, Catholic Kwandong University, Incheon, Republic of Korea
[5]Fertility Center, CHA Bundang Medical Center, CHA University School of Medicine, Seongnam, Republic of Korea

Corresponding Author:
Jin Young Park, MD, PhD
Department of Psychiatry
Yonsei University College of Medicine
Severance Hospital, Yonsei University Health System
50-1 Yonsei-ro, Seodaemun-gu
Seoul,
Republic of Korea
Phone: 82 2 2228 0972
Fax: 82 2 2227 6558
Email: empathy@yuhs.ac

Abstract

Background: Previous studies have revealed that most pregnant women rarely discuss informal information found on the internet with health professionals and have frequently expressed concerns for medical experts' reactions to the online information they shared, as well as the lack of time to consult the medical experts in general. To date, little information is available on the effect of individual differences in utilizing medical help-seeking strategies on their medical decisions during the perinatal period.

Objective: The objectives of this study were (1) to determine associations among perinatal women's medical help-seeking strategies, changes in medical decision making, and online health information utilization with a focus on the mediating effect of self-efficacy in perinatal health literacy on the intent to consult health professionals, and (2) to clarify these associations in perinatal women with two different medical problems: obstetric and mental health.

Methods: A total of 164 perinatal women aged 24 to 47 years (mean 34.64, SD 3.80) repeatedly completed the Problem Solving in Medicine and Online Health Information Utilization questionnaires to examine the moderating effect of two types of medical problems on their decision-making processes. To validate the hypothesized relationships in the proposed conceptual model encompassing obstetric and mental health problem-solving models, path analyses were performed.

Results: This study found that some perinatal women, who use an online informal medical help-seeking (OIMH) strategy, would be more likely to change their medical decisions based only on internet-based information without consulting health professionals ($P<.001$), compared to other women using different medical help-seeking strategies. Particularly, this concern is significantly prevalent when encountering obstetric problems during the perinatal period (obstetric problem-solving: $P<.001$; mental health problem-solving: $P=.02$). Furthermore, perinatal women with mental health issues using the OIMH strategy showed a significant difference in intent to consult health professionals based on online health information when the medical problem they had to solve was different (obstetric problem-solving: $P=.94$; mental health problem-solving: $P=.003$).

Conclusions: Despite the positive mediating effects of perinatal women's enhanced health literacy on the intent to discuss personal medical issues with health professionals based on online health information, the strategy used is of fundamental importance for understanding their help-seeking and decision-making processes during the perinatal period. Beyond a short consultation to steer patients quickly and authoritatively towards an obstetric doctor's choice of action, it is recommended in this study that obstetricians attempt to provide their patients with needed context for the information found online. To fully explain this information with an open mind, they should actively develop or support information and communications technology (ICT)-based health information services.

(J Med Internet Res 2020;22(3):e14095) doi:10.2196/14095

KEYWORDS

perinatal care; obstetrics; mental health; information seeking behavior; help-seeking behavior; self efficacy; health literacy; consultation; decision making; internet

Introduction

Background

Medical help-seeking can be defined as the ability to actively seek help from others to cope with medical problems or painful experiences [1-3]. The help-seeking process is a transition from the personal domain (being aware of personal needs, thoughts, and feelings) to the interpersonal domain (being willing to share and disclose one's needs to others) [2,3]. Internet health-seeking complements rather than replaces in-person health-seeking [4,5]. Particularly, online health information has been used to fill an informational void that promotes self-efficacy and affects medical decisions and actions about how to treat an illness or other medical conditions; this information can be discussed with friends, family members, and health care providers [6]. Physical and mental health professionals have serious concerns about the quality of online health information [6], and women in the perinatal period from conception through pregnancy to childbirth and the postpartum stage become more overwhelmed with false or misleading health information from multiple online sources, which in turn causes them to feel frightened and anxious rather than informed and empowered [7-11] and vise versa [12]. Even though internet searching allows pregnant women with the need for empirical medical information specific to their own situations to readily access experienced women's birth stories on social media and provides social support to anxious women at the same time [13], women with a less experienced or robust capacity to make sense of conflicting health information from informal online sources may challenge their pre-existing ideations and aspirations of birth processes and address issues that they have not yet considered [11,14]. Hence, the provision of appropriate online health information can be considered an important step toward patient education and empowerment [6,15].

When it comes to dealing with difficulty assessing the accuracy, reliability, and credibility of relevant medical information on the internet, the evidenced-based information provided by health professionals will play a critical role in guiding perinatal women's medical decisions on the right course of treatment [16]. For prenatal women, the association between trust in health providers and medical information sources should be carefully considered. Unlike many chronic disease patients who maintain the primacy of information acquisition from their doctors [17], mothers with lower levels of trust in health providers use more informal sources such as friends, family, other parents, the internet, and alternative medicine providers and do not regard their doctors as the main source of medical information [18]. Furthermore, most pregnant women have rarely discussed informal information found on the internet with health professionals [19], having frequently expressed concerns on medical experts' negative reactions to the online information they brought to the outpatient clinic [20] as well as on the lack of time to consult them [21,22]. Without either perceiving self-efficacy in perinatal health literacy or contacting health professionals as the last bastion of medical decision based on online health information, women who solve medical problems by online informal consultation or search results can be more likely to make the wrong medical decisions, fail to receive appropriate medical treatments, or be confronted with new decision-making situations [23]. In this regard, this study focused on the association between perinatal women's intentions of online informal medical help-seeking for health problems and their medical decision-making processes based on health information from the internet.

To address this issue, individual differences in using medical help-seeking strategies should be considered because the majority of pregnant women (83.7%) tend to use multiple information sources whose conflicting information increases anxiety and leads those with no chronic diseases to decide not to use medicines more than those with chronic diseases [24]. Based on the Chen and colleagues' research model [25], this study categorizes perinatal women into four types of health information seekers with different medical help-seeking strategies, considering the formality (formal: querying doctors, medical experts, and health care professionals vs informal: querying family members, friends or relatives, and experienced people with previous medical problem-solving experiences) and channel (online: email, internet medical forum, instant messenger, and websites or online blogs vs nononline: face-to-face visit and communication) of the information source: (1) nononline formal medical help-seeking (NFMH), (2) online formal medical help-seeking (OFMH), (3) nononline informal medical help-seeking (NIMH), and (4) online informal medical help-seeking (OIMH). Further, when encountering two different obstetric and mental health problems, whether perinatal women with the four distinct medical help-seeking strategies will make medical decisions and change their decisions in a different manner, particularly based on online health information, has yet to be determined by comparing the proposed research model within the same Korean sample.

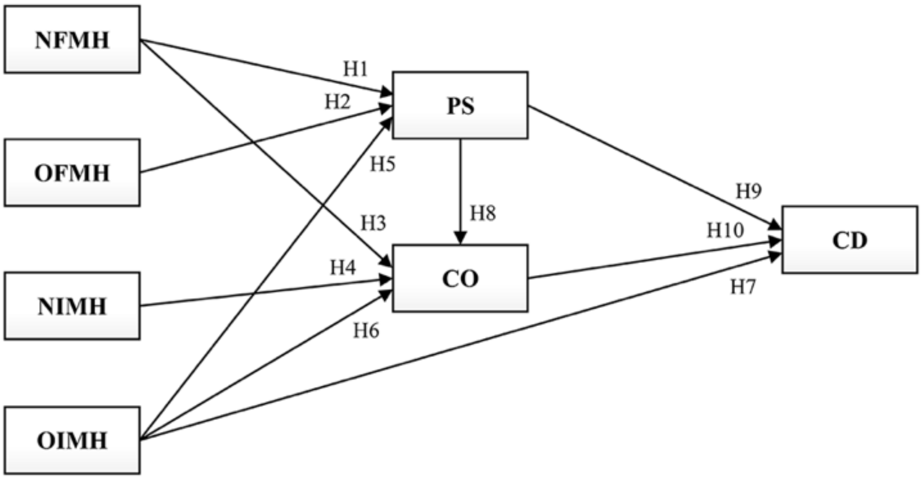
Objectives

Given the potential impact of the information on a patient's medical decision-making process that will lead to a change in their health-seeking behaviors, the personal, cognitive, and social skills that determine an individual's motivation and ability to obtain, evaluate, and use health care information is of great importance for making appropriate medical decisions and promoting and maintaining good health [26]. The objective of this study was to develop and validate a medical problem-solving model to (1) determine associations among perinatal women's medical help-seeking strategies, changes in medical decision making, and online health information use with a focus on the mediating effect of self-efficacy in perinatal health literacy on the intent to consult health professionals and (2) clarify these associations in perinatal women with two different medical problems (obstetric and mental health).

Research Model and Hypotheses

As shown in Figure 1, this study proposes a medical problem-solving model with a 2-step mediation approach that includes promoting self-efficacy in perinatal health literacy (PS) and consulting others (CO), both based on online health information perinatal women obtained, as potential mediators of the effects of perinatal women's medical help-seeking strategies on changes in their medical decisions based on online health information (CD). Furthermore, the moderating role of two types of medical problems (obstetric vs mental health) will be further examined in this conceptual research model. According to the model, the following hypotheses are stated in the literature review.

Figure 1. Medical problem-solving model. NFMH: nononline formal medical help-seeking; NIMH: nononline informal medical help-seeking; OFMH: online formal medical help-seeking; OIMH: online informal medical help-seeking; PS: promoting self-efficacy in perinatal health literacy based on online health information; CO: consulting others (health professionals) based on online health information; CD: changing their medical decisions based on online health information.



In terms of the formality of medical help-seeking strategies, patients who prefer to seek medical help from formal sources have good health literacy [27]. Particularly, patients seeking online formal medical help achieve better health literacy [28] than those seeking nononline formal medical help in that the OFMH strategy is beneficial to compare, cross-check, and evaluate health-related information from multiple reliable information sources by looking for consistent results. However, as far as patients' trust in their doctors is concerned, perinatal women with high trust levels will promote self-efficacy in perinatal health literacy based on their doctors' guidelines, while those with low trust levels will do so based on online health information because they do not regard their doctors as the main source of medical information [18]. Thus, we posit the following hypotheses:

- H1/H2: Perinatal women's strategies of (1) NFMH and (2) OFMH will be positively associated with PS.

Additionally, a recent study of patient samples, not specifically consisting of perinatal women, revealed that patients employing an NFMH strategy when facing medical problems had a significantly positive association with consulting others about relevant issues based on the online health information they obtained [25]. In the perinatal period, women who have maintained a long-term doctor-patient relationship through rapport building can be motivated to directly address their medical concerns sourced from the internet in order to reduce anxiety before and after a visit to the health professional [19,20,29-34]. As a facilitating factor in the decision to communicate with health professionals about difficulties in the perinatal period [35], a woman's social network (eg, partner, friends, relatives, family members, and experienced mothers) will influence the help-seeking process by encouraging professional help-seeking, participating in the decision-making process, and initiating help-seeking [36]. Accordingly, the following hypotheses are proposed:

- H3/H4: Perinatal women's strategies of (3) NFMH and (4) NIMH will be positively associated with CO.

Among the various benefits of using the internet, providing perinatal women with personal and emotional support from other women in similar situations is of unquestionable importance [13], since they are vulnerable to undergoing unknown, unpredictable events. In fact, finding people with the same medical problems is not easy online, but it will be helpful to share their concerns and doubts with other women and take positive energy from one another in online forums and social media platforms [37], thereby reducing anxiety and increasing confidence in making medical decisions for themselves and their babies after such internet use [20,21]. In addition, the internet can help expert users identify more questions they wanted to ask a health professional, be more involved in the decision-making process, make better decisions, and feel they had more control over their decisions than novice users [21]. Hence, we postulate the following hypotheses:

- H5: The perinatal women's strategy of OIMH will be positively associated with PS.
- H6: The perinatal women's strategy of OIMH will be positively associated with CO.
- H7: The perinatal women's strategy of OIMH will be positively associated with CD.

According to a continuum of health literacy including basic or functional health literacy, communicative/interactive health literacy, and critical health literacy, there is a need to focus more on developing the skills and confidence to make well-informed health choices than on limiting the transmission of information [38]. The internet can empower people, including pregnant women, to communicate with their health professionals and play a critical role in making decisions [21]. However, the association between promoting self-efficacy in perinatal health literacy and changing perinatal women's medical decisions might interfere with their anxiety levels. First-time pregnant women show a significant negative association between the level of maternal anxiety and feelings of control during labor [39]. As fears can be caused by negative moods and stories told by others about alarming information, diseases, and child-related problems in primiparas as well as negative experiences of a previous pregnancy, childbirth, and a baby's health and care in multiparas, it will not be easy to change their views to align with the online health information they obtained without communicating with health professionals [40]. Therefore, we hypothesize the following:

- H8: PS will be positively associated with CO.
- H9: PS will be negatively associated with CD.
- H10: CO will be positively associated with CD.

Methods

Participant Recruitment

To test the proposed hypotheses, we administered an internet-based questionnaire built with Google Forms (Google Inc). Perinatal women aged 19 to 59 years who were being prepared for pregnancy, in fertility treatment, pregnant, or postpartum (up to 2 years after delivery) and had the ability to read, understand, and respond to the questionnaire items were eligible to be enrolled in this study. Those who failed to meet the inclusion criteria were not listed in our survey pool. In total, 168 women responded to the questionnaire for a 50.5% response rate, as the total survey pool consisted of 333 patients from the Fertility Center and/or Department of Obstetrics Gynecology, CHA Bundang Women's Hospital, South Korea. This study was approved by the institutional review board of CHA Bundang Medical Center, CHA University.

The questionnaire was delivered to the participants' mobile phones via a Google Forms link embedded in a short message service or multimedia messaging service text message or chat in KakaoTalk (Kakao Corp). Participants were allowed to access the survey link during a 2-month period from November 1, 2018, to December 31, 2018. To identify participants who met the criteria, a unique identification (ID) number was assigned to each patient with the survey link, and all patients were instructed to enter the preassigned ID number into the Google Form. Before receiving access to the online questionnaires, patients were asked to read and understand the aims of this study and relevant information, confirming that participating in the survey would be considered the equivalent of giving informed consent.

Measurement Instruments

To assess the associations between perinatal women's strategies of medical help-seeking to solve obstetric and mental health problems, medical decision making, and their online medical information use, the Problem Solving in Medicine and Online Health Information Utilization questionnaires developed by Chen and colleagues [25] were implemented twice in this study. The wording of the original questionnaire items was slightly modified to reflect the context of medical problems to be solved.

The Problem Solving in Medicine questionnaire consisted of 2 constructs with 6 observed variables measuring 27 latent variables: medical help-seeking (NFMH: 4 items, OFMH: 4 items, NIMH: 5 items, and OIMH: 5 items) and health information searching (nononline health information search [4 items] and online health information search [5 items]). The Online Health Information Utilization questionnaire included 2 constructs with 3 observed variables measuring 12 latent variables: medical decision making (PS: 4 items and CO: 4 items) and changes in medical decision making (CD: 4 items) based on online health information. All items were rated on a 7-point Likert scale ranging from 1=strongly disagree to 7=strongly agree. The reliability, validity (Table 1), and descriptive statistics (Table 2) were assessed for the measurement instruments.

Pregnancy-related anxiety was measured with a 10-item scale [41] that assessed the extent to which women felt concerned about their health, their baby's health, labor and delivery, and caring for a baby. In this study, responses were made on a 5-point Likert scale ranging from 1=not at all to 5=very much. Items on the pregnancy-related anxiety scale were as follows: (1) I am confident of having a normal childbirth (reverse-coded), (2) I think my labor and delivery will go normally (reverse-coded), (3) I have a lot of fear regarding the health of my baby, (4) I am worried that the baby could be abnormal, (5) I am afraid that I will be harmed during delivery, (6) I am worried about how the baby is growing and developing inside me, (7) I am worried about losing the baby, (8) I am worried about having a hard or difficult labor and delivery, (9) I am worried about taking care of a new baby, and (10) I am worried about developing medical problems during my pregnancy.

Table 1. The internal reliability and convergent validity of the measurement instruments.

Factor and item	Factor loading	Cronbach alpha	Composite reliability[a]	Average variance extracted
NFMH[b]				
NFMH1	0.84	.80	0.86	0.67
NFMH2	0.79	—	—	—
NFMH3	0.83	—	—	—
OFMH[c]				
OFMH1	0.84	.85	0.85	0.58
OFMH2	0.80	—	—	—
OFMH3	0.76	—	—	—
OFMH4	0.64	—	—	—
NIMH[d]				
NIMH2	0.85	.83	0.87	0.68
NIMH4	0.72	—	—	—
NIMH5	0.90	—	—	—
OIMH[e]				
OIMH1	0.71	.86	0.86	0.55
OIMH2	0.66	—	—	—
OIMH3	0.72	—	—	—
OIMH4	0.82	—	—	—
OIMH5	0.78	—	—	—
PS[f]				
PSOHI1	0.68	.86	0.89	0.67
PSOHI2	0.89	—	—	—
PSOHI3	0.87	—	—	—
PSOHI4	0.82	—	—	—
CO[g]				
COOHI1	0.59	.78	0.83	0.62
COOHI2	0.89	—	—	—
COOHI4	0.85	—	—	—
CD[h]				
CDOHI1	0.83	.93	0.94	0.80
CDOHI2	0.91	—	—	—
CDOHI3	0.93	—	—	—
CDOHI4	0.91	—	—	—

[a] Composite reliability is defined for each factor by the ratio of the total variance (squared sum of standardized factor loadings for items) to the observed variance (squared sum of standardized factor loadings for items + sum of error variances). As Cronbach alpha is sensitive to the number of items in the scale and tends to underestimate the internal consistency reliability except under restricted assumptions of equal common factor loadings and uncorrelated measurement errors, composite reliability is measured as the second check for internal consistency.

[b] NFMH: nononline formal medical help-seeking.

[c] OFMH: online formal medical help-seeking.

[d] NIMH: nononline informal medical help-seeking.

[e] OIMH: online informal medical help-seeking.

[f] PS: promoting self-efficacy in health literacy based on online health information (PSOHI).

[g] CO: consulting others based on online health information (COOHI).

[h] CD: changing medical decisions based on online health information (CDOHI).

Table 2. Descriptive statistics and discriminant validity of the measurement instruments.

Factor	Mean (SD)	NFMH[a]	OFMH[b]	NIMH[c]	OIMH[d]	PS[e]	CO[f]	CD[g]
NFMH	5.44 (1.26)	0.82[h]	—	—	—	—	—	—
OFMH	3.44 (1.56)	0.30	0.76[h]	—	—	—	—	—
NIMH	4.82 (1.41)	0.25	0.13	0.83[h]	—	—	—	—
OIMH	2.95 (1.43)	0.09	0.57	0.30	0.74[h]	—	—	—
PS	4.29 (1.11)	0.19	0.30	0.18	0.27	0.82[h]	—	—
CO	4.84 (1.07)	0.25	0.19	0.25	0.24	0.33	0.79[h]	—
CD	4.09 (1.28)	0.05	0.08	0.15	0.26	0.03	0.28	0.90[h]

[a] NFMH: nononline formal medical help-seeking.

[b] OFMH: online formal medical help-seeking.

[c] NIMH: nononline informal medical help-seeking.

[d] OIMH: online informal medical help-seeking.

[e] PS: promoting self-efficacy in health literacy based on online health information.

[f] CO: consulting others based on online health information.

[g] CD: changing medical decisions based on online health information.

[h] Diagonal elements represents the square roots of the average variance extracted.

Kaisor-Meyer-Olkin and Bartlett sphericity tests, conducted to measure the sampling adequacy, showed that the samples met the criteria for factor analysis (0.83, *P*<.001). According to the results of an exploratory factor analysis using SPSS Statistics 18.0 software (IBM Corp) with varimax rotation on these 27 items, online health information search and nononline health information search were not clearly defined by the intended items and were excluded from the data analysis. We eliminated the following 4 items with loading values less than 0.50: NFMH4 (When I have an obstetric or mental health problem, I will ask the pharmacist of a nearby pharmacy store for information), NIMH1 (When I have an obstetric or mental health problem, I will seek help from a drugstore), NIMH3 (When I have an obstetric or mental health problem, I will go to the temple praying to god for advice), and COOHI3 (I will discuss relevant issues with my family or friends based on the health information on the internet). On the basis of the results of the exploratory factor analysis, the factor CO, which is an abbreviation for consulting others including a doctor (COOHI2), family members or friends (COOHI3), and other medical experts (COOHI4), did not include the item COOHI3, indicating that perinatal women of this study sample did not consider their family and friends as an important reference when making their medical decisions.

Results

Participant Characteristics

For the data analysis, 4 out of the 168 respondents were excluded (2.4%): 2 did not enter their given IDs and 2 did not have easy access to the internet when needed. A total of 164 perinatal women were included ranging in age from 24 to 47 years, with a mean age of 34.64 (SD 3.80) years. Perinatal women showed a mild to moderate level of fear and anxiety about their own health, their baby’s health, and labor and delivery outcomes (mean 2.85 [SD 0.69]), and the majority of them were aged from 30 to 49 years (150/164, 91.5%). As presented in Table 3, the sociodemographic and clinical characteristics of this study sample were documented as well as relevant information-seeking data more in detail.

Table 3. Sociodemographic and clinical characteristics of this study sample (n=164).

Characteristics	Participants, n (%)
Gender	
Female	164 (100)
Age of perinatal women in years	
19-29	14 (8.5)
30-39	134 (81.7)
40-49	16 (9.8)
Age of their partners in years	
19-29	8 (4.9)
30-39	116 (70.7)
40-49	39 (23.8)
50-59	1 (0.6)
Marital status	
Unmarried	2 (1.2)
Married	162 (98.8)
Highest level of education completed	
High school	14 (8.5)
Technical college (2-3 years)	37 (22.6)
Undergraduate degree (bachelor’s: 4-5 years)	76 (46.3)
Postgraduate degree (master’s)	32 (19.5)
Postgraduate degree (PhD)	5 (3.0)
Religious status	
Nonreligious	69 (42.1)
Christian (Protestant)	49 (29.9)
Christian (Catholic)	27 (16.5)
Buddhist	19 (11.6)
Professional status	
Unemployed (housewife)	72 (43.9)
Employed	92 (56.1)
Parity	
Nullipara	8 (4.9)
Primipara	120 (73.2)
Multipara	36 (22.0)
Fetal plurality	
Before birth	23 (14.0)
Singleton	131 (79.9)
Twins	8 (4.9)
Singleton + twins	2 (1.2)
Number of children	
Before birth	23 (14.0)
1	97 (59.1)
2	39 (23.8)
3	4 (2.4)

Characteristics	Participants, n (%)
4	1 (0.6)
Number of sources for searching pregnancy/delivery information	
One	13 (7.9)
Multiple	151 (92.1)
Most important aspect of pregnancy/delivery information seeking	
Accuracy and reliability of information	115 (70.1)
Diversity of information	13 (7.9)
Recency of information	4 (2.4)
Use of information	3 (1.8)
Ease of use of media	25 (15.2)
Interactivity of media	4 (2.4)
Digital device used for information seeking (multiple response question)	
Computer (desktop/laptop)	34 (16.8)
Tablet	7 (3.5)
Smartphone	161 (79.7)

Measurement Model

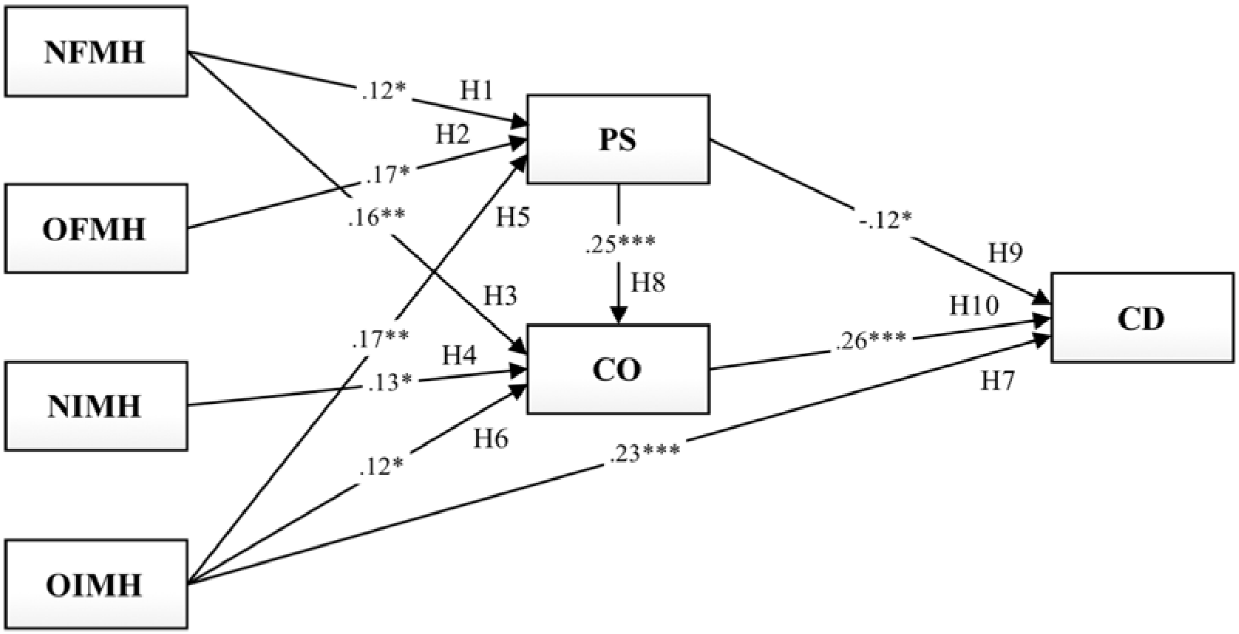
A confirmatory factor analysis performed with SPSS Amos 18.0 software (IBM Corp) indicated that the measurement model provided an acceptable fit to the data ($\chi^2_{278.00}$=686.03, normed chi-square [χ^2/df]=2.47, comparative fit index [CFI] 0.92, Tucker-Lewis index [TLI] 0.90, standardized root mean square residual [SRMR] 0.06, and root mean square error of approximation [RMSEA] 0.07). All values with an adequate level of construct validity were achieved by the recommended fit indices of the measurement model: χ^2/df<3.0 [42], CFI>0.90 [43], TLI>0.90 [44], SRMR<0.08 [45], and RMSEA<0.08 [46].

The reliability and validity results (Tables 1 and 2) indicated that the questionnaire items had acceptable internal consistency with Cronbach alpha>.70 and composite reliability values>0.60, convergent validity values with factor loadings (all significant at *P*<.001) and average variance extracted values>0.50, and discriminant validity with the square roots of the average variance extracted values being higher than the correlations between the factors.

Structural Model and Hypothesis Testing

To validate the hypothesized relationships in the proposed conceptual model (Figure 1), path analyses were performed. As mentioned previously, the same-fit indices of the measurement model were used to assess the goodness of fit of structural models. The results of the path analysis for pregnant women's medical help-seeking strategies, decision making, and any changes based on online information showed that the conceptual model had satisfactory levels of fit indices: $\chi^2_{5.00}$=5.07, *P*=.41, χ^2/df=1.01, CFI 1.00, TLI 1.00, SRMR 0.02, and RMSEA 0.01. As predicted in H1, H2, and H5, the strategies of NFMH (H1: β=.12, *P*=.02), OFMH (H2: β=.17, *P*=.01), and OIMH (H5: β=.17, *P*=.009) had positive effects on PS. Consistent with H3, H4, and H6, the strategies of NFMH (H3: β=.16, *P*=.003), NIMH (H4: β=.13, *P*=.02), and OIMH (H6: β=.12, *P*=.03) had positive effects on CO. Among those strategies for medical help-seeking, OIMH (H7: β=.23, *P*<.001) appeared as the only significant determinant of CD. H8, H9, and H10 were also supported by the results as CO was positively affected by PS (H8: β=.25, *P*<.001); however, CD was negatively affected by PS (H9: β=–.12, *P*=.03) and positively affected by CO (H10: β=.26, *P*<.001; Figure 2).

Figure 2. Medical problem-solving model with standardized path coefficients (**P*<.05, ***P*<.01, ****P*<.001). NFMH: nononline formal medical help-seeking; NIMH: nononline informal medical help-seeking; OFMH: online formal medical help-seeking; OIMH: online informal medical help-seeking; PS: promoting self-efficacy in perinatal health literacy based on online health information; CO: consulting others (health professionals) based on online health information; CD: changing their medical decisions based on online health information.



In addition, the other structural model for the moderating effect of different types of medical problems that pregnant women might encounter showed a satisfactory model fit: $\chi^2_{10.00}$=10.34, *P*=.41, χ^2/df=1.03, CFI 1.00, TLI 1.00, SRMR 0.02, and RMSEA 0.01. The results of the multiple group path analysis (Table 4) revealed that the only significant difference was found in 2 paths, H6_1 and H6_2, for obstetric and mental health problem-solving models, respectively (OIMHCO; critical ratio [CR] 2.01, significant at *P*<.05 [if CR>1.96]).

In the obstetric problem-solving model, the following hypotheses were rejected: H1_1 (β=.10, *P*=.20), H3_1 (β=.10, *P*=.18), H5_1 (β=.12, *P*=.17), H6_1 (β=.01, *P*=.94), and H9_1 (β=–.11, *P*=.15). As hypothesized, H2_1 (β=.21, *P*=.02) and H4_1 (β=.18, *P*=.02) were supported. In the mental health problem-solving model, the following hypotheses were rejected: H1_2 (β=.14, *P*=.07), H2_2 (β=.10, *P*=.29), H4_2 (β=.09, *P*=.24), and H9_2 (β=–.13, *P*=.10). Unlike the results of hypothesis testing for the obstetric problem-solving model, H3_2 (β=.20, *P*=.007), H5_2 (β=.24, *P*=.01), and H6_2 (β=.22, *P*=.003) were supported. In both models, the findings indicate that online health information was considered an important reference only for those with an OIMH strategy (H7_1: β=.27, *P*<.001; H7_2: β=.19, *P*=.02) when changing their medical decisions. Based on the online health information use, PS had positive effects on CO (H8_1: β=.26, *P*<.001; H8_2: β=.23, *P*=.002); furthermore, CO had positive effects on CD (H10_1: β=.25, *P*=.001; H10_2: β=.28, *P*<.001; Figure 3).

Table 4. Summary of the hypothesis testing results.

Model and hypotheses	β [a]	*P* value	SE[b]	CR[c]	Supported
Medical problem solving (n=328)					
H1: NFMH[d]→PS[h]	.12	.02	0.05	2.25	Yes
H2: OFMH[e]→PS	.17	.01	0.05	2.56	Yes
H3: NFMH→CO[i]	.16	.003	0.04	2.99	Yes
H4: NIMH[f]→CO	.13	.02	0.04	2.36	Yes
H5: OIMH[g]→PS	.17	.009	0.05	2.61	Yes
H6: OIMH→CO	.12	.03	0.04	2.12	Yes
H7: OIMH→CD[j]	.23	<.001	0.05	4.16	Yes
H8: PS→CO	.25	<.001	0.05	4.72	Yes
H9: PS→CD	–.12	.03	0.07	–2.16	Yes
H10: CO→CD	.26	<.001	0.07	4.72	Yes
Obstetric problem solving (n=164)					
H1_1: NFMH→PS	.10	.20	0.07	1.27	No
H2_1: OFMH→PS	.21	.02	0.06	2.38	Yes
H3_1: NFMH→CO	.10	.18	0.06	1.34	No
H4_1: NIMH→CO	.18	.02	0.05	2.29	Yes
H5_1: OIMH→PS	.12	.17	0.07	1.36	No
H6_1: OIMH→CO[k]	.01	.94	0.06	0.07	No
H7_1: OIMH→CD	.27	<.001	0.07	3.58	Yes
H8_1: PS→CO	.26	<.001	0.07	3.42	Yes
H9_1: PS→CD	–.11	.15	0.09	–1.43	No
H10_1: CO→CD	.25	.001	0.10	3.24	Yes
Mental health problem solving (n=164)					
H1_2: NFMH→PS	.14	.07	0.07	1.83	No
H2_2: OFMH→PS	.10	.29	0.07	1.05	No
H3_2: NFMH→CO	.20	.007	0.06	2.68	Yes
H4_2: NIMH→CO	.09	.24	0.06	1.17	No
H5_2: OIMH→PS	.24	.01	0.07	2.50	Yes
H6_2: OIMH→CO[k]	.22	.003	0.06	2.96	Yes
H7_2: OIMH→CD	.19	.02	0.07	2.38	Yes
H8_2: PS→CO	.23	.002	0.08	3.12	Yes
H9_2: PS→CD	–.13	.10	0.09	–1.64	No
H10_2: CO→CD	.28	<.001	0.09	3.46	Yes

aβ: standardized coefficient.

[b] SE: standard error.

[c] CR: critical ratio.

[d] NFMH: nononline formal medical help-seeking.

[e] OFMH: online formal medical help-seeking.

[f] NIMH: nononline informal medical help-seeking.

[g] OIMH: online informal medical help-seeking.

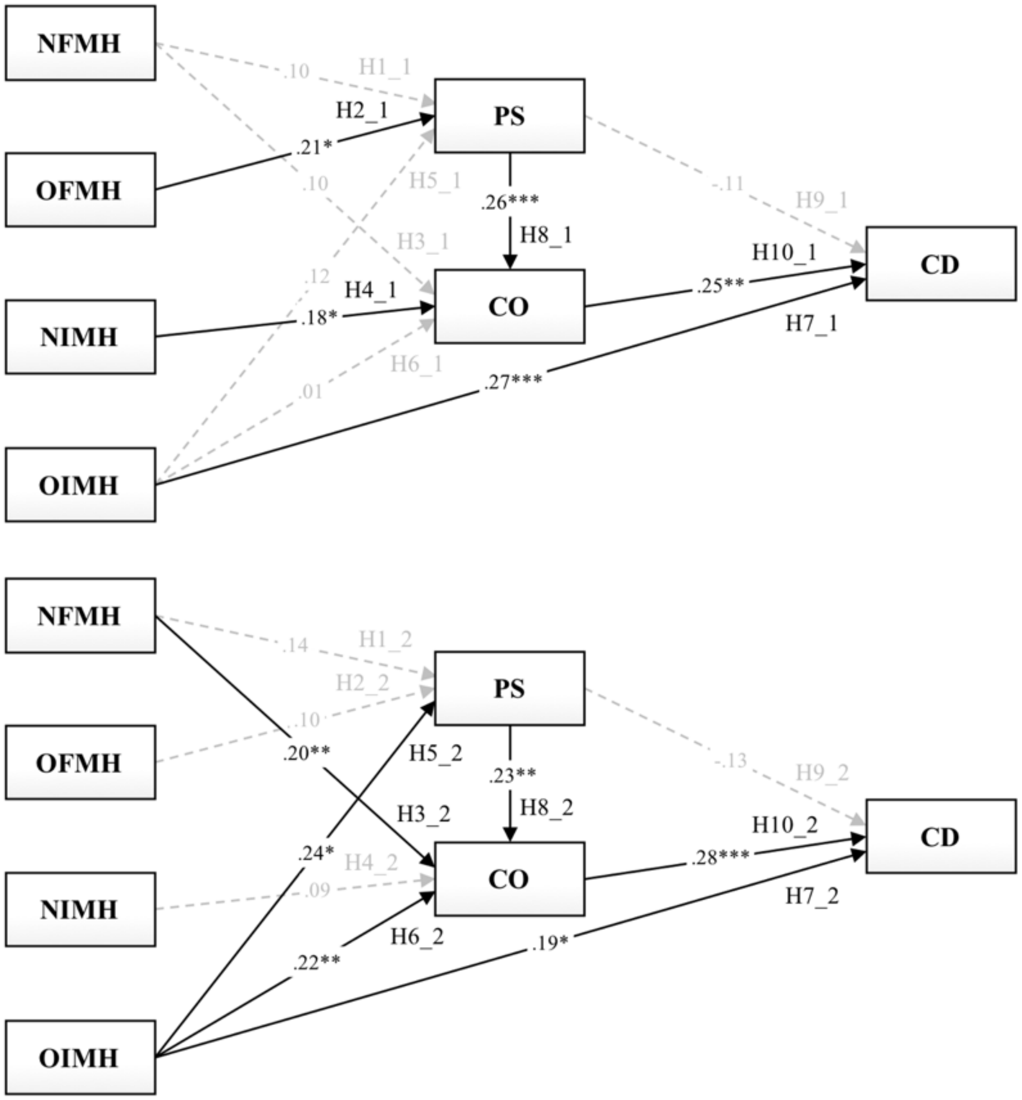
[h] PS: promoting self-efficacy in health literacy based on online health information.

[i] CO: consulting others based on online health information.

[j] CD: changing medical decisions based on online health information.

[k] A multigroup path analysis was performed between H6_1 and H6_2.

Figure 3. Obstetric (top) and mental health (bottom) problem-solving models with standardized path coefficients (**P*<.05, ***P*<.01, ****P*<.001). NFMH: nononline formal medical help-seeking; NIMH: nononline informal medical help-seeking; OFMH: online formal medical help-seeking; OIMH: online informal medical help-seeking; PS: promoting self-efficacy in perinatal health literacy based on online health information; CO: consulting others (health professionals) based on online health information; CD: changing their medical decisions based on online health information.



Finally, mediation analyses were conducted by examining the direct and indirect effects of perinatal women's medical help-seeking strategies and online health information use with CO and CD, with a focus on PS. It was revealed in the entire conceptual model that PS partially mediates the relation between CO and the strategies of NFMH (β=.03, *P*=.01) and OIMH (β=.04, *P*=.005), fully mediates the relation between OFMH and CO (β=.04, *P*=.003), and fully mediates the relation between NFMH and CD (β=.03, *P*=.049). In the obstetric problem-solving model, PS fully mediated the relation between the strategy of OFMH and CO (β=.06, *P*=.01), and in the mental health problem-solving model, PS partially mediated the relation between OIMH and CO (β=.06, *P*=.006).

Discussion

Principal Findings

As the goodness-of-fit indices between the measurement and structural models and the significance of all the hypothesized paths were statistically confirmed, the proposed conceptual model is expected to extend the understanding of the medical decision-making process in perinatal women with the 4 different medical help-seeking strategies. However, the decision-making process among perinatal women relied on which medical problem they would encounter, obstetrics or mental health, because there were substantial differences in the paths to self-efficacy in perinatal health literacy and intent to consult health professionals from their different medical help-seeking strategies, illustrated in the path diagrams (Figure 3). In the medical problem-solving model, the medical problem type had a partial moderating effect between OIMH and CO (H6_1 vs H6_2). Contrary to when the perinatal women using the OIMH strategy needed to solve mental health problems, no significant direct path from their adoption of the OIMH strategy to the intent to CO was found when they needed to solve obstetric problems. Furthermore, the direct paths from perinatal women's adoption of the OIMH strategy to CD were significant in both obstetric and mental health problem-solving models.

When encountering obstetric problems, perinatal women with the OIMH strategy were more likely to change their medical decisions based on numerous conflicting and confusing online information sources without consulting health professionals compared with those with the same strategy to solve mental health problems. In this study, the majority of the perinatal women (151/164, 92.1%) preferred to search pregnancy/delivery information from multiple sources to compensate for a lack of accuracy and reliability of information, and only 4 out of the remaining 13 women (30.8%) preferred to obtain the information from doctors, nurses, and other health professionals, indicating they did not consider their doctors to be a main source of medical information, showed lower levels of trust in medical experts [18], and used multiple information sources whose conflicting information increases anxiety [24]. Considering that perinatal women were not willing to discuss the pregnancy-related information obtained from the internet with health professionals in the previous studies [16,19,31], they will be more exposed to the risks of either applying treatment that is not evidence-based or missing the right time to initiate the appropriate medical treatment with the best chance to achieve better pregnancy and delivery outcomes.

As evidenced by the findings of previous studies [21,35], perinatal women who use the NIMH strategy to solve obstetric problems tended to validate the online health information by communicating with health professionals. Lagan and colleagues [21] reported that almost all (96.2%) discussed the internet-based information with their partner or husband, and 70.8% discussed the information with at least one health professional, reflecting that the presence of a partner or relatives facilitates perinatal women's communication with health professionals about their perceptions of being at risk [35]. Despite the associated support, informal help from partners, friends, and family members has disadvantages such as stigma or inappropriate support caused by their lack of relevant medical knowledge; therefore, making or changing medical decisions following suggestions provided by those with insufficient knowledge can be unhelpful or even harmful [47]. Meanwhile, those preferring to use the NFMH strategy did not make or change their medical decisions regarding obstetric problems by referring to online health information. As 70.1% of this study sample (115/164) considered accuracy and reliability to be the most important criteria in their evaluation when seeking pregnancy/delivery information, it could be inferred by our findings that the internet was considered a supplementary information source before or after contacting health professionals as a primary information source.

Our study revealed a positive relationship between self-efficacy in perinatal health literacy and intent to consult health professionals in both models. Once the level of perinatal women's health literacy increases, they can be more easily encouraged to identify questions they wish to ask a health professional, be involved in the decision-making process, make better decisions, and exert more control over their decisions [21]. Searching online health information before paying a visit to their health professionals and after their consultation [30] is common as a means to deal with doubts and navigate other women's pregnancy-related decisions, which can also result in increased confusion, anxiety, and fright [48]. With the perinatal women's concerns on the lack of time to sufficiently discuss personal medical issues with health professionals [21,22] and the professionals' reactions to the information they shared [20], a number of previous studies found that the majority of women believed that health professionals should suggest suitable and reliable internet websites [20,21] where they could find relevant information with appropriate context and avoid erroneous or misleading claims [48]. As revealed in the obstetric problem-solving model, the OFMH strategy was positively related to PS, which reflects the implications of other studies.

When it comes to solving mental health problems, perinatal women showed a different attitude toward online information use. Those using the NFMH strategy disclosed their intent to discuss relevant issues with mental health professionals without the precondition of enhanced self-efficacy in perinatal health literacy. Generally, psychiatrists and other mental health professionals, whose goals are to provide each patient with the opportunity to discuss problems with a sympathetic listener by helping to establish a rapport between the patient and doctor, educate and motivate the patient, and take a clinical examination of the mental state, depend on a psychiatric interview. In comparison with the significance of the direct and indirect paths from NFMH to CO (ie, NFMHPS, NFMHPSCO, and NFMHCO) in the two different medical problem-solving models shown in Figure 3 and Table 4, the findings of this study suggest that what women expected from the interaction with obstetricians or psychiatrists during their clinic visits would be different during the perinatal period. Accordingly, these findings can well explain why it is important for health professionals not to discourage their patients to share obstetric concerns and treatment suggestions retrieved from the internet.

Limitations and Future Direction

There are several limitations of this study that should be considered in future research. The sample size and its demographic composition limited our opportunity to examine the possible age difference, relationship with one's doctor, number of births, and history of experienced obstetric and mental health problems because this study focused on the entire perinatal period, encompassing conception, pregnancy, childbirth, and postpartum. Given the small sample size, it would result in an insufficient statistical power, particularly as reflected by significantly low standardized path coefficients. For a more accurate estimate, a larger sample size would be required in the future studies. In terms of age distribution, the proportion of participants aged 19 to 29 years (14/164, 8.5%) was considerably less than that of those aged 30 to 39 years (134/164, 81.7%). Given the age differences of people obtaining online health information [49], the possibility that younger women may differ in how they make and change their medical decisions by using online health information compared with older women cannot be ruled out. As this study sample, enrolled from a large hospital-based sample of patients, was willing to complete an online questionnaire, the participants might be more likely to have a good rapport with their doctors or be internet-savvy patients accessing the internet more frequently for solving medical problems than the others who did not so. Finally, whether perinatal women have given birth to at least one child or had a history of obstetric and mental health problems may be associated with their intent to either share their own experiences with other mothers in trouble or trust and follow the doctor's instructions only. Taken together, there is a need to control for these determinants in future studies.

Conclusions

Despite the positive mediating effects of perinatal women's enhanced health literacy on the intent to discuss personal medical problems with health professionals, the findings from this study revealed that some women using an OIMH strategy would be more likely to change their medical decisions not by consulting health care professionals but by referring to online health information compared with other women using different medical help-seeking strategies. Particularly, this concern would be prevalent when encountering participants with obstetric problems during the perinatal period. Beyond a short consultation to quickly and authoritatively steer patients toward an obstetric doctor's choice of action, it is recommended that obstetricians provide their patients with needed context for information found online. To fully explain this with an open mind, they need to actively develop or support information and communications technology–based health information services such as mobile apps and chatbots.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C0911).

Conflicts of Interest

None declared.

References

1. Featherstone B, Broadhurst K. Engaging parents and carers with family support services: What can be learned from research on help-seeking? Child Fam Soc Work 2003 Nov;8(4):341-350. [doi: 10.1046/j.1365-2206.2003.00289.x]
2. Rickwood D, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. Aust eJ Advancem Mental Health 2005 Dec;4(3):218-251. [doi: 10.5172/jamh.4.3.218]
3. Cornally N, McCarthy G. Help-seeking behaviour: a concept analysis. Int J Nurs Pract 2011 Jun;17(3):280-288. [doi: 10.1111/j.1440-172X.2011.01936.x] [Medline: 21605269]
4. Conrad P, Stults C. The internet and the experience of illness. In: Bird CE, Conrad P, editors. Handbook of Medical Sociology, 6th Edition. Nashville: Vanderbilt University Press; 2010:179-191.
5. Slauson-Blevins KS, McQuillan J, Greil AL. Online and in-person health-seeking for infertility. Soc Sci Med 2013 Dec;99:110-115. [doi: 10.1016/j.socscimed.2013.10.019] [Medline: 24355477]
6. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [doi: 10.1089/cpb.2004.7.497] [Medline: 15667044]
7. Fenwick J, Hauck Y, Downie J, Butt J. The childbirth expectations of a self-selected cohort of Western Australian women. Midwifery 2005 Mar;21(1):23-35. [doi: 10.1016/j.midw.2004.07.001] [Medline: 15740814]
8. Carolan M. Health literacy and the information needs and dilemmas of first-time mothers over 35 years. J Clin Nurs 2007 Jun;16(6):1162-1172. [doi: 10.1111/j.1365-2702.2007.01600.x] [Medline: 17518891]
9. Lagan BM, Sinclair M, Kernohan WG. What is the impact of the Internet on decision-making in pregnancy? A global study. Birth 2011 Dec;38(4):336-345. [doi: 10.1111/j.1523-536X.2011.00488.x] [Medline: 22112334]
10. Fleming SE, Vandermause R, Shaw M. First-time mothers preparing for birthing in an electronic world: internet and mobile phone technology. J Reprod Infant Psychol 2014 Mar 19;32(3):240-253. [doi: 10.1080/02646838.2014.886104]
11. Sanders RA, Crozier K. How do informal information sources influence women's decision-making for birth? A meta-synthesis of qualitative studies. BMC Pregnancy Childbirth 2018 Jan 10;18(1):21 [FREE Full text] [doi: 10.1186/s12884-017-1648-2] [Medline: 29316887]
12. Dako-Gyeke P, Aikins M, Aryeetey R, Mccough L, Adongo PB. The influence of socio-cultural interpretations of pregnancy threats on health-seeking behavior among pregnant women in urban Accra, Ghana. BMC Pregnancy Childbirth 2013 Nov 19;13(1):211. [doi: 10.1186/1471-2393-13-211]
13. Bjelke M, Martinsson A, Lendahls L, Oscarsson M. Using the Internet as a source of information during pregnancy: a descriptive cross-sectional study in Sweden. Midwifery 2016 Sep;40:187-191. [doi: 10.1016/j.midw.2016.06.020] [Medline: 27450590]
14. Seibold C. Young single women's experiences of pregnancy, adjustment, decision-making and ongoing identity construction. Midwifery 2004 Jun;20(2):171-180. [doi: 10.1016/S0266-6138(03)00057-3] [Medline: 15177861]

15. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371. [Medline: 15224975]
16. Bert F, Gualano MR, Brusaferro S, De Vito E, de Waure C, La Torre G, et al. Pregnancy e-health: a multicenter Italian cross-sectional study on Internet use and decision-making among pregnant women. J Epidemiol Community Health 2013 Dec 1;67(12):1013-1018. [doi: 10.1136/jech-2013-202584] [Medline: 24072743]
17. Knoepke CE, Matlock DD. Preliminary development of an informational media use measure for patients with implanted defibrillators: toward a model of social-ecological assessment of patient education and support. Health Soc Work 2017 Nov 01;42(4):199-206 [FREE Full text] [doi: 10.1093/hsw/hlx023] [Medline: 28575348]
18. Eller N. Right from the source: trust in health providers, information sources, and vaccine hesitancy. 2017. URL: https://www.semanticscholar.org/paper/Right-from-the-source%3A-Trust-in-health-providers%2C-Eller/7f23a06a66b2fd855d6b9ac71582b612d41d0cf6 [accessed 2020-01-07]
19. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
20. Slomian J, Bruyère O, Reginster JY, Emonts P. The internet as a source of information used by women after childbirth to meet their need for information: A web-based survey. Midwifery 2017 May;48:46-52. [doi: 10.1016/j.midw.2017.03.005] [Medline: 28324809]
21. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]
22. Lagan BM, Sinclair M, Kernohan WG. A Web-based survey of midwives' perceptions of women using the Internet in pregnancy: a global phenomenon. Midwifery 2011 Apr;27(2):273-281. [doi: 10.1016/j.midw.2009.07.002] [Medline: 19700228]
23. Lim AS, Stewart K, Abramson MJ, Ryan K, George J. Asthma during pregnancy: the experiences, concerns and views of pregnant women with asthma. J Asthma 2012 Jun;49(5):474-479. [doi: 10.3109/02770903.2012.678024] [Medline: 22530982]
24. Hämeen-Anttila K, Nordeng H, Kokki E, Jyrkkä J, Lupattelli A, Vainio K, et al. Multiple information sources and consequences of conflicting information about medicine use during pregnancy: a multinational Internet-based survey. J Med Internet Res 2014;16(2):e60 [FREE Full text] [doi: 10.2196/jmir.2939] [Medline: 24565696]
25. Chen Y, Li C, Liang J, Tsai C. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. J Med Internet Res 2018 Dec 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]
26. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Prom Int 2000;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
27. Smith CL, Shochet IM. The impact of mental health literacy on help-seeking intentions: results of a pilot study with first year psychology students. Int J Mental Health Promotion 2011 Jan;13(2):14-20. [doi: 10.1080/14623730.2011.9715652]
28. James DC, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mhealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Dec 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
29. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 1;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
30. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
31. Larsson M. A descriptive study of the use of the Internet by women seeking pregnancy-related information. Midwifery 2009 Feb;25(1):14-20. [doi: 10.1016/j.midw.2007.01.010] [Medline: 17408822]
32. Gao L, Larsson M, Luo S. Internet use by Chinese women seeking pregnancy-related information. Midwifery 2013 Jul;29(7):730-735. [doi: 10.1016/j.midw.2012.07.003] [Medline: 22958935]
33. Huberty J, Dinkel D, Beets MW, Coleman J. Describing the use of the internet for health, physical activity, and nutrition information in pregnant women. Matern Child Health J 2013 Oct;17(8):1363-1372. [doi: 10.1007/s10995-012-1160-2] [Medline: 23090284]
34. Song H, Cramer EM, McRoy S, May A. Information needs, seeking behaviors, and support among low-income expectant women. Women Health 2013;53(8):824-842. [doi: 10.1080/03630242.2013.831019] [Medline: 24215275]
35. Rance S, McCourt C, Rayment J, Mackintosh N, Carter W, Watson K, et al. Women's safety alerts in maternity care: is speaking up enough? BMJ Qual Saf 2013 Apr;22(4):348-355. [doi: 10.1136/bmjqs-2012-001295] [Medline: 23417732]
36. Fisher VM. Help-seeking for depression in rural women: a community portrait. 2005. URL: https://scholarscompass.vcu.edu/etd/681/ [accessed 2020-01-07]
37. Gholami K, Mohammadi S. Information interaction of pregnant women in Ninisite. Human Information Interaction 2015;1(4):305-318 [FREE Full text]
38. Renkert S, Nutbeam D. Opportunities to improve maternal health literacy through antenatal education: an exploratory study. Health Promot Int 2001 Dec;16(4):381-388. [Medline: 11733456]
39. Cheung W, Ip W, Chan D. Maternal anxiety and feelings of control during labour: a study of Chinese first-time pregnant women. Midwifery 2007 Jun;23(2):123-130. [doi: 10.1016/j.midw.2006.05.001] [Medline: 17055624]

40. Melender H. Experiences of fears associated with pregnancy and childbirth: a study of 329 pregnant women. Birth 2002 Jun;29(2):101-111. [doi: 10.1046/j.1523-536x.2002.00170.x] [Medline: 12051188]
41. Rini CK, Dunkel-Schetter C, Wadhwa PD, Sandman CA. Psychological adaptation and birth outcomes: the role of personal resources, stress, and sociocultural context in pregnancy. Health Psychol 1999 Jul;18(4):333-345. [doi: 10.1037//0278-6133.18.4.333] [Medline: 10431934]
42. Marsh H, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: first and higher order factor models and their invariance across groups. Psychol Bull 1985;97(3):562 [FREE Full text] [doi: 10.1037/0033-2909.97.3.562]
43. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990 Mar;107(2):238-246. [Medline: 2320703]
44. 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]
45. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling Multidiscipl J 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
46. Browne M, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 1993;154:136. [doi: 10.1177/0049124192021002005]
47. Griffiths KM, Crisp DA, Barney L, Reid R. Seeking help for depression from family and friends: a qualitative analysis of perceived advantages and disadvantages. BMC Psychiatry 2011 Dec 15;11:196 [FREE Full text] [doi: 10.1186/1471-244X-11-196] [Medline: 22171567]
48. 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]
49. Sharit J, Taha J, Berkowsky RW, Czaja SJ. Seeking and resolving complex online health information: age differences in the role of cognitive abilities. 2016 Sep 15 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2016; Los Angeles p. 1-5. [doi: 10.1177/1541931213601001]

Abbreviations

CD: changing their medical decisions based on online health information
CFI: comparative fit index
CO: consulting others (health professionals) based on online health information
CR: critical ratio
KHIDI: Korea Health Industry Development Institute
NFMH: nononline formal medical help-seeking
NIMH: nononline informal medical help-seeking
OFMH: online formal medical help-seeking
OIMH: online informal medical help-seeking
PS: promoting self-efficacy in perinatal health literacy based on online health information
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 19.04.19; peer-reviewed by KS Lee, A Bhat, C Knoepke, E Neter; comments to author 03.10.19; revised version received 28.11.19; accepted 16.12.19; published 04.03.20.

Please cite as:
Chung K, Cho HY, Kim YR, Jhung K, Koo HS, Park JY
Medical Help-Seeking Strategies for Perinatal Women With Obstetric and Mental Health Problems and Changes in Medical Decision Making Based on Online Health Information: Path Analysis
J Med Internet Res 2020;22(3):e14095
URL: https://www.jmir.org/2020/3/e14095
doi:10.2196/14095
PMID:32130139

Original Paper

The Association Between Health Information Seeking on the Internet and Physician Visits (The Seventh Tromsø Study - Part 4): Population-Based Questionnaire Study

Kassaye Yitbarek Yigzaw[1], MSc, PhD; Rolf Wynn[2,3], MD, PhD; Luis Marco-Ruiz[1], MSc, PhD; Andrius Budrionis[1], MSc, PhD; Sunday Oluwafemi Oyeyemi[4], MD, MSc, MPH; Asbjørn Johansen Fagerlund[1], MSc, PhD; Johan Gustav Bellika[1,2], MSc, PhD

[1]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway
[2]Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway
[3]Division of Mental Health and Addictions, University Hospital of North Norway, Tromsø, Norway
[4]Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Johan Gustav Bellika, MSc, PhD
Department of Clinical Medicine
Faculty of Health Sciences
UiT The Arctic University of Norway
Postboks 6050 Langnes
Tromsø, 9037
Norway
Phone: 47 95748049
Email: johan.gustav.bellika@ehealthresearch.no

Abstract

Background: The internet is being widely used for seeking health information. However, there is no consensus on the association between health information seeking on the internet and the use of health care services.

Objective: We examined the association between health information seeking via the internet and physician visits. In addition, we investigated the association between online health information seeking and the decisions to visit and not to visit a physician.

Methods: We used the cross-sectional electronic health (eHealth) data of 18,197 participants from the seventh survey of the Tromsø Study (Tromsø 7). The participants were aged ≥40 years and living in Tromsø, Norway. We used logistic regression models to examine the association between online health information seeking and physician visits, the decision to visit a physician, and the decision not to visit a physician, with adjustment for the demographic status, socioeconomic status, and health status of the participants.

Results: The use of Web search engines was associated with a physician visit. However, the association was moderated by age, and the OR decreased as age increased. The ORs for the use of Web search engines were 1.99 (95% CI 1.94-2.02) and 1.07 (95% CI 1.03-1.12) at ages 40 and 80 years, respectively. The decision to visit a physician was associated with the use of Web search engines (OR 2.95, 95% CI 2.03-4.46), video search engines (OR 1.43, 95% CI 1.21-1.70), and health apps (OR 1.26, 95% CI 1.13-1.42). The association between social media use and the decision to visit a physician was moderated by gender. Women who used social media had 1.42 (95% CI 1.31-1.55) times higher odds of deciding to visit a physician, whereas the decision to visit a physician was not different between men who used social media and those who did not use social media. Conversely, the decision not to visit a physician was associated with the use of Web search engines (OR 2.78, 95% CI 1.92-4.18), video search engines (OR 1.27, 95% CI 1.07-1.51), social media (OR 1.28, 95% CI 1.10-1.49), and health apps (OR 1.20, 95% CI 1.07-1.35).

Conclusions: Health information found on the internet was positively associated with both the decision to visit a physician and the decision not to visit a physician. However, the association of health information seeking with the decision to visit a physician was slightly stronger than the association with the decision not to visit a physician. This could imply that the use of eHealth services is associated with a resultant increase in physician visits. In summary, our findings suggest that the internet serves as a supplement to health care services rather than as a replacement.

(J Med Internet Res 2020;22(3):e13120) doi:10.2196/13120

KEYWORDS

eHealth; internet; health care service; physician visit; Tromsø study; health information seeking; Web search engine; health app; social media; video search engine

Introduction

Access to the internet is currently widespread. In 2017, around the time the data for this study were collected, the internet coverage of Norwegian households was 97%. Internet access varies between 91% and 100% depending on household income. Ninety percent of Norwegians aged between 16 and 79 years use the internet on a daily basis, with similar internet usage between women (87%) and men (90%) [1].

Broad access to the internet has triggered a rapid growth in the use of the internet for health-related applications, such as accessing and transferring health information and receiving guidance and support [2-6]. Health information seeking is considered the most common and influential use of the internet [7-11]. Among Norwegian internet users, 62% used the internet for seeking health information, with a higher proportion of women and younger adults [1], which is consistent with the findings in other countries [8,12-14].

The perceived benefits of online health information seeking include widespread access to health information, convenience (ie, ease and speed), and anonymity [15-19]. These benefits are expected to enable individuals to play an active role in their health care, make better-informed decisions, and possibly improve health outcomes [18,20,21]. There are also concerns that the variable quality of online health information combined with limited ability to critically evaluate health information may contribute to negative outcomes, such as unnecessary physician visits, delays in seeking necessary medical care, change in treatments, and seeking alternative treatments that can be harmful [5,16,17,22-26].

In Nordic countries, including Norway, and many other European countries, including the United Kingdom, general practitioners (GPs) represent the basis of publicly funded health services. GPs typically have lists of patients who they are responsible for and see more or less regularly. GPs are expected to diagnose and treat a major proportion of patients and refer only those who need more specialized health services to hospitals for further assessment and treatment (ie, the “gatekeeper” function) [27]. The threshold to seek consultation with a GP may be relatively low, but many GPs are very busy and patients nevertheless have to decide whether their current problem warrants a physician visit [27].

Searching for health information is a complex process influenced by a range of factors. Lambert & Loiselle [28], in their review, found that the behavior of health information searching is often studied within the context of coping with illness, involvement in medical decision-making, or preventive behavior. Different theoretical frameworks have been applied in research on health information searching [28]. The framework of Miller [29] is one of the most cited frameworks, and a differentiation is made between those who actively engage with information (monitoring) and those who avoid such information (blunting) [30]. Psychological factors, including cognitive and emotional factors [31,32], might be central in the decision-making process for many health information searchers. The process of online health information searching may be powered by not only a need for information about some health- or illness-related topic but also different emotional states, such as anxiety [33], and people with health anxiety are more likely to search for health information [33]. Moreover, how people react to the health information they find online may also vary according to a range of factors, including emotional factors [31,32]. Although finding the required information can result in a positive emotional reaction, it may also result in confusion or increased anxiety [34]. Similarly, the decision to visit a physician or not to visit a physician after finding health information online is a complex process that could be influenced by many different factors, some of which will be examined in this study.

Health information seeking is known to highly influence subsequent health-related behavioral decisions. It has been reported that 46% of European adults [8] and 48% of American adults [4] who sought health information via the internet used the information they found to decide whether they needed to see a physician.

Prior studies have found different results regarding the question about whether online health information searching impacts physician visits. Some studies have suggested that searching for online health information can be conducive to increased physician visits. A proposed mechanism underlying this increase in visits is that some individuals may find it difficult to interpret complex and uncertain medical knowledge without the help of health care professionals [35-41]. Therefore, more access to health information via the internet may lead to more uncertainties in understanding health conditions and, consequently, to additional contact with health care professionals. Other studies have found that online health information searching reduces traditional health care service consumption [42,43], and some studies have found no such association [44]. In this study, the objective was to make an accurate assessment of the association between health information seeking and health care service use. We believe that this will be useful for estimating future health care service need in Norway and other countries and for informing future policies related to health care in general and electronic health (eHealth) services in particular.

Studies often capture the outcomes of health information seeking in terms of the decision to visit a physician as a single question that asks if the health information individuals found online led them to decide to visit a physician or not [40]. However, some studies had two questions that capture both outcomes [44], as a single question may not capture both outcomes, given that individuals may make both decisions at different time points in a study period. Therefore, a separate question is needed for each of the outcomes.

The seventh survey of the Tromsø Study (Tromsø 7) included a questionnaire with a wide variety of questions, such as those on the use of eHealth tools, use of health care services, socioeconomic status, and health status. We explored a subset of the data from this questionnaire in a series of four papers. In paper 1, we presented the main findings regarding the characteristics of the participants and their use of eHealth [45]. In paper 2, we studied how the presence of different illnesses influences the use of eHealth [46]. In paper 3, we examined the psychological and emotional outcomes with the use of eHealth [34]. Paper 4 (this paper) aims to study the association between health information seeking using eHealth tools and physician visits. This paper also investigates the associations of online health information seeking with decisions to visit and not to visit a physician.

Methods

The Study Site

The participants of the study are inhabitants of Tromsø municipality. Tromsø is a major city in north Norway with a population of around 75,000 inhabitants. The city of Tromsø is located in the sub-Arctic region at 69 degrees North. North Norway is a sparsely populated area with most inhabitants working within the public sector, including health care, education, and administration, and the service sector. Other important employment areas in north Norway are tourism, fishery, agriculture, and some industries.

The Tromsø Study

The Tromsø Study is a population-based longitudinal health study conducted by UiT The Arctic University of Norway, the National Health Screening Service, and others [47]. Tromsø 7 was conducted in 2015-2016 with a focus on inhabitants aged 40 years or above. However, this is the first time the Tromsø Study collected eHealth related information. Therefore, this paper, which is based on eHealth survey data, involves a cross-sectional study. A total of 21,083 individuals (11,074 women and 10,009 men) participated in the study, with a response rate of 65%.

Independent Variables

The variables used in this study were measures of health information seeking via Web search engines (ie, Google), video search engines (ie, YouTube), social media (ie, Facebook), and smartphone or tablet health apps. The variables were constructed from four questions regarding whether a participant had used each of the eHealth tools in the last year. The responses to the questions were "never," "once," "sometimes," and "often." We subsequently dichotomized the responses into "never" and "ever," where "ever" includes once, sometimes, and often.

The other independent variable was a dichotomous measure of whether a participant reported having one or more diseases in the last year. The variable was constructed from questions regarding whether a participant had diseases, such as high blood pressure, heart attack, heart failure, atrial fibrillation, angina, stroke, diabetes, kidney disease, bronchitis, asthma, cancer, rheumatoid arthritis, arthrosis, migraine, psychological problems, and chronic pain. The responses to the questions were "no," "yes," and "yes, previously."

Age, gender, education, household income, occupation, and self-reported health, which are known to influence health information-seeking behaviors, were also controlled for in this study. Age was converted into 10-year intervals, as a small age difference is not associated with a large change in the outcome variables.

Dependent Variables

In the Tromsø Study, there were three questions examining different aspects of health service use. The first dependent variable was a dichotomous measure of whether the participant had a physician visit in the last year. The variable was constructed from questions that asked whether a person visited a GP, emergency care practitioner, psychologist, or psychiatrist in the last year.

The second dependent variable measures the decision to visit a physician following health information seeking on the internet. The variable was constructed from a question that asked whether a participant decided to visit a physician according to the information found on the internet. The response to the question was "never," "once," "sometimes," or "often." We dichotomized the responses into "never" and "ever," where "ever" includes once, sometimes, and often.

The third dependent variable measures the decision not to visit a physician following health information seeking on the internet. The variable was constructed from a question that asked whether a participant decided not to visit a physician according to the information found on the internet. The response to the question was "never," "once," "sometimes," or "often." We dichotomized the responses into "never" and "ever," where "ever" includes once, sometimes, and often.

The physician visit variable measures whether a participant visited a physician in the last year. It is known that people seek health information for a wide variety of purposes, including deciding whether to visit a physician, preparing for a physician appointment, and reassurance, second opinion, and expanding knowledge on the information received from health care providers after a physician visit [15,16,18,48]. In other words, a participant may seek health information before or after a physician visit. Even when a participant seeks health information before a physician visit, the health information the participant obtains may not necessarily be used for deciding whether to visit a physician.

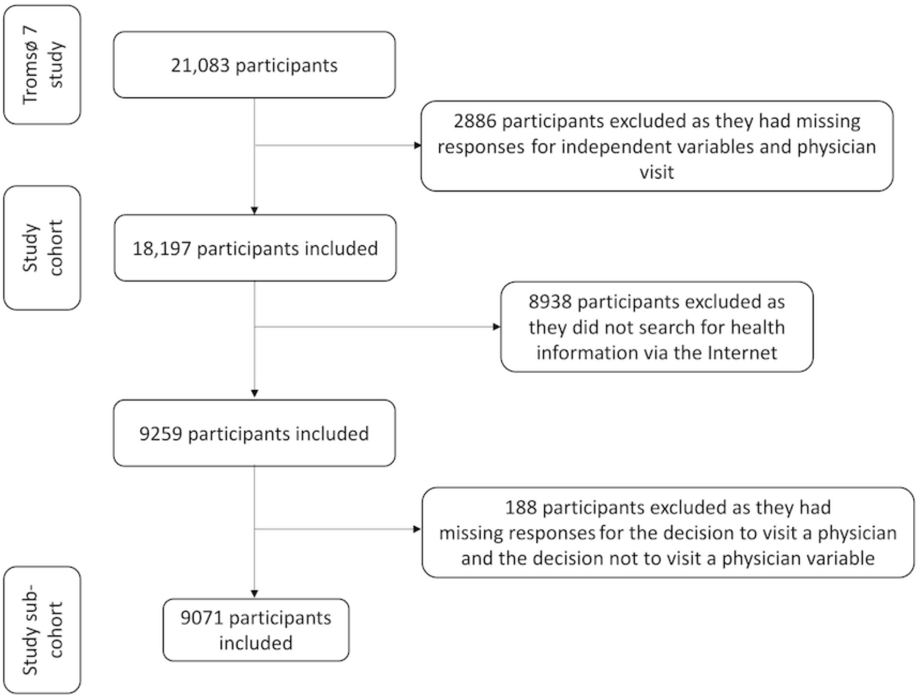
The variable of the decision to visit a physician measures whether a participant decided to visit a physician because of the health information the participant read online. Similarly, the variable of the decision not to visit a physician solely measures whether a participant decided not to visit a physician because of the health information the participant read online. Therefore, a subset of actual physician visits can be associated with decisions to visit made after reading online health information.

If a participant never decides to visit a physician according to health information read on the internet, it does not necessarily mean the participant used the information to decide not to visit a physician and vice versa. The participant may have sought the health information for other purposes, may not have found the information useful, or may not have understood the information well enough to base decisions on it.

Study Sample

Figure 1 shows the study sample selection workflow. Of the 21,083 participants in Tromsø 7, we excluded 2886 participants who had missing responses for one or more of the variables, such as age, gender, education, household income, occupation, Web search engine use, video search engine use, social media use, health app use, diseases, self-reported health status, and physician visit. The final study sample consisted of 18,197 participants (9251 women and 8946 men).

Figure 1. The study sample selection workflow.



Of the 18,197 participants included in this study, 9259 participants sought health information via one or more eHealth tools. Of these 9259 participants, we subsequently excluded 188 participants with missing information regarding the variables of the decisions to visit and not to visit a physician. As a result, a subcohort that consisted of 9071 participants (5110 women and 3961 men) was created.

Statistical Analysis

Descriptive statistics were calculated to summarize the dependent and independent variables. Multivariable logistic regression models were fitted for the following three outcome variables of interest: (1) physician visit; (2) decision to visit a physician according to health information read on the internet; and (3) decision not to visit a physician according to health information read on the internet. The first model was fitted on the whole cohort dataset (n=18,197), and the other two models were fitted on the subcohort dataset (n=9071).

We used purposeful selection of independent variables for each of the logistic regression models as proposed by Hosmer et al [49]. First, we fitted the model with only one independent variable. Second, we fitted a multivariable model with all independent variables whose *P* values were <.25 in the previous step. Third, we iteratively checked whether variables not significant (Wald test and .05 alpha value) in the multivariable model provided important adjustment to other variables that remained in the model. Important adjustment was measured by a change in coefficients by more than 20% between the multivariable models with and without the variable. Variables that provided important adjustment were added back into the multivariable model. Fourth, to ensure we did not miss any important variables, we added independent variables not added in the second step, one at a time to the multivariable model, using the Wald test to verify the effect of each additional variable at an alpha value of .05. Thereafter, we added all independent variables that were significant to the multivariable model. Fifth, we explored possible interactions of age, gender, and education with the use of eHealth tools. We also explored possible interactions between age and disease, age and self-reported health, age and household income, education and household income, education and disease, education and self-reported health, household income and disease, self-reported health and disease, and household income and self-reported health. We added the interactions, one at a time, to the multivariable model at the end of step four and tested significance using a likelihood ratio test and an alpha value of .1. Following separate analysis of each of the interaction terms, we added each significant interaction term to the model at the end of step four and removed interaction terms that were not significant with the Wald test and an alpha value of .05. Thereafter, we fitted the model excluding the nonsignificant interactions. Sixth, we checked the fitness of the model with regard to the data using the Hosmer-Lemeshow goodness-of-fit test.

We tested statistical significance at an alpha level of .05. All analyses were conducted using R statistical software (version 3.4.0; R Project for Statistical Computing, Vienna, Austria). All the ORs reported in the paper are adjusted ORs.

Ethics

Tromsø 7 was approved by the North Norway Regional Committee for Medical and Health Research Ethics (REK Nord, reference 2014/940). All participants provided written informed consent for Tromsø 7.

Results

Participants' Characteristics

Characteristics of the study participants are summarized in Table 1. Of the 18,197 study participants, 9251 (50.84%) were women and 8948 (49.16%) were men. The mean age of the overall study sample (n=18,197) was 56.38 years, with 51.10% (9298) of participants having a college education. About three-fourths of the participants (73.22%, 13,323/18,197) had one or more diseases, 50.88% (9259/18,197) sought health information via one or more eHealth tools, and 80.38% (14,627/18,197) had a physician visit in the last year.

Table 1. Characteristics of the study sample.

Variable	Cohort (n=18,197)		Subcohort (n=9071)	
	Count	Percentage	Count	Percentage
Gender				
Female	9251	50.84	5110	56.33
Male	8946	49.16	3961	43.67
Age, years	56.38[a]	11.00[b]	52.60[a]	9.25[b]
Education				
Primary or secondary	3831	21.05	970	10.69
Upper secondary	5068	27.85	2280	25.13
Less than 4 years of college	3650	20.06	2086	23.00
Four years or more of college	5648	31.04	3735	41.18
Household income (US)[c]				
<39 000 US	2132	11.72	568	6.26
39 000 – 61 000 US	3725	20.47	1632	18.00
61 000 – 83 000 US	3242	17.81	1519	16.74
83 000 – 111 000 US	4396	24.16	2449	27.00
>111 000 US	4702	25.84	2903	32.00
Occupation				
Full-time work	11,145	61.25	6461	71.23
Part-time work	1453	7.99	783	8.63
Unemployed	122	0.67	70	0.77
Housekeeping	99	0.54	29	0.32
Retired	3720	20.44	912	10.05
Student/military service	55	0.30	39	0.43
Disability and other family welfare benefits	1603	8.81	777	8.57
Disease				
Yes	13,323	73.22	6688	73.73
No	4874	26.78	2383	26.27
Self-reported health status				
Very bad	54	0.29	27	0.30
Bad	909	4.99	508	5.60
Neither bad nor good	4583	25.19	2082	22.95
Good	9891	54.36	5003	55.15
Very good	2760	15.17	1451	16.00
Physician visit				
Yes	14,627	80.38	7496	82.64
No	3570	19.62	1575	17.36
Web search engine use				
Ever	8956	49.22	8773	96.71
Never	9241	50.78	298	3.29
Video search engine use				
Ever	900	4.95	875	9.65

Variable	Cohort (n=18,197)		Subcohort (n=9071)	
	Count	Percentage	Count	Percentage
Never	17,297	95.05	8196	90.35
Social media use				
Ever	1259	6.92	1223	13.48
Never	16,938	93.08	7848	86.52
App use				
Ever	2358	12.96	2299	25.34
Never	15,839	87.04	6772	74.66
eHealth use[d]				
Ever	9259	50.88	9071	100
Never	8938	49.12	0	0
Decision to visit a physician				
Ever	—[e]	—	2075	22.88
Never	—	—	6996	77.12
Decision not to visit a physician				
Ever	—	—	2093	23.07
Never	—	—	6978	76.93

[a] The value is mean age.

[b] The value is standard deviation of age.

[c] One US dollar is approximately 9 Norwegian kr.

[d] It includes the use of one or more tools, such as Web search engines, health apps, video search engines, and social media.

[e] Not applicable.

Among our subcohort of 9071 participants who sought health information using one or more eHealth tools, 2075 (22.88%) decided to visit a physician and 2093 (23.07%) decided not to visit a physician according to the information they found. Among these participants, 1047 (11.54%) decided both to visit a physician and not to visit a physician in the last year. In other words, 34.40% (3121/9071) of the subcohort used the health information as a basis for one or more decisions.

Actual Physician Visit (n=18,197)

The independent variables selected in the multivariable logistic regression model for a physician visit were age, gender, education, household income, occupation, disease, self-reported health, Web search engine use, video search engine use, social media use, and health app use. The interactions between gender and health app use, age and disease, and age and Web search engine use were statistically significant in the multivariable model. The likelihood ratio test (χ^2_3=31.89, *P*<.001) of the models with and without the interaction terms indicated that the interactions had statistically significant contributions to the model. The Hosmer-Lemeshow goodness-of-fit test (χ^2_8=12.57, *P*=.13) indicated that the model fitted with the data very well, which means the observed and predicted values had no statistically significant difference (*P*=.13).

As shown in Table 2, the model revealed that a physician visit was predicted by Web search engine use, age, gender, education, household income, occupation, and self-reported health. Health app use, video search engine use, and social media use did not predict a physician visit.

Table 2. Odds ratios and 95% confidence intervals in the logistic regression analysis of the association between online health information seeking and a physician visit (n=18,197).

Variable	OR (95% CI)	*P* value
Age (per 10-year interval)	1.27 (1.17-1.37)	<.001
Gender		
Female	1.00	
Male	0.60 (0.55-0.66)	<.001
Education		
Primary and secondary	1.00	
Upper secondary	1.02 (0.90-1.15)	.78
Less than 4 years of college	0.98 (0.85-1.12)	.75
Four years or more of college	0.78 (0.68-0.88)	<.001
Household income (US)[a]		
<39 000 US	1.00	
39 000 – 61 000 US	1.28 (1.08-1.52)	.005
61 000 – 83 000 US	1.32 (1.10-1.57)	.002
>83 000 US	1.23 (1.04-1.46)	.02
Occupation		
Full-time work	1.00	
Part-time work	1.26 (1.06-1.50)	.008
Unemployed	0.76 (0.49-1.20)	.23
Housekeeping	0.59 (0.36-1.02)	.047
Retired	1.00 (0.85-1.19)	.96
Student/military service	0.68 (0.37-1.31)	.22
Disability and other family welfare benefits	1.65 (1.34-2.05)	<.001
Web search engine use		
Never	1.00	
Ever	3.69 (2.33-5.83)	<.001
Video search engine use		
Never	1.00	
Ever	0.95 (0.77-1.17)	.63
Social media use		
Never	1.00	
Ever	0.89 (0.75-1.08)	.23
App use		
Never	1.00	
Ever	1.00 (0.84-1.20)	.98
Disease		
Never	1.00	
Ever	1.22 (0.79-1.88)	.38
Self-reported health		
Very bad	1.00	
Bad	0.36 (0.02-1.73)	.32
Neither bad nor good	0.20 (0.01-0.92)	.11

Variable	OR (95% CI)	*P* value
Good	0.12 (0.01-0.57)	.04
Very good	0.07 (0.004-0.33)	.009
Gender, male; app use, ever	1.34 (1.05-1.71)	.02
Age; disease, ever	1.16 (1.07-1.25)	<.001
Age; Web search engine use, ever	0.86 (0.79-0.93)	<.001

[a] One US dollar is approximately 9 Norwegian kr.

The statistically significant interaction between Web search engine use and age indicated that the association between Web search engine use and physician visit was moderated by age. In other words, the OR for Web search engine use was not constant over different ages. Therefore, the meaningful ORs of Web search engine use at different ages needed to be derived from the coefficients of Web search engine use, age, and the interaction term (see [49] for an elaborate description on how the ORs are derived). In general, Web search engine use had lower odds of a physician visit as age increased. For example, at age 40 years, those who used Web search engines had 1.99 (95% CI 1.94-2.02) times higher odds of a physician visit as compared with those who did not use Web search engines. However, at age 80 years, those who used Web search engines had 1.07 (95% CI 1.03-1.12) times higher odds of a physician visit.

Men had lower odds of a physician visit as compared with women. High household income was positively associated with higher odds of a physician visit as compared with household income less than 350,000 Norwegian kroner.

Those with 4 or more years of college education had 22% (95% CI 0.68-0.88) lower odds of a physician visit as compared with those who had a primary or secondary education. However, no difference regarding a physician visit was found for the other education levels.

A physician visit was both positively and negatively predicted depending on the type of occupation. It was positively predicted by part-time work (OR 1.26, 95% CI 1.06-1.50) and disability and other family welfare benefits (OR 1.65, 95% CI 1.34-2.05) as compared with full-time work. Those who listed housekeeping as their occupation had 41% (95% CI 0.36-1.02) lower odds of a physician visit.

Decision to Visit a Physician (n=9071)

The independent variables selected in the multivariable logistic regression model for the decision to visit a physician were age, gender, education, occupation, self-reported health, Web search engine use, video search engine use, social media use, and health app use. The interaction between gender and social media use was statistically significant in the multivariable model. The likelihood ratio test (χ^2_1=4.96, *P*=.02) of the models with and without the interaction term indicated that the interaction had a statistically significant contribution to the model. The Hosmer-Lemeshow goodness-of-fit test (χ^2_8=5.2, *P*=.74) showed that the model's prediction of whether the participants decided to visit a physician according to the health information they read did not significantly differ (*P*=.74) from the actual values reported by the participants.

As shown in Table 3, the model revealed that the decision to visit a physician was predicted by Web search engine use, health app use, video search engine use, social media use, age, gender, education, and occupation.

Table 3. Odds ratios and 95% confidence intervals in the logistic regression analysis of the association between online health information seeking and the decision to visit a physician (n=9071).

Variable	OR (95% CI)	*P* value
Age (per 10-year interval)	0.73 (0.68-0.79)	<.001
Gender		
Female	1.00	
Male	0.87 (0.78-0.98)	.02
Education		
Primary and secondary	1.00	
Upper secondary	1.18 (0.97-1.43)	.11
Less than 4 years of college	1.25 (1.03-1.53)	.03
Four years or more of college	1.34 (1.11-1.62)	.003
Occupation		
Full-time work	1.00	
Part-time work	0.92 (0.76-1.10)	.36
Unemployed	0.83 (0.45-1.45)	.54
Housekeeping	0.95 (0.35-2.26)	.92
Retired	1.11 (0.87-1.42)	.40
Student/military service	0.86 (0.40-1.73)	.69
Disability and other family welfare benefits	1.39 (1.14-1.68)	<.001
Web search engine use		
Never	1.00	
Ever	2.95 (2.03-4.46)	<.001
Video search engine use		
Never	1.00	
Ever	1.43 (1.21-1.70)	<.001
Social media use		
Never	1.00	
Ever	1.43 (1.20-1.69)	<.001
App use		
Never	1.00	
Ever	1.26 (1.13-1.42)	<.001
Self-reported health		
Very bad	1.00	
Bad	1.43 (0.58-4.02)	.46
Neither bad nor good	1.45 (0.60-4.04)	.44
Good	1.27 (0.53-3.53)	.62
Very good	0.96 (0.40-2.70)	.94
Gender, male; social media use, ever	0.70 (0.51-0.96)	.03

Those who used Web search engines had 2.95 times higher odds of deciding to visit a physician as compared with those who did not use Web search engines (95% CI 2.03-4.46). The association of the decision to visit a physician with Web search engine use was higher than the association with other eHealth tools. Those who used health apps had 1.26 times higher odds (95% CI 1.13-1.42) and those who used video search engines had 1.43 times higher odds (95% CI 1.21-1.70) of deciding to visit a physician.

The statistically significant interaction between social media use and gender indicated that the association between the decision to visit a physician and social media use was moderated by gender. In other words, the OR of social media use was not the same for male and female participants. Therefore, the ORs of social media use for male and female participants needed to be derived from the coefficients of social media use, gender, and the interaction term (see [49] for an elaborate description on how ORs are derived). Women who used social media had 1.42 times higher odds of deciding to visit a physician as compared with those who did not use social media (95% CI 1.31-1.55), whereas among men, there was no difference between those who used social media and those who did not use social media.

Similarly, because of the interaction, the meaningful ORs for gender were different between a social media user and nonuser, and they needed to be derived from the coefficients of social media use, gender, and the interaction term. Men who used social media had 39% (95% CI 0.55-0.69) lower odds of deciding to visit a physician as compared with women who used social media, whereas men who did not use social media had 13% (95% CI 0.86-0.88) lower odds as compared with women. A 10-year age increment was associated with 27% (95% CI 0.68-0.79) lower odds of deciding to visit a physician according to the health information found on the internet.

Higher education positively predicted a physician visit. Participants with less than 4 years of college education had 1.25 (95% CI 1.03-1.53) times higher odds of deciding to visit a physician as compared with those having a primary or secondary education, whereas participants with more than 4 years of college education had 1.34 (95% CI 1.11-1.62) times higher odds. Those who received disability and other family welfare benefits had 1.39 (95% CI 1.14-1.68) times higher odds of deciding to visit a physician as compared with full-time workers.

Decision not to Visit a Physician (n=9071)

The independent variables selected in the multivariable logistic regression model for the decision not to visit a physician were age, gender, education, disease, Web search engine use, video search engine use, social media use, and health app use. No statistically significant interactions among the independent variables were found. The Hosmer-Lemeshow goodness-of-fit test (χ^2_8=6.74, *P*=.57) indicated that the model predicted the participants’ decisions not to visit a physician very well.

As shown in Table 4, the model revealed that a physician visit was predicted by Web search engine use, health app use, video search engine use, social media use, age, gender, education, and disease.

Those who used Web search engines had 2.78 times higher odds of deciding not to visit a physician as compared with those who did not use Web search engines (95% CI 1.92-4.18). The association of the decision not to visit a physician with Web search engine use was stronger than the association with other eHealth tools. Those who used health apps had 1.20 times higher odds (95% CI 1.07-1.35), those who used video search engines had 1.27 times higher odds (95% CI 1.07-1.51), and those who used social media had 1.28 times higher odds (95% CI 1.10-1.49) of deciding not to visit a physician.

A 10-year age increment was associated with 22% (95% CI 0.73-0.83) lower odds of deciding not to visit a physician according to the health information found on the internet. Men had 33% lower odds of deciding not to visit a physician as compared with women (95% CI 0.60-0.74).

Table 4. Odds ratios and 95% confidence intervals in the logistic regression analysis of the association between online health information seeking and the decision not to visit a physician (n=9071).

Variable	OR (95% CI)	*P* value
Age (per 10-year interval)	0.78 (0.73-0.83)	<.001
Gender		
Female	1.00	
Male	0.67 (0.60-0.74)	<.001
Education		
Primary and secondary	1.00	
Upper secondary	1.09 (0.90-1.33)	.37
Less than 4 years of college	1.29 (1.07-1.56)	.01
Four years or more of college	1.20 (0.99-1.44)	.06
Web search engine use		
Never	1.00	
Ever	2.78 (1.92-4.18)	<.001
Video search engine use		
Never	1.00	
Ever	1.27 (1.07-1.51)	.006
Social media use		
Never	1.00	
Ever	1.28 (1.10-1.49)	.001
App use		
Never	1.00	
Ever	1.20 (1.07-1.35)	.002
Disease		
No	1.00	
Yes	1.18 (1.05-1.32)	.006

The decision not to visit a physician was positively predicted by education. Participants with less than 4 years of college education had 1.29 times higher odds of deciding not to visit a physician as compared with those having a primary or secondary education (95% CI 1.07-1.56). Those who had one or more diseases in the past year had 1.18 times higher odds of deciding not to visit a physician as compared with those who did not have a disease (95% CI 1.05-1.32).

Discussion

Actual Physician Visit

This study examined the association between a physician visit and health information seeking using eHealth tools (ie, Web search engines, video search engines, social media, and health apps). Our results indicated that a physician visit was positively predicted by health information seeking on Web search engines, which confirms the findings of some prior studies [37-40]. The use of Web search engines had a stronger association with a physician visit possibly because online health information seeking often starts with a Web search engine (ie, Google) [17,40]. However, the association was moderated by age, where the OR decreased as age increased. A physician visit was also positively predicted by higher household income, female gender, and older age, which is in line with the results found in other studies [37,39,50]. A physician visit was both positively and negatively predicted depending on the type of occupation.

A previous study has shown that the difference in physician visits between highly educated and less educated individuals is gradually decreasing in Norway [51]. A recent Norwegian study on patients with diabetes found no difference (95% CI 0.44-5.59) regarding physician visits between highly educated and less educated individuals [44]. In general, our results were in line with existing findings on the lack of a difference in physician visits between highly educated and less educated individuals. However, we found that those who had 4 or more years of college education had lower odds of physician visits as compared with those who had primary and secondary education.

Decision to Visit a Physician

Of 9071 participants who sought health information online, 3121 (34.40%) used the information they read to decide whether they needed to visit a physician. This is lower than the number reported in previous studies [4,8,40]. The difference can possibly be attributed to the fact that the participants included in this study were older than the participants in the other studies. In addition, differences in the health care systems (ie, publicly and privately funded) of the countries could cause varying access to health care services and, consequently, influence the reasons for seeking health information on the internet. It is possible that tax-funded health systems with small or no payment from patients are conducive to the use of traditional face-to-face consultations and to a lower use of eHealth as compared with health systems that are dependent on the financial situation of individuals. A previous study found that 88% of Norwegians prefer to see their GPs face-to-face [52]. Hence, the most common reasons Norwegians seek health information via the internet could be for purposes other than to decide whether to visit a physician.

Studies have shown that people seek health information on the internet for many reasons including deciding whether to visit a physician; preparing for a physician appointment; and reassurance, second opinion, and expanding knowledge on the information received after a physician visit [15,16,18,48]. Thus, the results of our model on a physician visit provided general information on how health information seeking is associated with a physician visit. On the other hand, the models for the decisions to visit a physician and not to visit a physician provided information on how health information seeking is specifically associated with each of these decisions.

Similar studies often use a single variable that measures whether participants decide to visit a physician according to the health information they read on the internet [40]. However, our study used two variables that measure participants' decisions to visit a physician and not to visit a physician. Our results showed that of 9071 participants who sought health information using eHealth tools, 1047 (11.54%) decided both to visit a physician and not to visit a physician in the last year, a finding that would not have been captured by a single question.

Health information obtained from eHealth tools (ie, Web search engines, health apps, video search engines, and social media) positively predicted both the decision to visit a physician and the decision not to visit a physician. The positive associations of eHealth tools with both decisions would not have been revealed with a single model for both decisions. The association between social media use and the decision to visit a physician was moderated by gender. The association of Web search engine use with both decisions was stronger than the associations of other eHealth tools, which may be explained by the fact that online health information seeking often starts with a Web search engine (ie, Google) [17,40].

Each eHealth tool showed a slightly stronger association with the decision to visit a physician than with the decision not to visit a physician. In other words, the odds of the decision to visit a physician according to the health information read online was greater than the odds of the decision not to visit a physician. Therefore, health information seeking via the internet overall slightly increased physician visits, supporting our findings on the positive association between health information seeking via Web search engines and actual physician visits.

Women had higher odds of making a decision according to the health information they read on the internet as compared with men. Receiving disability and other family welfare benefits was also associated with higher odds of deciding to visit a physician as compared with full-time workers. This difference might be attributed to the disease conditions that led to the welfare benefit.

Participants with one or more diseases had higher odds of deciding not to visit a physician according to the health information they read on the internet as compared with those who did not have a disease. Difficulty to interpret medical knowledge found on the internet is considered to increase physician visits [37]. Another study revealed that individuals diagnosed with a medical condition had a positive association with higher knowledge from the health information they read as compared with those who were not diagnosed with a disease [34], which is probably because they were better equipped to understand the health information. Consequently, these participants were able to make a decision according to what they read. These results highlight the potential of eHealth for managing chronic conditions.

Being older and less educated were associated with lower odds of deciding to visit a physician and deciding not to visit a physician according to the health information found on the internet. In other words, older and less educated individuals had lower odds of using the health information they found online to make a decision. These characteristics are known to negatively predict the use of eHealth in general [17,40,45,53]. These participants' lower odds of decision-making according to online health information might be attributed to a lack of necessary skills for navigating the internet, doubts regarding the quality or relevance of the health information available on the internet, or a view that physicians know best [17,45]. These differences signal a need to provide educational support to these population groups in particular.

Many individuals are likely to search the internet for health information to prepare for a physician appointment that has already been decided [8]. Deciding to visit a physician or not to visit a physician according to online health information is a choice individuals may make in connection with each search. There can be many rational reasons to see a physician after finding health information online. For instance, the information can substantiate worries that people may have concerning symptoms or illnesses. In some cases, people can find information suggesting that they are at increased risk of having or developing some illnesses, such as cardiovascular disorders and diabetes [54]. Some disorders can also be screened or diagnosed by the patient with the help of information found online, and this may necessitate further evaluation and treatment by health professionals [55].

However, we also recognize that the decision to visit or not to visit a physician after obtaining online health information is multifactorial. The findings of this study can be understood in light of the theory by Kuhltau [31,32] regarding information searching. The Information Search Process Model describes information seeking in six phases. It sees the process as complex and takes into consideration both cognitive and emotional factors, recognizing that information searchers can experience a range of feelings throughout the process, including optimism, satisfaction, confusion, and disappointment.

In this study, overall, we found that online health information is associated with increased use of traditional health services (ie, visiting a physician). By drawing on the Information Search Process Model [31,32], we understand that simply finding the relevant information may not be the endpoint of the process and that many online health information searchers will have feelings, such as confusion and frustration, even after finding the information they were looking for online. The decision to visit a physician following online health information searching may therefore also, in part, be based on emotional factors. People with health anxiety are more likely to visit a physician after finding health information online [33], supporting the idea that emotions also play an important role in the decision-making process.

Implications

We examined the importance of some central factors, such as demographic characteristics and health status, in the association between online health information searching and the decision to visit or not to visit a physician. However, drawing on the theory about health information searching [28-32], we emphasized the complexity of our topic, pointing out that the decision to visit a physician or not to visit a physician is based on a range of factors that we lack information about, including cognitive factors and emotional factors.

We believe that online health information has its place as part of a well-functioning health service and that much can be done to further improve the availability of quality online health information that will benefit patients and the general public [56,57]. Although we believe that this study has made a great contribution to the literature, many questions remain, and they should be addressed in future studies.

Future Research

There is a need for more studies related to this topic. For instance, studies that include psychological data about participants could provide important insights into the relevance of cognitive and emotional factors regarding how health information searchers react with respect to seeing a physician following health information searching [33]. Further studies are needed to examine what aspects of information sources contribute to deciding either to visit a physician or not to visit a physician. Issues, such as emotional content, understandability, usability, and readability, could be of importance [56]. Future studies could also address how health information searching influences other health-related behaviors, such as posting about health and illness on social media. Studies combining data on health information seeking via the internet and outcomes of physician visits are also needed to investigate whether access to eHealth might be conducive to unnecessary physician visits.

Strengths and Limitations

This study is based on cross-sectional data, with a large sample size and relatively good response rate. The invitation for participation was sent to all inhabitants of Tromsø aged 40 years or above by mail, which contributes to the representativeness of the study sample. Although prior studies have shown that the reproducibility and validity of self-reported findings from the Tromsø Study are quite high [58,59], there is still a possibility of recall bias, which may affect the validity of the results. High proportions of nonrespondents were men and older individuals, and people with one or more diseases and those who had physician visits were overrepresented among the participants [60].

Both the decision to search for health information on the internet and the decision to visit a physician or not to visit a physician are likely to be influenced by a range of different factors, including cognitive factors and emotional factors. Similarly, how people react to the use of eHealth services and traditional face-to-face services will be influenced by many different factors. We have taken some of the factors, such as demographic factors and health status, into account in this study regarding the relationship between online health information searching and physician visits, but there was a lack of information regarding many other variables that might be of interest. For instance, there was a lack of data on health anxiety that may moderate the relationship between health information seeking and the decision of whether to visit a physician.

The cross-sectional study design did not make it possible to establish causality, and the results might be affected by unmeasured confounding variables. With the current widespread use of the internet on smartphones and tablets, the difference between the use of search engines and apps may be blurry for some participants, which may affect the results.

Conclusions

In this study, we examined the association between health information seeking on the internet with a physician visit, making a decision to visit a physician, and making a decision not to visit a physician. We found that both the decision to visit a physician and the decision not to visit a physician were positively predicted by searching for health information online. However, searching for health information on the internet was associated with a resultant increase in physician visits. The implication of this finding is that for our participants, overall, online health information did not replace or reduce the need for traditional face-to-face health services. As such, online health information does not, at least in our study, stand out as a means for saving resources or resolving the demands on traditional resource-strained health services.

Acknowledgments

We thank the organizers, management, and technical staff of Tromsø 7 for their valuable work in preparing and collecting the data. Above all, we thank the residents of Tromsø. Their willingness to participate was fundamental to our research. This research was made possible by a grant from the Research Council of Norway to the Norwegian Centre for E-health Research, University Hospital of North Norway (Grant No 248150/O70) and UiT The Arctic University of Norway. The publication charge for this article was funded by a grant from the Publication Fund of UiT The Arctic University of Norway.

Conflicts of Interest

None declared.

References

1. Statistics Norway. 2018. ICT usage in households URL: https://www.ssb.no/en/statbank/list/ikthus [accessed 2018-07-25]
2. Fox S. Pew Research Center. 2004. Older Americans and the Internet URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2004/PIP_Seniors_Online_2004.pdf.pdf [accessed 2018-07-25]
3. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki CE, Prokosch H, Wynn R. Informed citizen and empowered citizen in health: results from an European survey. BMC Fam Pract 2011;12:20 [FREE Full text] [doi: 10.1186/1471-2296-12-20] [Medline: 21496309]
4. Fox S, Duggan M. Pew Research Center. 2013. Health Online 2013 URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2020-02-06]
5. Cline R, Haynes K. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
6. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [doi: 10.3322/canjclin.53.6.356] [Medline: 15224975]
7. Andreassen HK, Wangberg SC, Wynn R, Sørensen T, Hjortdahl P. [Health-related use of the Internet in the Norwegian population]. Tidsskr Nor Laegeforen 2006 Nov 16;126(22):2950-2952 [FREE Full text] [Medline: 17117194]
8. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007 Apr 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
9. Wangberg S, Andreassen H, Kummervold P, Wynn R, Sørensen T. Use of the internet for health purposes: trends in Norway 2000-2010. Scand J Caring Sci 2009 Dec;23(4):691-696. [doi: 10.1111/j.1471-6712.2008.00662.x] [Medline: 19473317]
10. Rigby M, Koch S, Keeling D, Hill P, Alonso A. Europa. 2013. Developing a New Understanding of Enabling Health and Wellbeing in Europe Internet URL: https://ec.europa.eu/eip/ageing/library/developing-new-understanding-enabling-health-and-wellbeing-europeharmonising-health-and_en [accessed 2020-02-06]
11. Kummervold PE, Wynn R. Health information accessed on the internet: the development in 5 European countries. Int J Telemed Appl 2012;2012:297416-297413 [FREE Full text] [doi: 10.1155/2012/297416] [Medline: 23304133]
12. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
13. Fox S. Pew Research Center. 2011. The Social Life of Health Information Internet, 2011 URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2020-02-06]
14. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cogent Social Sciences 2017 Mar 13;3(1). [doi: 10.1080/23311886.2017.1302785]
15. Fox S, Lee R. Pew Research Center. 2000. The online health care revolution: How the Web helps Americans take better care of themselves URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2000/PIP_Health_Report.pdf.pdf [accessed 2015-03-20]
16. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
17. Chu JT, Wang MP, Shen C, Viswanath K, Lam TH, Chan SS. How, When and Why People Seek Health Information Online: Qualitative Study in Hong Kong. Interact J Med Res 2017 Dec 12;6(2):e24 [FREE Full text] [doi: 10.2196/ijmr.7000] [Medline: 29233802]
18. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. BMJ 2004 Mar 6;328(7439):564 [FREE Full text] [doi: 10.1136/bmj.328.7439.564] [Medline: 15001506]
19. Vambheim SM, Wangberg SC, Johnsen JK, Wynn R. Language use in an internet support group for smoking cessation: development of sense of community. Inform Health Soc Care 2013 Jan;38(1):67-78. [doi: 10.3109/17538157.2012.710685] [Medline: 22958060]
20. 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: 10.1177/014107680309600206] [Medline: 12562977]
21. Slev VN, Mistiaen P, Pasman HR, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: A meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]

22. Robinson TN, Patrick K, Eng TR, Gustafson D. An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. JAMA 1998 Oct 14;280(14):1264-1269. [doi: 10.1001/jama.280.14.1264] [Medline: 9786378]
23. 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]
24. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? BMJ 2014 Oct 14;349:g6178. [doi: 10.1136/bmj.g6178] [Medline: 25315514]
25. Wynn R, Oyeyemi SO, Johnsen JK, Gabarron E. Tweets are not always supportive of patients with mental disorders. Int J Integr Care 2017 Jul 11;17(3):149. [doi: 10.5334/ijic.3261]
26. Gabarron E, Serrano JA, Wynn R, Lau AY. Tweet content related to sexually transmitted diseases: no joking matter. J Med Internet Res 2014 Oct 06;16(10):e228 [FREE Full text] [doi: 10.2196/jmir.3259] [Medline: 25289463]
27. Llanwarne N, Newbould J, Burt J, Campbell JL, Roland M. Wasting the doctor's time? A video-elicitation interview study with patients in primary care. Soc Sci Med 2017 Mar;176:113-122 [FREE Full text] [doi: 10.1016/j.socscimed.2017.01.025] [Medline: 28135690]
28. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
29. Miller SM. Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. Journal of Personality and Social Psychology 1987;52(2):345-353. [doi: 10.1037/0022-3514.52.2.345]
30. Miller SM. Monitoring versus blunting styles of coping with cancer influence the information patients want and need about their disease. Implications for cancer screening and management. Cancer 1995 Jul 15;76(2):167-177. [doi: 10.1002/1097-0142(19950715)76:2<167::aid-cncr2820760203>3.0.co;2-k] [Medline: 8625088]
31. Kuhlthau C. Seeking Meaning: A Process Approach To Library And Information Services, 2nd Edition (Libraries Unlimited Guided Inquiry). Santa Barbara, CA: Libraries Unlimited; 2020.
32. Kuhlthau CC. The role of experience in the information search process of an early career information worker: Perceptions of uncertainty, complexity, construction, and sources. J Am Soc Inf Sci 1999;50(5):399-412. [doi: 10.1002/(sici)1097-4571(1999)50:5<399::aid-asi3>3.0.co;2-l]
33. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):613-618. [doi: 10.1089/cyber.2010.0425] [Medline: 21548797]
34. Budrionis A, Wynn R, Marco-Ruiz L, Yigzaw KY, Bergvik S, Oyeyemi S, et al. Impact of the Use of Electronic Health Tools on the Psychological and Emotional Well-Being of Electronic Health Service Users (The Seventh Tromsø Study - Part 3): Population-Based Questionnaire Study. J Med Internet Res 2020 Mar 05;22(3):e13118 [FREE Full text] [doi: 10.2196/13118]
35. Marco-Ruiz L, Bønes E, de la Asunción E, Gabarron E, Aviles-Solis JC, Lee E, et al. Combining multivariate statistics and the think-aloud protocol to assess Human-Computer Interaction barriers in symptom checkers. J Biomed Inform 2017 Oct;74:104-122 [FREE Full text] [doi: 10.1016/j.jbi.2017.09.002] [Medline: 28893671]
36. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google Is Here to Stay but Health Care Professionals Are Still Valued: An Analysis of Health Care Consumers' Internet Navigation Support Preferences. J Med Internet Res 2017 Jun 14;19(6):e210 [FREE Full text] [doi: 10.2196/jmir.7489] [Medline: 28615156]
37. Lee C. Does the internet displace health professionals? J Health Commun 2008;13(5):450-464. [doi: 10.1080/10810730802198839] [Medline: 18661387]
38. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]
39. Hsieh RW, Chen L, Chen T, Liang J, Lin T, Chen Y, et al. The Association Between Internet Use and Ambulatory Care-Seeking Behaviors in Taiwan: A Cross-Sectional Study. J Med Internet Res 2016 Dec 07;18(12):e319 [FREE Full text] [doi: 10.2196/jmir.5498] [Medline: 27927606]
40. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
41. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
42. Robinson TN. Community health behavior change through computer network health promotion: preliminary findings from Stanford Health-Net. Computer Methods and Programs in Biomedicine 1989 Oct;30(2-3):137-144. [doi: 10.1016/0169-2607(89)90065-5]
43. Spoelman WA, Bonten TN, de Waal MW, Drenthen T, Smeele IJ, Nielen MM, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Nov 09;6(11):e013166 [FREE Full text] [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]
44. Hansen AH, Broz J, Claudi T, Årsand E. Relations Between the Use of Electronic Health and the Use of General Practitioner and Somatic Specialist Visits in Patients With Type 1 Diabetes: Cross-Sectional Study. J Med Internet Res 2018 Nov 07;20(11):e11322 [FREE Full text] [doi: 10.2196/11322] [Medline: 30404766]

45. Wynn R, Oyeyemi S, Budrionis A, Marco-Ruiz L, Yigzaw K, Bellika JG. Electronic Health Use in a Representative Sample of 18,497 Respondents in Norway (The Seventh Tromsø Study - Part 1): Population-Based Questionnaire Study. JMIR Med Inform 2020 Mar 05;22(3):e13106 [FREE Full text] [doi: 10.2196/13106]
46. Marco-Ruiz L, Wynn R, Oyeyemi S, Budrionis A, Yigzaw KY, Bellika JG. Impact of Illness on Electronic Health Use (The Seventh Tromsø Study - Part 2): Population-Based Questionnaire Study. J Med Internet Res 2020 Mar 05;22(3):e13116 [FREE Full text] [doi: 10.2196/13116]
47. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njølstad I. Cohort profile: the Tromso Study. Int J Epidemiol 2012 Aug;41(4):961-967 [FREE Full text] [doi: 10.1093/ije/dyr049] [Medline: 21422063]
48. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015 Jan 08;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
49. Hosmer D, Lemeshow S, Sturdivant R. Applied Logistic Regression, 3rd Edition. Hoboken, NJ: John Wiley & Sons, Inc; 2013:978.
50. Nguyen HT, Nakamura K, Seino K, Vo VT. Association Between a Wider Availability of Health Information and Health Care Utilization in Vietnam: Cross-Sectional Study. J Med Internet Res 2017 Dec 18;19(12):e405 [FREE Full text] [doi: 10.2196/jmir.8328] [Medline: 29254911]
51. Vikum E, Bjørngaard JH, Westin S, Krokstad S. Socio-economic inequalities in Norwegian health care utilization over 3 decades: the HUNT Study. Eur J Public Health 2013 Dec;23(6):1003-1010. [doi: 10.1093/eurpub/ckt053] [Medline: 23729479]
52. Sørensen T, Andreassen H, Wangberg S. eHealth Research. 2013. E-helse i Norge 2013 URL: https://ehealthresearch.no/files/documents/Rapporter/NST-rapport_2014-02_e-helse_i_Norge_2013.pdf [accessed 2018-07-25]
53. Ginossar T. Predictors of Online Cancer Prevention Information Seeking Among Patients and Caregivers Across the Digital Divide: A Cross-Sectional, Correlational Study. JMIR Cancer 2016 Mar 09;2(1):e2 [FREE Full text] [doi: 10.2196/cancer.5108] [Medline: 28410177]
54. Stiglic G, Pajnkihar M. Evaluation of Major Online Diabetes Risk Calculators and Computerized Predictive Models. PLoS One 2015;10(11):e0142827 [FREE Full text] [doi: 10.1371/journal.pone.0142827] [Medline: 26560153]
55. Vetter S, Rossegger A, Elbert T, Gerth J, Urbaniok F, Laubacher A, et al. Internet-based self-assessment after the Tsunami: lessons learned. BMC Public Health 2011 Jan 07;11:18 [FREE Full text] [doi: 10.1186/1471-2458-11-18] [Medline: 21214894]
56. Beaunoyer E, Arsenault M, Lomanowska AM, Guitton MJ. Understanding online health information: Evaluation, tools, and strategies. Patient Educ Couns 2017 Feb;100(2):183-189. [doi: 10.1016/j.pec.2016.08.028] [Medline: 27595436]
57. Tian C, Champlin S, Mackert M, Lazard A, Agrawal D. Readability, suitability, and health content assessment of web-based patient education materials on colorectal cancer screening. Gastrointest Endosc 2014 Aug;80(2):284-290. [doi: 10.1016/j.gie.2014.01.034] [Medline: 24674352]
58. Engstad T, Bonaa KH, Viitanen M. Validity of self-reported stroke: The Tromso Study. Stroke 2000 Jul;31(7):1602-1607. [doi: 10.1161/01.str.31.7.1602] [Medline: 10884460]
59. Lundblad MW, Jacobsen BK. The reproducibility of self-reported age at menarche: The Tromsø Study. BMC Womens Health 2017 Aug 22;17(1):62 [FREE Full text] [doi: 10.1186/s12905-017-0420-0] [Medline: 28830397]
60. Statistics Norway. 2018. GPs and emergency primary health care URL: https://www.ssb.no/en/statbank/table/09535/ [accessed 2018-07-25]

Abbreviations

eHealth: electronic health
GP: general practitioner
Tromsø 7: the seventh survey of the Tromsø Study

Edited by G Eysenbach; submitted 12.12.18; peer-reviewed by MP Wang, T Ginossar, K Nakamura, D Sent, FC Wang; comments to author 08.04.19; revised version received 03.09.19; accepted 16.12.19; published 05.03.20.

Please cite as:
Yigzaw KY, Wynn R, Marco-Ruiz L, Budrionis A, Oyeyemi SO, Fagerlund AJ, Bellika JG
The Association Between Health Information Seeking on the Internet and Physician Visits (The Seventh Tromsø Study - Part 4): Population-Based Questionnaire Study
J Med Internet Res 2020;22(3):e13120
URL: https://www.jmir.org/2020/3/e13120
doi:10.2196/13120
PMID:32134387

Original Paper

Assessment of the Frequency of Online Searches for Symptoms Before Diagnosis: Analysis of Archival Data

Irit Hochberg[1,2], MD, PhD; Raviv Allon[2], BSc; Elad Yom-Tov[3,4], PhD

[1]Institute of Endocrinology, Diabetes, and Metabolism, Rambam Health Care Campus, Haifa, Israel
[2]Bruce Rappaport Faculty of Medicine, Technion - Israel Institute of Technology, Haifa, Israel
[3]Microsoft Research, Herzeliya, Israel
[4]Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa, Israel

Corresponding Author:
Elad Yom-Tov, PhD
Microsoft Research
13 Shenkar St
Herzeliya, 46733
Israel
Phone: 972 747111359
Email: eladyt@yahoo.com

Abstract

Background: Surveys suggest that a large proportion of people use the internet to search for information on medical symptoms they experience and that around one-third of the people in the United States self-diagnose using online information. However, surveys are known to be biased, and the true rates at which people search for information on their medical symptoms before receiving a formal medical diagnosis are unknown.

Objective: This study aimed to estimate the rate at which people search for information on their medical symptoms before receiving a formal medical diagnosis by a health professional.

Methods: We collected queries made on a general-purpose internet search engine by people in the United States who self-identified their diagnosis from 1 of 20 medical conditions. We focused on conditions that have evident symptoms and are neither screened systematically nor a part of usual medical care. Thus, they are generally diagnosed after the investigation of specific symptoms. We evaluated how many of these people queried for symptoms associated with their medical condition before their formal diagnosis. In addition, we used a survey questionnaire to assess the familiarity of laypeople with the symptoms associated with these conditions.

Results: On average, 15.49% (1792/12,367, SD 8.4%) of people queried about symptoms associated with their medical condition before receiving a medical diagnosis. A longer duration between the first query for a symptom and the corresponding diagnosis was correlated with an increased likelihood of people querying about those symptoms (rho=0.6; *P*=.005); similarly, unfamiliarity with the association between a condition and its symptom was correlated with an increased likelihood of people querying about those symptoms (rho=−0.47; *P*=.08). In addition, worrying symptoms were 14% more likely to be queried about.

Conclusions: Our results indicate that there is large variability in the percentage of people who query the internet for their symptoms before a formal medical diagnosis is made. This finding has important implications for systems that attempt to screen for medical conditions.

(J Med Internet Res 2020;22(3):e15065) doi:10.2196/15065

KEYWORDS

search engines; diagnosis; screening

Introduction

Online self-diagnosis of health conditions is a well-known phenomenon that has grown substantially with ease of access to medical information facilitated by the internet and mobile technologies [1,2]. A large survey found that more than one-third of Americans self-diagnose when they encounter a health problem [3], and another study indicated that about 70% of American adults consult the internet for a variety of medical information [4].

The prevalence of self-diagnosis is leading countries and large epidemiologic centers to use the available information for public health goals [5,6]. Epidemics such as influenza and dengue fever have been tracked by observing the number of people who query internet search engines for the symptoms of these diseases [7,8]. A recent study showed the potential of identifying serious medical conditions, such as cervical and ovarian cancers, from people's searches on online search engines [9]. These results suggest that search data could be used as a novel screening tool.

Nevertheless, utilizing search engines as an effective screening tool requires an accurate characterization of how people use search engines for self-diagnosis. In addition, conditions need to be independently characterized to understand the type and number of people who are searching for information and the common words used for these searches. It is currently not known how commonly people conduct an online search for their condition before diagnosis by a health professional. Determining this will provide an important indication of the percentage of people for whom online data screening is applicable and the diseases for which internet-based screening is effective. The purpose of this study was to characterize the prevalence and content of searches made by users before a medical diagnosis by a health care professional.

In this work, we analyzed data from search engine users who self-identified their diagnosis of a medical condition and traced back the data to determine how many of these instances could have been predicted by an earlier search for the signs and symptoms of the condition by that same user.

In addition, to better understand our search data results, we analyzed the association between the frequency of internet inquiries and the population's general knowledge regarding certain conditions and their symptoms. We hypothesized that people inquire more about symptoms that they do not recognize or cannot associate with a certain disease.

Methods

Search Data

We extracted all queries made on Microsoft Bing in English by people in the United States between May 1, 2017, and April 30, 2018. For each user, we recorded an anonymized username, the time and date of the query, and the text used in the query. We focused on 20 medical conditions that are known to have evident symptoms, not systematically screened, not usually diagnosed in asymptomatic individuals by usual medical tests, and generally diagnosed after the investigation of specific symptoms. To ensure statistical power and validity, we limited the analysis to conditions for which at least 75 people self-identified their condition. The 20 conditions used in this study are listed in Table 1.

The population of self-identifying users was defined as those people who made a diagnosis ascertainment query (DAQ), indicating that they had been formally diagnosed with 1 of the 20 conditions analyzed in this study (eg, “I was diagnosed with COPD” or “I have COPD”). Queries that indicated the possibility of such a condition (eg, “do I have COPD”) were excluded. Specifically, DAQs were defined as queries that matched the phrases “I have” or “diagnosed with” and the name of the condition and excluded queries that contained any of the phrases “do I have,” “can I have,” “I think I have,” “did I have,” “nurse,” “patient,” “cat,” “dog,” “wife,” “husband,” “son,” or “daughter.”

For each condition, we calculated the fraction of people who queried about a relevant symptom before their first mention of the condition in the DAQ. The list of relevant symptoms was defined by two authors (IH and RA, both medical doctors) and enhanced using the synonym list developed by Yom-Tov and Gabrilovich [10].

In addition, symptoms were mapped to their perceived Medical Severity Rank (MSR), a validated measure of their apparent importance to both medical specialists and laypeople [11]. MSR measures the urgency of a symptom as perceived by people, from a symptom that requires immediate urgent care (MSR=1) to one that can be disregarded (MSR=10).

DAQs do not usually provide an indication of when the diagnosis was made. To estimate whether the DAQs are typically made around the time of diagnosis or throughout a person's illness, we assessed if the time of the DAQ corresponded closely to the time of the first queries for hospitals, medical centers, or clinics. This analysis was conducted for each user indicating a diagnosis of 1 of the 3 malignant conditions (endometrial cancer, esophageal cancer, and lymphoma), where a hospital visit is usually required soon after the initial diagnosis is made.

Table 1. Conditions analyzed, symptoms associated with them, and the percentage of people who asked about the symptoms before their first query indicating that they have the condition.

Condition	Symptoms	People with symptoms[a]	
		n/N	%
Degenerative disc disease	Back pain, leg weakness, leg pain, leg numbness, leg tingling, loss of bowel control, and loss of bladder control	29/165	17.6
Chronic obstructive pulmonary disorder	Chronic cough, shortness of breath, dyspnea, recurrent pneumonia, wheezing, and dystonia	44/538	8.4
Menopause	Hot flash, night sweat, vaginal dryness, and alopecia	34/308	11.4
Heart failure	Shortness of breath, dyspnea, chronic cough, leg edema, leg swelling, rapid weight gain, and fatigue	116/896	12.9
Gout	Pain, tenderness, swelling, inflammation, and redness	520/2052	25.34
Ulcerative colitis	Diarrhea, abdominal pain, bloody bowel movement, rectal bleeding, tenesmus, lack of appetite, and fatigue	70/324	21.9
Bladder cancer	Blood in urine, hematuria, blood clots in urine, pain or burning sensation during urination, frequent urination, and not able to pass urine	33/302	11.3
Parkinson disease	Tremor and bradykinesia	144/3613	4.01
Endometrial cancer	Discharge and bleeding	25/110	22.7
Crohn disease	Diarrhea, blood in stool, fatigue, abdominal pain, cramping, mouth sores, reduced appetite, weight loss, and fistula	146/894	16.4
Angina pectoris or coronary heart disease	Chest pain, chest pressure, chest tightness, shortness of breath, and heartburn	239/775	30.8
Grave’s disease	Anxiety, irritability, heat sensitivity, increased sweating, weight loss, enlargement of thyroid or goiter, frequent bowel movements, diarrhea, bulging eyes, rapid heartbeat, rapid pulse, irregular heartbeat, and atrial fibrillation	112/479	23.4
Esophageal cancer	Difficulty or pain while swallowing solid food, vomiting, choking on food, heartburn, chest pressure, weight loss, coughing, and hoarseness	29/119	24.4
Lymphoma	Enlarged lymph nodes, night sweats, weight loss, intermittent fever, and fatigue	177/849	21.0
Plantar fasciitis	Foot pain and heel pain	10/212	5.2
Cellulitis	Painful area of skin, leg, foot, hand, or face; skin erythema or redness; skin edema; hot skin; and dropsy	4/278	1.8
Prostatitis	Painful, difficult, or frequent urination, blood in urine, groin pain, rectal pain, abdominal pain, low back pain, malaise, body aches, urethral discharge, and painful ejaculation or sexual dysfunction	12/103	12.6
Mastitis	Breast tenderness, pain or burning sensation, breast warmth or redness, breast swelling or thickening, breast lump, breast pain or burning, and malaise	10/90	12.2
Bell’s palsy	Facial paralysis on one side; drooping of the mouth to one side; asymmetrical mouth movement or smile; loss of blinking on one side; decreased or increased tearing; altered sense of taste; slurred speech; drooling; difficulty eating, drinking, or chewing; and pain or numbness behind the ear	6/168	3.6
Mononucleosis	Sore throat, malaise, headache, loss of appetite, myalgia, muscle pain, chills, and nausea	21/91	22.8

[a] Average=15.49% (1792/12,367).

Survey Data

To estimate whether laypeople recognize the investigated medical conditions and their symptoms, we conducted a survey using the crowdsourcing platform CrowdFlower. We randomly selected 104 actual condition and symptom pairs and created another 156 random pairs. The latter were created by randomly matching a condition and a symptom and then verifying that the symptom is not manifested in the selected condition. A total of 10 crowdsourced workers were asked to answer, for each of the 260 pairs, whether they recognized the name of the medical condition and whether they thought that the given symptom could be the sign of the condition.

Laypeople’s knowledge about certain conditions and their symptoms, as determined through the surveys, was compared with the rate of searches for these symptoms in our search data.

Statistics

Data analysis was conducted using MATLAB version 9.4.0. Spearman correlation was used to evaluate associations. The level of significance was set as 5% (*P*<.05). The Institutional

Review Board of the Technion—Israel Institute of Technology approved this study.

Results

On average, 618 people (mean SE 189) self-identified their diagnosis from 1 of the 20 analyzed conditions by making a DAQ. The users asked about each symptom, on average, 1.7 times. Table 1 lists the 20 conditions analyzed, the symptoms that were designated as being associated with that condition, and the percentage of people who queried about their symptoms before a formal diagnosis was made. On average, 15.49% (1792/12,367, SD 8.4%) of people queried about their symptoms before receiving a formal medical diagnosis.

There was a significant correlation between the percentage of people querying about a condition and the median number of days between the first symptom query and the DAQ (rho=0.60; *P*=.005; number of conditions=20). Thus, patients with conditions that had a longer duration between the onset of symptoms and the time of diagnosis were more likely to query for symptoms before they received a medical diagnosis.

We labeled each condition according to the lowest (most worrying) MSR for the related symptoms. We focused on the most worrying symptoms because of prior work, which shows that people most often recall the worst experience of pain [12,13], and thus, we hypothesized that people would be driven to search for the most worrying symptoms. A total of 4 conditions were excluded from this calculation because they had no symptoms for which Youngmann and Yom-Tov [11] provided an MSR. We found that more people inquired about the symptoms of a condition before a formal medical diagnosis (429/2328, 19.13%) for conditions that had an MSR≤2 as compared to those that had an MSR>2 (1382/10,107, 16.81%). Thus, people tended to query 14% more for diseases with more worrying symptoms (lower MSR) than those with less distressing symptoms.

A total of 4636 condition-symptom pairs were evaluated by crowdsourced workers in a survey aimed at estimating whether laypeople recognize the 20 investigated medical conditions and their symptoms. The responders reported recognizing the condition in 86% (3987/4636) of the pairs presented to them. The least recognized conditions were plantar fasciitis (40%, 16/40) and chronic obstructive pulmonary disorder (29%, 132/462). As described in the Methods section, our survey included actual disease-symptom pairs and sham disease-symptom pairs. For the real pairs, 86% (3987/4636) of responses correctly identified the pair as being associated. A similar level of success was observed for the sham pair, with 74% of the responses correctly indicating that the symptom was not associated with the disease.

Spearman correlation between the percentage of people querying for the symptom of a condition on Bing and the percentage of people who correctly recognized its symptoms in the survey was −0.47 (*P*=.08; number of conditions=14). This means that people tend to query more for conditions with less recognizable symptoms.

To estimate whether the DAQs give an accurate indication of the time of initial diagnosis, we compared the timing of the DAQ with the first query about a hospital in the three malignant conditions. These specific conditions were analyzed because a hospital visit is usually required soon after the initial diagnosis is made. Table 2 shows the median time from the first query for a hospital or clinic until the first DAQ, the percentage of queries for hospitals made before the DAQ, and the percentage of people querying for a hospital. Figure 1 shows the distribution of the number of days between a query for a hospital and a query indicating a formal diagnosis, for all three conditions. There was no significant difference in time between queries for hospitals and queries for clinics (*P*>.05; rank sum test). As shown in Table 2 and Figure 1, there is only a short lag between the queries for a hospital and queries indicating a formal medical diagnosis. Indeed, 67% of queries for hospitals were within a window of 2 weeks before or after the DAQ (47% within a week around the DAQ). Figure 1 shows a clear peak for the first hospital searches occurring on the same day as the DAQ (19% of queries). Thus, queries indicative of diagnosis, specifically DAQs, correspond to the time of the actual diagnosis.

Table 2. The time from the first query for a hospital and the first query indicating the condition, the percentage of times the hospital query occurred before the diagnosis query, and the percentage of people who queried for a hospital.

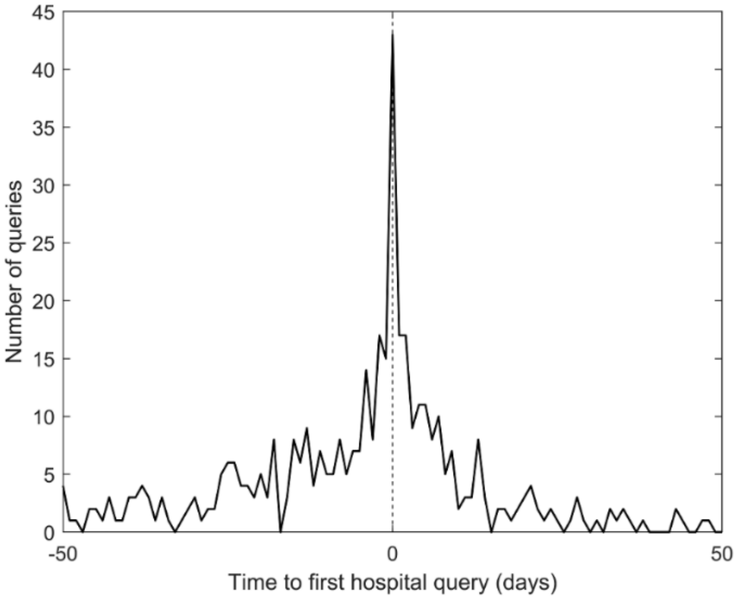
Condition	Median time from hospital query to diagnosis query (days)[a]	Queries for hospitals made before queries for diagnosis[b]		People who inquired about a hospital[c]	
		n/N	%	n/N	%
Endometrial cancer	2.5	46/72	64	72/110	65
Esophageal cancer	14.3	45/66	68	66/119	55
Lymphoma	6.6	304/478	63.6	478/849	56.3

[a] Average=7.8.

[b] Average=65% (395/616).

[c] Average=59% (616/1078).

Figure 1. The time from the first query about a hospital and the time of the first query indicating the condition. Negative times indicate that the first query to a hospital was made prior to the first query about the disease. The figure shows data for endometrial cancer, esophageal cancer, and lymphoma.



Discussion

Internet interventions for public health have increased dramatically in the past decade, with multiple interventions across a wide range of conditions and populations. Following rapid developments in the last few years, it has been suggested that studies on internet interventions in the current decade will determine both the effectiveness and potential of such online interventions on public health [14].

It is well known that people search for symptoms on the internet before consulting a medical professional [15]. However, there has been no detailed characterization of when or what people inquire about when symptoms appear.

We found that an average of 15.5% of people queried for their symptoms before receiving a medical diagnosis from a health care professional, and we found high variability in this figure among the 20 conditions analyzed in our study. For instance, only 1.8% of patients diagnosed with cellulitis searched for a painful red area of skin before a medical diagnosis was made, compared with 30.8% of patients diagnosed with coronary heart disease who searched for chest pain or shortness of breath before receiving a medical diagnosis.

Our results showed that more people search for symptoms when the time between that first search to a formal diagnosis is longer (Spearman rho=0.6). One possible explanation for this correlation is that a long period of diagnosis gives an opportunity for more people to ask online about the condition. Another possible explanation is that people who search the internet for symptoms tend to do more thinking and consulting before approaching a medical professional, thus delaying their diagnosis. If the latter is true, better information needs to be provided to information seekers when it is likely that they have an acute condition that needs prompt treatment, emphasizing the need to enable internet-based screening for certain conditions.

The finding that most people inquire about hospitals around the time of making queries that suggest a diagnosis supports our assumption that such queries are made around the time of actual diagnosis. This concurs with past studies [16] that found that the number of people querying for different types of cancer were correlated with the incidence of cancers, not prevalence.

The survey results show that the conditions we studied and the match between the condition and symptom are well recognized. However, some conditions were less recognized than others. By comparing our survey results and our search data results, we showed a negative correlation between the rate of inquiry and the knowledge about a condition, suggesting that the less people know about a condition, the more they query for its symptoms. This result was not statistically significant (*P*=.08), but there is clearly a pattern that could illuminate one of the basic motives for people to first approach the Web when symptoms occur.

Our study limitations are inherent in a search engine data study, including the dependence on the user's declaration of diagnosis, which assumes the exclusion of healthy people searching for diseases out of general curiosity. The main limitation is that most users do not declare the diagnosis in a search and therefore cannot be identified as patients. Moreover, users who self-declare their diagnosis are known to be a biased sample of patients, comprising relatively more females and younger people [17], a bias caused by a preference for query length. Thus, the prior rate of queries could be more heavily reflective of these population segments than that of the general population.

Although we have strived to find a comprehensive list of symptoms for each condition (including synonyms thereof), some symptoms could have been missed, especially colloquial references to symptoms, and these, if included, could have increased the reported percentage of searches for symptoms before the formal medical diagnosis. In addition, as our observation window is finite (1 year), people might have queried for their symptoms before the beginning of the data period.

Such searches might increase the reported fraction of people performing searches for disease by approximately 1/12, meaning that the average percentage of people conducting an online search for a disease would be approximately 16.8%.

Finally, a symptom could be related to multiple underlying conditions. Although we have tried to focus on conditions with clear and distinct symptoms, such cases could have skewed our estimate for the search rates.

Despite these shortcomings, the strength of the study is significant; it is based on a diverse cohort made available because most people in industrialized countries now have access to the internet and use search engines as their primary data source when seeking health-related information [4]. This, together with our new understanding of when and how people seek information on symptoms, may enable future systems for screening of serious medical conditions from internet data, in general, and search queries, in particular, thereby overcoming the barriers of health illiteracy, unfamiliarity with medical conditions, and difficult access to the health system. Our hope is that such systems will enable earlier diagnosis of many serious medical conditions.

Conflicts of Interest

EY is an employee of Microsoft, the owner of the Bing search engine.

References

1. Avery N, Ghandi J, Keating J. The 'Dr Google' phenomenon--missed appendicitis. N Z Med J 2012 Dec 14;125(1367):135-137. [Medline: 23321891]
2. Robertson N, Polonsky M, McQuilken L. Are my symptoms serious Dr Google? A resource-based typology of value co-destruction in online self-diagnosis. Australas Mark J 2014;22(3):246-256. [doi: 10.1016/j.ausmj.2014.08.009]
3. Kuehn BM. More than one-third of US individuals use the internet to self-diagnose. J Am Med Assoc 2013 Feb 27;309(8):756-757. [doi: 10.1001/jama.2013.629] [Medline: 23443421]
4. Fox S, Duggan M. Pew Internet - Pew Research Center. 2013. Health Online 2013 URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2020-01-13]
5. Paul MJ, Dredze M. Social monitoring for public health. Synthesis Lect Inf Concepts Retr Serv 2017;9(5):1-183. [doi: 10.2200/s00791ed1v01y201707icr060]
6. Yom-Tov E, Cox IJ, Lampos V. Learning About Health and Medicine from Internet Data. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 Presented at: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining; February 2 - 6, 2015; Shanghai, China p. 417-418. [doi: 10.1145/2684822.2697042]
7. 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]
8. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis 2011 May;5(5):e1206 [FREE Full text] [doi: 10.1371/journal.pntd.0001206] [Medline: 21647308]
9. Soldaini L, Yom-Tov E. Inferring Individual Attributes from Search Engine Queries and Auxiliary Information. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3 - 7, 2017; Perth, Australia p. 293-301. [doi: 10.1145/3038912.3052629]
10. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res 2013 Jun 18;15(6):e124 [FREE Full text] [doi: 10.2196/jmir.2614] [Medline: 23778053]
11. Youngmann B, Yom-Tov E. Anxiety and Information Seeking: Evidence From Large-Scale Mouse Tracking. In: Proceedings of the 2018 World Wide Web Conference. 2018 Presented at: WWW'18; April 23 - 27, 2018; Lyon, France p. 753-762. [doi: 10.1145/3178876.3186156]
12. Ariely D, Loewenstein G. When does duration matter in judgment and decision making? J Exp Psychol Gen 2000;129(4):508-523. [doi: 10.1037/0096-3445.129.4.508]
13. Ariely D, Kahneman D, Loewenstein G. Joint comment on 'When does duration matter in judgment and decision making?' (Ariely & Loewenstein, 2000). J Exp Psychol Gen 2000;129(4):524-529. [doi: 10.1037/0096-3445.129.4.524]
14. Bennett GG, Glasgow RE. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
15. Higgins O, Sixsmith J, Barry MM, Domegan C. ARAN - Access to Research at NUI Galway. 2011. A Literature Review on Health Information Seeking Behaviour on the Web: A Health Consumer and Health Professional Perspective URL: https://aran.library.nuigalway.ie/xmlui/handle/10379/4588 [accessed 2020-01-13]
16. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]
17. Yom-Tov E. Demographic differences in search engine use with implications for cohort selection. Inf Retrieval J 2019;22(6):570-580. [doi: 10.1007/s10791-018-09349-2]

Abbreviations

DAQ: diagnosis ascertainment query
MSR: Medical Severity Rank

Edited by G Eysenbach; submitted 17.06.19; peer-reviewed by R West, C Eickhoff; comments to author 03.10.19; revised version received 07.10.19; accepted 16.12.19; published 06.03.20.

Please cite as:
Hochberg I, Allon R, Yom-Tov E
Assessment of the Frequency of Online Searches for Symptoms Before Diagnosis: Analysis of Archival Data
J Med Internet Res 2020;22(3):e15065
URL: https://www.jmir.org/2020/3/e15065
doi:10.2196/15065
PMID:32141835

Original Paper

Surgeons’ Interactions With and Attitudes Toward E-Patients: Questionnaire Study in Germany and Oman

Ken Masters[1*], MA, PhD; Teresa Loda[2*], MSc; Jonas Johannink[3*], MD; Rashid Al-Abri[1*], BSc, MD, FRCS; Anne Herrmann-Werner[2*], MD, MME

[1]Medical Education and Informatics Department, College of Medicine and Health Sciences, Sultan Qaboos University, Al-Khoud, Oman
[2]Medical Department VI, Psychosomatic Medicine and Psychotherapy, University Hospital Tübingen, Tübingen, Germany
[3]Department of Surgery and Transplantation, University Hospital Tübingen, Tübingen, Germany
[*]all authors contributed equally

Corresponding Author:
Ken Masters, MA, PhD
Medical Education and Informatics Department
College of Medicine and Health Sciences
Sultan Qaboos University
Box 50
Al-Khoud, 123
Oman
Phone: 968 2414 3499
Email: itmeded@gmail.com

Abstract

Background: Doctors’ interactions with and attitudes toward e-patients have an overall impact on health care delivery.

Objective: This study aimed to gauge surgeons’ interactions with e-patients, their attitudes toward those e-patient activities, the possible impact on the delivery of health care, and the reasons behind those activities and attitudes.

Methods: We created a paper-based and electronic survey form based on pertinent variables identified in the literature, and from March 2018 to July 2018 we surveyed 49 surgeons in Germany and 59 surgeons in Oman, asking them about their interactions with and attitudes toward e-patients. Data were stored in Microsoft Excel and SPSS, and descriptive statistics, Pearson correlations, and chi-square tests were performed on the data.

Results: Of our sample, 71% (35/49) of the German surgeons and 56% (33/59) of the Omani surgeons communicated electronically with their patients. Although the German surgeons spent a greater percentage of Internet usage time on work-related activities (χ^2_{18}=32.5; *P*=.02) than the Omani surgeons, there were many similarities in their activities. An outstanding difference was that the German surgeons used email with their patients more than the Omani surgeons (χ^2_1=9.0; *P*=.003), and the Omani surgeons used social media, specifically WhatsApp, more than the German surgeons (χ^2_1=18.6; *P*<.001). Overall, the surgeons were equally positive about the most common e-patient activities such as bringing material from the internet to the consultation (mean 4.11, SD 1.6), although the German surgeons (mean 3.43, SD 1.9) were more concerned (*P*=.001) than the Omani surgeons (mean 2.32, SD 1.3) about the potential loss of control and time consumption (German: mean 5.10, SD 1.4 and Omani: mean 3.92, SD 1.6; *P*<.001).

Conclusions: The interactions show a high degree of engagement with e-patients. The differences between the German and the Omani surgeons in the preferred methods of communication are possibly closely linked to cultural differences and recent historical events. These differences may, moreover, indicate e-patients’ desired method of electronic communication to include social media. The low impact of surgeons’ attitudes on the activities may also result from a normalization of many e-patient activities, irrespective of the doctors’ attitudes and influences.

(J Med Internet Res 2020;22(3):e14646) doi:10.2196/14646

KEYWORDS

internet; e-patient; internet-informed patient; doctor-patient relationship; attitude; digital health; technology; e-physician; empowerment; physician; communication

Introduction

Background and Literature

The overall impact of the engaged and better-informed patient on the patient-doctor relationship can be positive or negative [1,2]. The internet, moreover, has introduced a new dimension to the engaged and better-informed patient (ie, the e-patient). In short, e-patients are patients who are “equipped, enabled, empowered and engaged in their health and health care decisions” [3] and who use “the Internet to gather information about a medical condition of particular interest to them” [3].

This is not the place to give a more detailed account of e-patient activities; nevertheless, it is useful to note that typical activities involve searching for medical and health-related information on the internet, joining Web-based patient discussion groups, communicating electronically with their physicians, accessing their electronic medical records, accessing their laboratory results, using personal health records, researching their physician’s electronic footprint, and other electronically based health-related activities [4-16].

The term *e-patient* has only recently gained traction in the medical literature. A PubMed/ Medical Literature Analysis and Retrieval System Online search by the authors for articles published before 2011 and referring to the term *e-patient* or *epatient* in the title or abstract revealed only 5 citations. The same search for the years 2011 to 2019 conducted in March 2019 revealed 24 citations.

Similar to the impact of the engaged patient before the advent of the internet, the e-patient’s activities’ impact on the patient-doctor relationship can vary, with some doctors reporting a worsening relationship and others reporting a strengthening of the relationship and improved health care [11,17-19].

In general, a more engaged patient can result in better health outcomes and lower costs [20,21]. In the domain of surgery, more informed patients have experienced less postoperative pain than uninformed patients [22]. It is known that e-patients consult the internet pre surgery and post surgery, and approximately 20% of e-patients consult websites recommended by their surgeon or doctor and rate those websites higher in quality than other sites they have consulted [12]. It is also known that e-patients use the internet to research their surgeons, although this is not always welcomed by surgeons [12].

There is strong evidence that doctors’ attitudes toward patients’ use of the internet, especially if they recommend sites to their patients, directly influence those patients’ usage; the quality of material found; and, by extension, the patient-doctor relationship [12]. Worryingly, however, international research shows that approximately 65% to 80% of e-patients *do not* share this information with their doctors, primarily because of their doctors’ negative attitudes toward the internet [19]. One can only speculate the harm to the patient-doctor relationship and the communication breakdown that could result from this.

Setting

In Germany and Oman, approximately 96% and 80% of the population has access to the internet, respectively [23,24]. A 2009 survey indicated that there were some 40 million e-patients in Germany [25]; based on internet usage in Germany from 2016, this figure can be calculated at approximately 49 million or higher [26]. This number represents a potentially great impact on the patient-doctor relationship.

There are currently no similar figures for e-patients in Oman. That said, the internet usage adoption rate in Oman, although behind Germany, has followed the trend of increasing usage seen in Germany and other countries [4]. Given this, there is reason to believe that e-patient activities will also follow similar trends. There is, however, always the possibility of differences in practice because of differences in culture, history, and other influencing factors, so one should not too easily make assumptions about one country’s practices based on practices in others.

Aim of the Study

This study focuses on surgeons’ attitudes toward e-patients in Germany and Oman. The aim of this study was to gauge surgeons’ interactions with e-patients, their attitudes toward those e-patient activities, the possible impact on the delivery of health care, and the reasons behind those activities and attitudes. This knowledge would give us some idea of the impact of e-patient activity on the patient-surgeon relationship in these 2 countries.

Methods

Variables

As part of the study was a comparison between the 2 countries, the first independent variable was the country in which the surgeon resided. Our selection of further independent variables to be studied was guided by the literature that had indicated possible predictors of attitudes toward e-patients. These included the doctors’ age and gender, amount of internet usage, and work-related time on the internet and the assumption that patients with chronic conditions who have access to the internet are more likely to be engaged in self-care and communicate electronically with doctors [2,9,13,16].

In addition, as there would be variations in usage and we wished to know if these variations might have an impact on attitudes, we asked questions about the specific internet sites that doctors visited.

The dependent variables were guided by the knowledge and activities described as typical e-patient activities, as listed above, also derived from the literature on the e-patient [4-16]. The last variables on attitudes toward the e-patient were questions derived from a study by Moick and Terlutter [27], described in more detail below.

Questionnaire

The description of the questionnaire design and delivery follows the Checklist for Reporting Results of Internet E-Surveys checklist (see Multimedia Appendix 1).

In Germany, the questionnaire was delivered to all surgeons from the General, Transplant, Visceral, Heart, and Orthopedic and Trauma Surgery Departments from the University Hospital, Tübingen, and the affiliated *Berufsgenossenschaftliche Unfallklinik*, Tübingen. The questionnaire was on paper and in an electronic format, using Google Forms. English fluency could be assumed among German doctors working in an academic environment, so the questionnaire was administered in English. Ethics approval was obtained from the University of Tübingen's Medical Ethics Committee (No. 001/2018BO2).

In Oman, the questionnaire was delivered to surgeons in the Department of Surgery at Sultan Qaboos University Hospital (including those surgeons affiliated to the department from the Oman Medical Association). The questionnaire was on paper and in an electronic format, using SurveyMonkey (SVMK Inc, San Mateo, CA). As all doctors in Oman need to be fluent in English, the questionnaire was delivered in English. Ethics approval was obtained from the Sultan Qaboos University College of Medicine Research Ethics Committee (MREC#1628).

In all cases, the information sheet and consent form contained the title and a brief description of the research project, names and contact details of the researchers, a brief statement about risks to the participants, confidentiality, storage of information (256-bit encryption), the voluntary nature of the participation, and permission to retain (or obtain) a copy of the informed consent form. All surgeons signed the informed consent form or checked an appropriate box on the electronic form. After the collection of the paper forms, the signed consent form was separated from the questionnaire and stored in a separate location.

For the core of the questionnaire design, we elected to use the relevant part of the survey form designed by Moick and Terlutter [27]. Moick and Terlutter's questionnaire was based on issues raised in the medical literature, and they determined it to be internally consistent. The questionnaire consists of 6 items about attitudes of online informed patients, ranging from 1 (*absolutely disagree*) to 7 (*absolutely agree*). As a double check, we inspected the literature that Moick and Terlutter had cited in the construction of their questionnaire. We did this to ensure that those sources did support the construction of their questions and accepted the questions as valid.

In addition to the questions from Moick and Terlutter, other literature [13] and surgeons were consulted to add further questions. The final version of the questionnaire is provided in Multimedia Appendix 2. The Web-based version was delivered on a single, scrollable screen so that the surgeons could review all their answers before submission.

Questionnaire Delivery

The surgeons were contacted through internal electronic mailing lists and WhatsApp groups and directed to the Web-based forms through URLs. Where surgeons preferred the paper form, these were delivered to them. No incentives were offered to the surgeons for the completion of the form. The data collection was performed from March 2018 to July 2018.

To maintain confidentiality and on the grounds that these are extremely easy to circumvent, no checks or preventative measures through cookies or internet protocol address identification were taken.

Data Analysis

All questionnaires were analyzed. Data were stored in Microsoft Excel 2016 and SPSS (version 25). Means, standard deviations, and frequencies were calculated. The data were normally distributed. Owing to this, Pearson correlations were run to examine correlations between the appropriate variables described above. To test for the differences between the German and Omani surgeons, chi-square tests and t tests for independent samples were conducted. Corrections for multiple testing were based on Bonferroni correction. (For the sake of brevity, the Results section speaks of the *German* and the *Omani* surgeons, although the reader should remember that this does not refer to their nationality but their location at the time of this study.)

Results

General

A total of 38% (49/128) German surgeons and 71% (59/83) Omani surgeons completed the survey. Of all the surgeons, only 6 Omani surgeons completed the survey on paper, so the number was too small to run any comparative statistical tests. In addition, the tables below show summary data only. Multimedia Appendix 3 contains charts with more details.

Demographics and Setting

We established the surgeons' age; gender; and whether the majority of their patients suffered from chronic, acute, or roughly the same types of condition. Table 1 summarizes these results.

Table 1. Surgeons' age, gender, and majority of patients, by country.

Category	Germany (n=49)	Oman (n=59)	Overall (n=108)	Statistics		
				t value (*df*)	Chi-square value (*df*)	P value
Age (years)				4.43 (104)	N/A[a]	.001
Mean (SD)	36.27 (8.83)	44.19 (9.59)	40.56 (10.07)			
Range	26-62	25-61	25-62			
Gender, n (%)				N/A	2.2 (2)	.33
Female	17 (34.7)	13 (22.0)	30 (27.8)			
Male	31 (63.3)	45 (76.3)	76 (70.4)			
Unknown gender	1 (2.0)	1 (1.7)	2 (1.9)			
Types of condition, n (%)				N/A	10.5 (2)	.01
Chronic	5 (10.2)	22 (37.3)	27 (25.0)			
Acute	10 (20.4)	8 (13.6)	18 (16.7)			
Both	34 (69.4)	29 (49.2)	63 (58.3)			

[a] Not applicable.

For most of the practices and attitudes given below, the figures will be viewed in light of the figures in Table 1.

Internet Usage

We measured the surgeon's internet usage, both broadly and more specifically, their knowledge and usage of sites and apps. When regarding the hours per day spent on the internet, there was no significant effect for origin (P=.67), age (P=.06), gender (P=.97), or condition type (P=.67). Table 2 shows the number of hours spent per day on the internet.

The German surgeons had a higher percentage of time (mean 60.04, SD 17.95) devoted to work-related activities (t_{106}=−3.72; P<.001) than the Omani surgeons (mean 45.34, SD 23.10). There was no difference between the work-related internet time and age (P=.15), gender (P=.21) or condition type (P=.91).

Delving further into work-related activities, we asked which sites and apps the surgeons knew about and used at least once per month. This would help to complete an overall picture of the surgeons' general familiarity with medically related websites and apps. (As can be seen from the questionnaire in Multimedia Appendix 1, examples of each of these categories were provided in case the subjects were not sure of what was meant by the category.) Tables 3 to 6 provide these figures.

Although there is similar knowledge of general references, databases, and journals, a significantly higher proportion of the German surgeons have knowledge about books, videos, networking sites, official sites, and magazines (Table 3).

Interestingly, however, when looking at the *usage* of these sites, most of these differences are reduced or even disappear. The use of databases is an exception, with usage by Omani surgeons far less than that by German surgeons (Table 4).

When viewing surgeon's knowledge of apps, again we see differences, with the German surgeons usually having greater knowledge than the Omani surgeons. A notable exception is Continuing Professional Development (CPD) apps (Table 5).

With app usage, most of the differences (including CPD) are removed, except for tools apps (Table 6).

Table 2. Hours spent per day on the internet.

Country	Hours						
	0	1-2	3-4	5-6	7-8	9-10	>10
Germany, n (%)	0 (0)	21 (42.9)	15 (30.6)	5 (10.2)	3 (6.1)	2 (4.1)	3 (6.1)
Oman, n (%)	0 (0)	23 (39.0)	24 (40.7)	9 (15.3)	2 (3.4)	1 (1.7)	0 (0)
Overall, n (%)	0 (0)	44 (40.7)	39 (36.1)	14 (13.0)	5 (4.6)	3 (2.8)	3 (2.8)

Table 3. Surgeons' knowledge of sites, by country.

Site	Germany (n=49), n (%)	Oman (n=59), n (%)	Total (N=108), n (%)	Statistics	
				Chi-square value (*df*)	*P* value
Books	46 (93.9)	42 (71.2)	88 (81.5)	8.4 (1)	.004
Videos	36 (73.5)	27 (45.8)	63 (58.3)	8.0 (1)	.005
General references	41 (83.7)	48 (81.4)	89 (82.4)	0.0 (1)	.90
Networking sites	22 (44.9)	10 (16.9)	32 (29.6)	9.7 (1)	.002
Official/institutional	32 (65.3)	24 (40.7)	56 (51.9)	6.1 (1)	.01
Databases	49 (100.0)	55 (93.2)	104 (96.3)	2.6 (1)	.11
Journals	44 (89.8)	50 (84.7)	94 (87.0)	0.3 (1)	.57
Magazines	36 (73.5)	11 (18.6)	47 (43.5)	32.0 (1)	.001

Table 4. Surgeons' use of sites at least once per month, by country.

Site	Germany (n=49), n (%)	Oman (n=59), n (%)	Total (N=108), n (%)	Statistics	
				Chi-square value (*df*)	*P* value
Books	33 (67.3)	29 (49.2)	62 (57.4)	2.9 (1)	.09
Videos	24 (49.0)	17 (28.8)	41 (38.0)	4.1 (1)	.04
General references	31 (63.3)	39 (66.1)	70 (64.8)	0.3 (1)	.58
Networking sites	9 (18.4)	4 (6.8)	13 (12.0)	3.2 (1)	.08
Official/institutional	10 (20.4)	15 (25.4)	25 (23.1)	0.5 (1)	.48
Databases	48 (98.0)	42 (71.2)	90 (83.3)	12.1 (1)	.001
Journals	31 (63.3)	36 (61.0)	67 (62.0)	0.0 (1)	.99
Magazines	9 (18.4)	7 (11.9)	16 (14.8)	0.7 (1)	.40

Table 5. Surgeons' knowledge of these apps, by country.

App Types	Germany (n=49), n (%)	Oman (n=59), n (%)	Total (N=108), n (%)	Statistics	
				Chi-square value (*df*)	*P* value
Monitoring	25 (51.0)	14 (23.7)	39 (36.1)	8.3 (1)	.004
Information	34 (69.4)	35 (59.3)	69 (63.9)	1.0 (1)	.33
Continuing Professional Development	20 (40.8)	36 (61.0)	56 (51.9)	4.8 (1)	.03
Tools	34 (69.4)	22 (37.3)	56 (51.9)	10.5 (1)	<.001
Videos	27 (55.1)	32 (54.2)	59 (54.6)	0.00 (1)	.99

Table 6. Surgeons' use of these apps at least once per month, by country.

App Types	Germany (n=49), n (%)	Oman (n=59), n (%)	Total (N=108), n (%)	Statistics	
				Chi-square value (*df*)	*P* value
Monitoring	8 (16.3)	5 (8.5)	13 (12.0)	1.5 (1)	.22
Information	27 (55.1)	26 (44.1)	53 (49.1)	1.1 (1)	.29
Continuing Professional Development	13 (26.5)	24 (40.7)	37 (34.3)	2.6 (1)	.12
Tools	23 (46.9)	15 (25.4)	38 (35.2)	5.2 (1)	.02
Videos	16 (32.7)	29 (49.2)	45 (41.7)	3.3 (1)	.07

E-Patient

We wanted to know if the surgeons engaged in the types of communication with e-patients that the literature had identified, their experience with e-patients, and their attitudes toward some of the implications of e-patient activities. The last set of questions would also allow a comparison with the data from the study by Moick and Terlutter [27].

We began our investigation of the surgeons' interactions and attitudes toward the e-patient by examining whether or not the surgeons were aware of the terminology. Of these surgeons, fewer German (23/49, 46%) than Omani (34/59, 57%) surgeons had heard of the term *e-patient*; however, this difference is not significant (χ^2_1=0.6; *P*=.23).

Of these surgeons, 71% (35/49) German and 55% (33/59) Omani surgeons communicated electronically with patients (χ^2_1=2.8; *P*=.10).

Given that electronic communication can take various forms, we wanted to see if there were differences in the methods of electronic communication between the 2 countries. Table 7 shows the method of communication used by the surgeons who communicate electronically with patients.

The figures for email communication, WhatsApp, and Twitter are significantly different, with the German doctors preferring email, and the Omani surgeons use both email and WhatsApp equally, and some Omani surgeons use Twitter. We should also note that there are several *Messenger* apps (eg, WhatsApp Messenger and Facebook Messenger), so the 2 items from that category might actually belong elsewhere, for example, WhatsApp or Facebook. Either way, this is a social media site, rather than email. These differences in methods of electronic communication are discussed in more detail in the Discussion section.

In addition to knowing how many surgeons communicate electronically with their patients, we wanted to know what percentage of their patients use email to communicate with them.

The results indicate that the German surgeons communicated with a larger percentage of their patients via email than the Omani surgeons (χ^2_9=25.1; *P*=.003). In addition, when looking on a metric level, for the German surgeons, email usage was strongly associated with their overall amount of internet usage (Pearson *r*=0.522; *P*<.001); this association was not found with the Omani surgeons.

Similarly, we wished to find out what percentage of their patients communicate with the surgeons via social media. The Omani surgeons communicated with a far greater amount of their patients via social media than the German surgeons (χ^2_{11}=48.6; *P*<.001).

An important aspect of the e-patient is the patient who brings material from the internet to the doctor. Of the surgeons, 89% (44/49) German and 84% (50/59) Omani surgeons indicated that patients bring information from the internet to the consultation (χ^2_{12}=18.9; *P*=.09).

Correspondingly, we wished to know how frequently surgeons recommend websites or apps to their patients. Of the surgeons, 40% (20/49) German and 49% (29/59) Omani surgeons recommended a website or app to their patients at least once per month (χ^2_1=3.5; *P*=.06).

For the German surgeons, the number of website recommendations was strongly associated with the amount of internet usage (Pearson *r*=0.326; *P*=.02).

Interestingly, the impact of these recommendations on the patients appeared minimal, as there was no association between the number of recommendations and the frequency that patients are bringing information from the internet to the consultation (*P*=.06).

Of the specific websites or apps recommended by the German surgeons, 19 sites were mentioned, but no site was recommended more than once. From the Omani surgeons, 13 sites were mentioned. The 3 most recommended sites were Google (11), Medscape, and YouTube (3 each). All other sites were recommended only once. (Of interest, the top 3 Omani sites [Google, Medscape, and YouTube] were not recommended at all by the German surgeons.)

Table 7. Surgeons' methods of electronic communication with patients, by country.

Method	Germany (n=35), n (%)	Oman (n=33), n (%)	Total (N=68), n (%)	Statistics	
				Chi-square value (*df*)	*P* value
Email	32 (91.4)	21 (63.6)	53 (77.9)	9.0 (1)	.003
Social media					
WhatsApp	1 (2.9)	21 (63.6)	22 (32.4)	18.6 (1)	<.001
Twitter	0 (0.0)	7 (21.2)	7 (10.3)	6.3 (1)	.008
Facebook	1 (2.9)	5 (15.2)	6 (8.8)	2.2 (1)	.10
Instagram	0 (0.0)	2 (6.1)	2 (2.9)	1.7 (1)	.19
Messenger	0 (0.0)	2 (6.1)	2 (2.9)	1.7 (1)	.19
Other	1 (2.9)	0 (0.0)	1 (1.5)	1.5 (1)	.32

Attitudes Toward Patients Bringing Material From the Internet

In addition to surgeons' behaviors, we wanted to know about their attitudes toward e-patient behaviors, particularly regarding the patient-doctor relationship. Table 8 shows results from statements beginning with "If a patient brought some health-related information to a consultation..." and based on a Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree).

The surgeons from both the countries felt more positive than negative about patients' bringing health-related information from the internet to the consultation, and there was no difference between the 2 countries on this question. There was an inverse effect with age for Germany, but there was no other effect by age (*P*=.11), gender (*P*=.69), or condition type (*P*=.52). There was no effect between the answers to this question and the number of electronic interactions with patients via email or the percentage of patients with whom they communicate via email (*P*=.14).

The surgeons from both the countries were prepared to correct wrong or incomplete information, and there was no difference between the 2 countries on this question and no associations with age (*P*=.69), gender (*P*=.95), or condition type (*P*=.95).

Although the surgeons from neither country felt very strongly about the loss of control, there was a difference between the 2 countries on this question, with the German surgeons feeling more strongly about this. There were no associations by condition type (*P*=.71) but an overall association by gender, with males feeling more strongly about this issue (*P*=.005).

The German surgeons felt more strongly than the Omani surgeons about time-consuming consultations, and there was an inverse association with age among the Omani doctors (*r*=-0.282; *P*=.003).

Generally, the surgeons felt that the resultant communication would lead to an improvement in the patient-doctor relationship, and there was no difference between the 2 countries on this question and no effect by age (*P*=.99), gender (*P*=.52), or condition type (*P*=.97).

The surgeons were disinclined to prescribe different medications, and there was no difference between the 2 countries on this question and no effect by age (*P*=.34), gender (*P*=.06), or condition type (*P*=.83).

Finally, because surgeons' attitudes can be associated with the amount of interaction with e-patients, we looked for any associations between the most common activities and the answers to the above 6 questions in general and on the country level. Of all these, the only association was between the percentage of patients bringing in information from the internet (Pearson *r*=0.278; *P*=.004) and the likelihood that the surgeons would prescribe a desired medication.

Table 8. Summary of responses to attitude statements.

Item	Germany (n=49), mean (SD)	Oman (n=59), mean (SD)	Overall (n=108), mean (SD)	Statistics	
				t value (*df*)	*P* value
I think it is generally positive.	4.20 (1.5)	4.03 (1.7)	4.11 (1.6)	−0.56 (105)	.58
I am prepared to correct wrong, incomplete, and misunderstood information.	5.12 (1.7)	5.03 (1.3)	5.10 (1.8)	−0.11 (104)	.91
I sometimes feel I might lose authority and control.	3.43 (1.9)	2.32 (1.3)	2.82 (1.7)	−3.53 (106)	.001
I expect a more time-consuming patient visit than with uninformed patients.	5.10 (1.4)	3.92 (1.6)	4.45 (1.7)	−4.03 (106)	<.001
The physician-patient relationship will be improved by better communication.	4.41 (2.0)	4.81 (1.8)	4.63 (1.9)	1.09 (105)	.28
I would be more likely to prescribe a desired medication than if the patients were uninformed.	3.08 (1.7)	3.53 (1.9)	3.32 (1.8)	1.27 (105)	.21

Discussion

Principal Findings

This paper reported the e-patient–related activities and attitudes of surgeons at 2 sites in Germany and Oman. There are many similarities between the 2 groups; indeed, many of these similarities can be found in other studies. There are, however, differences between the 2 sites, particularly in the methods of communication between the surgeons and patients. The discussion below will explore some of these similarities and differences in light of the literature.

As we based a large portion of our questionnaire on the one produced by Moick and Terlutter [27] (who studied general practitioners, orthopedists, and dermatologists in Germany), we shall refer frequently to their paper as a point of comparison, although the other literature will also have a bearing on our results.

Internet Usage

The surgeons in this study generally used the internet far more than the doctors in Moick and Terlutter's study [27], and the percentage of work-related time on the internet was also higher (52.0% as opposed to 25.3%). Given that the work of Moick and Terlutter [27] was published in 2012 and earlier studies also generally show lower figures of usage [13], this difference may be a reflection of the fact that overall internet usage is growing around the world.

Overall, the German surgeons spent a greater percentage of their online time on work-related activities than the Omani surgeons. Earlier studies of internet usage by doctors [13] showed differences associated with gender, but even those figures were not always consistent. Our study also shows few differences between males and female activities.

Surgeons' Knowledge, Use, and Recommendation of Sites and Apps

As was noted in the Introduction section, patient engagement has led to better health outcomes [20-22], and so it is important to consider the level of engagement afforded by the surgeons with their patients. Several studies have found that doctors recommend sites to patients, and this has been happening for many years [13], and the surgeons in this study confirm this practice.

With the surgeons in this study, we wished to have a more detailed knowledge of their baseline knowledge and usage of various sites and apps. Although the aims of the study did not require that we conduct detailed statistical analyses on the various sites and apps, we were able to gather a greater sense that, overall, the surgeons' knowledge and use of the internet were extensive. This provided important contextual information so that we could be surer that any lack of site recommendation for patients would not be simply because of a lack of knowledge or awareness of these sites.

In this light, among the Omani surgeons, the high number of Google recommendations (11/29, 38%) is of great concern. Unsurprisingly, patients have long been using these general search engines as a starting point to health sites [11], and there is a need for patients to have more guided searches. For some time, it has been suggested that surgeons need to be taught about the use of the internet, including useful patient education sites and site evaluation [4,10,28], in a similar way that they are taught other skills and knowledge. It is not merely knowledge of these sites (as these surgeons have indicated that they have this knowledge), but rather transferring that knowledge into useful guiding information for their patients. This should be addressed in the Omani surgeons' education.

Communication With the E-Patient

From our perspective, among the most interesting results were the figures on electronic communication with patients. For several years, studies have shown that, in spite of some reservations, many physicians are satisfied with email communication with patients and have continued to use email as a standard method of communication [13]. (Most studies have shown, however, that the actual percentage of patients whom doctors email is low [13].) In our study, given that the German surgeons used the internet for work-related activities more than the Omani surgeons, the German surgeon's higher electronic communication figures (71% vs 56%) were not surprising.

Noteworthy in this study, however, was the great difference in the *methods* of communication between the 2 countries. Of the surgeons who communicated electronically with their patients, the German surgeons almost used only email (91%), whereas social media usage for patient communication was virtually zero. The Omani figures are very different: although 64% of the Omani doctors used email, an equal number used WhatsApp, and overall, social media tools were used far more than standard email.

It would be dangerous to speculate too deeply on the Omani figures, and this could be the subject of later research. There are, however, 2 possible reasons based on sociological and recent historical differences between the 2 countries:

1. In general, Omani society is very close knit, and it is not unusual for doctors and patients to be related. Approximately 52% of Omani marriages are consanguineous, and more than 75% of these marriages are with first cousins [29]. Relatives are frequently recognized by common surnames related to tribes and region of origin. Within these social structures, sharing of private telephone numbers is reasonably common. As WhatsApp is based on telephone numbers and many Omani doctors use the same phone for professional and personal reasons, their patients will already have access to their numbers and their WhatsApp accounts, and it is not uncommon for patients to contact doctors on more than one social media platform simultaneously. It should be noted, moreover, that of the Omani surgeons in the department, 47% (39/83) surgeons are Omani nationals, so the impact of this social characteristic is uncertain. This would be an area to be explored in more research.
2. Although Oman was not affected by the 2011 Arab Spring to the same extent as many other countries in the region, the power of social media became obvious during that time [30,31]. Since then, social media, especially, WhatsApp and Facebook, have become a part of daily life for people living in the Arab world, including Oman [32,33].

There may be a third reason: an overall growth in patient communication with social media. Currently, although there is international recognition of the role of social media in medical practice [5,34], its usage is frequently in the form of informal patient communities, just as the Web was used for patient communities before social media. It is possible that the sociological and historical elements outlined above have merely provided an early impetus—indicating a coming change—and the fact that just as email communication by doctors was very low barely 10 to 15 years ago [13] and has increased over time, social media communication between patients and doctors may also increase in the future.

The implications for medical education are important: currently, medical education communication teaching still focuses on face-to-face communication teaching, in spite of the fact that there has long been a call for email communication to be explicitly taught [4]. Just as email communication with patients grew mostly because of demand from patients, social media communication may also do so. There are already calls for social media communication to be included in communication teaching [4], and this study indicates the urgency of that call. Part of this training will have to include careful management, as email is far less intrusive than social media, and the impact on physician's time (and resultant physician burnout) will be potentially devastating if social media communication is not managed correctly.

Irrespective of the differences between the German and Omani surgeons, both groups show that more than half of the surgeons communicate electronically with their patients, thereby affording the opportunity for Web-based patient engagement; as indicated above, the literature has shown a positive association between patient engagement and better health outcomes [20-22], so we anticipate that this level of interaction will have a positive impact on overall health care delivery.

E-Patients Bringing Information From the Internet

From our study, the high percentage of surgeons from Germany and Oman reporting patients bringing information to them from the internet is consistent with the literature [13]. The percentage of patients who bring material from the internet remains relatively low, although higher than earlier studies [13], and so, these numbers may increase in the future. In the Introduction section of this paper, reference was made to 65% to 80% of patients who seek information on the internet and do not discuss it with their doctors, primarily because of their doctors' attitudes toward internet-based information [19,35]. These attitudes are explored in a little more detail below.

Other literature has found an association between recommendations from the doctor and patients bringing information from the internet [14], but we did not find that association in our study. Similarly, previous studies have indicated that when patients with chronic diseases have access to the internet, they have used it for seeking health-related material more than patients with acute conditions [16]. Again, however, this was not reflected in our study. Part of the reason for not finding these trends in our study may be a normalization of e-patient activities. As the percentage of patients searching for information on the internet increases and *e-patient* activities become normal patient activities, recommendations from doctors may have less impact on the numbers of patients searching for material, and patients with acute conditions will search for material as much as patients with chronic conditions. It is to be expected, however, that doctors still have a role in determining *which* sites are visited. Either way, these results emphasize the need to teach doctors how to cope with this phenomenon [28].

Attitudes Toward Patients Bringing Material From the Internet

The comparison of our results with those of Moick and Terlutter's study [27] shows many similarities. These surgeons were generally positive about patients bringing material from the internet and were prepared to correct patients' incorrect information.

Indications from the literature are that patients are increasingly requesting prescriptions by name, having found the information from both advertisements in traditional media and also from social media and other electronic sources [14], and this undoubtedly places pressure on doctors. Similar to Moick and Terlutter's sample [27], however, these surgeons were somewhat reluctant to change their prescription based on the patients' findings. Where the surgeons were willing to change prescriptions, this change was positively associated with the percentage of patients bringing information. Whether or not this association is causal and the exact nature of the possible causality would need to be studied further. It may be that the sheer number of requests leads to changes or it may be that surgeons who are more open to having their patients bringing information are correspondingly more open to changing their prescriptions based on patients' desires. This is entirely plausible, given that 2 medications may have almost the same impact, and if a patient really does have a preference for a particular medication, and there is no harm to come from it, then changing the prescription could be easily accommodated.

These surgeons envisaged an improvement in the patient-doctor relationship caused by the better-informed patient. This is a controversial discussion point in the literature, in which some studies have found a negative impact, whereas others have found results similar to ours and have seen that the more engaged patient has resulted in improved health care delivery [1,20,21]. Qualitative studies [1] have found that part of the reason for the positive feeling is that physicians can deal with more important and profound issues rather than getting bogged down in trivial explanations.

The differences between the German and Omani figures, however, are also noteworthy. The Omani surgeons were significantly less concerned than the German surgeons and Moick and Terlutter's sample about the loss of control. Other studies have also indicated that even when physicians are generally positive about well-informed patients, sometimes arguing with patients over irrational points can lead to a fear of lack of trust in the doctor's ability [1]. Again, this lack of fear among the Omani surgeons may have to do with sociological environment in which these surgeons function, and one may hypothesize that the somewhat relaxed social strata in Oman may lead to less concern about losing control.

The other difference was the fear of loss of time; again, the figures show that the Omani surgeons were far less concerned than the German surgeons and those in Moick and Terlutter's sample and also in other studies [1]. This, also, may be related to cultural differences, as research indicates that Arab culture can be considered *polychronic*, having a more flexible approach to time and appointments than one may find in the German culture [36-39]. The fact that the Omani surgeons' fear over loss of time was negatively associated with age may indicate changes in attitudes of the younger generation of doctors in Oman.

The Larger Context of Medical Education and Participatory Medicine

This paper focuses on the surgeons working with e-patients. As this is already a lengthy paper, it would not be wise to broaden it much further, although it is necessary to look a little at the broader context.

Reference has already been made to the implications of these findings for medical and health education. In short, these findings reinforce and then extend the notion that health education needs to be tailored to meet the demands of the e-patient, specifically on the effective management of electronic communication in health care.

In addition, moreover, there is the entire field of participatory medicine, in which patients move away from *compliance* to active participation in their health care [40], further trending also toward a *role convergence* between the patient and doctor [2]. But the changes are not occurring with patients only, as doctors are also changing. Many previous studies on physicians' use of technologies showed variation within the demographic indicators of age and gender, although even then there were indications that the differences were not always clear-cut [13]. In this study, the differences are further blurred, and other factors, such as culture or historical events, appear to have a greater impact on the differences. We would argue that this is to be expected as these technologies become mainstream, and are no longer used only by early adopters. Indeed, reflecting on these results, the literature shows a reduction in earlier reports of doctors lagging behind patients in embracing the internet to a mirroring of e-patient and the rise and development of the *empowered physician* (*e-physician*) [41] who will perform these tasks as a normal part of their work.

The activities and especially the positive attitudes of the surgeons in this study appear to show a great opportunity for increased patient participation and engagement in their health care, and this bodes well for future positive health care delivery. Previous studies have indicated that doctors' attitudes strongly impact patient activities, but this has not been strongly supported in this study. It may be that e-patients are now less impacted by doctors' attitudes than before or it may be that there is a tipping point where a change from negative to positive is the main difference, after which the patients take greater control irrespective of small differences in surgeons' positive attitudes.

Limitations of the Study

Although the study was conducted in 2 countries, the centers were localized to the 2 hospitals associated with the researchers' universities. That said, the figures for Germany conform to many of the trends found elsewhere, and Oman is a relatively small country with few surgical centers and only 2 medical schools. A second limitation is the sample size of a total of 108 surgeons.

Conclusions

Recognizing that doctors' electronic interactions with patients impact health care delivery, this study has examined German and Omani surgeons' interactions with and attitudes toward e-patients. We have seen that, overall, these doctors are comfortable with many e-patient activities and that the doctors' attitudes do not have a significant impact on the e-patient activities. This may be because the e-patients see these activities as part of their normal lives and are performed irrespective of doctors' opinions about the activities. Either way, the approach by these surgeons has created the opportunity for greater patient participation and engagement, and the literature indicates that this opportunity should have a great positive impact on health care delivery.

In addition, however, there were differences between the 2 countries, most notably in the methods of electronic communication, with the German surgeons using mainly email and the Omani surgeons' heavy use of social media. Further research needs to be performed to determine the extent to which this difference results from cultural differences and recent historical events; it is also possible that the difference is an early indication of e-patients wishing to shift communication from email to social media.

Beyond medical practice, there are implications for medical education, and medical schools need to ensure that medical students receive comprehensive training on working with the e-patient, including appropriate electronic communication with patients.

Acknowledgments

The authors would like to thank all the physicians who participated in this study. The authors also wish to thank reviewers of previous versions of this paper. This study had no external funding or sponsorships.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys checklist and corresponding actions taken in the paper.
[DOCX File , 28 KB - jmir_v22i3e14646_app1.docx]

Multimedia Appendix 2
Questions for physician questionnaire.
[DOCX File , 199 KB - jmir_v22i3e14646_app2.docx]

Multimedia Appendix 3
More detailed charts.
[DOCX File , 75 KB - jmir_v22i3e14646_app3.docx]

References

1. Sommerhalder K, Abraham A, Zufferey MC, Barth J, Abel T. Internet information and medical consultations: experiences from patients' and physicians' perspectives. Patient Educ Couns 2009 Nov;77(2):266-271. [doi: 10.1016/j.pec.2009.03.028] [Medline: 19411157]
2. Buetow S, Jutel A, Hoare K. Shrinking social space in the doctor-modern patient relationship: a review of forces for, and implications of, homologisation. Patient Educ Couns 2009 Jan;74(1):97-103. [doi: 10.1016/j.pec.2008.07.053] [Medline: 18789627]
3. Ferguson T, e-Patient Scholars Working Group. Society for Participatory Medicine. e-patients: how they can help us heal healthcare URL: https://participatorymedicine.org/e-Patient_White_Paper_with_Afterword.pdf [accessed 2019-05-07] [WebCite Cache ID 78BrqptQD]
4. Masters K, Ng'ambi D, Todd G. 'I Found it on the Internet': preparing for the e-patient in Oman. Sultan Qaboos Univ Med J 2010 Aug;10(2):169-179. [Medline: 21509226]
5. Colineau N, Paris C. Talking about your health to strangers: understanding the use of online social networks by patients. New Rev Hypermedia Multimed 2010 Apr;16(1-2):141-160. [doi: 10.1080/13614568.2010.496131]
6. 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 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
7. O'Kane MJ. Direct patient access to test results: implications for the laboratory. Ann Clin Biochem 2015 Sep;52(Pt 5):525-526. [doi: 10.1177/0004563215590452] [Medline: 25995286]
8. Lee BS, Walker J, Delbanco T, Elmore JG. Transparent electronic health records and lagging laws. Ann Intern Med 2016 Aug 02;165(3):219-220. [doi: 10.7326/M15-2827] [Medline: 27214266]
9. Gee PM, Paterniti DA, Ward D, Soederberg Miller LM. e-Patients perceptions of using personal health records for self-management support of chronic illness. Comput Inform Nurs 2015 Jun;33(6):229-237. [doi: 10.1097/CIN.0000000000000151] [Medline: 25899440]
10. Masters K. Preparing medical students for the e-patient. Med Teach 2017 Jul;39(7):681-685. [doi: 10.1080/0142159X.2017.1324142] [Medline: 28532256]
11. Fox S, Duggan M. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2013. Health Online 2013 URL: http://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2018-12-23] [WebCite Cache ID 74sbsnhYG]
12. Fraval A, Chong YM, Holcdorf D, Plunkett V, Tran P. Internet use by orthopaedic outpatients - current trends and practices. Australas Med J 2012;5(12):633-638 [FREE Full text] [doi: 10.4066/AMJ.2012.1530] [Medline: 23382767]
13. Masters K. For what purpose and reasons do doctors use the internet: a systematic review. Int J Med Inform 2008 Jan;77(1):4-16. [doi: 10.1016/j.ijmedinf.2006.10.002] [Medline: 17137833]
14. Makovsky Health.: Kelton; 2015. Makovsky 2015 Survey: Initial Data URL: http://www.makovsky.com/news/fifth-annual-pulse-of-online-health-survey-2/ [accessed 2016-05-22]
15. Ted Talks. The Netherlands: TEDxMaastricht; 2011. Dave deBronkart: Meet e-Patient Dave URL: https://www.ted.com/talks/dave_debronkart_meet_e_patient_dave?language=en [accessed 2019-05-07] [WebCite Cache ID 78BnXllGs]
16. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
17. Wolters Kluwer. 2012. Survey: consumers show high degree of trust in online health information, report success in self-diagnosis URL: http://wolterskluwer.com/company/newsroom/news/health/2012/05/survey-consumers-show-high-degree-of-trust-in-online-health-information-report-success-in-self-diagnosis.html [accessed 2019-05-08] [WebCite Cache ID 78Bo8nhUE]
18. Fage-Butler AM, Nisbeth Jensen M. Medical terminology in online patient-patient communication: evidence of high health literacy? Health Expect 2016 Jun;19(3):643-653 [FREE Full text] [doi: 10.1111/hex.12395] [Medline: 26287945]
19. Klinar I, Balazin A, Barsić B, Tiljak H. Identification of general characteristics, motivation, and satisfaction of internet-based medical consultation service users in Croatia. Croat Med J 2011 Aug 15;52(4):557-565 [FREE Full text] [doi: 10.3325/cmj.2011.52.557] [Medline: 21853551]
20. 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]
21. Shortell SM, Poon BY, Ramsay PP, Rodriguez HP, Ivey SL, Huber T, et al. A multilevel analysis of patient engagement and patient-reported outcomes in primary care practices of accountable care organizations. J Gen Intern Med 2017 Jun;32(6):640-647 [FREE Full text] [doi: 10.1007/s11606-016-3980-z] [Medline: 28160187]

22. Schmidt M, Eckardt R, Scholtz K, Neuner B, von Dossow-Hanfstingl V, Sehouli J, PERATECS Group. Patient empowerment improved perioperative quality of care in cancer patients aged ≥ 65 years - a randomized controlled trial. PLoS One 2015;10(9):e0137824 [FREE Full text] [doi: 10.1371/journal.pone.0137824] [Medline: 26378939]
23. Internet World Stats. 2019. Internet in Europe Stats URL: https://www.internetworldstats.com/stats4.htm [accessed 2019-05-07] [WebCite Cache ID 78BpXNLgL]
24. Internet World Stats. 2019. Internet Usage in the Middle East URL: https://www.internetworldstats.com/stats5.htm [accessed 2019-05-07] [WebCite Cache ID 78Bq0sbtB]
25. Schachinger A. Der digitale Patient: Analyse eines neuen Phänomens der partizipativen Vernetzung und Kollaboration von Patienten im Internet. 1st ed. Baden-Baden: Nomos Verlagsgesellschaft; 2014.
26. DESTATIS. 2017. DESTATIS: Statistisches Bundesamt URL: https://www.destatis.de/DE/Startseite.html [accessed 2012-06-23] [WebCite Cache ID 68dGWNwIt]
27. Moick M, Terlutter R. Physicians' motives for professional internet use and differences in attitudes toward the internet-informed patient, physician-patient communication, and prescribing behavior. Med 2 0 2012 Jul 06;1(2):e2 [FREE Full text] [doi: 10.2196/med20.1996] [Medline: 25075230]
28. Herrmann-Werner A, Weber H, Loda T, Keifenheim KE, Erschens R, Mölbert SC, et al. 'But Dr Google Said…' - training medical students: how to communicate with e-patients. Med Teach 2019 Dec;41(12):1434-1440. [doi: 10.1080/0142159X.2018.1555639] [Medline: 30707847]
29. Islam MM, Dorvlo AS, Al-Qasmi AM. The pattern of female nuptiality in Oman. Sultan Qaboos Univ Med J 2013 Feb;13(1):32-42 [FREE Full text] [Medline: 23573380]
30. Bakri N. New York Times. 2011 Feb 28. Protests in Oman Spread From Port City to Capital URL: https://www.nytimes.com/2011/03/01/world/middleeast/01oman.html [accessed 2020-02-19]
31. Worrall J. Middle East Policy Council. 2012. Oman: The 'Forgotten' Corner of the Arab Spring URL: https://www.mepc.org/oman-forgotten-corner-arab-spring [accessed 2019-02-19]
32. Arab Social Media Influencers Summit. Dubai: Times News Service; 2015. Arab Social Media Report URL: http://www.govtpracticewpp.com/~/media/wppgov/files/arabsocialmediareport-2015.pdf [accessed 2019-05-07] [WebCite Cache ID 78BtcHHbO]
33. Times of Oman. 2018 Jun 2. Social media user numbers in Oman up by over 600,000 since 2017 URL: https://timesofoman.com/article/135630 [accessed 2019-02-19]
34. Thompson MA, Ahlstrom J, Dizon DS, Gad Y, Matthews G, Luks HJ, et al. Twitter 101 and beyond: introduction to social media platforms available to practicing hematologist/oncologists. Semin Hematol 2017 Oct;54(4):177-183. [doi: 10.1053/j.seminhematol.2017.06.006] [Medline: 29153077]
35. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
36. Lipson JG, Meleis AI. Issues in health care of Middle Eastern patients. West J Med 1983 Dec;139(6):854-861 [FREE Full text] [Medline: 6364575]
37. Feghali E. Arab cultural communication patterns. Int J Intercult Relations 1997 Aug;21(3):345-378. [doi: 10.1016/S0147-1767(97)00005-9]
38. Billing TK, Bhagat RS, Lammel A, Leonard KM, Ford Jr DL, Brew F, et al. 18th International Congress of the International Association for Cross-Cultural Psychology. 2009. Temporal orientation and its relationships with organizationally valued outcomes: results from a 14 country investigation URL: https://scholarworks.gvsu.edu/iaccp_papers/45/ [accessed 2019-02-19]
39. Nydell MK. Understanding Arabs: A Contemporary Guide to Arab Society. 5th ed. Boston: Nicholas Brealey; 2012.
40. deBronkart D. The patient's voice in the emerging era of participatory medicine. Int J Psychiatry Med 2018 Nov;53(5-6):350-360. [doi: 10.1177/0091217418791461] [Medline: 30114957]
41. Mesko B, Győrffy Z. The rise of the empowered physician in the digital health era: viewpoint. J Med Internet Res 2019 Mar 26;21(3):e12490 [FREE Full text] [doi: 10.2196/12490] [Medline: 30912758]

Abbreviations

CPD: Continuing Professional Development

Edited by G Eysenbach; submitted 09.05.19; peer-reviewed by R Terlutter, K Venance Basil, M El Tantawi; comments to author 31.05.19; revised version received 23.07.19; accepted 04.02.20; published 09.03.20.

Please cite as:
Masters K, Loda T, Johannink J, Al-Abri R, Herrmann-Werner A
Surgeons' Interactions With and Attitudes Toward E-Patients: Questionnaire Study in Germany and Oman
J Med Internet Res 2020;22(3):e14646
URL: http://www.jmir.org/2020/3/e14646/
doi:10.2196/14646
PMID:32149714

Original Paper

General Practitioners’ Perceptions of Whether Teleconsultations Reduce the Number of Face-to-face Visits in the Catalan Public Primary Care System: Retrospective Cross-Sectional Study

Francesc López Seguí[1,2], MSc; Josep Vidal-Alaball[3,4], MD, PhD; Marta Sagarra Castro[5], MD; Anna García-Altés[2,6], PhD; Francesc García Cuyàs[7], MD, PhD

[1]TIC Salut Social, Ministry of Health, Barcelona, Spain
[2]Centre for Research in Health and Economics, Pompeu Fabra University, Barcelona, Spain
[3]Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain
[4]Unitat de Suport a la Recerca de la Catalunya Central, Fundacio´ Institut Universitari per a la recerca a l'Atencio´ Prima` ria de Salut Jordi Gol i Gurina, Sant Fruito´s de Bages, Spain
[5]Centre d’Atenció Primària Capellades, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Barcelona, Spain
[6]Agency for Healthcare Quality and Evaluation of Catalonia (AQuAS), Catalan Ministry of Health, Barcelona, Spain
[7]Sant Joan de Déu Hospital, Catalan Ministry of Health, Barcelona, Spain

Corresponding Author:
Josep Vidal-Alaball, MD, PhD
Health Promotion in Rural Areas Research Group
Gerència Territorial de la Catalunya Central
Institut Català de la Salut
Carrer Pica d'Estats, 13-15
Sant Fruitós de Bages, 08272
Spain
Phone: 34 93 693 0040
Email: jvidal.cc.ics@gencat.cat

Abstract

Background: eConsulta is a teleconsultation service involving general practitioners (GPs) and patients. It is part of the information system belonging to Catalonia’s primary care service. It has been in operation since the end of 2015 in conjunction with face-to-face consultations with Primary Care Teams as one of the services offered in the patient’s Personal Health Folder.

Objective: This study aimed to assess the ability of using eConsulta to reduce the number of face-to-face visits to Primary Care Teams.

Methods: Using 13 categories proposed by the researchers, 18 GPs from the Central Catalonia Health Region retrospectively classified 2268 cases managed with eConsulta and indicated whether, in their opinion, the teleconsultations reduced the number of face-to-face visits.

Results: There was broad consensus among the GPs that eConsulta has the potential to resolve patient queries for every type of consultation. eConsulta avoided the need for a face-to-face visit in 87.9% of cases. In addition, the GPs reported that the ease of access increased the demand for health care support in 27.7% of cases; otherwise, the patient would not have initiated the queries. Therefore, based on the equation (88% x [1-28%]), eConsulta could replace 63%-88% of conventional appointments. The most frequent uses of the teleconsultation service were for management of test results (35.2%), medical enquiries (16.0%), and the management of repeat prescriptions (12.2%). On average, the teleconsultations consisted of a mean 1.57 messages (SD 0.54 messages); 45.9% (1040/2268) of the teleconsultations consisted of 1 message, and the majority of the remaining teleconsultations consisted of 2-5 interactions. The patient initiated 60.0% (1361/2268) of the teleconsultations.

Conclusions: Based on the GPs’ perceptions, eConsulta could replace 63%-88% of conventional appointments. Therefore, asynchronous teleconsultations between practitioners and patients in primary care could avoid interactions that have limited added clinical value.

(J Med Internet Res 2020;22(3):e14478) doi:10.2196/14478

KEYWORDS

telemedicine; remote consultation; primary care; general practitioners

Introduction

Teleconsultation in the Context of the Public Health System in Catalonia

The Catalan health system provides publicly financed universal health coverage that is free at the point of access, thus ensuring that everyone who lives in Catalonia has the right to health care. It is a mixed health model funded through taxes, with equal access to a wide range of benefits offered by a single publicly available network of health resources, not all of which are publicly owned. The network includes a range of organizations (eg, mutual societies, foundations, health consortiums) that have historically provided health care. One in four individuals in Catalonia has additional private health insurance, which does not forfeit their right to public health care [1].

Catalonia is a pioneer in the use of Information and Communication Technologies (ICT) in public health care. Over the last decade, the adoption of teleconsultations has overtaken the use of other forms of telemedicine [2]. Of the many existing projects, eConsulta is particularly noteworthy. It is an asynchronous teleconsulting service involving general practitioners (GPs) and members of the public connected to the electronic medical history systems of public primary health care. eConsulta was designed to complement face-to-face contact with Primary Care Teams (PCT) in Catalonia. The service was introduced in 2015 and was phased in gradually until 2017, when it became established as a service available to all PCTs. At present, over 90% of the teams have used the tool [3]. Despite the uptake at the institutional level, its use in everyday medical practice is still growing. Interactions via eConsulta are intended to replace low value-added GP visits such as the collection of laboratory results, issuing of repeat prescriptions, and routine follow-up consultations for existing patients who, due to improvements in their symptoms or positive test results, do not require the GP to conduct a physical examination.

From the patient's point of view, eConsulta is one of the services offered in their Personal Health Folder, a personal digital space that provides access to personal health information, allowing them to make enquiries and perform specific actions. Following a secure authentication process, patients can access an interface that allows them to submit their query and attach files, if necessary. The portal can be customized and keeps a record of previous teleconsultations.

Comparison of Teleconsultation and Face-to-face Consultations

A comprehensive study comparing the use of eConsulta and face-to-face visits has not yet been conducted in Catalonia. A survey carried out during the initial stages of the intervention showed that 70% of GPs saw it as "a tool which had the potential to reduce the number of face-to-face visits" [4]. Moreover, no conclusive relationship between the use of teleconsultations and a reduction in the frequency of face-to-face visits has been identified in studies published internationally [5]. What is clear is that the uptake of teleconsultation is still very low and needs to increase if it is to have any real impact on primary care workload and costs [6-8]. Relatively recent studies from similar interventions such as AskMyGP or eConsult, both part of the United Kingdom's National Health Service system, have produced mainly qualitative evidence that seems to indicate that their performance meets expectations in terms of access to the health care system but not in terms of patient autonomy [9] and the role of a local champion, or a clinician with a good understanding of the tool and an interest in using it, is key for the intervention to be perceived as useful in the context of routine practice and therefore successful. However, because it is often necessary to combine the service with face-to-face or telephone consultations, it is perceived as an additional administrative burden for doctors [10,11]. To identify the impact of teleconsultation use on professional workload, we must first increase their use [12]. Teleconsultations involving interactions between GPs and hospital specialists have been subject to more extensive investigation and have produced positive results in terms of reducing waiting times and improving coordination with hospitals [13]. They are also well accepted by users [14,15].

Study Aim

This study aimed to assess the impact of the use of eConsulta on the number of face-to-face visits to PCTs.

Methods

Participants

The study was conducted in the Central Catalonia Health Region, a large, mainly rural area that also includes major cities such as Manresa, Igualada, Vic, Solsona, and Berga. The region's total population exceeds 500,000 people. In this area, 173 GPs had used eConsulta at least once, but most of them had tried it only a few times. The 20 GPs who most frequently used eConsulta and accounted for nearly 70% of the total number of teleconsultations were asked to participate in the study. The invitation was accepted by 18 of the 20 GPs. The study data comes from the administrative data of health provider organizations and covers the period April 8, 2016 to August 18, 2018.

The Territorial Administration Office of the Central Catalonia Region of the Catalan Institute of Health provided the GPs with a register of eConsulta interactions undertaken during the study period. The 18 participating GPs only had access to their own data, for reasons of confidentiality. The text in the subject line and message body was analyzed after the data were anonymized.

Variables in the Analysis

Each GP recorded three pieces of information for each of their interactions: the type of interaction according to the 13 author-proposed categories (Multimedia Appendix 1); whether they believed a face-to-face visit was avoided, which was defined as the absence of the need for a face-to-face visit following the consultation; and whether they believed the patient would have requested a face-to-face visit had eConsulta not been available. The latter was used as an approximate measure of the possible increased demand resulting from the ease of access to a GP. This subjective information was cross-referenced with information registered by the information systems, which is shown in Table 1. With regard to the ID, it refers to the number of teleconsultations, not patients; therefore, the object of the analysis cannot be inferred as the number of participating patients but the number of interactions. With reference to "Message order," the interlocutor is inferred based on who initiated the teleconsultations and the order of the messages. Thus, if a teleconsultation is initiated by the patient, it is assumed that the messages that follow alternate between the GP and the patient, although it is possible that either may have written more than one message in succession

Table 1. Example of an anonymized administrative record.

ID	Initiated by	Title	Date	Message order	Message
306	Patient	Test message	04/13/2017	1	Good morning XXXXX, it's an honor to be the first person to use this service. Cheers!
306	Patient	Test message	04/14/2017	2	Good morning XXXXXX, I hope you find the service useful. Goodbye!

The statistical programs Epi Info v.7.2.2.1 (Division of Health Informatics & Surveillance Center for Surveillance, Epidemiology & Laboratory Services, Atlanta, GA), SPSS v.8 (IBM Corp, Armonk, NY) and R v.3.6.1 (R Project) were used for the statistical analyses. The results were considered significant with $P<.05$. The study was approved by the Ethical Committee for Clinical Research at the Foundation University Institute for Primary Health Care Research Jordi Gol i Gurina (registration number P18/023).

Results

Descriptive Analysis of the Sample of Teleconsultations

A total of 3559 messages corresponding to 2268 teleconsultations were analyzed (mean 1.57 messages per teleconsultation, SD 0.54). The patient initiated 60.0% (1361/2268) of the teleconsultations, and a significant proportion consisted of a single message that did not generate a response (1040/2268, 45.9%). The remainder consisted of mostly 2-5 interactions (see Table 2). Regarding the format, the messages were composed of a title and text body that were on average 17 and 250 characters long, respectively. Texts of messages written by the GPs were slightly longer than those written by patients (mean 280 vs 190 characters, respectively).

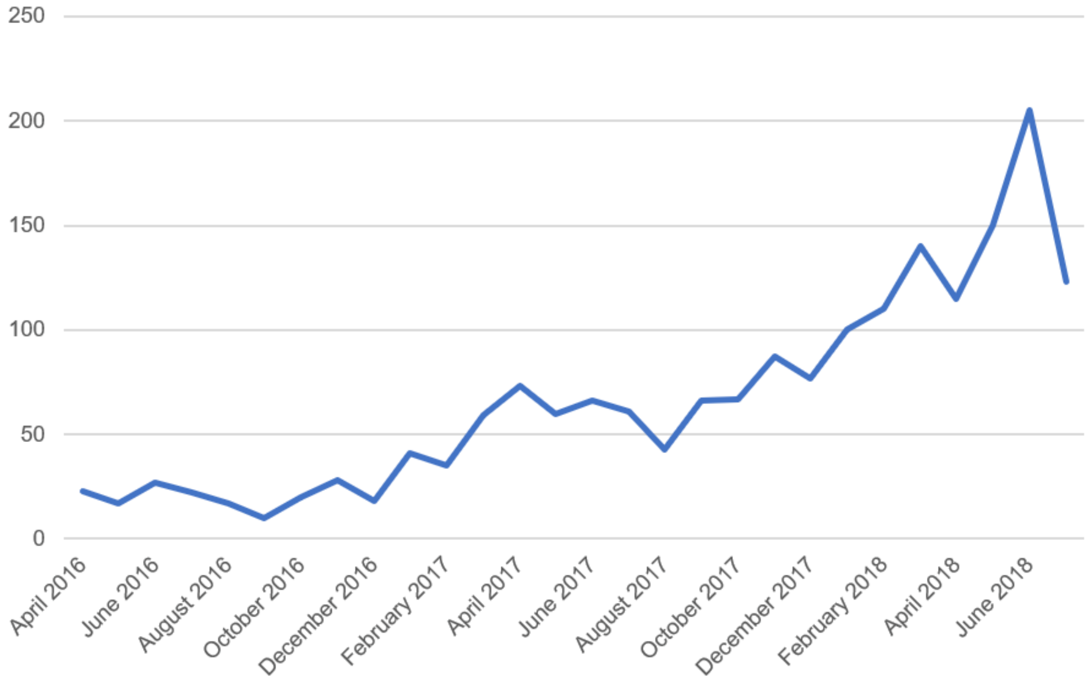
Table 2. Frequency of each number of messages per teleconsultation.

Number of messages per teleconsultation	n
1	1040
2	1177
3	40
4	8
5	3

The number of teleconsultations fluctuated throughout the year, showing an upward trend in the use of the tool over time, with fluctuations in usage corresponding to the months of greater (winter) and lesser (summer holidays) health care activity (Figure 1).

Most messages were sent to individuals, with one notable exception. A GP sent the same message to multiple patients simultaneously, with the following text: "The anti-flu campaign is about to begin. It is recommended that anyone aged over 60 and those suffering from a respiratory infection, heart or kidney disease or diabetes ought to be vaccinated from 23-10-17. See attached file. Ask your nurse for an appointment. Sincerely, Dr XXXX".

Figure 1. Number of teleconsultations per month during the study period.



Types of Teleconsultations

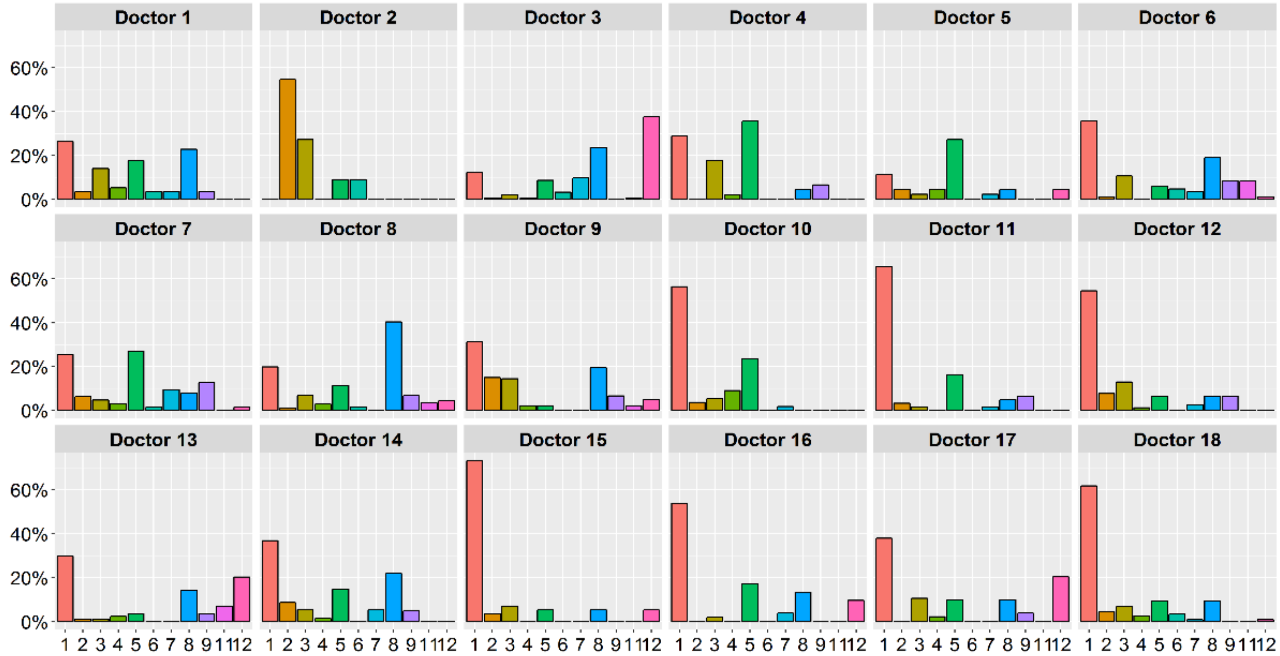
The information in Table 1 was supplemented by linking each case with an anonymous ID and the collection of data from the GPs in the form of a template that included three questions. In response to the question “Did eConsulta avoid a face-to-face visit?”, the GP replied in the affirmative for 88% of the cases. In terms of the types of enquiries, the most common were related to the management of test results, clinical enquiries, and repeat prescriptions, while the least frequent were requests for clinical reports and sick notes and queries related to vaccinations and the use of anticoagulants. Because 112 conversations were not classified correctly, they were excluded from the analysis, leaving 2156 conversations (see Table 3). Errors and test messages constituted some 7% of the total, which demonstrates the experimental nature of the intervention. The subsequent analysis excludes messages corresponding to categories 11 (messages sent in error), 12 (other), and 13 (test messages).

Finally, in terms of the type of teleconsultations engaged in by each GP (Figure 2), while Type 1 was the most frequent, the GPs (numbered 1 to 18 in the figures) favored a specific purpose. For example, see the use by GPs 2, 5, and 8 in Figure 2.

Table 3. Number of teleconsultations, by type.

Type of consultation	n (%)
1. Management of test results	758 (35.2)
2. Temporary disability management	113 (5.2)
3. Arranging an appointment	160 (7.4)
4. Requesting a clinical report/sick note	37 (1.7)
5. Repeat prescription	262 (12.2)
6. Vaccinations	21 (0.97)
7. Other administrative issues	67 (3.1)
8. Medical enquiries	345 (16.0)
9. Issues regarding medicines	79 (3.7)
10. Queries regarding anticoagulants	1 (0.0)
11. Messages sent in error	45 (2.1)
12. Other	144 (6.7)
13 Test messages	124 (5.8)

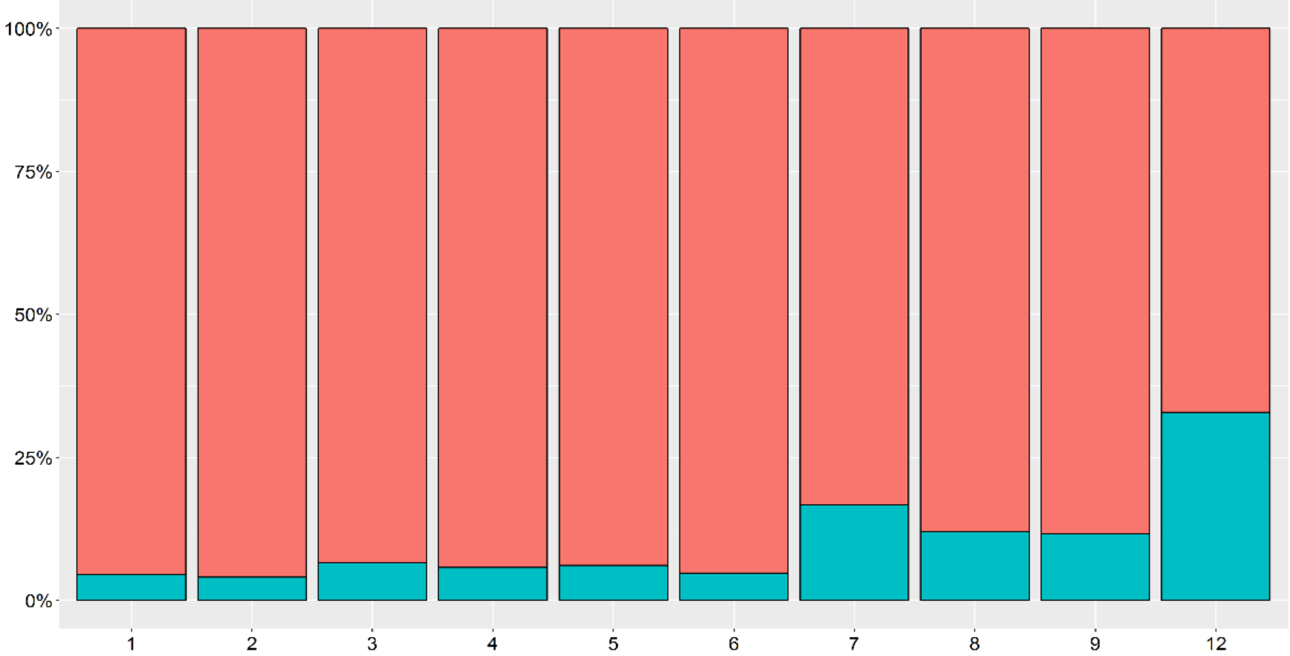
Figure 2. Type of teleconsultation by professional.



Does eConsulta Reduce the Number of Face-to-face Visits?

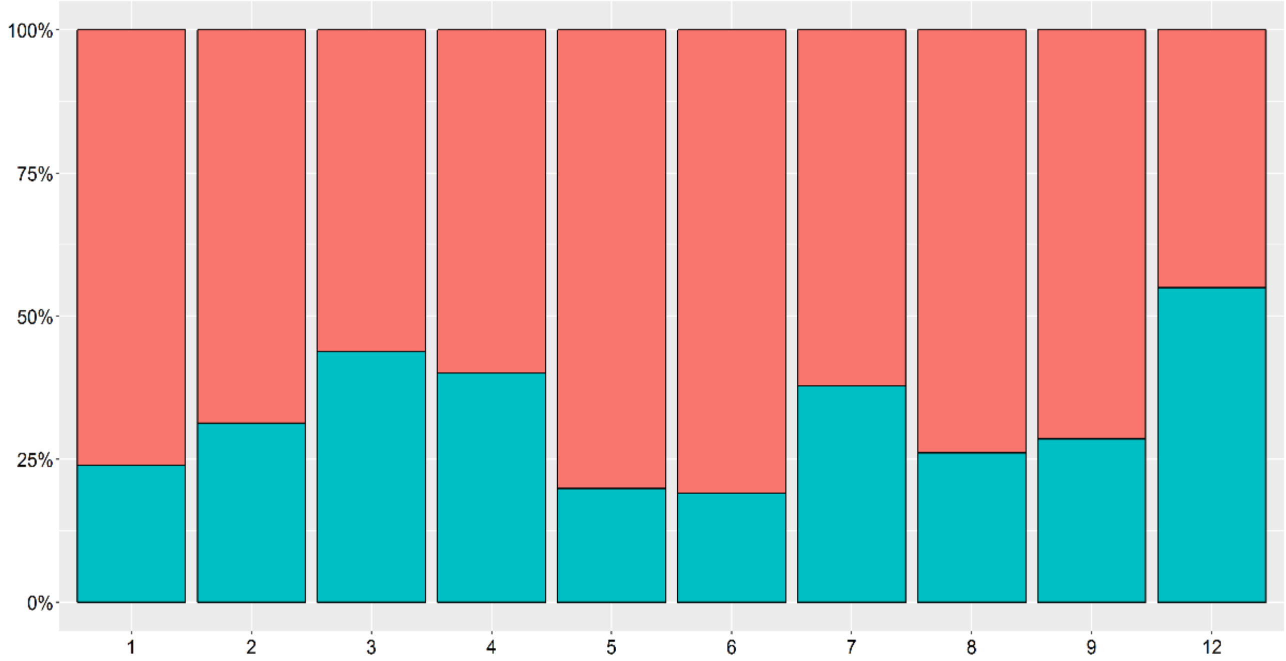
In response to the question “Has the online consultation avoided a face-to-face visit?”, GPs answered yes for 87.9% (1918/2180) of the teleconsultations. A breakdown of the results (Figure 3) shows that the ability to decrease the number of face-to-face visits (mean 0.89, SD 0.08) is largely uniform in terms of the type of consultation. The teleconsultations around which the GPs were the least decisive corresponded to the “Other” category, although there was no indication as to why they fail to avoid a face-to-face visit.

Figure 3. Results of whether each type of teleconsultation avoided a face-to-face visit (upper bar=Yes). Refer to Table 3 for expansions of numbers used on the x-axis.



In response to the question “In the absence of a service like eConsulta, would the patient have made a face-to-face visit?”, GPs answered yes for 72.2% (1574/2180) of the teleconsultations, suggesting that the ease of contact with the GP reduced demand for 27.7% (604/2180) of the cases. In the analysis by type of consultation (Figure 4), tool-facilitated ease of contact occurred mainly for type 12 consultations (Other: mean 0.67, SD 0.12). Therefore, based on the equation (88% x [1-28%]), telemedicine could replace 63%-88% of conventional appointments, which supports the findings of other studies [11].

Figure 4. If eConsulta was not available, whether the patient would have visited the general practitioner's surgery (upper bar=yes), by type of visit. Refer to Table 3 for expansions of numbers used on the x-axis.



Other Analyzed Relationships

Because teleconsultations consisting of a single message were so frequent, we performed additional analyses with these data. They were related primarily to the "Management of test results" category. These correspond to GPs who provide test results that do not require a specific comment (Multimedia Appendix 2). These tests are known as "complementary" in the clinical setting, since they complement the clinical assessment and it makes sense to analyze them outside the GP office as they are normal and of no added clinical value.

Finally, the relationships between the ability to resolve an issue, who initiated the interaction, and the message length were studied. First, who initiated the teleconsultation did not determine its degree of resolution, since the frequency of resolution by teleconsultations was not statistically different between consultations initiated by the patient or GP (*P*=.045; Table 4). With regard to message length, we analyzed whether longer conversations are better able to resolve issues. Although the message length might serve as an indicator of the degree of complexity of the problem, very short messages may not be sufficiently descriptive to resolve the query. The messages were divided equally into three length categories: short, medium, and long. Message length had a statistically significant impact, with longer messages increasing the likelihood of the consultation being satisfactorily concluded (odds ratio 1.66, 95% CI 1.32-2.07; *P*<.001; Multimedia Appendix 3).

Table 4. Degree of resolution in avoiding face-to-face visits, by initiator.

Initiator of the teleconsultation	The teleconsultation avoided a face-to-face visit	%
General practitioner	No	11.87
	Yes	88.12
Patient	No	12.56
	Yes	87.43

Discussion

This study offers useful pointers for policy making since it suggests that eConsulta is a tool that can reduce the number of face-to-face consultations in a primary care setting. The study also provides information about the visit types for which eConsulta is most likely to be used by both patients and GPs. Although a recent study suggested that the topics suitable for teleconsultations will vary depending on the clinical settings and on the individuals who use the tool [16], the current study found that eConsulta is mainly used for the management of test results, to resolve clinical problems, and for queries related to repeat prescriptions. It is worth noting that the second most frequent reason for sending a message was for a medical enquiry, demonstrating that there is demand for a non-face-to-face means to resolve health issues.

This study has several limitations. First, despite efforts to systematize the approach, the evaluation was purely subjective. However, despite the lack of a quantitative approach, it seems logical that GPs, in the absence of a conflict of interest, can realistically classify the intervention's ability to avoid a face-to-face visit. Second, eConsulta is still in the early stages, and GPs are still experimenting with the service, as demonstrated by the high proportion of test messages, message errors, and messages that failed to receive a response. This analysis represents the first steps in the use of the tool. Third, using evaluations by GPs who use the tool most introduces significant bias into the evaluation, since they may not be representative of the population. Future studies ought to assess the impact of the intervention on objective measures.

Although it is challenging to incorporate teleconsultations into the daily clinical workflow, as reported by other authors [17,18], the bottleneck in the deployment of the intervention may not originate with the GPs but instead with the patients. Few patients make use of their Personal Health Folder, the portal through which the eConsulta service is available. As a result, it would be worthwhile to investigate the relationship between face-to-face visits and the use of ICT tools in a more general sense. Future research should include the role of telephone consultations, another type of non-presential service available to GPs, to assess the impact of one type of consultation on the other.

Acknowledgments

We are grateful for the collaboration of the GPs from Central Catalonia of the Catalan Health Institute who participated in this study. The study was conducted with the support of the Secretary of Universities and Research of the Department of Business and Knowledge of the Generalitat de Catalunya.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reasons that patients and general practitioners use eConsulta.
[DOCX File , 16 KB - jmir_v22i3e14478_app1.docx]

Multimedia Appendix 2
Teleconsultations that consisted of a single message, by type.
[PNG File , 44 KB - jmir_v22i3e14478_app2.png]

Multimedia Appendix 3
According to message length, whether the patient would have visited the general practitioner’s office if eConsulta had not been available (left bar=Yes).
[PNG File , 41 KB - jmir_v22i3e14478_app3.png]

References

1. García-Altés A. Desigualdades Socioeconómicas en Salud. Health Policy Papers CRES 2017:1-11.
2. Fundació TIC Salut Social. Informe Mapa de Tendències. URL: https://ticsalutsocial.cat/serveis/observatori/informe-de-tendencies-tic/ [accessed 2020-02-18]
3. Agència d'Avaluació i Qualitat de l'Assistència Sanitària. Central de Resultats 2018. Barcelona: Agència d'Avaluació i Qualitat de l'Assistència Sanitària; 2018.
4. Solans O. Econsulta, desplegando el modelo de atención no presencial en Cataluña. I.S informática y salud (125) 2017:64.
5. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tulledge-Scheitel SM, et al. Impact of patient portal secure messages and electronic visits on adult primary care office visits. Telemed J E Health 2014 Mar;20(3):192-198 [FREE Full text] [doi: 10.1089/tmj.2013.0097] [Medline: 24350803]
6. Huygens MWJ, Swinkels ICS, Verheij RA, Friele RD, van Schayck OCP, de Witte LP. Understanding the use of email consultation in primary care using a retrospective observational study with data of Dutch electronic health records. BMJ Open 2018 Jan 21;8(1):e019233 [FREE Full text] [doi: 10.1136/bmjopen-2017-019233] [Medline: 29358442]
7. Valverde-Albacete J, Folkvord F, Hocking L, Lupiañez-Villanueva F, Altenhofer M, Devaux A, et al. Benchmarking Deployment of EHealth among General Practitioners. Luxembourg: European Union; 2018:1-184.
8. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Nov 22;7(11):e016901 [FREE Full text] [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]
9. Liddy C, Afkham A, Drosinis P, Joschko J, Keely E. Impact of and Satisfaction with a New eConsult Service: A Mixed Methods Study of Primary Care Providers. J Am Board Fam Med 2015 May 08;28(3):394-403 [FREE Full text] [doi: 10.3122/jabfm.2015.03.140255] [Medline: 25957372]
10. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. Br J Gen Pract 2017 Nov 06;68(666):e1-e8. [doi: 10.3399/bjgp17x693509]

11. Cowie J, Calveley E, Bowers G, Bowers J. Evaluation of a Digital Consultation and Self-Care Advice Tool in Primary Care: A Multi-Methods Study. IJERPH 2018 May 02;15(5):896. [doi: 10.3390/ijerph15050896]
12. 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. Health Serv Deliv Res 2018 Jun;6(20):1-200. [doi: 10.3310/hsdr06200]
13. Caffery LJ, Farjian M, Smith AC. Telehealth interventions for reducing waiting lists and waiting times for specialist outpatient services: A scoping review. J Telemed Telecare 2016 Sep 29;22(8):504-512. [doi: 10.1177/1357633x16670495]
14. Tuot DS, Liddy C, Vimalananda VG, Pecina J, Murphy EJ, Keely E, et al. Evaluating diverse electronic consultation programs with a common framework. BMC Health Serv Res 2018 Oct 24:1-18. [doi: 10.1186/s12913-018-3626-4]
15. Keely E, Williams R, Epstein G, Afkham A, Liddy C. Specialist Perspectives on Ontario Provincial Electronic Consultation Services. Telemed J E Health 2019 Jan 10;25(1):3-10 [FREE Full text] [doi: 10.1089/tmj.2018.0012] [Medline: 29746210]
16. Dhruva SS. Implementation of an EConsult System with Patient Navigation. Journal of Health Care for the Poor and Underserved 2019;30(1):28-39. [doi: 10.1353/hpu.2019.0004]
17. Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Sep;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]
18. Mold F, Hendy J, Lai Y, de Lusignan S. Electronic Consultation in Primary Care Between Providers and Patients: Systematic Review. JMIR Med Inform 2019 Dec 3;7(4):e13042. [doi: 10.2196/13042]

Abbreviations

GP: general practitioner.
ICT: Information and Communication Technologies.
PCT: Primary Care Team.

Edited by G Eysenbach; submitted 25.04.19; peer-reviewed by F Saigí-Rubió, J Guanyabens, MA Mayer; comments to author 09.08.19; revised version received 09.09.19; accepted 03.02.20; published 16.03.20.

Please cite as:
López Seguí F, Vidal-Alaball J, Sagarra Castro M, García-Altés A, García Cuyàs F
General Practitioners' Perceptions of Whether Teleconsultations Reduce the Number of Face-to-face Visits in the Catalan Public Primary Care System: Retrospective Cross-Sectional Study
J Med Internet Res 2020;22(3):e14478
URL: http://www.jmir.org/2020/3/e14478/
doi:10.2196/14478
PMID:32175914

Original Paper

Identifying Lung Cancer Risk Factors in the Elderly Using Deep Neural Networks: Quantitative Analysis of Web-Based Survey Data

Songjing Chen[1], PhD; Sizhu Wu[1], PhD

Institute of Medical Information and Library, Chinese Academy of Medical Sciences / Peking Union Medical College, Beijing, China

Corresponding Author:
Songjing Chen, PhD
Institute of Medical Information and Library
Chinese Academy of Medical Sciences / Peking Union Medical College
No 3, Yabao Road, Chaoyang District
Beijing
China
Phone: 86 01052328761
Email: chen.songjing@imicams.ac.cn

Abstract

Background: Lung cancer is one of the most dangerous malignant tumors, with the fastest-growing morbidity and mortality, especially in the elderly. With a rapid growth of the elderly population in recent years, lung cancer prevention and control are increasingly of fundamental importance, but are complicated by the fact that the pathogenesis of lung cancer is a complex process involving a variety of risk factors.

Objective: This study aimed at identifying key risk factors of lung cancer incidence in the elderly and quantitatively analyzing these risk factors' degree of influence using a deep learning method.

Methods: Based on Web-based survey data, we integrated multidisciplinary risk factors, including behavioral risk factors, disease history factors, environmental factors, and demographic factors, and then preprocessed these integrated data. We trained deep neural network models in a stratified elderly population. We then extracted risk factors of lung cancer in the elderly and conducted quantitative analyses of the degree of influence using the deep neural network models.

Results: The proposed model quantitatively identified risk factors based on 235,673 adults. The proposed deep neural network models of 4 groups (age ≥65 years, women ≥65 years old, men ≥65 years old, and the whole population) achieved good performance in identifying lung cancer risk factors, with accuracy ranging from 0.927 (95% CI 0.223-0.525; *P*=.002) to 0.962 (95% CI 0.530-0.751; *P*=.002) and the area under curve ranging from 0.913 (95% CI 0.564-0.803) to 0.931(95% CI 0.499-0.593). Smoking frequency was the leading risk factor for lung cancer in men 65 years and older. Time since quitting and smoking at least 100 cigarettes in their lifetime were the main risk factors for lung cancer in women 65 years and older. Men 65 years and older had the highest lung cancer incidence among the stratified groups, particularly non–small cell lung cancer incidence. Lung cancer incidence decreased more obviously in men than in women with smoking rate decline.

Conclusions: This study demonstrated a quantitative method to identify risk factors of lung cancer in the elderly. The proposed models provided intervention indicators to prevent lung cancer, especially in older men. This approach might be used as a risk factor identification tool to apply in other cancers and help physicians make decisions on cancer prevention.

(J Med Internet Res 2020;22(3):e17695) doi:10.2196/17695

KEYWORDS

deep learning; lung cancer; risk factors; aged; primary prevention

Introduction

Background

Lung cancer is one of the most dangerous malignant tumors, with the fastest-growing morbidity and mortality, especially in the elderly. With the rapid growth of the elderly population in recent years, lung cancer prevention and control are becoming much more important than ever before. Non–small cell lung cancer (NSCLC) is the most common type of lung cancer [1].

Lung cancer pathogenesis is a complex process involving various risk factors. Factors such as smoking [2,3], secondhand smoke [4], high levels of air pollution exposure [5], and drinking water that has a high level of arsenic [6,7] can increase the risk of occurrence of lung cancer. The relationship between these risk factors and lung cancer incidence is an urgent research problem.

In high-income countries, a combination of early diagnosis, screening, and treatment has been effective in increasing population-based survival for certain cancers [8-10]. Many lung cancer screening-related studies have been conducted recently. In the United States, the National Lung Screening Trial was conducted to investigate the possibility that low-dose computed tomography (CT) could reduce lung cancer mortality [11]. Zahnd and Eberth found that use of CT screening was higher than in earlier estimates using 2017 Behavioral Risk Factor Surveillance System (BRFSS) survey data [12]. The US Preventive Services Task Force recommended annual screening of individuals at high risk of lung cancer aged 55 to 80 years who have a 30–pack-year smoking history and currently smoke or had quit within the past 15 years [13]. Berkowitz and colleagues used 2012 BRFSS data to develop multilevel small-area estimate mixed models to generate county-level estimates for 6 smoking status categories (current, some days, every day, former, ever, and never) [14].

Machine learning algorithms are being used more widely for lung cancer screening, detection, diagnosis, and other related research. Luna and colleagues used random forest as an accurate machine learning method to identify known and new predictors of symptomatic radiation pneumonitis, which is a radiotherapy dose-limiting toxicity for locally advanced NSCLC [15]. Palani and Venkatalakshmi used a fuzzy clustering method to predict lung cancer through continuous monitoring using a new internet of things and to improve health care by providing medical instructions [16]. A K-means clustering algorithm, based initially on 400 cancer and non–cancer patients' data, was developed to identify relevant and nonrelevant lung cancer data for early detection of lung cancer [17]. Liu and colleagues used multivariable logistic regression to assess the relationship between body mass index and respiratory conditions, asthma, and chronic obstructive pulmonary disease (COPD) based on BRFSS data [18]. A series of machine learning methods were applied to classify lung cancer patients' survival, including linear regression, decision trees, gradient boosting machines, support vector machines, and a custom ensemble [19]. Deep learning methods were previously rarely used to identify lung cancer risk factors, but their use has become more common recently. Cha and colleagues studied a deep convolutional neural network model to detect operable lung cancer with chest radiographs [20]. Deep learning algorithms might aid fully automated lung cancer detection even at very low effective radiation doses of 0.11 mSv [21]. Hosny and colleagues provided evidence that a convolutional neural network might be used for mortality risk stratification based on standard-of-care CT images from NSCLC patients [22].

Objective

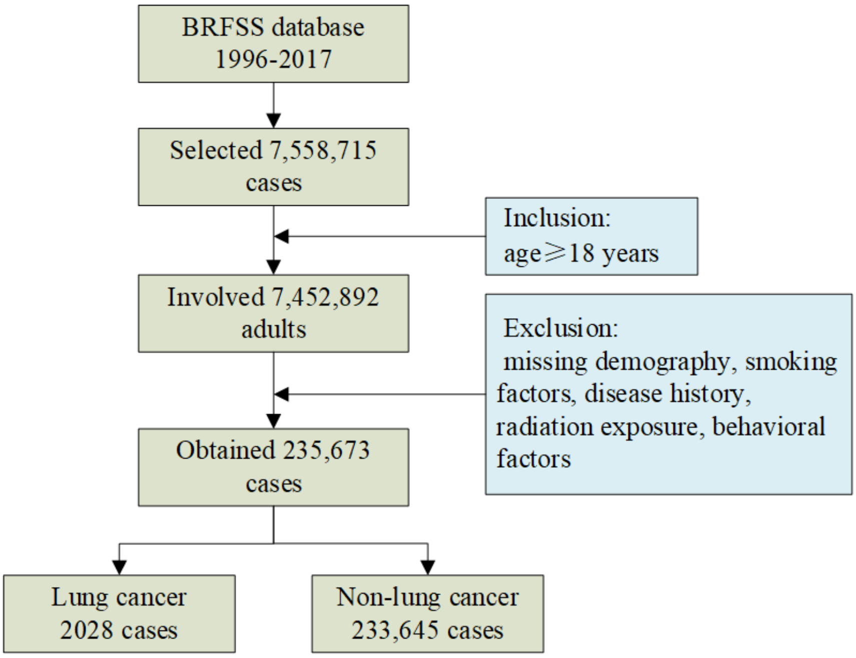
This study aimed at identifying key risk factors of lung cancer incidence in a stratified elderly population and quantitatively analyzing the risk factors' degree of influence using a deep neural network (DNN) method. Using Web-based survey data, we focused on multidisciplinary risk factors, such as smoking habit, disease history, radiation exposure, behavioral risk, environmental risk, and medical demographics. Our main research problems were how to find the leading causative factors of lung cancer incidence from complex related risk factors and to quantitatively analyze their degree of influence. Our results could help physicians in preventing lung cancer and taking effective measures for early detection.

Methods

Data Source

We obtained lung cancer risk factors from the BRFSS [23], an open access source from the US Centers for Disease Control and Prevention. BRFSS collects survey data from US residents about their health-related risk behaviors, chronic health conditions, use of preventive services, and so on. In this study, we used lung cancer behavioral health risk data of 235,673 adults from all 50 US states between 1996 and 2017. The flowchart in Figure 1 shows the data selection process.

Figure 1. Data selection flowchart. BRFSS: Behavioral Risk Factor Surveillance System.



Lung cancer has many causative factors, including age 65 years and older, body mass index, education, smoking habit, personal history of cancer, family history of cancer, CT or computerized axial tomography (CAT) scan, asthma history, and COPD history. Table 1 lists some relevant survey questions from the BRFSS questionnaire that we used to collect data for this study.

Table 1. Lung cancer risk factors assessed by the Behavioral Risk Factor Surveillance System questionnaire.

Risk factors	Description
Age	Age ≥65 years? (yes/no)
Body mass index	Level 1: <18.5 kg/m^2; 2: 18.5-24.9 kg/m^2; 3: 25.0-29.9 kg/m^2; 4: ≥30.0 kg/m^2
Education	Level of education completed (level 1: Did not graduate from high school; 2: Graduated from high school; 3: Attended postsecondary or technical school; 4: Graduated from postsecondary or technical school)
Smoked at least 100 cigarettes	Smoked at least 100 cigarettes in your entire life (yes/no; 1 pack contains 20 cigarettes)
Smoking frequency	Level 1: Every day; 2: Some days; 3: Not at all
Smoking start age	How old were you when you first started to smoke cigarettes regularly? (Age in years)
Smoking intensity	How many cigarettes do you smoke each day? (Number of cigarettes/day)
Smoking quit attempts	During the past 12 months, have you stopped smoking for 1 day or longer? (yes/no)
Time since quitting	How long has it been since you last smoked a cigarette? (1: Within the past month; 2: Within the past 3 months; 3: Within the past 6 months; 4: Within the past year; 5: Within the past 5 years; 6: Within the past 10 years; 7: 10 years or more; 8: Never smoked regularly)
E-cigarette use	Have you ever used an e-cigarette or other electronic vaping product, even just one time? (yes/no)
E-cigarette use frequency	Do you now use e-cigarettes or other electronic vaping products every day, some days, or not at all? (1: Every day; 2: Some days; 3: Not at all)
Chronic obstructive pulmonary disease (COPD) history	History of COPD (yes/no)
Asthma history	History of asthma (yes/no)
Cancer history	Personal history of cancer (yes/no)
Family history of cancer	Family history of cancer (yes/no)
Computed tomography (CT) or computerized axial tomography (CAT) scan	In the last 12 months, did you have a CT or CAT scan? (yes/no)

Participants who were 65 years and older accounted for about 35.01% (82,503/235,673) of the survey population and those aged 18 to 64 years accounted for 64.99% (153,170/235,673). By sex, 53.99% (127,262/235,673) were women and 46.01% (108,411/235,673) were men.

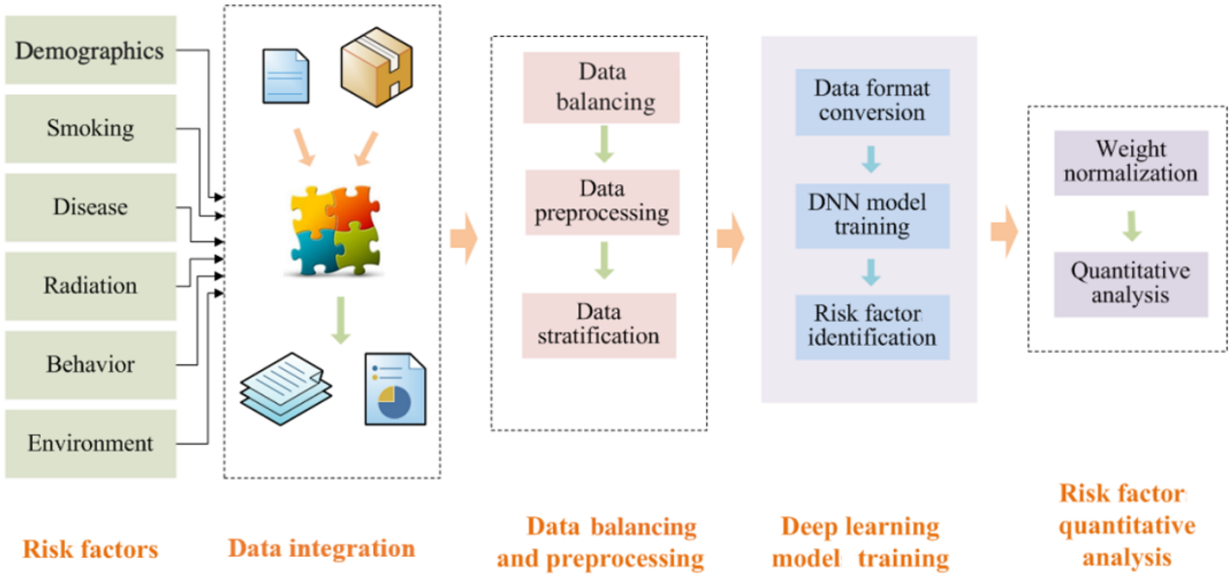
We derived environmental risk factors from the open access website of the US Environmental Protection Agency [24], including air pollutants and drinking water. According to the investigation date, we linked environmental data with risk factors from the BRFSS.

Data Analysis

Overview

In this study, we employed a DNN model to identify risk factors for lung cancer in the elderly. The DNN model had a multiple-layer structure and powerful data expression ability. In particular, in training models based on the large dataset, DNN had high calculation accuracy. First, we integrated the data on medical demographics, smoking habit, disease history, radiation exposure, behavioral risk, and other aspects. Second, since the number of cases of lung cancer was much lower than that of non–lung cancers, we balanced the data. Then we preprocessed these balanced data. Third, we trained DNN models by leveraging the stratified data of the elderly population. We extracted the stratified risk factors through DNN models. Fourth, we developed a quantitative analysis of the degree of effect of the risk factors in elderly patients. Therefore, the whole study comprised 4 sections: data integration, data balancing and preprocessing, training of DNN models, and quantitative analysis of risk factors, as Figure 2 shows.

Figure 2. Schematic diagram of lung cancer risk factor identification in the elderly. DNN: deep neural network.



Data Integration

Lung cancer incidence is caused by multiple risk factors [25-27], particularly in the elderly [28]. We integrated these risk factors, including medical demographics, smoking, disease history, radiation exposure, behavioral risk, and environmental risk. Medical demographic factors were age, sex, body mass index, and education level. Smoking factors were smoking intensity, age when starting to smoke, smoking frequency, time since quitting, e-cigarette use, secondhand smoke exposure, and other smoking habits. Disease history referred to COPD history, asthma history, personal cancer history, and family history of cancer. Radiation exposure involved radiotherapy of the breast or chest, CT or CAT medical imaging examination, and occupational exposure to asbestos, radon, and arsenic. We also took into account dietary and exercise habits and other behavioral risk factors.

Data Balancing and Preprocessing

The ratio of lung cancer to non–lung cancer cases was about 1:115. When studying the pathogenesis of lung cancer, this situation could cause a data imbalance problem. Therefore, we used the synthetic minority oversampling technique (SMOTE) [29] to solve the imbalance problem. SMOTE is based on the K-nearest neighbor algorithm to simulate the minority sample. We then added these simulated samples to the whole dataset.

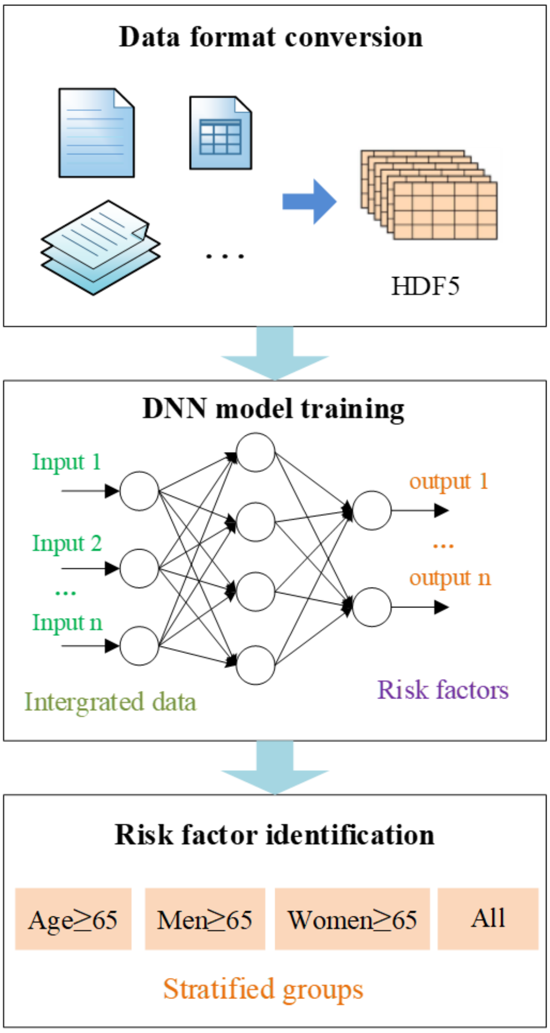
At the same time, the integrated data had vacancy value, incompleteness, and other problems. We therefore preprocessed the data using techniques such as vacancy value filling and noise data smoothing. We used multiple imputation [30] to fill in missing values. We conducted singular value decomposition [31] to reduce data noise in the data preprocessing stage.

We divided the preprocessed data into 4 groups: those aged 65 years and older (age ≥65 years), women aged 65 years and older (women ≥65 years), men aged 65 years and older (men ≥65 years), and the whole population.

Deep Neural Network Model Training

By leveraging the weights of the DNN models, we quantified the degree of influence of risk factors on lung cancer incidence in the elderly (Figure 3). First, we converted the data format into hierarchical data format version 5 (HDF5) [32] in the 4 stratified groups (age ≥65 years, women ≥65 years, men ≥65 years, and the whole group) separately. HDF5 is recognized by Convolutional Architecture for Fast Feature Embedding (Caffe) [33], an open source general deep learning framework. Second, we used the Caffe framework to train DNN models based on the stratified groups in sequence. We input integrated data through an input layer, and then computed the weight values of different risk factors in a hidden layer. We obtained key risk factors using weight values through the output layer of the DNN model. Third, we extracted different risk factors of the stratified groups according to their DNN models.

Figure 3. Deep learning model training process. DNN: deep neural network; HDF5: hierarchical data format version 5.



The DNN model of the group aged 65 years and older consisted of 3 layers: the input layer, hidden layer, and output layer. This model included 1 input layer, 3 hidden layers, and 1 output layer. Layer-to-layer was fully connected. In other words, any neuron in the *i*th layer must be connected to any neuron in the *(i+1)*th layer. Therefore, there was a linear relationship where $z = \sum w_i x_i + b$, plus an activation function, $\sigma(z)$. We used a rectified linear unit function, given in Equation 1 (Figure 4), as an activation function to improve model expression ability. We used 10-fold cross-validation to test algorithm accuracy. We divided the data of the group aged 65 years and older into 10 parts. We rotated them to use 9 of them as a training dataset and 1 as a test dataset for DNN model training.

Figure 4. Data analysis equations.

$$f(x)=\begin{cases}0, & x\le 0\\ x, & x>0\end{cases} \qquad (1)$$

$$\alpha^{l}=\sigma(\mathbf{W}^{l}\alpha^{l-1}+\mathbf{b}^{l}) \qquad (2)$$

$$w_{Normalized}=\frac{w-w_{Min}}{w_{Max}-w_{Min}} \qquad (3)$$

The output results α^L of the DNN model can be deduced from Equation 2 (Figure 4), where W is the weight matrix between the hidden layer and the output layer, which represents the influence of risk factors on lung cancer incidence; L is the number of layers and variable *l* is 2 to L; and b is the bias vector. The numbers of input nodes and output nodes relied on the number of input and output factors, and the hidden-layer number was determined by data size. We set a value of 30 for the input nodes, 3 for the hidden layers, and 9 for the output nodes. In this way, we constructed the DNN model of the group aged 65 years and older. We used the same network structure to train the DNN models of the other 3 stratified groups separately.

Risk Factor Quantitative Analysis

We normalized the weight (*w*) using Equation 3 (Figure 4) to extract key risk factors of lung cancer occurrence. The value of normalized weight ($w_{\text{Normalized}}$) was between 0 and 1. w_{Min} is the minimum value of weight, and w_{Max} is the maximum value of weight. We developed a quantitative analysis of different risk factors in the 4 groups. Because weights represented the degree of influence of risk factors on lung cancer occurrence, we compared the weights of risk factors to identify targeted factors among the 4 stratified groups.

Results

Risk Factor Weights

Figure 5 shows the weights of risk factors in the 4 stratified groups obtained using DNN models. Though leveraging weights of DNN models, we quantitatively analyzed the degree of the risk factors' influence on lung cancer incidence in the elderly. Table 2 shows the values of weights and odds ratios (95% CI) of these main risk factors.

Figure 5. Normalized weights of risk factors in the stratified groups. BMI: body mass index; CAT: computerized axial tomography; COPD: chronic obstructive pulmonary disease; CT: computed tomography; PM2.5: fine particulate matter with a diameter ≤2.5 μm.

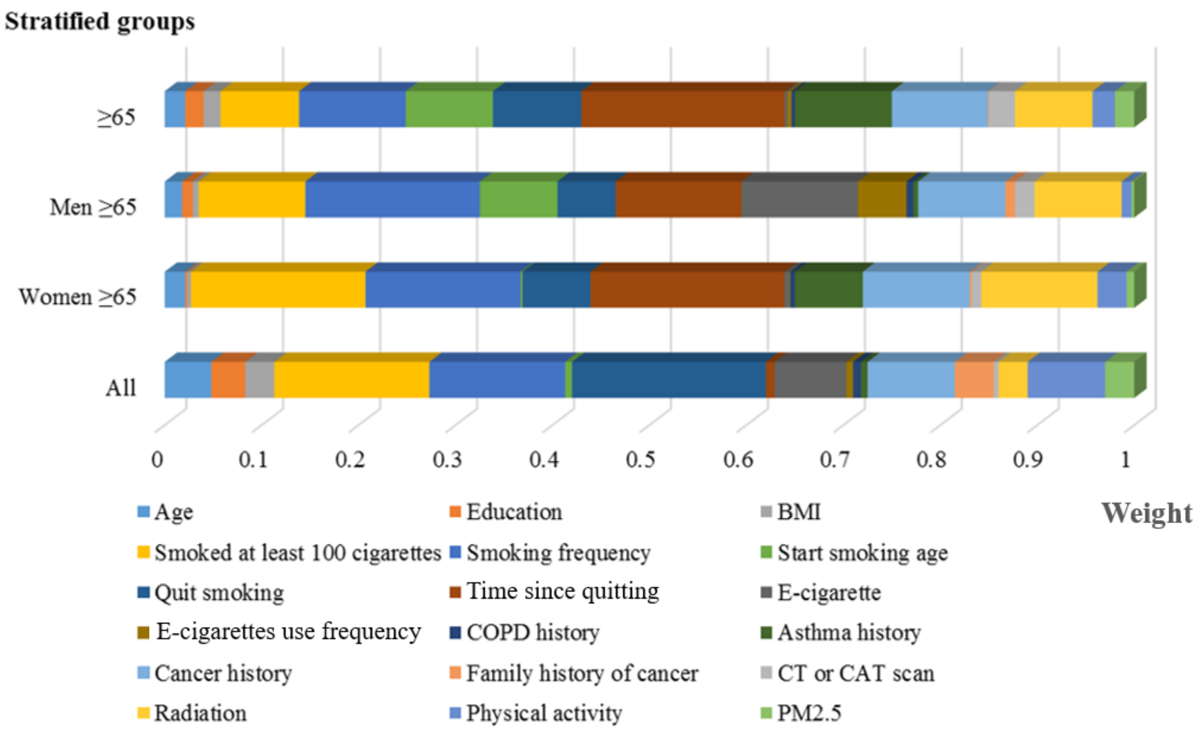


Table 2. Normalized weight values and odds ratios (95% CI) of the main risk factors in the 4 population groups.

Risk factors	Population aged ≥65 years		Men aged ≥65 years		Women aged ≥65 years		All age groups	
	Weight	Odds ratio (95% CI)	Weight	Odds ratio (95% CI)	Weight	Odds ratio (95% CI)	Weight	Odds ratio (95% CI)
Time since quitting	0.21	1.422 (0.806-1.095)	0.13	1.587 (0.776-0.998)	0.20	1.590 (0.927-1.358)	0.009	1.109 (0.993-1.322)
Smoking frequency	0.11	1.312 (0.796-0.998)	0.18	1.625 (0.866-1.097)	0.16	1.536 (1.106-1.427)	0.14	1.370 (1.352-1.701)
Cancer history	0.099	1.295 (0.876-1.027)	0.09	1.387 (1.239-1.667)	0.11	1.442 (0.951-1.356)	0.09	1.271 (0.852-1.201)
Smoking quit attempts	0.091	1.253 (0.933-1.201)	0.06	1.273 (1.413-1.702)	0.07	1.368 (1.127-1.406)	0.20	1.405 (0.995-1.381)
Lifetime smoking of ≤100 cigarettes	0.081	1.239 (1.336-1.587)	0.11	1.506 (0.681-0.937)	0.18	1.588 (1.237-1.601)	0.16	1.387 (1.225-1.611)
Asthma history	0.08	1.303 (1.029-1.403)	0.005	1.095 (0.962-1.329)	0.07	1.381 (0.953-1.317)	0.007	1.112 (0.961-1.406)
Radiation	0.08	1.224 (1.550-1.781)	0.09	1.291 (0.983-1.307)	0.12	1.453 (1.302-1.759)	0.03	1.190 (0.952-1.357)
E-cigarette use	0.023	1.025 (0.766-0.934)	0.12	1.539 (1.112-1.406)	0.005	1.135 (0.897-1.309)	0.074	1.239 (0.851-1.307)
Physical activity	0.023	1.132 (0.983-1.246)	0.01	1.170 (0.851-1.209)	0.03	1.280 (0.991-1.308)	0.08	1.268 (1.131-1.670)

Effect of Risk Factors on Lung Cancer

Those aged 65 years and older were more sensitive to how long ago former smokers had quit and smoking frequency, which were related to smoking. This correlation was more obvious in men aged 65 years and older. Those aged 65 years and older who had quit smoking for a short time or smoked more every day were prone to lung cancer.

Smoking frequency was the leading risk factor for lung cancer in men aged 65 years and older. As Table 2 shows, the weights of smoking frequency and time since quitting were 0.18 and 0.13, respectively, in this group of men. The weight of smoking frequency was 38.5% higher than the weight of time since quitting. The top 4 risk factors of men aged 65 years and older (smoking frequency, time since quitting, use of e-cigarettes, and having smoked at least 100 cigarettes in their lifetime) were all associated with smoking. These smoking-related risk factors had a greater influence than other risk factors on men who were 65 years and older. Men in this age group who actively quit smoking were more likely to avoid lung cancer.

Time since quitting and smoking at least 100 cigarettes over their lifetime were the main risk factors for lung cancer occurrence in women aged 65 years and older. As Table 2 shows, the weight of time since quitting was 0.20 in this group of women, which was 11.1% greater than the weight of having smoked at least 100 cigarettes (0.18). The top 3 relevant risk factors were associated with smoking habit factors in women aged 65 years and older: time since quitting, having smoking at least 100 cigarettes, and smoking frequency. Therefore, smoking-related risk factors had a greater influence than other risk factors on women in this age group.

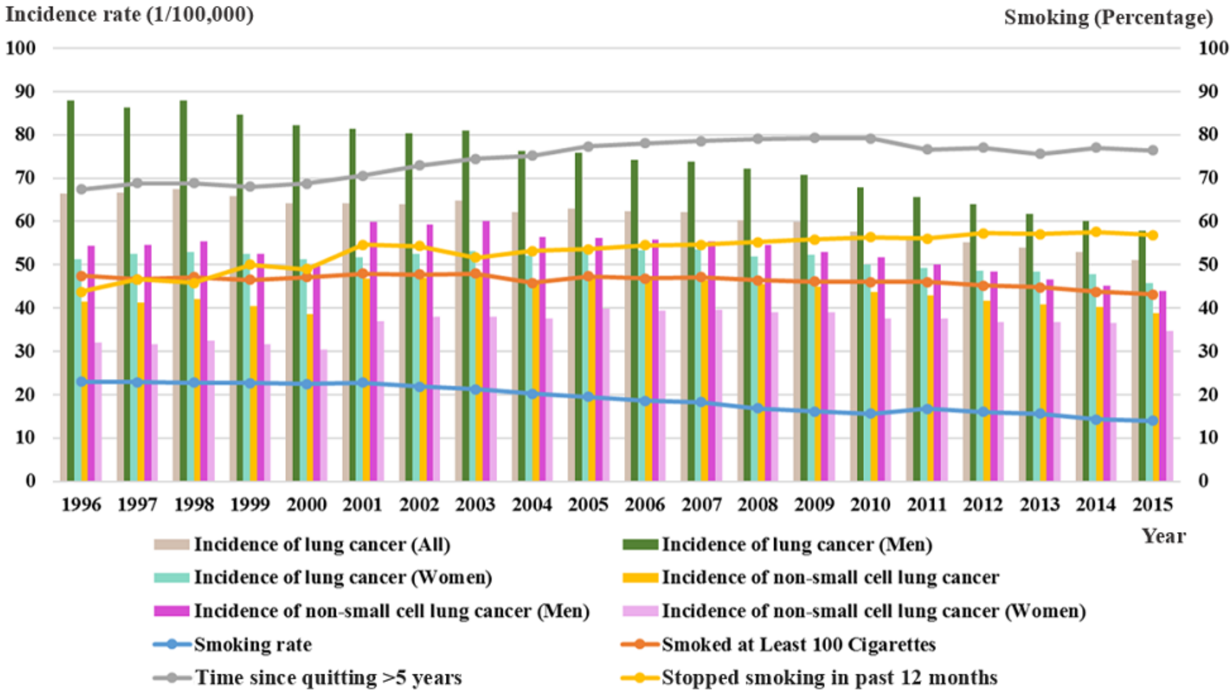
Cancer history ranked in the top risk factors in the 4 stratified groups, which may suggest that cancer history played an important role in the incidence of lung cancer [34,35]. Women aged 65 years and older were more sensitive to radiation exposure than were other groups. Physical activity was the fifth risk factor in the whole group.

Association Between Smoking and Lung Cancer Incidence

Men aged 65 years and older had the highest lung cancer incidence in these stratified groups, especially the incidence of NSCLC. We compared the incidence rate of lung cancer, NSCLC, and small cell lung cancer among all ages, under 65 years, and 65 years and older. NSCLC incidence in men 65 years and older was 286 cases per 100,000 people between 2011 and 2015, which was higher than that of women aged 65 years and older (203 per 100,000). Therefore, controlling smoking in men age 65 years and older could be more effective in preventing lung cancer.

Lung cancer incidence decreased much more rapidly in men than in women with a decline in smoking rate, as Figure 6 shows. The smoking rate curve shows that the number of smokers decreased between 1996 and 2015, from 23% to 14% (a decrease of about 39.1 percentage points). Smoking rate has declined continuously in recent years. Figure 6 also shows that the incidence of lung cancer in men declined from 88 per 100,000 in 1996 to 58 per 100,000 in 2015, a reduction of 34.1 percentage points. As a result, lung cancer incidence had decreased along with smoking rate declining in men.

Figure 6. Relationship between smoking and lung cancer incidence, 1996-2015.



Accuracy of Deep Neural Network Models

Table 3 summarizes the performance of the 4 DNN models. The proposed models had good accuracy and area under the receiver operating characteristic curve (AUROC), using the whole group as a baseline to reveal lung cancer incidence in elderly patients. Accuracies were 96.2% (95% CI 0.530-0.751, *P*=.002) for age 65 years and older, 94.3% (95% CI 0.459-0.643, *P*=.015) for men 65 years and older, and 93.2% (95% CI 0.437-0.689, *P*=.003) for women 65 years and older, which were higher than the whole group accuracy of 92.7% (95% CI 0.223-0.525, *P*=.002). Moreover, AUROCs were 0.931 (95% CI 0.499-0.593) for age 65 years and older, 0.927 (95% CI 0.506-0.681) for men 65 years and older, and 0.926 (95% CI 0.543-0.782) for women 65 years and older, performing better than the whole group at 0.913 (95% CI 0.564-0.803). This proposal model efficiently output identified risk factors, which was timesaving.

Table 3. Performance of the 4 DNN models.

Model	Accuracy (95% CI)	AUROC[a] (95% CI)	*P* value[b]
≥65 years	0.962 (0.530-0.751)	0.931(0.499-0.593)	.002
Men ≥65 years	0.943 (0.459-0.643)	0.927 (0.506-0.681)	.015
Women ≥65 years	0.932 (0.437-0.689)	0.926 (0.543-0.782)	.003
All	0.927 (0.223-0.525)	0.913 (0.564-0.803)	.002

[a] AUROC: area under the receiver operating characteristic curve.

[b] *P*<.05 was considered to indicate statistical significance.

Discussion

Principal Findings

We developed, to our knowledge, the first deep learning classification model to quantitatively identify corresponding risk factors for lung cancer for stratified groups of elderly people. By leveraging the weights of the DNN models, we identified risk factors for lung cancer in the elderly and quantitatively analyzed the risk factors' degree of influence. The proposed DNN models of 4 groups (age ≥65 years, women ≥65 years, men ≥65 years, and the whole population) achieved good performance in identifying lung cancer risk factors, with accuracy ranging from 0.927 (95% CI 0.223-0.525, *P*=.002) to 0.962 (95% CI 0.530-0.751, *P*=.002) and AUROCs ranging from 0.913 (95% CI 0.564-0.803) to 0.931 (95% CI 0.499-0.593). The proposed models had a fast training speed and high accuracy and efficiency compared with logistic regression [18] and previous models for targeted identification of lung cancer risk factors [12,36-40].

In recent years, the deep learning method has been applied more frequently in lung cancer detection and prediction due to its advantages of high accuracy and fast computing speed. Hosny and colleagues used deep learning networks to predict mortality risk stratification of patients with NSCLC [22]. Cha and colleagues found that a deep learning method had high diagnostic performance in detecting operable lung cancer with chest radiographs [20]. The DNN model, which we proposed to extract risk factors, could also be applied to provide intervention indicators for lung cancer prevention and carry out targeted intervention measures.

Through integrating multidisciplinary data, we the employed the DNN method to identify key lung cancer risk factors in the elderly. We computed quantitative weights of different risk factors in a stratified population to deduce their degrees of influence on lung cancer incidence. Our results showed that DNN models identified specific risk factors of targeted elderly people. People who were 65 years or older were more sensitive to time since quitting and smoking frequency, especially in men in this age group: smoking frequency was the leading causative risk factor for lung cancer in men 65 years and older. Time since quitting and smoking at least 100 cigarettes over a lifetime were the main risk factors for lung cancer in women 65 years and older. Men 65 years and older had the highest lung cancer incidence in these stratified groups. Lung cancer incidence decreased more obviously in men than in women with a decline in smoking rate. Cancer history played an important role in the incidence of lung cancer. Taking part in more physical activities to enhance physical quality might reduce lung cancer incidence [41,42]. Smoking-related factors (eg, smoking frequency, time since quitting, smoking at least 100 cigarettes) were important risk factors for lung cancer in elderly patients. Risk factors such as smoking-related factors, exercise, and cancer history were intervention indicators in preventing lung cancer. Tammemagi and colleagues found that smokers aged 65 to 80 years were a high-risk group who might benefit from low-dose CT lung cancer screening [43]. Chen and colleagues found that regional application of effective primary cancer prevention strategies on smoking, poor diet, and other modifiable risk factors had a vast potential to reduce the burden of cancer and disparities in China [9]. These suggested that interventional measures targeting the main risk factors might be possible to prevent lung cancer occurrence.

Comparison With Prior Work

Previously, researchers conducted several models to identify lung cancer risk factors [36-40]. Table 4 shows a comparison of our model with previous models. Compared with previous models, our proposed model identified risk factors for lung cancer in the elderly with high accuracy and AUROC. Our model used data from a larger population, more lung cancer occurrence-related risk factors, and a more efficient identification algorithm than previous models. Our DNN models had faster training speeds than previous models when training on the same scale of big data, which could save a lot of time. Moreover, we balanced and preprocessed the data before training the DNN models, which was helpful to improve model accuracy effectively.

Table 4. Comparison of our model with previous models for identifying lung cancer risk factors.

Model	Population	Method	Risk factors	Accuracy	AUROC[a]
Our model	235,673	Deep neural network	As listed in the Results section	0.927	0.913
Panayiotis, 2016 [36]	25,486	Dynamic Bayesian network	Demographics, smoking status, family history of cancer, cancer history, comorbidities related to lung cancer, occupational exposures, and low-dose computed tomography screening outcomes	0.65	0.75
Wang, 2019 [37]	961	Conditional Gaussian Bayesian network	Age, sex, level of education, region, urbanization, diagnosis-based factors, prior utilization factors, prescription factors	0.67	N/A[b]
Ankit, 2012 [38]	70,132	Decision tree	Age, birthplace, cancer grade, diagnostic confirmation, farthest extension of tumor, type of surgery performed, reason for no surgery, order of surgery and radiation therapy, scope of regional lymph node surgery	0.863	0.91
Xie, 2014 [39]	1703	Artificial neural network	41 risk factors: age, education level, marital status, income status, smoking, alcohol drinking, coffee intake, etc	0.838	N/A
Kaviarasi, 2019 [40]	321	Gaussian classifier	Age, sex, radiation sequence with surgery, first malignant primary indicator, radiation, etc	N/A	0.881

[a] AUROC: area under the receiver operating characteristic curve.

[b] Not available.

Some aspects of our results were similar to the results of these previous studies. In our results, smoking was the leading cause of lung cancer in the elderly. This view was consistent with the reported literature [2,44-46]. Nevertheless, we focused on some original findings in stratified groups of older people.

Limitations

This study had several limitations. First, we mainly focused on modifiable risk factors of lung cancer in the elderly. In the future, we should validate these identified modifiable risk factors using a simulated intervention process to prevent lung cancer. Second, because we used open survey data, we did not obtain the participants' genetic and dietary factors. We are matching the data to source region now and we will analyze lung cancer risk factors by region in the future.

Conclusions

This study demonstrated a quantitative method to identify risk factors for lung cancer in the elderly. The proposed models provided intervention indicators to prevent lung cancer, especially in older men, which could be used with effective intervention methods to reduce lung cancer incidence in the elderly and improve their life quality in their later years. This approach might be used as a risk factor identification tool in other cancers and help physicians make decisions on cancer prevention.

Acknowledgments

This study was supported by the General Project on Humanities and Social Science Research of the Ministry of Education of China under grant no 19YJC870002, the National Key R&D Program of China under grant no 2016YFC0901602, and the Medical and Health Technology Innovation Project of the Chinese Academy of Medical Sciences under grant no 2019-I2M-2-002.

Conflicts of Interest

None declared.

References

1. U.S. National Library of Medicine. MedlinePlus. Non-small cell lung cancer. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health; 2019. URL: https://medlineplus.gov/ency/article/007194.htm [accessed 2019-06-20]
2. Schuller HM. The impact of smoking and the influence of other factors on lung cancer. Expert Rev Respir Med 2019 Aug;13(8):761-769. [doi: 10.1080/17476348.2019.1645010] [Medline: 31311354]
3. Park SK, Cho LY, Yang JJ, Park B, Chang SH, Lee K, Scientific Committee, Korean Academy of Tuberculosis and Respiratory Diseases. Lung cancer risk and cigarette smoking, lung tuberculosis according to histologic type and gender in a population based case-control study. Lung Cancer 2010 Apr;68(1):20-26. [doi: 10.1016/j.lungcan.2009.05.017] [Medline: 19545930]
4. Hahn EJ, Hooper M, Riker C, Butler KM, Rademacher K, Wiggins A, et al. Lung cancer worry and home screening for radon and secondhand smoke in renters. J Environ Health 2017;79(6):8-13 [FREE Full text] [Medline: 29135198]
5. Yang W, Zhao H, Wang X, Deng Q, Fan W, Wang L. An evidence-based assessment for the association between long-term exposure to outdoor air pollution and the risk of lung cancer. Eur J Cancer Prev 2016 May;25(3):163-172. [doi: 10.1097/CEJ.0000000000000158] [Medline: 25757194]
6. Lamm SH, Boroje IJ, Ferdosi H, Ahn J. Lung cancer risk and low (50 g/L) drinking water arsenic levels for US counties (2009-2013)-a negative association. Int J Environ Res Public Health 2018 Jun 07;15(6):1-21 [FREE Full text] [doi: 10.3390/ijerph15061200] [Medline: 29880761]
7. Cheng M, Chiu H, Tsai S, Chen C, Yang C. Calcium and magnesium in drinking-water and risk of death from lung cancer in women. Magnes Res 2012;25(3):112-119 [FREE Full text] [doi: 10.1684/mrh.2012.0318] [Medline: 23073359]
8. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, CONCORD Working Group. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 2018 Mar 17;391(10125):1023-1075 [FREE Full text] [doi: 10.1016/S0140-6736(17)33326-3] [Medline: 29395269]
9. Chen W, Xia C, Zheng R, Zhou M, Lin C, Zeng H, et al. Disparities by province, age, and sex in site-specific cancer burden attributable to 23 potentially modifiable risk factors in China: a comparative risk assessment. Lancet Glob Health 2019 Feb;7(2):e257-e269 [FREE Full text] [doi: 10.1016/S2214-109X(18)30488-1] [Medline: 30683243]
10. Lobach DF, Johns EB, Halpenny B, Saunders T, Brzozowski J, Del Fiol G, et al. Increasing complexity in rule-based clinical decision support: the symptom assessment and management intervention. JMIR Med Inform 2016 Nov 08;4(4):e36 [FREE Full text] [doi: 10.2196/medinform.5728] [Medline: 27826132]
11. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 4;365(5):395-409 [FREE Full text] [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]
12. Zahnd WE, Eberth JM. Lung cancer screening utilization: a behavioral risk factor surveillance system analysis. Am J Prev Med 2019 Aug;57(2):250-255. [doi: 10.1016/j.amepre.2019.03.015] [Medline: 31248742]
13. Moyer VA, U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2014 Mar 4;160(5):330-338. [doi: 10.7326/M13-2771] [Medline: 24378917]
14. Berkowitz Z, Zhang X, Richards TB, Peipins L, Henley SJ, Holt J. Multilevel small-area estimation of multiple cigarette smoking status categories using the 2012 Behavioral Risk Factor Surveillance System. Cancer Epidemiol Biomarkers Prev 2016 Oct;25(10):1402-1410 [FREE Full text] [doi: 10.1158/1055-9965.EPI-16-0244] [Medline: 27697795]
15. Luna JM, Chao H, Diffenderfer ES, Valdes G, Chinniah C, Ma G, et al. Predicting radiation pneumonitis in locally advanced stage II-III non-small cell lung cancer using machine learning. Radiother Oncol 2019 Apr;133:106-112. [doi: 10.1016/j.radonc.2019.01.003] [Medline: 30935565]
16. Palani D, Venkatalakshmi K. An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification. J Med Syst 2018 Dec 18;43(2):21. [doi: 10.1007/s10916-018-1139-7] [Medline: 30564924]

17. Ahmed K, Emran AA, Jesmin T, Mukti RF, Rahman MZ, Ahmed F. Early detection of lung cancer risk using data mining. Asian Pac J Cancer Prev 2013;14(1):595-598 [FREE Full text] [doi: 10.7314/apjcp.2013.14.1.595] [Medline: 23534801]
18. Liu Y, Pleasants RA, Croft JB, Lugogo N, Ohar J, Heidari K, et al. Body mass index, respiratory conditions, asthma, and chronic obstructive pulmonary disease. Respir Med 2015 Jul;109(7):851-859 [FREE Full text] [doi: 10.1016/j.rmed.2015.05.006] [Medline: 26006753]
19. Lynch CM, Abdollahi B, Fuqua JD, de Carlo AR, Bartholomai JA, Balgemann RN, et al. Prediction of lung cancer patient survival via supervised machine learning classification techniques. Int J Med Inform 2017 Dec;108:1-8 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.09.013] [Medline: 29132615]
20. Cha MJ, Chung MJ, Lee JH, Lee KS. Performance of deep learning model in detecting operable lung cancer with chest radiographs. J Thorac Imaging 2019 Mar;34(2):86-91. [doi: 10.1097/RTI.0000000000000388] [Medline: 30802232]
21. Schwyzer M, Ferraro DA, Muehlematter UJ, Curioni-Fontecedro A, Huellner MW, von Schulthess GK, et al. Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks - initial results. Lung Cancer 2018 Dec;126:170-173. [doi: 10.1016/j.lungcan.2018.11.001] [Medline: 30527183]
22. Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. PLoS Med 2018 Nov;15(11):e1002711 [FREE Full text] [doi: 10.1371/journal.pmed.1002711] [Medline: 30500819]
23. US Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System.: U.S. Department of Health & Human Services; 2019. URL: https://www.cdc.gov/brfss/ [accessed 2020-01-27]
24. United States Environmental Protection Agency. Environmental Data database. 2019. URL: https://www.epa.gov/
25. Samet JM, Avila-Tang E, Boffetta P, Hannan LM, Olivo-Marston S, Thun MJ, et al. Lung cancer in never smokers: clinical epidemiology and environmental risk factors. Clin Cancer Res 2009 Sep 15;15(18):5626-5645 [FREE Full text] [doi: 10.1158/1078-0432.CCR-09-0376] [Medline: 19755391]
26. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc 2008 May;83(5):584-594 [FREE Full text] [doi: 10.4065/83.5.584] [Medline: 18452692]
27. Zheng S, Jabbour SK, O'Reilly SE, Lu JJ, Dong L, Ding L, et al. Automated information extraction on treatment and prognosis for non-small cell lung cancer radiotherapy patients: clinical study. JMIR Med Inform 2018 Feb 01;6(1):e8 [FREE Full text] [doi: 10.2196/medinform.8662] [Medline: 29391345]
28. Im Y, Park HY, Shin S, Shin SH, Lee H, Ahn JH, et al. Prevalence of and risk factors for pulmonary complications after curative resection in otherwise healthy elderly patients with early stage lung cancer. Respir Res 2019 Jul 04;20(1):136 [FREE Full text] [doi: 10.1186/s12931-019-1087-x] [Medline: 31272446]
29. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002 Jun 01;16(6):321-357. [doi: 10.1613/jair.953]
30. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res 1999 Mar;8(1):3-15. [Medline: 10347857]
31. Kalman D. A singularly valuable decomposition: the SVD of a matrix. Coll Math J 1996 Jan;27(1):2. [doi: 10.2307/2687269]
32. The HDF group. HDF5 format. 2019. URL: https://support.hdfgroup.org/HDF5/
33. Caffe. Caffe deep learning framework. 2019. URL: http://caffe.berkeleyvision.org/
34. Jatoi I, Anderson WF, Miller AB, Brawley OW. The history of cancer screening. Curr Problems Surg 2019 Apr;56(4):138-163. [doi: 10.1067/j.cpsurg.2018.12.006]
35. Zhou H, Huang Y, Qiu Z, Zhao H, Fang W, Yang Y, et al. Impact of prior cancer history on the overall survival of patients newly diagnosed with cancer: a pan-cancer analysis of the SEER database. Int J Cancer 2018 Oct 01;143(7):1569-1577 [FREE Full text] [doi: 10.1002/ijc.31543] [Medline: 29667174]
36. Petousis P, Han SX, Aberle D, Bui AAT. Prediction of lung cancer incidence on the low-dose computed tomography arm of the National Lung Screening Trial: a dynamic Bayesian network. Artif Intell Med 2016 Sep;72:42-55 [FREE Full text] [doi: 10.1016/j.artmed.2016.07.001] [Medline: 27664507]
37. Wang K, Chen J, Wang K. Medical expenditure estimation by Bayesian network for lung cancer patients at different severity stages. Comput Biol Med 2019 Mar;106:97-105. [doi: 10.1016/j.compbiomed.2019.01.015] [Medline: 30708222]
38. Ankit A, Sanchit M, Ramanathan N, Lalith P, Aloi C. Lung cancer survival prediction using ensemble data mining on SEER data. Sci Programming 2012;20(1):29-42 [FREE Full text] [doi: 10.3233/SPR-2012-0335]
39. Xie N, Hu L, Li T. Lung cancer risk prediction method based on feature selection and artificial neural network. Asian Pac J Cancer Prev 2014;15(23):10539-10542 [FREE Full text] [doi: 10.7314/apjcp.2014.15.23.10539] [Medline: 25556505]
40. Kaviarasi R, Gandhi RR. Accuracy enhanced lung cancer prognosis for improving patient survivability using proposed Gaussian classifier system. J Med Syst 2019 May 24;43(7):201. [doi: 10.1007/s10916-019-1297-2] [Medline: 31127444]
41. Granger CL, McDonald CF, Berney S, Chao C, Denehy L. Exercise intervention to improve exercise capacity and health related quality of life for patients with Non-small cell lung cancer: a systematic review. Lung Cancer 2011 May;72(2):139-153. [doi: 10.1016/j.lungcan.2011.01.006] [Medline: 21316790]
42. Arbane G, Tropman D, Jackson D, Garrod R. Evaluation of an early exercise intervention after thoracotomy for non-small cell lung cancer (NSCLC), effects on quality of life, muscle strength and exercise tolerance: randomised controlled trial. Lung Cancer 2011 Feb;71(2):229-234. [doi: 10.1016/j.lungcan.2010.04.025] [Medline: 20541832]

43. Tammemägi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, et al. Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. PLoS Med 2014 Dec;11(12):e1001764 [FREE Full text] [doi: 10.1371/journal.pmed.1001764] [Medline: 25460915]
44. Guo NL, Wan Y. Pathway-based identification of a smoking associated 6-gene signature predictive of lung cancer risk and survival. Artif Intell Med 2012 Jun;55(2):97-105 [FREE Full text] [doi: 10.1016/j.artmed.2012.01.001] [Medline: 22326768]
45. Giuliani ME, Liu G, Xu W, Dirlea M, Selby P, Papadakos J, et al. Implementation of a novel electronic patient-directed smoking cessation platform for cancer patients: interrupted time series analysis. J Med Internet Res 2019 Apr 09;21(4):e11735 [FREE Full text] [doi: 10.2196/11735] [Medline: 30964445]
46. Wraith D, Mengersen K. Assessing the combined effect of asbestos exposure and smoking on lung cancer: a Bayesian approach. Stat Med 2007 Feb 28;26(5):1150-1169. [doi: 10.1002/sim.2602] [Medline: 16779874]

Abbreviations

AUROC: area under the receiver operating characteristic curve
BRFSS: Behavioral Risk Factor Surveillance System
Caffe: Convolutional Architecture for Fast Feature Embedding
CAT: computerized axial tomography
COPD: chronic obstructive pulmonary disease
CT: computed tomography
DNN: deep neural network
HDF5: hierarchical data format version 5
NSCLC: non–small cell lung cancer
SMOTE: synthetic minority oversampling technique

Edited by G Eysenbach; submitted 04.01.20; peer-reviewed by A Brown, W Clinton; comments to author 18.01.20; revised version received 19.01.20; accepted 22.01.20; published 17.03.20.

Please cite as:
Chen S, Wu S
Identifying Lung Cancer Risk Factors in the Elderly Using Deep Neural Networks: Quantitative Analysis of Web-Based Survey Data
J Med Internet Res 2020;22(3):e17695
URL: http://www.jmir.org/2020/3/e17695/
doi:10.2196/17695
PMID:32181751

Original Paper

Impact of Illness on Electronic Health Use (The Seventh Tromsø Study - Part 2): Population-Based Questionnaire Study

Luis Marco-Ruiz[1], MSc, PhD; Rolf Wynn[2,3], MD, PhD; Sunday Oluwafemi Oyeyemi[4], MD, MSc; Andrius Budrionis[1], MSc, PhD; Kassaye Yitbarek Yigzaw[1], MSc, PhD; Johan Gustav Bellika[1,2], MSc, PhD

[1]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway
[2]Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway
[3]Division of Mental Health and Addictions, University Hospital of North Norway, Tromsø, Norway
[4]Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Johan Gustav Bellika, MSc, PhD
Department of Clinical Medicine
Faculty of Health Sciences
UiT The Arctic University of Norway
Postboks 6050 Langnes
Tromsø, 9037
Norway
Phone: 47 95748049
Email: johan.gustav.bellika@ehealthresearch.no

Abstract

Background: Patients who suffer from different diseases may use different electronic health (eHealth) resources. Thus, those who plan eHealth interventions should take into account which eHealth resources are used most frequently by patients that suffer from different diseases.

Objective: The aim of this study was to understand the associations between different groups of chronic diseases and the use of different eHealth resources.

Methods: Data from the seventh survey of the Tromsø Study (Tromsø 7) were analyzed to determine how different diseases influence the use of different eHealth resources. Specifically, the eHealth resources considered were use of apps, search engines, video services, and social media. The analysis contained data from 21,083 participants in the age group older than 40 years. A total of 15,585 (15,585/21,083; 73.92%) participants reported to have suffered some disease, 10,604 (10,604/21,083; 50.29%) participants reported to have used some kind of eHealth resource in the last year, and 7854 (7854/21,083; 37.25%) participants reported to have used some kind of eHealth resource in the last year and suffered (or had suffered) from some kind of specified disease. Logistic regression was used to determine which diseases significantly predicted the use of each eHealth resource.

Results: The use of apps was increased among those individuals that (had) suffered from psychological problems (odds ratio [OR] 1.39, 95% CI 1.23-1.56) and cardiovascular diseases (OR 1.12, 95% CI 1.01-1.24) and those part-time workers that (had) suffered from any of the diseases classified as others (OR 2.08, 95% CI 1.35-3.32). The use of search engines for accessing health information increased among individuals who suffered from psychological problems (OR 1.39, 95% CI 1.25-1.55), cancer (OR 1.26, 95% CI 1.11-1.44), or any of the diseases classified as other diseases (OR 1.27, 95% CI 1.13-1.42). Regarding video services, their use for accessing health information was more likely when the participant was a man (OR 1.31, 95% CI 1.13-1.53), (had) suffered from psychological problems (OR 1.70, 95% CI 1.43-2.01), or (had) suffered from other diseases (OR 1.43, 95% CI 1.20-1.71). The factors associated with an increase in the use of social media for accessing health information were as follows: (had) suffered from psychological problems (OR 1.65, 95% CI 1.42-1.91), working part time (OR 1.35, 95% CI 0.62-2.63), receiving disability benefits (OR 1.42, 95% CI 1.14-1.76), having received an upper secondary school education (OR 1.20, 95% CI 1.03-1.38), being a man with a high household income (OR 1.67, 95% CI 1.07-2.60), suffering from cardiovascular diseases and having a high household income (OR 3.39, 95% CI 1.62-8.16), and suffering from respiratory diseases while being retired (OR 1.95, 95% CI 1.28-2.97).

Conclusions: Our findings show that different diseases are currently associated with the use of different eHealth resources. This knowledge is useful for those who plan eHealth interventions as they can take into account which type of eHealth resource may be used for gaining the attention of the different user groups.

(J Med Internet Res 2020;22(3):e13116) doi:10.2196/13116

KEYWORDS

eHealth; internet; search engines; mobile apps; social media

Introduction

Background

This is the second paper of a series of 4 that studies electronic Health (eHealth) consumption using the data gathered by the seventh survey of the Tromsø Study (Tromsø 7). In the study by Wynn et al (part 1) [1], we present main findings regarding characteristics of the participants and their use of eHealth. In this second paper, we focus on understanding how long-term or chronic diseases influence the choice of one eHealth resource over another. In the study by Budrionis et al (part 3) [2], we examine outcomes of the use of eHealth, and in the study by Yigzaw et al (part 4) [3], we study how eHealth consumption influences actual doctor visits.

The overall aim of the series was to provide a clearer overview of the characteristics of eHealth users and their interaction with the health care sector. As a matter of fact, the health care sector in many developed economies is facing challenges that include aging populations, lack of workforce, and insufficient coordination among caregivers and services [4-7]. At the same time, the use of information and communication technology is increasing among citizens [8-10]. In the United States, 84% of the population has access to the internet [8]. In Norway, 85% of the population uses the internet on a daily basis [11]. The increase in the use of technology is also powered by the broad access to mobile phones and tablets. In 2012, 85% of US adults owned a mobile phone and 31% had used it to look for health information [12]. The broad adoption of smartphones and ubiquitous access to the internet have led to a steady increase in the use of technology that may be used for health purposes, such as search engines, social media, and Web-based video services [13-17].

eHealth can be understood as the “intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the internet and related technologies” [18]. Access to technology allows citizens to easily access health information and monitor their health status with, for example, mobile apps. It is known that chronic conditions can influence the use of the internet for seeking health information [9,19-22]. The appropriate use of technology has the potential to improve patients’ health and make them more knowledgeable about their condition [23-25]. However, the vast amount of health-related information available on the internet also includes irrelevant information and misinformation [26-28]. Typical challenges for patients with chronic disorders when looking for health information on the internet are finding appropriate online resources and filtering online health information [29].

Interventions and Challenges of Electronic Health

Currently, many health trusts are promoting eHealth interventions [30-34]. These interventions focus on eHealth resources such as mobile apps, social media, video services, and search engines on the internet, among others [8-10]. These interventions have focused on improving health care by guiding health consumers to the most appropriate service [35-37], improving treatment adherence [30,31], or involving patients in shared decision making [38]. A strong focus on these interventions has been set on long-term and chronic diseases such as cardiovascular diseases, cancer, and psychological problems [31,33,39].

Some studies have shown that technology can improve treatment adherence for chronic patients [30,31]. Examples are the positive impact of mobile apps and social media on the management of chronic diseases such as diabetes and epilepsy [30,40]. Most studies have focused on studying the effect of eHealth resources on the management of a condition. However, another important aspect that has been less explored is how a particular condition predisposes to the use of one type of eHealth resource over another. This knowledge is important to decide what type of eHealth resource is the most appropriate for every eHealth intervention. However, to our knowledge, the importance of diseases for eHealth use has not been explored with a sufficient sample size to find which eHealth resources are preferred by different groups of chronic patients.

Tromsø 7 included a questionnaire about the use of eHealth. In a series of 4 papers, we explore data from the Tromsø Study questionnaire analyzing the relationships among eHealth use and other demographic and clinical variables. The large sample size of the Tromsø 7 offers the opportunity to compare eHealth preferences in different patient groups.

Methods

The Tromsø Study

The Tromsø Study is a longitudinal population-based study conducted in the municipality of Tromsø, Norway, since 1974 [41,42]. Its original purpose was to determine the reasons for the high mortality due to cardiovascular diseases in Norway. However, over time, it has expanded, and currently, it covers many different diseases such as mental disorders, cancer, and osteoporosis, to name a few [42]. The study is funded directly by the Norwegian Government. The study is conducted by the University of Tromsø in collaboration with the Norwegian Institute of Public Health and others [42]. The most recent version is Tromsø 7 , comprising the years 2015 and 2016 [42]. The Tromsø Study focuses on a range of chronic diseases and conditions. In Tromsø 7, people aged 40 years or older were included, which provided a sample of 21,083 participants accounting for 64.69% (21,083/32,591) of the total invited. A personal invitation was mailed to all residents in Tromsø aged 40 years or older [41,42] together with a paper-based questionnaire and a link to an electronic questionnaire. Those who chose to participate could complete the questionnaire in paper or electronically at home. Alternatively, they could do so when they attended the study center, where they were also included in other tests. Those who did not respond to the initial mailed invitation were mailed a follow-up reminder.

Part 1 of this series of papers has already presented the characteristics of the participants in Tromsø 7 [1].

Questionnaire

The questionnaire in Tromsø 7 included data regarding many diseases, symptoms, and lifestyle and contained in total more than 300 questions. Examples of the data included are dietary habits; medication; sleeping patterns; socioeconomic status; education; work; and, the most relevant for this study, the use of eHealth resources.

The eHealth questions were selected based on a review of prior literature and with a particular focus on prior studies involving Norwegian participants. As there were strict limits on the number of items (because of the overall size of the questionnaire), only the main questions regarding eHealth services were included (as described below).

The Tromsø 7 questionnaire completed by participants contained several blocks of information. In this study, we focused on a subset of the information contained in the questionnaire. The information considered in this study is as follows:

- Demographics: including questions about age, gender, education, household income (expressed in Norwegian kroner [kr] and US dollars), lifestyle, and occupation.
- Groups of diseases: the participant suffers or has suffered from a cardiovascular disease (high blood pressure, heart attack, heart failure, atrial fibrillation, angina, and stroke), respiratory disease (bronchitis and asthma), cancer, psychological problems, or other disorders (rheumatoid arthritis, arthrosis, diabetes, kidney disease, migraine, and chronic pain). Participants could choose any of the specific diseases available in the questionnaire (ie, more than 1 disease if relevant). We grouped specific diseases before the statistical analyses were performed.
- Emotional: live with a spouse and support from friends.
- Use of eHealth resources: participants were asked: “How often during the last year have you used the following internet-services for information and advice on health and disease issues: Applications (‘Apps’) for smart phone or tablet?, Search engines (like Google)?, Social media (like Facebook)?, Video services (like YouTube)?” For each question, the participants could answer either “never,” “once,” “a few times,” or “often.” The participants who answered that they had used minimum 1 of the eHealth services were thereafter asked: “If you during the last year have used internet-services for information and advice on health and disease issues, based on the information you found on the internet: Have you decided to go to the doctor?, Have you decided not to go the doctor?, Have you discussed the information with a doctor?, Have you changed your medication without consulting a doctor?, Have you been unsure whether the treatment you have received is correct?, Have you decided to seek out complementary or alternative treatment?, Have you made lifestyle changes?, Have you felt anxiety?, Have you felt reassured?, Have you felt more knowledgeable?, Have you felt more confused?” For each of the questions, the participants could answer either “never,” “once,” “a few times,” or “often.” All the questions and response-options have been published on the Tromsø Study website [43].

Statistical Analysis

We used multivariable logistic regression to determine which variables influenced the use of eHealth resources. We proceeded in 2 steps. First, a general model predicting the use of any type of eHealth resources was estimated using the whole dataset. For this, we defined a binary variable that indicated if the participant had used any of the eHealth resources or none. The use of each type of eHealth resource (mobile apps, search engines, video services, or social media) was analyzed separately by regressing the dependent variable that represented each type of eHealth resource with the independent variables previously presented. For the second step, we used the subset of patients that had some of the diseases under study (independent of whether or not they used some eHealth resource). In this way, we identified the specific variables that most strongly influenced the use of each type of eHealth resource.

Multimedia Appendix 1 shows the diseases and the eHealth resources considered. In addition, it also shows the demographic variables included in models. Age was treated as a continuous variable. Household income and education were treated as ordinal variables that represented increasing degrees of the feature represented. Occupation was represented as a categorical variable. The groups of diseases considered were coded as dichotomous variables that represented the presence (value=1) or absence (value=0) of any of the diseases included in the group. Similarly, the use and nonuse of different dichotomous resources (mobile apps, search engines, Web videos, and social media) were coded as another dichotomous variable (use=1 and nonuse=0). The sex of the participant was also represented by 0=woman and 1=man. We studied the interactions between age, sex, occupation, education, household income, and the diseases included in the study. All the independent variables were included for the estimation of every model.

Observations with missing data were excluded from the analysis when any of the missing variables (dependent or independent) needed for calculating each logistic regression model were missing. The reader should note that this caused a variation in the total sample available for each specific model, but the procedure maximized the amount of data available for the estimation of each model. This is a common practice to increase the robustness of the statistical model (pairwise exclusion) [44]. We adjusted for covariates by including possible confounders and interactions in the logistic regression models [45]. Models were then simplified excluding nonsignificant variables and interactions. Deviance analysis was performed to check that the models were significant in predicting the use of eHealth resources. All analyses were 2-sided, and *P* values were considered statistically significant at a level of <.05.

Ethics

The Regional Committee for Medical and Health Research Ethics approved the study (REK Nord, reference 2014/940). All participants provided written informed consent.

Results

User Statistics

This section presents the results of the statistical analysis performed on Tromsø 7 data to establish which variables influence the use of each type of eHealth resource. First, this section presents the analysis of eHealth resources as a combined variable that represents any type of eHealth resource (apps, search engines, video services, or social media). Second, this section presents the results of analyzing the relationship between different disease groups and the use of specific eHealth resources for those individuals that suffered from at least one disease. Some interactions are available in Multimedia Appendix 2.

Multimedia Appendix 1 shows the data regarding the demographic characteristics of the sample selected. A total of 15,585 (8565 men, 7020 women) out of 21,083 (73.92%) participants (had) suffered from some kind of disease, 10,604 out of 21,083 (50.29%) participants reported to have used some kind of eHealth resource in the last year, and 7854 out of 21,083 (37.25%) participants reported both to have used some kind of eHealth resource in the last year (apps, search engines, Web videos, or social media) and suffered (or had suffered) from some kind of specified disease. By disease group, of the total 21,083 participants, 34.00% (7169/21,083) participants (had) suffered from some cardiac disease, 7.76% (1636/21,083) from cancer, 12.91% (2723/21,083) from psychological problems, 12.99% (2738/21,083) from respiratory diseases, and 52.69% (11,109/21,083) from any of the diseases included in the others group. More details about the demographic characteristics can be seen in part 1 of this series of studies. In addition, Multimedia Appendix 1 displays the use of eHealth resources per patient group.

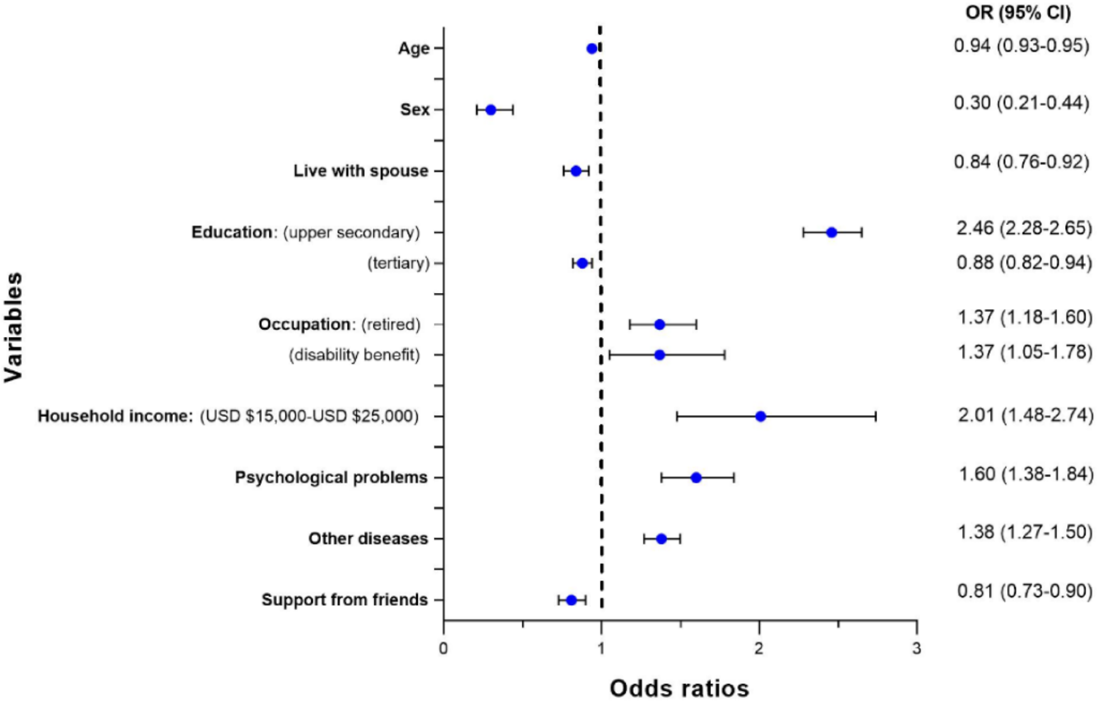
Study of the Use of Electronic Health Resources in General

The first model estimated the use of any eHealth resource. The sample size after removing the respondents that had any missing values was 18,578 individuals. We found that various groups of diseases have a significant effect on the use of eHealth resources. Moreover, different diseases are related to the use of different types of eHealth resources.

Attending to the odds ratios (ORs) in Multimedia Appendix 2, it is possible to see that an increment in age (OR 0.94, 95% CI 0.93-0.95), being a man (OR 0.3, 95% CI 0.21-0.44), living with a spouse (OR 0.84, 95% CI 0.76-0.92), receiving support from friends (OR 0.81, 95% CI 0.73-0.90), and having received college education for less than 4 years (OR 0.88, 95% CI 0.82-0.94) were associated with a decrease in the use of eHealth resources in general.

Figure 1 shows the forest plot summarizing the significant variables that predicted the use of eHealth resources in general (apps, search engines, videos, or social media). The full result of the analysis is available in Multimedia Appendix 2. Having received education of upper secondary school (OR 2.46, 95% CI 2.28-2.65), being retired (OR 1.38, 95% CI 1.18-1.60), receiving a disability benefit (OR 1.37, 95% CI 1.05-1.78), having a household income between US $15,000-$25,000 (OR 2.01, 95% CI 1.48-2.74), suffering from psychological problems (OR 1.60, 95% CI 1.38-1.84), and suffering from any of the diseases contained in the group named *other diseases* (OR 1.38, 95% CI 1.27-1.50) were associated with an increase in the use of eHealth resources in general.

Figure 1. Forest plot for the logistic regression for the use of any electronic health resource.



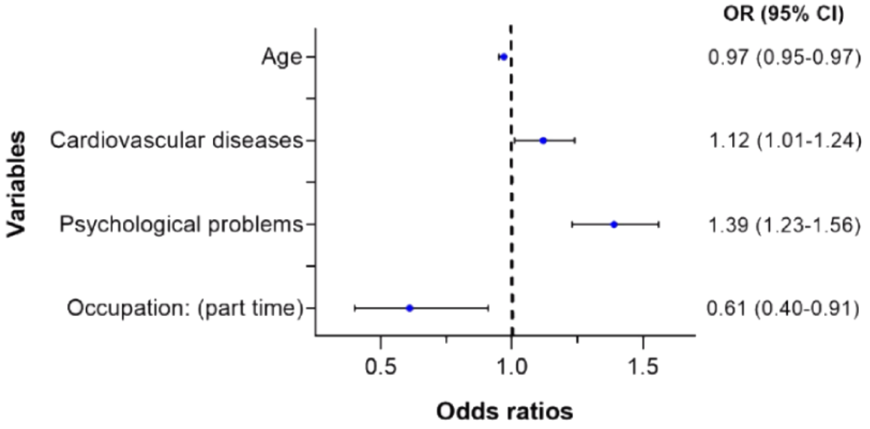
Study of the Use of Specific Electronic Health Resources

For studying the effect of each disease on the use of eHealth resources, we selected a subsample containing all the participants that suffered from any of the diseases previously presented (n=15,585). Observations containing missing data were only excluded if any of the variables needed for the regression analysis were missing.

Study of the Use of Mobile Apps

The sample size used by the statistical software after removing the observations missing any of the variables used by the mobile apps regression model was 15,321 individuals. Figure 2 summarizes the significant disease groups and demographic characteristics related to the users of apps contained in the subsample. In addition, it contains the ORs from the regression model predicting the use of mobile apps and the influence of each independent variable. The full result of the analysis is available in Multimedia Appendix 3.

Figure 2. Forest plot for the logistic regression for the use of mobile apps.



A total of 2 main factors were associated with a decrease in the use of mobile apps: age and part-time workers. As age increased, there was a decreasing use of mobile apps (OR 0.97, 95% CI 0.95-0.97). In addition, those included in the work group representing part-time employees were associated with a decrease in the use of mobile apps for accessing health information (OR 0.61, 95% CI 0.40-0.91).

There were 2 main diseases that were associated with an increase in the use of apps for accessing health information: psychological problems (OR 1.39, 95% CI 1.23-1.56) and cardiovascular diseases (OR 1.12, 95% CI 1.01-1.24).

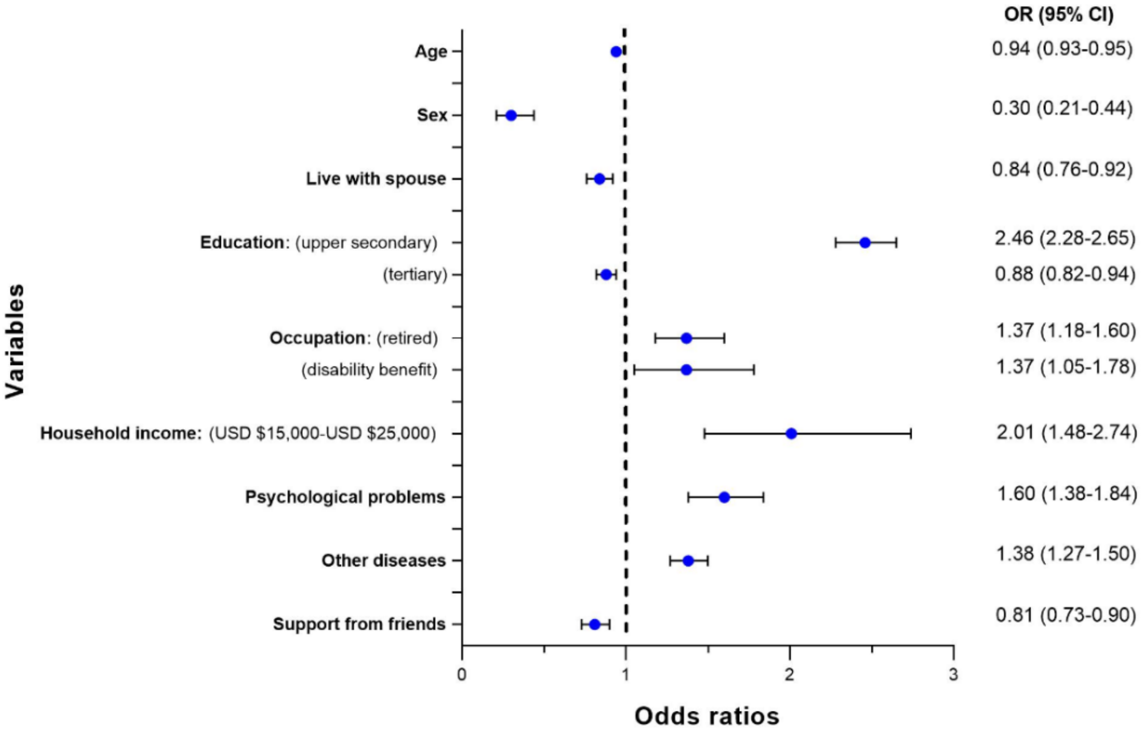
Suffering from any of the diseases contained in the group *other diseases* did not have a significant influence over the use of mobile apps.

Study of the Use of Search Engines

The sample size used by the statistical software after removing the observations missing any of the variables used by the search engines model was 13,610 individuals. Figure 3 summarizes the significant disease groups and demographic characteristics related to the users of search engines contained in the subsample. In addition, it contains the ORs from the regression model predicting the use of Web search engines and the influence of each independent variable. The full result of the analysis is available in Multimedia Appendix 4.

From the logistic regression model, it is possible to interpret that having an educational level of upper secondary education (OR 2.54, 95% CI 2.33-2.77), having a household income of US $15,000-$25,000 (OR 2.57, 95% CI 1.86-3.60), suffering from psychological problems (OR 1.39, 95% CI 1.25-1.55), suffering from cancer (OR 1.26, 95% CI 1.11-1.44), suffering from some of the diseases included in the group *other diseases* (OR 1.27, 95% CI 1.13-1.42), or being retired (OR 1.31, 95% CI 1.07-1.59) contributed to increasing the use of Web search engines for health information. Increasing age (OR 0.94, 95% CI 0.93-0.95), being a man (OR 0.32, 95% CI 0.21-0.50), living with the spouse (OR 0.82, 95% CI 0.73-0.92), having less than 4 years of college education (OR 0.85, 95% CI 0.79-0.92), having support from friends (OR 0.80, 95% CI 0.71-0.90), and having a household income between US $55,100-$75,000 (OR 0.74, 95% CI 0.62-0.87) were associated with a decrease in the use of Web search engines for accessing health information.

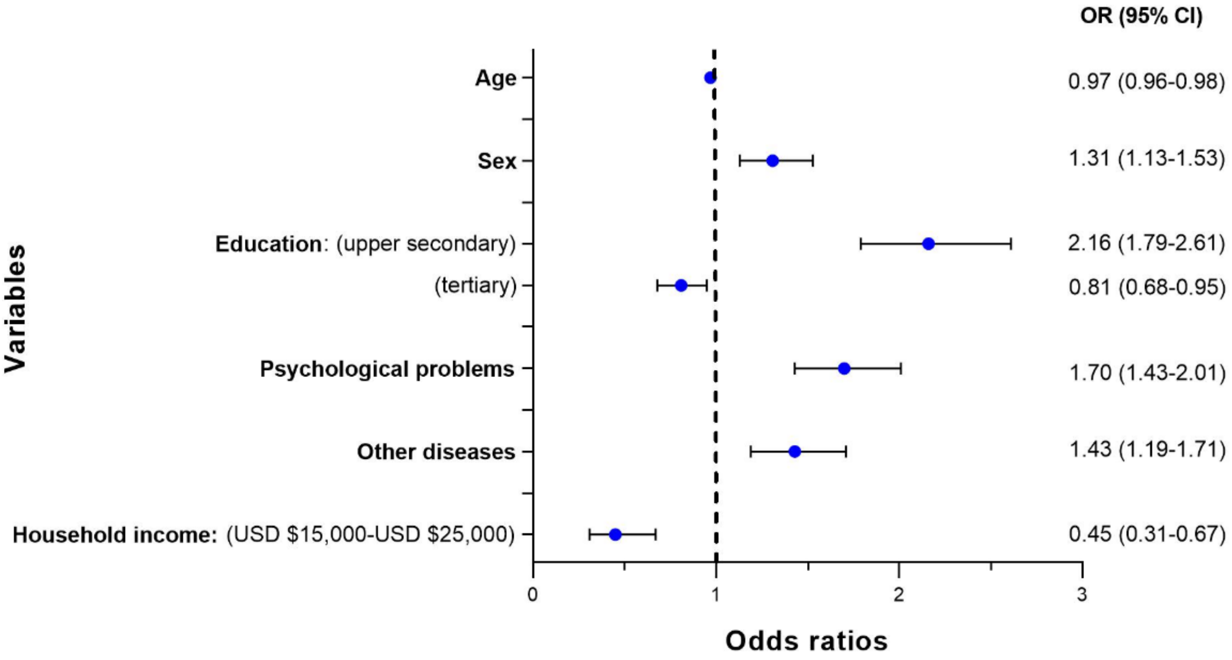
Figure 3. Forest plot for the logistic regression for the use of search engines.



Study of the Use of Video Services

The sample size used by the statistical software after removing the observations missing any of the variables used by the model for video services was 14,724 individuals. Figure 4 summarizes the significant disease groups and demographic characteristics related to the users of video services contained in the subsample. In addition, it contains the ORs from the regression model predicting the use of video services and the influence of each independent variable. The full result of the analysis is available in Multimedia Appendix 5.

Figure 4. Forest plot for the logistic regression for the use of video services.



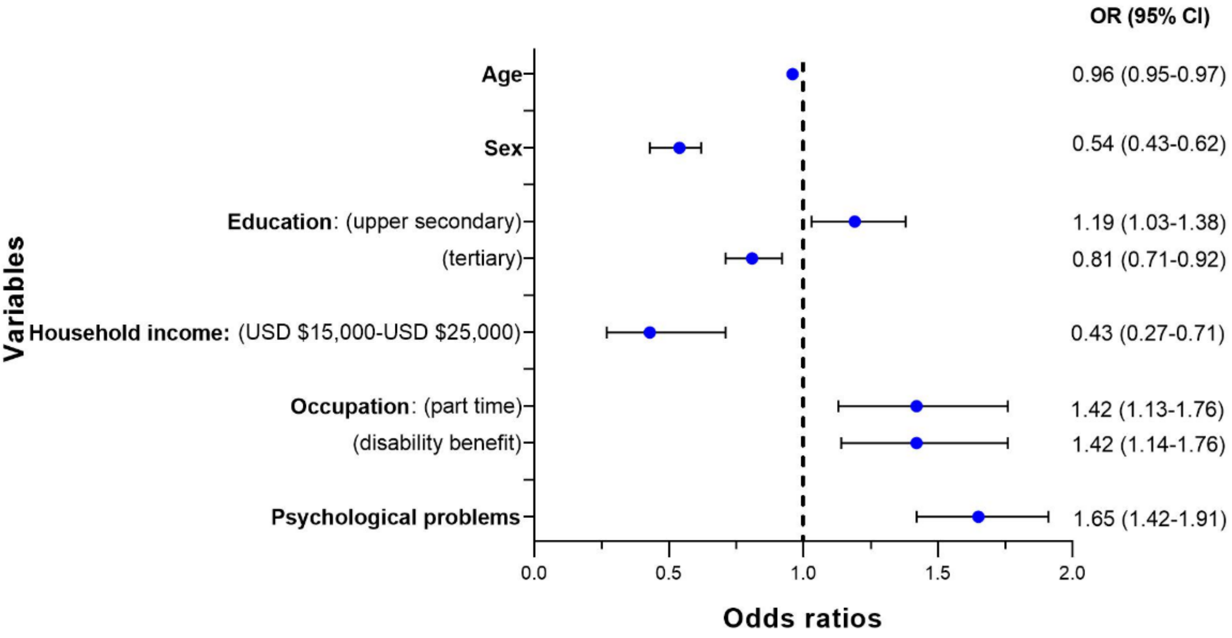
Having an educational level of upper secondary school (OR 2.16, 95% CI 1.79-2.61), being a man (OR 1.31, 95% CI 1.13-1.53), suffering from psychological problems (OR 1.70, 95% CI 1.43-2.01), and suffering from any of the diseases contained in the group of *others* (OR 1.43, 95% CI 1.19-1.71) were associated with an increase in the use of video services for accessing health information. Increasing age (OR 0.97, 95% CI 0.96-0.98), having an education of less than 4 years of college (OR 0.81, 95% CI 0.68-0.95), and having a household income of US $15,000-$25,000 (OR 0.45, 95% CI 0.31-0.67) were associated with a decrease in the use of video services for accessing health information.

Study of the Use of Social Media

The sample size used by the statistical software after removing the observations missing any of the variables used by the model for social media was 14,514 individuals.

Figure 5 summarizes the significant disease groups and demographic characteristics related to the users of social media contained in the subsample. In addition, it contains the ORs from the regression model predicting the use of social media and the influence of each independent variable. The full result of the analysis is available in Multimedia Appendix 6.

Figure 5. Forest plot for the logistic regression for the use of social media.



Having a part-time job (OR 1.42, 95% CI 1.13-1.76), receiving a disability benefit (OR 1.42, 95% CI 1.14-1.76), having an education level of upper secondary school (OR 1.19, 95% CI 1.03-1.38), and suffering from psychological problems (OR 1.65, 95% CI 1.42-1.91) were associated with an increase in the use of social media for accessing health information.

Higher age (OR 0.96, 95% CI, 0.95-0.97), being a man (OR 0.54, 95% CI, 0.43-0.62), being in the group of those with a household income of US $15,000-$25,000 (OR 0.43, 95% CI 0.27-0.71), and having an education level of less than 4 years of college (OR 0.81, 95% CI 0.71-0.92) were associated with a decrease in the use of social media for accessing health information.

Table 1 shows a summary with the associations that were significant regressing disease groups with eHealth resources, some variables have been omitted for clarity.

Table 1. Summary of the association of electronic resources and disease groups.

Disease group	Mobile apps	Search engines	Video services	Social media
Cardiovascular diseases	OR[a] 1.12, 95% CI 1.01-1.24	—[b]	—	—
Cancer	—	OR 1.26, 95% CI 1.11-1.44	—	—
Psychological problems	OR 1.39, 95% CI 1.23-1.56	OR 1.39, 95% CI 1.25-1.55	OR 1.70, 95% CI 1.43-2.01	OR 1.65, 95% CI 1.42-1.91
Respiratory problems	—	Significant when interacting with a household income of US $55,100 – $75,000[c]	—	Significant when interacting with occupation=retired[c]
Other diseases	Significant when interacting with part-time work[c]	OR 1.27, 95% CI 1.13-1.42	OR 1.43, 95% CI 1.19-1.71	—

[a] OR: odds ratio.

[b] The association between the disease and the electronic resource is not statistically significant.

[c] The interaction is significant. OR for interactions are available in multimedia appendices.

Discussion

Summary of Evidence

This study sheds light on the use of various eHealth resources by patients that suffer from diverse conditions. To our knowledge, this is the first study covering the relationship between ranges of different health conditions and varying preferences for different eHealth resources. As depicted in Table 1, our results show that, in general, different diseases are associated with the use of different eHealth resources.

In general, lower socioeconomic class (SES) positively predicted the use of eHealth resources. In addition, for cardiovascular and respiratory diseases, the interaction with lower SES caused an additional increase in the use of social media and search engines, respectively.

Previous studies have shown that people suffering from chronic illness are more likely to search for health information on the internet [9,14]. However, these studies focused on the use of internet for health in general, whereas this study analyzes in depth the use of each type of resource by each group of patients.

Our study shows that long-term and chronic diseases significantly influence the use of eHealth resources. This is consistent with prior research reporting that internet users living with a chronic disease are more likely to gather information using the internet [9,14,19,46]. In addition, this study adds to this knowledge by showing that the influence on each specific eHealth resource varies depending on the medical condition.

Prior literature has shown that most online health information searching starts at a search engine [9]. Furthermore, in our dataset, search engines were the most frequently used eHealth service among those suffering from some disease (7468/15,585; 47.92%), followed by apps (1982/15,585; 12.72%), social media (1145/15,585; 7.35%), and video services (767/15,585; 4.92%). However, in this study, we found that when the use of different eHealth resources is studied independently over the whole dataset, patients with different diseases appear to have variations in preferences regarding different eHealth resources. We believe that these differences in preferences can be partially explained by the availability and popularity of various eHealth resources for the different patient groups—which again might depend in part on characteristics of the different patient groups. For instance, there are many popular apps available for the management of psychological problems [32,39,47,48], such as sleeping problems, anxiety, and depression. In contrast to other chronic disorders, the apps available for psychological problems might even help cure a problem (ie, sleeping problem)—this is obviously not the case, for instance, with cancer or cardiovascular diseases.

Patients with psychological problems were likely to use all the eHealth resources under study (apps, search engines, videos, and social media). Previous studies have reported psychological variables as predictors of health-related internet use [49]. Internet videos have been reported to benefit patients with mental illness [47]. Moreover, internet- and apps-based interventions have showed that beyond helping those with psychological problems, they can act as an attractor for those in need for help [39]. In addition, social media has been found to be beneficial by decreasing the distress of people with schizophrenia [50]. Our results suggest that all the eHealth resources covered may be used for providing health information to people with psychological disorders. Determining which psychological disorders respond better to each of the resources remains a future task.

Prior research has shown that patients with cardiovascular diseases constitute one group that benefits, in part through improved disease management, from telemedicine and eHealth interventions [42]. We found that patients who had cardiovascular diseases were associated with a preference for mobile apps (OR 1.12) and social media (if they had high SES; OR 3.39; see Multimedia Appendix 6). Our study complements previous findings by showing that mobile apps and social media might be the most appropriate eHealth resources for interventions for providing health information to patients with cardiovascular problems.

Concerning respiratory diseases, SMS messages, WhatsApp, and Facebook have been mentioned as useful tools for receiving health information about chronic obstructive pulmonary disease (COPD) and also for communicating with a doctor [51]. However, in our study, respiratory diseases alone were not significantly associated with a preference for any specific eHealth resource. Only for the subgroups of retired participants and participants with medium-high household income, was it possible to determine that social media or search engines, respectively, were preferred by individuals suffering from respiratory diseases. A possible explanation for the lack of significance of eHealth resources may be the lack of impact of eHealth interventions on patients with this type of diseases. In fact, the Cochrane reviews in the studies by McCabe et al and Marcano et al did not find any statistical significance in the use of mobile technology for the management of people with COPD and asthma, respectively [52,53]. Future works should focus on specific respiratory diseases to determine if these findings are applicable to all of them or whether there are differences across patients with different respiratory conditions.

Cancer was a significant predictor of the use of search engines, that is, general searches for health information on the internet. Cancer was not associated with the use of other eHealth resources. Previous studies have already shown that the use of eHealth among cancer patients is extensive [33]. Our results complement these studies by helping to understand which particular eHealth resource should be used for cancer. Currently, there is a high availability of eHealth resources for cancer. For example, there are many mobile apps concerning cancer [54]. However, the analyses of Bender et al [54] and Giunti et al [55] show that information apps about cancer are much more common than disease management ones. This is aligned with our findings. We believe that this points out that the demand for eHealth resources that provide information is higher than the demand for eHealth resources for disease management among cancer patients. Therefore, eHealth interventions should focus on providing information by, for example, pointing patients to high-quality websites about cancer.

Patients that suffered from conditions included in the group of *other diseases* (arthrosis, rheumatoid arthritis, diabetes, kidney disease, migraine, or chronic pain) were more likely to use internet videos, search engines, and mobile apps (if they were part-time workers) as eHealth resources but less likely to use social media. Some of the diseases contained in our generic group (*other diseases*) have been reported to benefit from the use of smartphone apps [30-32]. Our findings suggest that for those diseases, apps (significant only for part-time workers), in addition to videos and search engines, are appropriate resources to provide eHealth.

For patients that suffer from conditions contained in the group of *other diseases*, in some cases, our findings are not consistent with the previous literature. Prior studies have reported differences in the benefits of technology for the different diseases contained in this group. For example, mobile apps have been reported as inadequate for patients with chronic diseases [56]. For patients with chronic pain, Merolli et al found that social media was beneficial [57]; also, Hou et al found that mobile apps had a small [31] or no improvement at all on the self-management of diabetes [30]. A possible cause for the contradictions of our results with some of the previous studies is that the diseases contained in this group are very heterogeneous. Therefore, our conclusions for this group should be taken with caution. Future research should examine if differences exist across these diseases and the preference for different eHealth resources of those affected by them.

Limitations

There are several limitations in this study that should be considered. The logistic regression model for search engines is not a robust model because the residual in the analysis of deviance is significant. This makes sense because, as stated in the Introduction section, search engines are the first input for searching information [58]. Thus, there may be very disparate factors that influence the use of search engines that are unavailable to us.

Another limitation is that, to our knowledge, it is unknown if users in the higher age groups differentiate well in their responses to the questionnaire between accessing health information from a browser on the mobile phone, an app, videos, or social media. Those with limited eHealth literacy may confuse one with another, which may lead to a high variability in the results of this part of the study. For example, search engines could be used by participants to find other resources such as social media or videos; therefore, the large use of search engines could be misleading, and some of the participants considered as search engine users may in fact have used other types of eHealth resources. In our study, we have only analyzed the use of different eHealth services (or channels), and the actual content of these services is not known.

Psychological problems is a crude categorization, and the group is likely to be quite heterogenous—which again may influence the outcome in terms of eHealth services used. Unfortunately, we do not have more detailed information about the types of psychological problems of the participants. However, it is reasonable to assume that most participants who had such problems had less serious psychological problems as these are the most prevalent in the general population. It is also likely that there is a participation bias in that those who suffered from the most serious psychological problems (ie, psychosis and severe depression) did not participate in the study as participation required a relatively large effort (completing a long questionnaire and attending a study center for more testing).

We lack information about the current availability of different eHealth resources for different patient groups. It is, therefore, difficult to infer from our findings whether increased use of a particular type of eHealth service may be related to a higher availability of a particular type of service.

Our results indicate which eHealth resources are more commonly used by people in each disease group. This does not necessarily mean that this eHealth resource is the best one for patients suffering from that disease. Some other factors that are relevant when planning eHealth services for different patient groups are the characteristics of the diseases and the users' health and the users' eHealth literacy level. Nevertheless, even in those cases where the most frequently used resource for a group of patients may not be the optimal one, our results can be used to reach those patients in the first place and redirect them to the optimal eHealth resource for a specific intervention.

As indicated in the study by Wynn et al (part 1), although the population in Tromsø may be representative of the Norwegian population, caution should be taken when extrapolating the findings to other populations [1]. This is the first time eHealth questions have been included in the Tromsø Study, and these items have, therefore, not been formally tested for validity and reliability—this will be a future task of the eHealth study group.

Conclusions

Our findings show that different diseases influence the use of different eHealth resources. This is an important finding for health organizations to plan eHealth interventions more effectively by taking into account which type of eHealth resource should be used for each patient group. It is not clear why people with specific illnesses currently seem to favor specific eHealth resources, and it may be related to the current availability of high-quality information on different resources. However, certain eHealth resources may be better suited to specific patient groups. For instance, social media is experienced as the most useful eHealth resource for people with psychological problems. Further studies are needed to examine the underlying reasons why different patient groups prefer one type of eHealth resource over another.

Acknowledgments

The authors would like to thank the organizers, management, and the technical staff of Tromsø 7 for their valuable work in preparing and collecting the data. Above all, the authors would like to thank the residents of Tromsø; their willingness to participate is fundamental to this research. This research was made possible by a grant from the Research Council of Norway to the Norwegian Centre for eHealth Research, University Hospital of North Norway, Grant No 248150/O70, and by UiT The Arctic University of Norway. The publication charges for this study have been funded by a grant from the Open Access Publication Fund of UiT The Arctic University of Norway.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic data and diseases related to the participants that used any electronic health resource.

[DOCX File , 65 KB - jmir_v22i3e13116_app1.docx]

Multimedia Appendix 2

Logistic regression for the use of any electronic health resource.

[DOCX File , 88 KB - jmir_v22i3e13116_app2.docx]

Multimedia Appendix 3

Logistic regression for mobile apps.

[DOCX File , 88 KB - jmir_v22i3e13116_app3.docx]

Multimedia Appendix 4

Logistic regression for search engines.

[DOCX File , 92 KB - jmir_v22i3e13116_app4.docx]

Multimedia Appendix 5

Logistic regression for video services.

[DOCX File , 85 KB - jmir_v22i3e13116_app5.docx]

Multimedia Appendix 6

Logistic regression for social media.

[DOCX File , 93 KB - jmir_v22i3e13116_app6.docx]

References

1. Wynn R, Oyeyemi S, Budrionis A, Yigzaw K, Bellika J. Electronic Health Use in a Representative Sample of 18,497 Respondents in Norway (The Seventh Tromsø Study - Part 1): Population-Based Questionnaire Study. JMIR Med Inform 2020 Mar 05;22(3):e13106 [FREE Full text] [doi: 10.2196/13106]
2. Budrionis A, Wynn R, Marco-Ruiz L, Yigzaw K, Bergvik S, Oyeyemi S. Impact of the Use of Electronic Health Tools on the Psychological and Emotional Well-Being of Electronic Health Service Users (The Seventh Tromsø Study - Part 3): Population-Based Questionnaire Study. J Med Internet Res 2020 Mar 05;22(3):e13118 [FREE Full text] [doi: 10.2196/13118]
3. Yigzaw KY, Wynn R, Marco RL, Budrionis A, Oyeyemi S, Johansen FA. The Association Between Health Information Seeking on the Internet and Doctor Visits (The Seventh Tromsø Study - Part 4): Population-Based Questionnaire Study. J Med Internet Res 2020 Mar 05;22(3):e13120 [FREE Full text] [doi: 10.2196/13120]
4. Rigby M, Koch S, Keeling D, Hill P, Alonso A, Maeckelberghe E. European Commission. 2013 Jan 31. Developing a New Understanding of Enabling Health and Wellbeing in Europe URL: https://ec.europa.eu/eip/ageing/library/developing-new-understanding-enabling-health-and-wellbeing-europeharmonising-health-and_en [accessed 2019-11-13]
5. Institute of Medicine, Roundtable on Evidence-Based Medicine. In: Olsen LA, Aisner D, McGinnis JM, editors. The Learning Healthcare System: Workshop Summary. Washington (DC): National Academies Press; 2007.
6. Djulbegovic B, Guyatt GH, Ashcroft RE. Epistemologic inquiries in evidence-based medicine. Cancer Control 2009 Apr;16(2):158-168. [doi: 10.1177/107327480901600208] [Medline: 19337202]
7. Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? Br Med J 2014 Jun 13;348:g3725 [FREE Full text] [doi: 10.1136/bmj.g3725] [Medline: 24927763]
8. Pew Research Center. 2015 Jun 26. Americans' Internet Access: 2000-2015 URL: http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/ [accessed 2018-05-07]
9. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health Online 2013 URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2019-11-15]
10. Pew Research Center. 2013 Jan 1. Majority of Adults Look Online for Health Information URL: http://www.pewresearch.org/fact-tank/2013/02/01/majority-of-adults-look-online-for-health-information/ [accessed 2018-05-07]
11. SSB: Forside. ICT Usage in Households URL: https://www.ssb.no/statistikkbanken/selecttable/hovedtabellHjem.asp?KortNavnWeb=ikthus&CMSSubjectArea=teknologi-og-innovasjon&PLanguage=1&checked=true [accessed 2018-05-08]
12. Fox S, Duggan M. Pew Research Center. 2012 Nov 8. Mobile Health 2012 URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2018-05-08]
13. Smith A, Anderson M. Pew Research Centre. 2018 Mar 1. Social Media Use in 2018 URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2018-05-08]
14. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007 Apr 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
15. Kummervold PE, Wynn R. Health information accessed on the internet: the development in 5 European countries. Int J Telemed Appl 2012;2012:297416 [FREE Full text] [doi: 10.1155/2012/297416] [Medline: 23304133]
16. Gabarron E, Lau AY, Wynn R. Is there a weekly pattern for health searches on Wikipedia and is the pattern unique to health topics? J Med Internet Res 2015 Dec 22;17(12):e286 [FREE Full text] [doi: 10.2196/jmir.5038] [Medline: 26693859]
17. Wangberg S, Andreassen H, Kummervold P, Wynn R, Sørensen T. Use of the internet for health purposes: trends in Norway 2000-2010. Scand J Caring Sci 2009 Dec;23(4):691-696. [doi: 10.1111/j.1471-6712.2008.00662.x] [Medline: 19473317]
18. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
19. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013 Jan 30;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
20. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. J Am Med Assoc 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
21. Wagner TH, Baker LC, Bundorf MK, Singer S. Use of the internet for health information by the chronically ill. Prev Chronic Dis 2004 Oct;1(4):A13 [FREE Full text] [Medline: 15670445]
22. Sørensen T, Andreassen H, Wangberg S. Norwegian Centre for E-health Research. 2014. E-helse i Norge 2013 Prosjektrapport URL: https://www.researchgate.net/publication/303186485_E-helse_i_Norge_2013_Prosjektrapport?channel=doi&linkId=5738847c08ae9f741b2bcc37&showFulltext=true [accessed 2019-11-15]
23. 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 7;4(2):e33 [FREE Full text] [doi: 10.2196/mhealth.5114] [Medline: 27056639]
24. van Kerkhof LW, van der Laar CW, de Jong C, Weda M, Hegger I. Characterization of apps and other e-tools for medication use: insights into possible benefits and risks. JMIR Mhealth Uhealth 2016 Apr 6;4(2):e34 [FREE Full text] [doi: 10.2196/mhealth.4149] [Medline: 27052946]

25. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki CE, Prokosch H, Wynn R. Informed citizen and empowered citizen in health: results from an European survey. BMC Fam Pract 2011 Apr 16;12:20 [FREE Full text] [doi: 10.1186/1471-2296-12-20] [Medline: 21496309]
26. 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]
27. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? Br Med J 2014 Oct 14;349:g6178. [doi: 10.1136/bmj.g6178] [Medline: 25315514]
28. Gabarron E, Serrano JA, Wynn R, Lau AY. Tweet content related to sexually transmitted diseases: no joking matter. J Med Internet Res 2014 Oct 6;16(10):e228 [FREE Full text] [doi: 10.2196/jmir.3259] [Medline: 25289463]
29. 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 2;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
30. Sheehy S, Cohen G, Owen KR. Self-management of diabetes in children and young adults using technology and smartphone applications. Curr Diabetes Rev 2014;10(5):298-301. [doi: 10.2174/1573399810666141010113050] [Medline: 25311195]
31. 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 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
32. Escoffery C, McGee R, Bidwell J, Sims C, Thropp EK, Frazier C, et al. A review of mobile apps for epilepsy self-management. Epilepsy Behav 2018 Apr;81:62-69. [doi: 10.1016/j.yebeh.2017.12.010] [Medline: 29494935]
33. Slev VN, Mistiaen P, Pasman HR, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
34. 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]
35. Col N, Correa-de-Araujo R. Chapter 27 - Consumers and Clinical Decision Support. In: Greenes RA, editor. Clinical Decision Support: The Road To Broad Adoption. Phoenix, AZ, USA: Academic Press; 2014:741-769.
36. NHS Direct wales. Check Your Symptoms URL: https://www.nhsdirect.wales.nhs.uk/selfassessments/# [accessed 2019-11-13]
37. Marco-Ruiz L, Bønes E, de la Asunción E, Gabarron E, Aviles-Solis JC, Lee E, et al. Combining multivariate statistics and the think-aloud protocol to assess Human-Computer Interaction barriers in symptom checkers. J Biomed Inform 2017 Oct;74:104-122 [FREE Full text] [doi: 10.1016/j.jbi.2017.09.002] [Medline: 28893671]
38. Quaglini S, Shahar Y, Peleg M, Miksch S, Napolitano C, Rigla M, et al. Supporting shared decision making within the MobiGuide project. AMIA Annu Symp Proc 2013;2013:1175-1184 [FREE Full text] [Medline: 24551401]
39. Harrer M, Adam SH, Fleischmann RJ, Baumeister H, Auerbach R, Bruffaerts R, et al. Effectiveness of an internet- and app-based intervention for college students with elevated stress: randomized controlled trial. J Med Internet Res 2018 Apr 23;20(4):e136 [FREE Full text] [doi: 10.2196/jmir.9293] [Medline: 29685870]
40. Gabarron E, Årsand E, Wynn R. Social media use in interventions for diabetes: rapid evidence-based review. J Med Internet Res 2018 Aug 10;20(8):e10303 [FREE Full text] [doi: 10.2196/10303] [Medline: 30097421]
41. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njølstad I. Cohort profile: the Tromso Study. Int J Epidemiol 2012 Aug;41(4):961-967 [FREE Full text] [doi: 10.1093/ije/dyr049] [Medline: 21422063]
42. Jacobsen BK. University of Tromsø – The Arctic University of Norway. The Tromsø Study URL: https://academic.oup.com/ije/article/41/4/961/683871 [accessed 2019-11-13]
43. Tromsø Study. URL: http://tromsoundersokelsen.uit.no/tromso/ [accessed 2019-11-06]
44. Field A, Miles J, Field Z. Discovering Statistics Using R. London: Sage Publications Ltd; 2012.
45. Diggle P. Analysis Of Longitudinal Data. Second Edition. Cornwall: Oxford University Press, Usa; 2013.
46. Fox S, Duggan M. Pew Research Center. 2013. The Diagnosis Difference URL: http://www.pewinternet.org/2013/11/26/the-diagnosis-difference/ [accessed 2019-11-13]
47. Williams A, Fossey E, Farhall J, Foley F, Thomas N. Recovery after psychosis: qualitative study of service user experiences of lived experience videos on a recovery-oriented website. JMIR Ment Health 2018 May 8;5(2):e37 [FREE Full text] [doi: 10.2196/mental.9934] [Medline: 29739737]
48. Smith GC, Egbert N, Dellman-Jenkins M, Nanna K, Palmieri PA. Reducing depression in stroke survivors and their informal caregivers: a randomized clinical trial of a Web-based intervention. Rehabil Psychol 2012 Aug;57(3):196-206 [FREE Full text] [doi: 10.1037/a0029587] [Medline: 22946607]
49. van Uden-Kraan C, Drossaert C, Taal E, Seydel E, van de Laar M. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
50. Välimäki M, Athanasopoulou C, Lahti M, Adams CE. Effectiveness of social media interventions for people with schizophrenia: a systematic review and meta-analysis. J Med Internet Res 2016 Apr 22;18(4):e92 [FREE Full text] [doi: 10.2196/jmir.5385] [Medline: 27105939]

51. Ojeda IC, Calderon JC, Jove OL, Guerreros A, Plaza KJ, Cano JA, et al. What kind of information and communication technologies do patients with COPD prefer to use? A cross-sectional study in Latin America. Chron Respir Dis 2018 Aug;15(3):286-295 [FREE Full text] [doi: 10.1177/1479972317741895] [Medline: 29152999]
52. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 May 23;5:CD011425 [FREE Full text] [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
53. Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013 Nov 27(11):CD010013 [FREE Full text] [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
54. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
55. Giunti G, Giunta DH, Guisado-Fernandez E, Bender JL, Fernandez-Luque L. A biopsy of Breast Cancer mobile applications: state of the practice review. Int J Med Inform 2018 Feb;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
56. Lee Y, Cui Y, Tu M, Chen Y, Chang P. Mobile health to maintain continuity of patient-centered care for chronic kidney disease: content analysis of apps. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e10173 [FREE Full text] [doi: 10.2196/10173] [Medline: 29678805]
57. Merolli M, Gray K, Martin-Sanchez F, Lopez-Campos G. Patient-reported outcomes and therapeutic affordances of social media: findings from a global online survey of people with chronic pain. J Med Internet Res 2015 Jan 22;17(1):e20 [FREE Full text] [doi: 10.2196/jmir.3915] [Medline: 25616273]
58. Fox S. Pew Research Center. 2006 Oct 29. Part 2. A Typical Search for Health Information URL: http://www.pewinternet.org/2006/10/29/part-2-a-typical-search-for-health-information/ [accessed 2019-11-13]

Abbreviations

eHealth: electronic health
kr: Norwegian Kroner
OR: odds ratio
Tromsø 7: seventh survey of the Tromsø Study

Edited by G Eysenbach; submitted 12.12.18; peer-reviewed by YL Lee, A Martinez-Millana, J Hewitt, C Gibson, Y Wang, FC Wang; comments to author 14.04.19; revised version received 09.08.19; accepted 26.09.19; published 05.03.20.

Please cite as:
Marco-Ruiz L, Wynn R, Oyeyemi SO, Budrionis A, Yigzaw KY, Bellika JG
Impact of Illness on Electronic Health Use (The Seventh Tromsø Study - Part 2): Population-Based Questionnaire Study
J Med Internet Res 2020;22(3):e13116
URL: https://www.jmir.org/2020/3/e13116
doi:10.2196/13116
PMID:32134390

Original Paper

Hackathons as Stepping Stones in Health Care Innovation: Case Study With Systematic Recommendations

Akira-Sebastian Poncette[1,2,3], MD; Pablo-David Rojas[3], MSc, PhD; Joscha Hofferbert[3], MSc; Alvaro Valera Sosa[3,4], MSc; Felix Balzer[1,2], MD, MSc, PhD; Katarina Braune[3,5], MD

[1]Einstein Center Digital Future, Berlin, Germany

[2]Department of Anesthesiology and Intensive Care Medicine, Charité – Universitätsmedizin Berlin, (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health), Berlin, Germany

[3]Hacking Health, Berlin, Germany

[4]Department of Architecture in Health, Berlin Institute of Technology, Berlin, Germany

[5]Department of Pediatric Endocrinology and Diabetes, Charité – Universitätsmedizin Berlin, (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health), Berlin, Germany

Corresponding Author:
Akira-Sebastian Poncette, MD
Department of Anesthesiology and Intensive Care Medicine
Charité – Universitätsmedizin Berlin
(corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health)
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30450 ext 50
Email: Akira-sebastian.poncette@charite.de

Abstract

Background: Until recently, developing health technologies was time-consuming and expensive, and often involved patients, doctors, and other health care professionals only as passive recipients of the end product. So far, users have been minimally involved in the ideation and creation stages of digital health technologies. In order to best address users' unmet needs, a transdisciplinary and user-led approach, involving cocreation and direct user feedback, is required. In this context, hackathon events have become increasingly popular in generating enthusiasm for user-centered innovation.

Objective: This case study describes preparatory steps and the performance of a health hackathon directly involving patients and health care professionals at all stages. Feasibility and outcomes were assessed, leading to the development of systematic recommendations for future hackathons as a vehicle for bottom-up innovation in health care.

Methods: A 2-day hackathon was conducted in February 2017 in Berlin, Germany. Data were collected through a field study. Collected field notes were subsequently discussed in 15 informal meetings among the research team. Experiences of conducting two further hackathons in December 2017 and November 2018 were included.

Results: In total, 30 participants took part, with 63% (19/30) of participants between 25 and 34 years of age, 30% (9/30) between 35 and 44 years of age, and 7% (2/30) younger than 25 years of age. A total of 43% (13/30) of the participants were female. The participation rate of medical experts, including patients and health care professionals, was 30% (9/30). Five multidisciplinary teams were formed and each tackled a specific health care problem. All presented projects were apps: a chatbot for skin cancer recognition, an augmented reality exposure-based therapy (eg, for arachnophobia), an app for medical neighborhood connectivity, a doctor appointment platform, and a self-care app for people suffering from depression. Patients and health care professionals initiated all of the projects. Conducting the hackathon resulted in significant growth of the digital health community of Berlin and was followed up by larger hackathons. Systematic recommendations for conducting cost-efficient hackathons ($n \leq 30$) were developed, including aspects of community building, stakeholder engagement, mentoring, themes, announcements, follow-up, and timing for each step.

Conclusions: This study shows that hackathons are effective in bringing innovation to health care and are more cost- and time-efficient and potentially more sustainable than traditional medical device and digital product development. Our systematic recommendations can be useful to other individuals and organizations that want to establish user-led innovation in academic hospitals by conducting transdisciplinary hackathons.

(J Med Internet Res 2020;22(3):e17004) doi:10.2196/17004

KEYWORDS

digital health; transdisciplinary research; hackathon; technological innovation; patient-centered care; social media

Introduction

Research in health care is expensive, time-consuming, and does not always ensure the development and implementation of sustainable and appropriate technologies that best address the needs and requirements of patients, doctors, and health care professionals [1]. A transdisciplinary approach, together with direct end-user feedback, may benefit the cost-efficient development of innovative health technologies [2]. In this context, hackathons have become an increasingly popular venue for health care institutions to generate enthusiasm for innovation.

The term *hackathon* derives from the words *hack* and *marathon*. In this context, *hacking* refers to intensive collaborative computer programming. Since the late 1990s, the concept of gathering experts into teams to foster collaboration and solve pressing problems has become increasingly popular. Initially, hackathons were highly targeted at those working in tech [3] but, more recently, hackathons have also found a niche in the medical field and academic literature [4]. Since 2010, hundreds of health hackathons have been documented worldwide, with most of them being held in the United States [5].

The question remains whether hackathons are an effective method to accelerate the creation of novel medical technology. The Consortium for Affordable Medical Technologies (CAMTech), based at the Massachusetts General Hospital's Global Health department, recently published the outcomes of 12 hackathons from 2012 to 2015 in India, Uganda, and the United States [6]. The projects initiated through these events were often followed up afterward and have reached pilot-testing stages, started clinical trials, or even resulted in the formation of new companies. The health hackathon model, including preceding priming activities and targeted postevent support, were rated as a reliable source of solutions to challenges in health care.

However, these results mainly speak for hackathons organized either in the United States or by people from the United States. Research on hackathons in Europe is rare or almost nonexistent. A possible reason may be that the organization of a hackathon requires considerable effort and costs, especially for organizations without any previous experience in this area, such as academic hospitals [7]. Furthermore, hackathons in Europe, especially in Germany, are relatively recent and unknown to most health care providers and operators.

This case study analyzed the preparation and performance of a transdisciplinary health hackathon conducted in Berlin, Germany. Feasibility and outcomes were assessed, leading to the development of systematic recommendations for future health hackathons.

Methods

Study Setting

The 2-day hackathon was conducted over a weekend in February 2017 in Berlin, Germany. The organizing team consisted of 10 people: a core organizing team of three people (ASP, AVS, and JH) and seven volunteers. All members of the organizing team were members of the nonprofit organization Hacking Health.

Hackathon Resources

Our first most valuable resource was a nonprofit organization called Hacking Health, founded and operated in Montreal, Canada, since 2012 [8]. Hacking Health has organized health hackathons worldwide, mostly across Canada, the United States, the Netherlands, France, and eventually in Germany in 2017 with the hackathon presented here.

Prior to this event, the German chapter of Hacking Health was founded by three volunteers in Berlin. Supported by Hacking Health Canada, they were provided with all relevant resources for building a successful digital health community free of charge. The newly founded Hacking Health Berlin chapter then organized various local events, from workshops about biomedical technology to talks about digital health. Since then, the chapter and its network have grown gradually in Berlin [9].

A further helpful resource was the Health Hackathon Handbook by MIT Hacking Medicine, which is available online for free [10] and was used as a guide for organizing the hackathon studied here. Founded at the Massachusetts Institute of Technology (MIT) in Boston, USA, MIT Hacking Medicine aims at accelerating medical innovation by carrying out health hackathons, workshops, or networking gatherings all over the world [11].

Hackathon Preparation

The theme of the hackathon was chosen to be specific, on the one hand, to target physicians and other health care professionals and, on the other hand, to be as broad as possible to include participants from any professional background. By choosing the theme *Patient Care Goes Digital*, we intentionally focused on digital health and patient-centered care, thereby excluding the fitness and lifestyle sector.

In order to reach a wide multidisciplinary community that included medical professionals and patients, we released targeted announcements 10 weeks prior to the event via social networks: Facebook, Meetup, and Twitter. Additionally, leading senior physicians and several resident physicians were contacted personally. In promoting the event, the general concept of hackathons was explained as an "open transdisciplinary workshop," since many potential participants—especially from the medical sector—were still unfamiliar with the term *hackathon*. Additionally, 3 weeks prior to the hackathon, a prehackathon event was organized as an introduction, which was not mandatory for participation at the hackathon.

To participate, applicants were asked to fill out an online questionnaire and provide a brief description of their motivation to participate, their background, and their skill set using a Web-based, team-building platform called *Sparkboard* [12]. Team building was started online 2 weeks prior to the hackathon and was moderated by a member from the organizing team. Participants were encouraged to upload project ideas or *pain points* they would like to work on and develop solutions for onto Sparkboard in advance. A solution approach was not mandatory.

Hackathon Event

The hackathon event was started on a Saturday morning and opened with an inspirational keynote speech on the dynamics of digital health and instructions about hackathon-related practices, in order to increase participants' understanding of the complexity of health care-related challenges. The majority of participants, especially physicians and health care professionals, had little previous experience with digital health- or innovation-related events and were used to more rigidly structured organizations and practices. After the opening keynote speech, participants were given the opportunity to pitch their project idea within 1 minute and find members for their team with the right skill set to help in developing the solution. Most of the hackathon was then spent on further developing ideas and prototypes (ie, *hacking*) and on preparing the final pitch and demonstration.

Throughout the event, all teams were supported by mentors from the local academic hospital Charité−Universitätsmedizin Berlin (doctors and psychologists), the Berlin Institute of Technology (engineers, developers, and architects), the software company ThoughtWorks (developers), and the Hasso-Plattner-Institute for Design Thinking (designers), as well as by patients living with chronic conditions and entrepreneurs from Berlin-based health care startups. Exchange of expertise and assistance was encouraged between teams.

On Sunday afternoon, the hackathon concluded with a public pitch session lasting 5 minutes each. An expert jury panel voted to select two winning teams. The multidisciplinary jury consisted of five members, including a physician with an entrepreneurial background, a designer, two developers, and a patient with a chronic condition. The jury criteria included the following:

1. Innovation potential and feasibility of the idea.
2. Execution of the idea at the hackathon.
3. Multidisciplinary composition of the team.
4. Design and user experience of the prototype.
5. Presentation of the project.

Instead of awarding monetary prizes, we collaborated with Berlin's leading design-thinking studios to enable the winning teams to continue their project work supported by a team of experts and design-thinking workshops, of which both teams took part in about 3 months after the hackathon.

Data Collection and Analysis

Data was collected through a field study in the context of the hackathon event in February 2017. Collected field notes were subsequently discussed during informal meetings within the research team between February 2017 and February 2019. Additionally, experiences of conducting two further hackathons in December 2017 and November 2018 were included. All authors were involved in either one, two, or all three hackathons. In total, 15 meetings were held, with the goals to develop systematic recommendations for conducting a health hackathon on the basis of this presented hackathon and to examine the impact of a smaller, cost-efficient hackathon on the digital health community of Berlin as a large city.

The research team consisted of a resident anesthesiologist with expertise in intensive care medicine, geriatrics, and digital health (ASP); a resident pediatric endocrinologist and patient advocate with expertise in digital health and patient-centric care (KB); a health care entrepreneur with a professional background in psychology (JH); an architect with training in evidence-based design research for health care (AVS); a professor for digital health, who is a consultant anesthesiologist and computer scientist (FB); and a microbiologist trained in human-centered design and innovation (PDR). Assumptions were consistently challenged in the transdisciplinary team setup.

Results

Hackathon Demographics and Projects

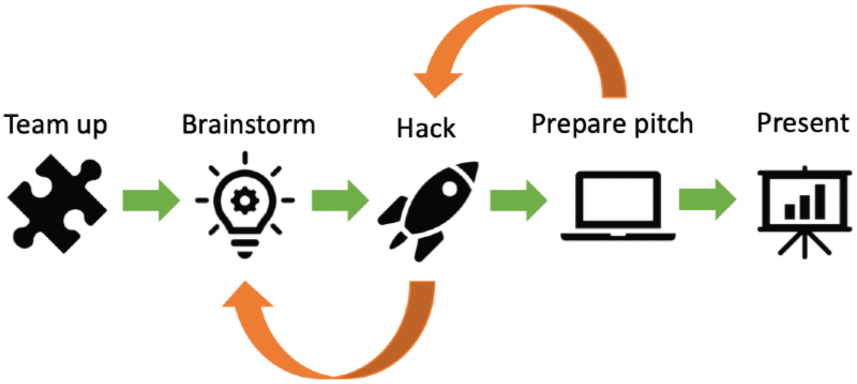
In total, 30 participants took part in the hackathon in February 2017. Their age varied across different groups, with 63% (19/30) of attendees between 25 and 34 years of age, 7% (2/30) younger than 25 years of age, and 30% (9/30) between 35 and 44 years of age. A total of 43% (13/30) of our participants were female. The participation rate of medical experts—physicians and health care professionals, as well as patients as experts for their disease management in everyday life—was 30% (9/30); the majority of the remaining participants were developers and designers.

Prior to the hackathon, 14 projects were uploaded to Sparkboard by participants to sign up and form teams. The jury selected five projects based on the previously established criteria, which applied later on at the hackathon. The teams consisted of 6 members each with a multidisciplinary background. All presented projects were mobile apps: a chatbot for skin cancer recognition; a mobile phone-powered, augmented reality exposure-based therapy (eg, for patients suffering from arachnophobia); an app for medical neighborhood connectivity to improve the experience of patients who have to visit several doctors; a doctor appointment platform; and a self-care app for patients suffering from depression. Patients, physicians, or other health care professionals initiated all of the projects.

Hackathon Flow

After multidisciplinary team formation, project ideas were discussed in brainstorm sessions within the teams, often involving techniques using post-it notes (eg, Venn diagrams) [13]. During this phase, mentors with medical expertise were recruited to explore the pain point and solution approach in relation to several health care actors (eg, physician, nurse, or patients). Afterward, participants carried on with their *hacking* and explored solution approaches with the support of information technology- or design-focused mentors. As a result of this iterative process, the teams developed prototypes (eg, a software mock-up) demonstrating their ideas and solutions. Often in parallel to hacking, one of the team members started working on their pitch. Finally, solutions were presented at a plenum including the jury, ideally showing a working prototype. As sketched in Figure 1, in contrast to our expectations, teams often moved back and forth between brainstorming, hacking, and preparing the pitch.

Figure 1. Hackathon flow, from teaming up to demonstration of the project pitch and working prototype (green arrows). In contrast to our expectations, teams often jumped back and forth between brainstorming, hacking, and preparing the pitch (orange arrows).

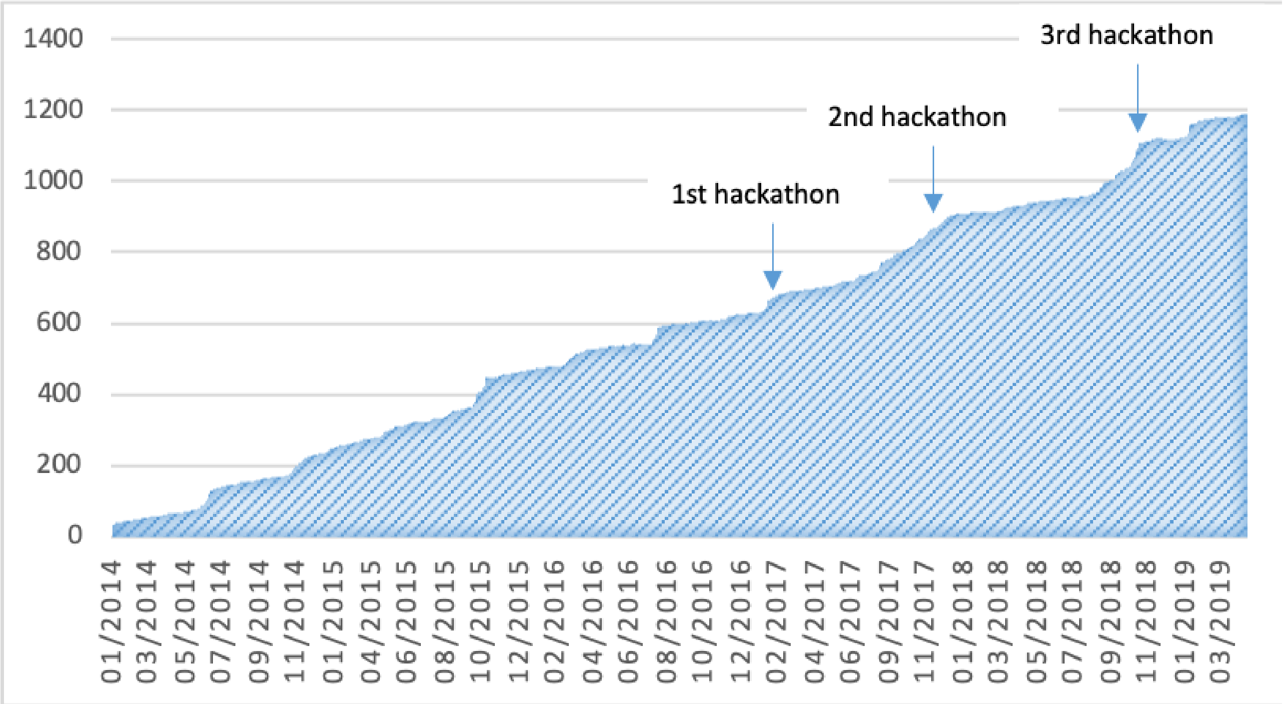


Performance of the First Hackathon

Following this hackathon with 30 participants in February 2017, the Berlin Institute of Health (BIH), a scientific institution for translation and precision medicine, partially funded further hackathons in December 2017 and November 2018. Both larger hackathons included about 75 active participants each, whereas more than 300 participants took part in the opening and closing ceremonies. All three hackathons were unique, self-contained events.

Regarding impact on the digital health community of Berlin, relevant growth of the Meetup group Hacking Health Berlin could be recorded, especially before the hackathons (see Figure 2) [9]. The average growth of the community is 0.6 (SD 1.4) new members per day. A total of 4 weeks prior and 2 weeks after the first, second, and third hackathons, the average growth was 1.2 (SD 2.0), 1.3 (SD 1.2), and 2.0 (SD 2.6) new members per day, respectively.

Figure 2. Total number of members of the Meetup group Hacking Health Berlin from January 2014 to March 2019. The three hackathons are marked with arrows [9].

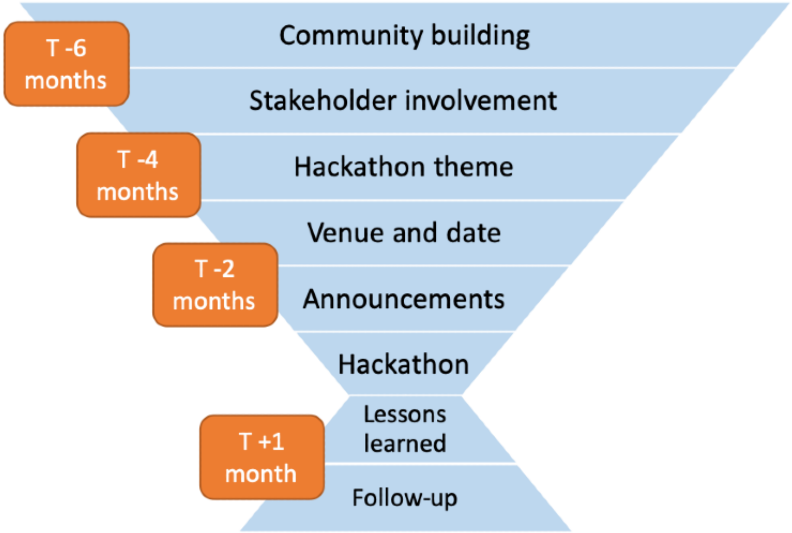


Systematic Recommendation for Conducting a Cost-Efficient Health Hackathon

Overview

The developed systematic recommendation includes eight steps (see Figure 3). In preparation for the hackathon, five steps have to be undertaken (ie, community building, stakeholder involvement, selection of the hackathon theme, venue and date, and announcements); after the hackathon, two steps are recommended (ie, lessons learned and follow-up). The timing for each step was chosen to work for the organization of a hackathon with no more than 30 participants; depending on the available community, the preparation of larger hackathons (ie, >100 participants) should be started at least 2 months earlier. In our experience, smaller hackathons are much more cost-efficient, with up to €100 per capita, compared to larger events with approximately €1000 spent per capita.

Figure 3. Systematic recommendation for conducting a cost-efficient hackathon with 30 or fewer participants. The time (T) to action is visualized in orange boxes. The hackathon occurs at T0.



Community Building and Stakeholder Involvement

The intention or drive to conduct a hackathon derives from the interest among a specific group of people. An existing community of at least 100 people is recommended. In this context, the community is defined as a stable and loyal network of people in an online or offline context, hence, potential participants that can be contacted online or addressed at related networking gatherings. Strategies to grow a digital health community include organizing regular small events or gatherings, advertising these events locally and on social media, and building up a member database (eg, using mailing lists or social media groups). This step should be achieved at least 6 months prior to the hackathon.

A successful hackathon also depends on stakeholder and expert involvement, people who are ideally members of the organizing team. However, senior physicians or developers might be unable to commit to organizing or participating in an entire hackathon [14]. Therefore, the participation of experienced advisors and mentors is highly recommended and their recruitment should be performed well in advance, due to their profession-related constraints. The establishment of a local network may reduce personnel expenses, as mentors, keynote speakers, or jury members often provide their support free of charge and do not have to travel far in order to attend the event. In particular, these volunteers might be more attracted to smaller *bootstrapping* events than larger events with a more corporate character.

A local network is essential for the acquisition of financial resources. In our experience, the time spent cold-contacting potential sponsors should be better invested in the growth of the network. Rewarding options for sponsors may include advertising banners at the event, the opportunity to present as a keynote speaker, challenge cocreation, or branded giveaways to participants; sponsors may even send recruiters to the event to scout talent. In our opinion, it has proven to be helpful to create a brochure for potential sponsors explaining different sponsorship options and *packages* (eg, platinum, gold, or silver sponsorship).

Hackathon Theme

The hackathon theme is an essential selling point to medical professionals and patients. Along these lines, the theme of a hackathon, especially when held for the first time, should be general enough to attract a broad spectrum of potential participants. Following hackathons may be more specific to certain medical or problem areas. In both cases, it is important to highlight how a health-related hackathon may be different from hackathons that are more general or more specific to other subject areas.

When selecting a theme, it is important that the corresponding expertise—ideally provided by a patient or practicing physician—is represented by at least one expert as part of the organizing team. Once a theme is chosen, specific challenges should be defined, matching the theme. We recommend three to five challenges in total. As an alternative to a specific theme, only challenges with different themes may be chosen. In this case, we recommend no more than three challenges.

Venue and Timing

In contrast to larger hackathons, it is much easier to find a fitting venue for smaller events (n≤30), as only a larger room and an *escape room* are needed. When choosing the venue, factors such as light, air quality, and acoustics should be of importance, taking into account that participants are spending most of their time in that venue, mostly in teams sitting at tables. Choosing a big room for many teams has the advantage of fostering team spirit. However, especially at the beginning of the hackathon where conversation is most important, it can get very noisy. Smaller escape rooms for quick team meetings might solve this problem.

Event organization for a hackathon is mostly similar to organizing conferences. However, we observed three major differences. Firstly, we recommend leaving it to the participants when to stop working on their projects in the evening. Hence, a warm snack should be prepared at night for teams who prefer to stay longer. Secondly, from our experience, intense collaborative work burns lots of calories and participants usually consume more food than average conference attendees. Thus, a warm lunch and dinner should be prepared in a sufficient quantity, instead of small snacks or cold dishes only. Lastly, a snack bar should be ready at all times with fruits, nuts, vegetables, coffee, tea, juice, and water. We recommend hiring a professional caterer for events of 50 people and above. Alternatively, more affordable options such as local delivery services may be used for smaller hackathons instead.

Announcements and Call for Application

Depending on the size of the hackathon, announcements for application or registration should be made 2-4 months prior to the event via all available channels, including social media (eg, Twitter, LinkedIn, Meetup, and Facebook), hospitals' internal blackboards, mailing lists, and cold emails to various stakeholders, which may be shared among their staff. Recurrent announcements may be biweekly, more condensed before deadlines, and justifiable through new information (eg, new challenges and partners) or approaching deadlines for applications. For larger hackathons, it has proven successful to extend the application deadline at least once, with the first deadline being 2 weeks prior to the event.

In contrast to conferences, a successful hackathon highly depends on the quality of participants, their preparation, and their composition (eg, equally represented professions). Although it is time-consuming, it is advisable to select the participants individually on the basis of a short questionnaire and to ensure that all disciplines and skill sets are represented sufficiently. The following participants should be selected, in order to include the desired disciplines: practicing physicians or health care professionals; developers, engineers, and designers, with further detailed descriptions of their skills, especially for the developers; patients, with descriptions of their conditions and experience (eg, in advocacy); and other professionals, including architects, other researchers, and business economists.

To break the ice, it is important to inform potential participants about what a hackathon actually is. Moreover, the word *hack* has a negative connotation to several individuals. One way to prevent this would be to use an alternative word such as *datathon* or *collaborathon* instead [15].

Hackathon

Whereas a hackathon often starts and ends with long opening and closing ceremonies that include keynote speeches on recent challenges in health care, panel discussions, or workshops, these activities are optional. Factors that significantly improve the outcome of a hackathon are as follows:

1. The acknowledgement of the expertise of the participants.
2. The presentation of the hackathon challenges.
3. The introduction of mentors.
4. The incitement of team spirit.

At a hackathon, participants are eager to solve entrenching problems, hence, less than 30 minutes for the welcoming remarks, an opening keynote speech, and introduction of the challenges is sufficient. Following the introduction, all team leaders should pitch their project, without using slides, in 1 minute. To keep within the timeline, an audio signal (eg, playing a jingle) interrupting the pitch after the time is up has been proven helpful (see Table 1).

Team building should be initiated by the organizers prior to the event, depending on team constellations and their multidisciplinary backgrounds. By starting prior to the event, optimal team compositions can be achieved, and most of the time can be spent on hacking. Mentors should fill in a profile that can be presented to the teams in order to find the right mentor faster, and mentors should be given acknowledgment. Occasionally, teams might need a space to think and work in silence. This can be achieved by providing teams with “do not disturb” signs. For relaxation, activities like yoga or meditation may be offered.

Either on Saturday evening or Sunday morning, participants may be given the opportunity to get feedback from a professional pitching expert. Each team should get at least 10 minutes of mentoring time, including the 3-minute final pitch.

Table 1. Example schedule of a 2-day weekend hackathon.

Time of day	Schedule items and events	
	Day 1: Saturday	Day 2: Sunday
Morning		
07:00		Coffee and breakfast
08:00		Pitch clinic (10 minutes for each team)
08:30	Registration opening Coffee and breakfast	
09:30	Welcome	
09:40	Keynote speech	Hacking and mentoring
09:50	Introduction of the challenges	
10:00	60-second pitches (no slides)	
10:15	Final team formation[a]	
10:30	Teams will be assigned hacking space	
11:00	Hacking and mentoring	
Noon		
12:15		Drinks and canapés[b]
12:30	Lunch	
12:45		Group photo
Afternoon		
13:00	Hacking and mentoring	
14:00	Yoga session	Submission deadline
14:05		Jury briefing
14:40		Keynote speech[b]
15:00	Hacking and mentoring	Public pitching of demos (3 minutes each with 2 minutes of Q&A)[b]
17:30		Drinks and canapés[b]
Evening and night		
18:00	Dinner	Announcement of winners with award ceremony[b]
19:00	Hacking	After-hours drinks and networking[b]
21:00	Optional hacking through the night	
00:00	Midnight snack	

[a] Team formation may also be started online 2 weeks prior to the event.

[b] These events are open to the public, while all others are only open to active hackathon participants and mentors.

Lessons Learned and Follow-Up

A debriefing meeting among the organizing team a few days after the hackathon should be planned in advance to discuss and document optimization potential for future hackathons.

To improve the performance of the hackathon, a survey should be handed out in person and sent to the participants after the hackathon [16]. Questions may include demographic data such as age or professional background of the participants, feedback on the hackathon, as well as rating scores representing their levels of confidence in starting a health care project before and after the hackathon. To ensure sufficient return, the survey can be combined with incentives, such as access to videos and photos of the hackathon or a discount to future hackathons or other events. Participants should be followed up 3, 6, and 12 months after the event by email (eg, with a survey) to determine whether the teams have continued working on their projects.

Discussion

Principal Findings

This study describes and analyzes a small and cost-efficient health hackathon with 30 participants. In total, five multidisciplinary teams were formed, and each team tackled a specific health care problem. Performing the hackathon resulted in a significant growth of the digital health community in Berlin and the execution of subsequent larger hackathons.

By including the results from the subsequent larger hackathons, a systematic recommendation for conducting cost-efficient hackathons (n≤30) was developed. These recommendations include aspects of community building, hackathon theme, announcements, and timing for each step.

Hackathon Performance

Compared with other health hackathons, the final number of participants of our health hackathon was smaller [5]. We purposefully limited the number of participants, subsequently reducing logistic tasks, costs, and workload for organizing staff.

From our point of view, the greatest potential and value of hackathons lies in providing an opportunity for people to meet and collaborate throughout the event and at mid- to long-term time points after the event. A hackathon puts experts' brains into the right gear and inspires them to think in an unconventional fashion. This does not necessarily have to result in a prototype immediately; however, it may be achievable in the near future. The impact of a hackathon has been described in various ways [12,13]. Silver et al pointed out that the best metric for impact is to measure how many teams continue to work on their solution after the event [12]. Other proposed metrics include the diversity of skill sets and the number of teams that have been able to receive financial support or start a business after the event.

In our study, to measure the impact of a hackathon, we used the metrics of (1) growth of the digital health community and (2) subsequent events (eg, hackathons) that followed the initial hackathon. The former can be easily measured by creating mailing lists or social media groups. These metrics could reflect the impact of the hackathon better than the immediate results in the form of prototypes developed by the participants.

Hackathon Team Building and Flow

Generally, team building at hackathons is a delicate topic. Firstly, an ideal team should consist of several disciplines, including developers, designers, entrepreneurs, health care professionals, and patients [7]. Secondly, the optimal size of a team is crucial. Based on studies of problem solving in groups, a total number of 5 people is recommended [17], whereas some hackathons allow larger team sizes of up to 8 people per team [16,18]. Lastly, a team may be formed by the organizers or by the participants themselves prior to or at the hackathon. We encourage organizers to define multidisciplinary teams of 5-7 participants in advance, respecting individual needs and backgrounds regarding the proposed challenge and team composition.

We divided the hackathon flow into the following steps: team up, brainstorm, hack, prepare pitch, and present. Using a systems approach, which was developed by MIT Hacking Medicine, the steps *brainstorm* and *hack* can be further divided into four phases: identification of the problem, description from the perspective of different health care actors, alteration as a solution approach, and implementation for early user feedback [19]. The advantage of using the systems approach at a hackathon may be that it follows a structured path to product development, forcing the participants to identify relevant pain points, highlighting it from different perspectives, and focusing the final pitch toward the implementation of the new solution into clinical routine. A project template to be filled in by the participants prior to the hackathon may be useful and may shorten the two phases of identification and description, leaving more time for transdisciplinary work on the solution approach.

Recommendations for Future Hackathons

In this work, we developed systematic recommendations for organizing a small hackathon with minimal effort and resources. A major difference in comparison to larger hackathons is the reduced number of logistic challenges, giving the organizers more flexibility. In regard to timing, the actual planning for the hackathon may begin only 3 months prior to the event for smaller hackathons, compared to larger ones where planning has to start 6 months in advance [7,14].

Other available recommendations for hackathons include the extended hackathon model [16]. Wang et al developed this model over a course of eight hackathons by including seminars and workshops about design thinking, hardware prototyping, or business plan development into a weekend hackathon. Notably, the number of participants was also rather small per event, ranging from 18 to 55 participants per hackathon. However, the overall size of this hackathon series could discourage potential academic hospitals from conducting transdisciplinary events themselves, due to the high effort involved.

Limitations

Our study is limited by the choice of the methodology and is possibly biased by the fact that most of the researchers were also involved in conducting the hackathon. The systematic recommendations were developed in a transdisciplinary approach based on three hackathons. Further hackathons applying these recommendations should be used to validate stated findings. To increase the significance, interview protocols or survey results from hackathon participants should be included in future research.

Conclusions

A hackathon may break down the barriers between technical experts—who are able to build innovative technologies—and clinicians or patients—who know best which solution is sustainable—by physically bringing both groups together in one space and closing gaps in language and character.

With this study we were able to show that small hackathons are an effective way to bring innovation to health care and may be more cost- and time-efficient in the long run than larger hackathons. The systematic recommendations are useful for everyone who wants to bring innovation and a fresh breeze into the rigid structures of academic hospitals through transdisciplinary hackathons.

Acknowledgments

The authors would like to thank Yannik Schreckenberger, driving founder of the Hacking Health Berlin chapter, and Annie Lamontagne of Hacking Health Canada for their dedication to supporting the hackathon organization. The authors would also like to thank all volunteers and members of the nonprofit organization Hacking Health and its regional chapter in Berlin. The hackathons were financially supported by Barmer, B Braun, BCG Digital Ventures, Berlin Partner, BIH, bunch.ai, Deloitte, Designit, FTWK, MIAMED, NorthWest, Research2Guidance, Roche, T Systems, Technologie Stiftung Berlin, Pfizer, and 42Berlin. We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Funds of Charité−Universitätsmedizin Berlin.

Authors' Contributions

ASP had the idea to utilize the concept of the hackathon to foster innovation in health care. The study was conceived by ASP and KB. ASP and KB conducted data acquisition and analysis, supported by AVS, FB, JH, and PDR. ASP wrote the manuscript with support from KB. All authors critically reviewed and approved the manuscript.

Conflicts of Interest

None declared.

References

1. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: New estimates of drug development costs. J Health Econ 2003 Mar;22(2):151-185. [doi: 10.1016/S0167-6296(02)00126-1] [Medline: 12606142]
2. Shah SG, Robinson I, AlShawi S. Developing medical device technologies from users' perspectives: A theoretical framework for involving users in the development process. Int J Technol Assess Health Care 2009 Oct;25(4):514-521. [doi: 10.1017/S0266462309990328] [Medline: 19845981]
3. OpenBSD. Hackathons URL: http://www.openbsd.org/hackathons.html [accessed 2019-06-04]
4. Walker A, Ko N. Bringing medicine to the digital age via hackathons and beyond. J Med Syst 2016 Apr;40(4):98. [doi: 10.1007/s10916-016-0461-1] [Medline: 26872780]
5. Bhandari A, Hayward M, Luminary Labs. MIT Hacking Medicine. Health hackathon database URL: http://hackingmedicine.mit.edu/health-hackathon-database/ [accessed 2017-06-08]
6. Olson KR, Walsh M, Garg P, Steel A, Mehta S, Data S, et al. Health hackathons: Theatre or substance? A survey assessment of outcomes from healthcare-focused hackathons in three countries. BMJ Innov 2017 Feb;3(1):37-44 [FREE Full text] [doi: 10.1136/bmjinnov-2016-000147] [Medline: 28250965]
7. Wang JK, Roy SK, Barry M, Chang RT, Bhatt AS. Institutionalizing healthcare hackathons to promote diversity in collaboration in medicine. BMC Med Educ 2018 Nov 20;18(1):269 [FREE Full text] [doi: 10.1186/s12909-018-1385-x] [Medline: 30458759]
8. Chowdhury J. Technology Innovation Management Review. 2012 Jul. Hacking health: Bottom-up innovation for healthcare URL: https://timreview.ca/sites/default/files/article_PDF/Chowdhury_TIMReview_July2012_1.pdf [accessed 2020-02-02]
9. Meetup. Hacking Health Berlin URL: https://www.meetup.com/de-DE/Hacking-Health-Berlin/ [accessed 2019-05-03]
10. Healthcare Hackathon Handbook. Cambridge, MA: MIT Hacking Medicine; 2016 Mar 27. URL: http://hackingmedicine.mit.edu/assets/Health_Hackathon_Handbook.pdf [accessed 2020-02-02]
11. MIT Hacking Medicine. URL: http://hackingmedicine.mit.edu/ [accessed 2020-02-02]
12. Sparkboard. URL: https://hhberlin.sparkboard.com/ [accessed 2017-06-08]
13. Kernbach S, Nabergoj AS. Visual design thinking: Understanding the role of knowledge visualization in the design thinking process. In: Proceedings of the 22nd International Conference Information Visualisation (iV 2018). New York, NY: IEEE; 2018 Presented at: 22nd International Conference Information Visualisation (iV 2018); July 10-13, 2018; Fisciano, Italy p. 362 URL: https://tinyurl.com/ryytn2b [doi: 10.1109/iv.2018.00068]
14. Silver JK, Binder DS, Zubcevik N, Zafonte RD. Healthcare hackathons provide educational and innovation opportunities: A case study and best practice recommendations. J Med Syst 2016 Jul;40(7):177 [FREE Full text] [doi: 10.1007/s10916-016-0532-3] [Medline: 27277278]
15. Aboab J, Celi LA, Charlton P, Feng M, Ghassemi M, Marshall DC, et al. A "datathon" model to support cross-disciplinary collaboration. Sci Transl Med 2016 Apr 06;8(333):333ps8 [FREE Full text] [doi: 10.1126/scitranslmed.aad9072] [Medline: 27053770]
16. Wang JK, Pamnani RD, Capasso R, Chang RT. An extended hackathon model for collaborative education in medical innovation. J Med Syst 2018 Oct 17;42(12):239. [doi: 10.1007/s10916-018-1098-z] [Medline: 30328518]
17. Hoegl M. Smaller teams–better teamwork: How to keep project teams small. Bus Horiz 2005 May;48(3):209-214. [doi: 10.1016/j.bushor.2004.10.013]
18. Angelidis P, Berman L, Casas-Perez MD, Celi LA, Dafoulas GE, Dagan A, et al. The hackathon model to spur innovation around global mHealth. J Med Eng Technol 2016;40(7-8):392-399 [FREE Full text] [doi: 10.1080/03091902.2016.1213903] [Medline: 27538360]

19. Gubin TA, Iyer HP, Liew SN, Sarma A, Revelos A, Ribas J, et al. A systems approach to healthcare innovation using the MIT Hacking Medicine model. Cell Syst 2017 Jul 26;5(1):6-10 [FREE Full text] [doi: 10.1016/j.cels.2017.02.012] [Medline: 28750199]

Abbreviations

BIH: Berlin Institute of Health
CAMTech: Consortium for Affordable Medical Technologies
DFG: German Research Foundation
MIT: Massachusetts Institute of Technology

Edited by E Borycki; submitted 14.11.19; peer-reviewed by K Paik, T Muto, C Fernández; comments to author 01.01.20; revised version received 13.01.20; accepted 17.01.20; published 24.03.20.

Please cite as:
Poncette AS, Rojas PD, Hofferbert J, Valera Sosa A, Balzer F, Braune K
Hackathons as Stepping Stones in Health Care Innovation: Case Study With Systematic Recommendations
J Med Internet Res 2020;22(3):e17004
URL: http://www.jmir.org/2020/3/e17004/
doi:10.2196/17004
PMID:32207691

Review

Risks and Benefits of Web-Based Patient Narratives: Systematic Review

Daniel Drewniak[1], PhD; Andrea Glässel[1,2], PhD; Martina Hodel[1], PhD; Nikola Biller-Andorno[1], MD, PhD

[1]Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

[2]Institute of Health Sciences, Zurich University of Applied Sciences, Winterthur, Switzerland

Corresponding Author:
Daniel Drewniak, PhD
Institute of Biomedical Ethics and History of Medicine
University of Zurich
Winterthurerstrasse 30
Zurich, 8006
Switzerland
Phone: 41 44 634 40 82
Email: daniel.drewniak@ethik.uzh.ch

Abstract

Background: Patient narratives are illustrative, individual accounts of patients' experiences with certain health conditions. Web-based patient narratives have become widely available on the internet and in social media, as part of electronically available patient decision aids or Web-based databases. In recent years, the role of patient narratives as a source of information, insight, and support for both health care users and providers has increasingly been emphasized. Although the potential impact of patient stories has high immediate plausibility, it is of interest to know if this impact can be captured in quantitative studies.

Objective: This review aimed to evaluate whether research-generated Web-based patient narratives have quantifiable risks or benefits for (potential) patients, relatives, or health care professionals.

Methods: We searched the following databases from August 2017 to March 2019: Medical Literature Analysis and Retrieval System Online, PsycInfo, Sociological Abstracts, Web of Science, and EMBASE. Titles and abstracts of the retrieved studies were reviewed and assessed for the inclusion and exclusion criteria. Papers were included if they studied the following: (1) (potential) patients, relatives, or health care professionals; (2) the effects of Web-based patient narratives that were generated scientifically (eg, through qualitative research methods); and (3) were quantitative studies. Furthermore, 2 authors independently performed an assessment of the quality of the included studies using a validated checklist.

Results: Of 4226 documents, 17 studies met the inclusion criteria. The studies investigated 10 different sources of Web-based patient narratives. Sample sizes ranged from 23 to 2458. The mean score of the quality assessment was 82.6 (range 61-100). Effects regarding five different purposes were identified as follows: provide information, engage, model behavior, persuade, and comfort. We found positive effects in every category and negative effects in one category (persuade). Several of the reported effects are rather small or were identified under specific experimental conditions.

Conclusions: Patient narratives seem to be a promising means to support users in improving their understanding of certain health conditions and possibly to provide emotional support and have an impact on behavioral changes. There is limited evidence for beneficial effects on some outcomes. However, narratives are characterized by considerable heterogeneity and the investigated outcomes are hardly comparable with each other, which makes the overall judgment difficult. As there are numerous possible measures and purposes of narratives, quantifying the impact of Web-based patient narratives remains a challenge. Future research is needed to define the optimal standards for quantitative approaches to narrative-based interventions.

(J Med Internet Res 2020;22(3):e15772) doi:10.2196/15772

KEYWORDS

patient experiences; personal narratives as topic; systematic review

Introduction

Background

In their recent report, the *Lancet Global Health Commission* calls for an improved integration of patient experiences in the evaluation of health care systems, including experiences about competent care, health care utilization, or confidence in the health care system [1]. Such experiences can be collected by using tools such as patient satisfaction surveys. Although quantitative data about patient experiences are essential measures for the accountability and improvement of health care systems, they fall short of capturing a more comprehensive picture of how patients experience health care encounters or illnesses [1,2].

Patient narratives are illustrative accounts of individual patients' experiences with a certain illness [3] and are available on social media sites, in patient decision aids, and on databases such as the *Database of Individual Patients' Experiences* (DIPEx). There is neither a clear definition of what constitutes a narrative nor any guidance on the length or content [4], which may lead to conflicting research results about the effects of patient narratives because of insufficient operationalization of the term [3].

Patient narratives are a promising tool that can support people in coping with their illness [5], serve as a resource for preparing health care decisions [6], or help identify questions for physicians [7]. Characteristically, narratives can retrospectively structure actions in ways that convey perceived causality; they are nonlinear and powerful in making sense of complex, emergent phenomena [8]. Furthermore, stories transport images and emotions, which makes them evocative and memorable. Most people recall stories better than statistical information expressed in graphs or numbers [2,8].

Several qualitative studies report that illness narratives enjoy high acceptance among other patients [9]. Furthermore, positive effects of personal health and illness experiences, including improvements in decision making [10,11] or addressing information needs [12], were identified in qualitative studies. Narratives have a high potential to add unknown insights into patient-focused issues, which can only be provided by a person who has the respective lived experience. For example, as a World Health Organization report states, "qualitative methods help to present narratives that broadly reflect the gendered social norms about parent-child relations. They also provide 'lived experiences' from ageing populations about how satisfied they are with the life they have lived" [2].

On the contrary, there are also serious concerns about the use of patient narratives because they are powerful message formats [13] and are suspected to override statistical information [14,15]. The concern is that patients' decision-making regarding treatment options could be based on personal experiences of a few, whereas statistical data remain largely ignored [2,14,15]. Furthermore, patients' experiences presented on the Web may contain unbalanced or misleading messages, which may lead to a manipulation of choices in favor of a particular health care option [16]. A study among mothers of children with genetic disorders, eg, found that several parents put more trust on online communities than on their physicians [17]. Such findings are especially problematic when stories in such communities are biased.

In recent years, internet platforms, patient blogs, and fora have become important means for individuals to seek information relevant to health, including information describing how other individuals live with illnesses. Such websites often provide biomedical information but lack information on wider experiences [18] or the experiences are not systematically collected, analyzed, and presented [19]. Therefore, in this review, we focused on studies that used established scientific methodologies to elicit the stories [8,20].

Objectives

This systematic review aimed to evaluate whether research-generated Web-based patient narratives have quantifiable risks or benefits for patients, relatives, or health care professionals. Patient narratives are understood as immediate personal experience reports.

Methods

Search Strategy

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines [21]. To identify relevant studies, the databases Medical Literature Analysis and Retrieval System Online, PsycInfo, Sociological Abstracts, Web of Science, and EMBASE were searched from August 2017 to March 2019. A search term was developed and was adjusted to the different databases. The search terms were tested and evaluated by the study team. In addition, the search strategy was discussed and evaluated with a member of Cochrane Switzerland and with an employee from the University library who specialized in systematic reviews. The search terms were adjusted based on the discussion and recommendations. The search terms consisted of the following: [Narration: narration, personal narratives, narrative medicine, anecdot*, testimonial*] + [Databases: internet, bibliographic database, online, Web based] + [Participants: patient, health care personnel, relative*, caregivers] + [Study: Surveys and Questionnaires, controlled clinical trials, cohort studies] (Multimedia Appendix 1).

Selection Criteria

Titles and abstracts of the retrieved studies were reviewed and assessed for inclusion and exclusion criteria independently by all members of the study team (DD, AG, MH, and NB). Researchers were trained in applying the predefined selection criteria. Nonagreements were discussed until consensus was reached. Papers were included if they (1) studied (potential) patients (with or without an established diagnosis or condition), relatives (or other nonrelated informal caregivers), or health care professionals; (2) studied the effects of Web-based patient narratives (real experiences or fictional stories; collections or single narratives; presentation as text or audio or video clips) that were generated scientifically (eg, through qualitative research methods and not just stories put selectively on the Web with a view to their human interest for marketing purposes);

and (3) were quantitative studies such as surveys and questionnaires, observational studies, nonrandomized controlled trials (non-RCTs), RCTs, comparative effectiveness research, cohort studies, or longitudinal studies. We excluded studies that used qualitative study designs such as interview studies, focus groups, or ethnographic studies and studies that were neither published in English or German. Studies that used narratives that were not generated by a scientific method were also excluded (eg, unmoderated blogs or fora). Furthermore, we excluded studies published before 2000 and studies that examined narratives not Web-based. We made no restrictions on the inclusion of studies regarding content, context, length, or depth of the narratives. We decided to focus on Web-based narratives as we felt the range would have been too broad to allow for meaningful comparisons had we included narratives available in different media (books, leaflets, newspapers, etc).

Quality Assessment

A protocol was written about all the steps of data collection and analysis, including selection of studies and extraction of content. Researchers were trained in applying the predefined selection criteria. Overall, 3 researchers reviewed and assessed all studies (DD, AG, and MH), whereas nonagreements were discussed with a fourth independent expert (NB). Evaluation tools designed for conventional systematic reviews typically assess the quality of RCTs. However, the diversity of research designs and outcome measures of the included studies required the use of a tool that is able to systematically appraise disparate evidence stemming from different study types. Therefore, 2 authors (DD and AG) independently performed an assessment of the quality of the included studies using the checklist proposed by Hawker et al [22]. This validated checklist consists of nine evaluation sections: abstract and title, introduction and aims, method and data, sampling, data analysis, ethics and bias, results, transferability or generalizability, and implications and usefulness. Each section was assessed by giving a score ranging from 1 to 4 (4=good, 3=fair, 2=poor, and 1=very poor), resulting in a potential score range of 9 to 36. Similar to the Appraisal of Guidelines for Research and Evaluation II instrument [23], we calculated sum scores for each section and an overall score, scaled as a percentage of the maximum possible score over all sections:

> obtained score−minimum possible score/maximum possible score−minimum possible score×100

Data Extraction

A data elicitation form was developed and applied systematically to all publications included in the review by 1 author (DD). The form includes information about background characteristics (authors, year of publication, and location), study characteristics (aim, sample size, participants, and study design), narrative (type of narrative and degree of exposure), study measures (attitudes and beliefs, psychometric scales, and preferences), and a summary of findings.

Data Synthesis

We extracted study results as they were reported in the results section of the publications. The analysis was based on the comparison of study details using descriptive statistics and text. The analysis was mainly focused on the identification of similarities and differences between the findings of the individual studies. As the study aims, designs, and findings were too heterogeneous, a meta-analysis was not conducted.

The specific outcomes of the studies were grouped using the taxonomy proposed by Shaffer and Zikmund-Fisher [3]. As several of the included studies provided few details about the content of the narratives, the studies were grouped around the purpose of the narrative. According to Shaffer and Zikmund-Fisher [3], five different purposes of narratives can be described. As most of the studies focus on (potential) patients rather than on relatives or health care professionals, the Shaffer and Zikmund-Fisher [3] taxonomy is suitable for our review. The purposes and their possible outcomes as proposed by Shaffer and Zikmund-Fisher [3] are described in Table 1.

Table 1. Purposes of narratives.

Purpose	Possible outcomes
Inform	• Increased knowledge • Improved affective forecasting
Engage	• Greater engagement • Greater transportation (increased depth of processing) • Greater time spent with materials
Model behavior	• Increased participation in health care decisions • Increased shared decision making • Altered behavioral intentions • Increased uptake of target behaviors
Persuade	• Altered behavioral intentions • Increased uptake of target behaviors
Comfort	• Reduced psychological distress • Reduced anxiety

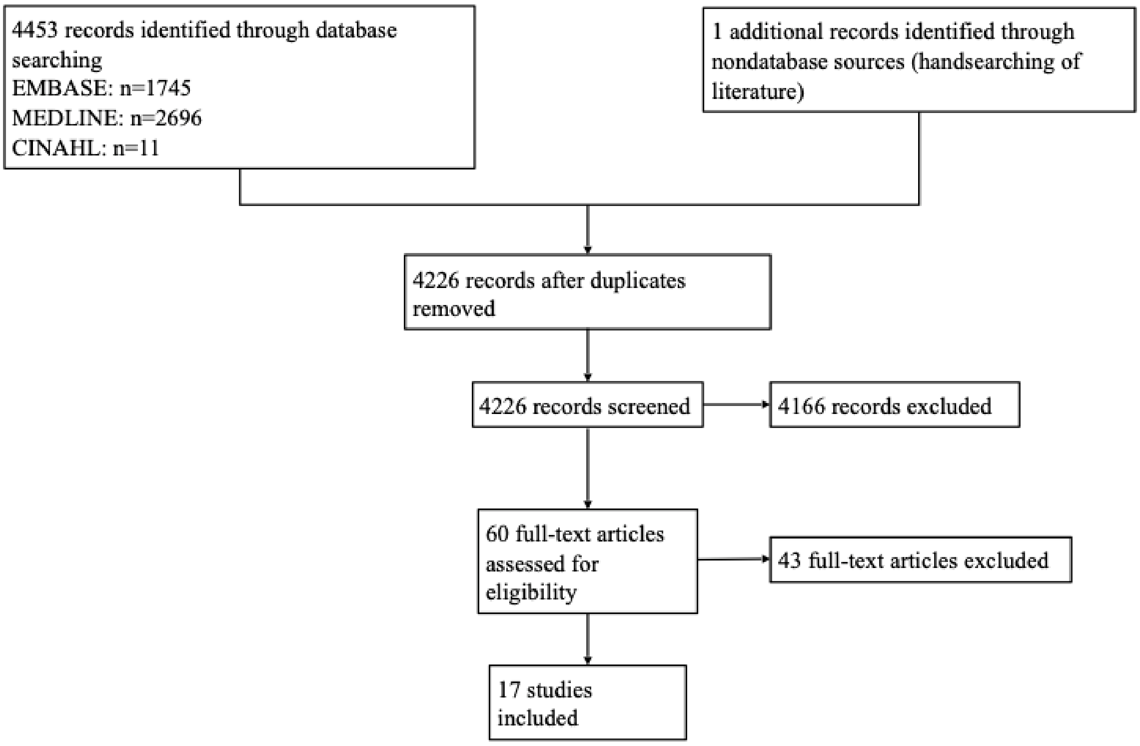
The definition of *effective* and *preference-sensitive* decisions proposed by Wennberg et al [24,25] was applied to assign the outcomes of the included studies to *risks* and *benefits* categories: outcomes were assigned to the *risk* category when they were *preference sensitiv*e. In *preference-sensitive* decisions, the best decision for an individual is unclear because of two reasons: the evidence for specific treatments is inadequate and firm conclusions about risk-to-benefit ratios cannot be drawn and the risk-to-benefit ratio might be clear, but it depends on the patients' values [24,25]. Outcomes were assigned to the *benefit* category when they were *effective* following the definition by Wennberg et al [24,25]. In these cases, the best decision is clear to practitioners and patients. The clinical evidence of harms and benefits is known, and compared with the benefits, the harms are minimal. In *effective* decisions, there is a widespread consensus among clinicians and patients about known and favorable risk-to-benefit ratios [24,25]. The outcomes of the included studies were assigned to a *no-effect* category when the corresponding studies reported experimental conditions inferior to the control group or no statistically significant effects (significance level chosen by the individual study). In descriptive studies, thresholds such as significance levels are not available. Therefore, outcomes of descriptive studies that were mentioned by ≤50% of the participants were also assigned to the *no-effect* category.

Results

Literature Search

Our search strategy identified 4226 documents. Of these, 60 documents potentially fulfilled the inclusion criteria of the study and were assessed in full text. After assessing the full texts, 43 more studies were excluded for specific reasons, including, eg, the study did not focus on systematically generated narratives or the narratives were not Web-based. There were 95.50% (4036/4226) agreements among the raters. Finally, 17 studies were included in the analysis (Figure 1).

Figure 1. Study flowchart.



Description of Included Studies

The studies were taken from Germany (n=5), the United States (n=6), the United Kingdom (n=4), the Netherlands (n=1), and Switzerland (n=1) and covered the period from 2000 to 2018 (Table 2). They investigated 10 different Web sources. The Web source and their specifications are shown in Multimedia Appendix 2.

Table 2. Characteristics of the included studies.

Authors (year)	Country applied	Sample size	Name of database	Degree of exposure to the narrative (eg, length of stay on a website)
Aardoom et al [26] (2014)	Netherlands	311	Proud2Bme	Mean time in months since first website visit: 19.8. Participants indicating to visit the website every day to several times a day: 189/247 (76.5%)
Allam et al [27] (2015)	Switzerland	157	ONESELF	Mean visits to the website: 53.68 (SD 93.07)
Betsch et al [14] (2011)	Germany	385 (study 1: 72; study 2: 313)	Online bulletin board	NR[a] (paper-and-pencil version of an online bulletin board)
Betsch et al [15] (2013)	Germany	458	Simulated website similar to the website *patientslikeme*	Mean time in minutes to complete the study: 9.94 (SD 3.49)
Brunette et al [28] (2015)	United States	39	Let's Talk About Smoking	NR
Engler et al [18] (2016)	Germany	23	DIPEx[b]	NR
Giesler et al [29] (2017)	Germany	212	DIPEx	Mean time in minutes on the intervention website: 42.21 (SD 45.64, median 26)
Newman et al [30] (2009)	United Kingdom	37	DIPEx	NR (paper-and-pencil survey)
Shaffer et al [31] (2013)	United States	302	Web decision aid	Mean time in minutes on the intervention website: 5.38 (SD 2.37); mean time in minutes on the control website: 4.92 (SD 2.03)
Shaffer et al [32] (2013b)	United States	56	Web decision aid	Mean time in seconds on different pages with text narratives: 5.00-67.28; mean time in seconds on different pages with video narratives: 15.11-117.19
Shaffer et al [33] (2014)	United States	200	Web decision aid	Length of narrative video: approximately 1 hour
Schweier et al [34] (2014)	Germany	571	lebensstil-aendern	Website usage in the intervention group: 46.1% (119/258); website usage in the control group: 7.0% (22/313)
Snow et al [35] (2016)	United Kingdom	88	DIPEx	Expected time to complete the module: 20 min. No time limits were set. Participants could watch the videos multiple times
Sullivan et al [36] (2018)	United States	2125 (study on acid reflux: 1070; study on high blood pressure: 1055)	Simulated prescription drug websites	All participants were exposed to the video. Participants that viewed the entire video: 94.86% (1015/1070) (acid reflux) and 98.66% (1041/1055) (high blood pressure). Participants that replayed the video: 7.5% (acid reflux) and 6.8% (high blood pressure)
Winterbottom et al [37] (2012)	United Kingdom	1694 (study 1: 578; study 2: 1116)	Web decision aid	NR
Wise et al [38] (2008)	United States	353	Comprehensive Health Enhancement Support System	No directives for the frequency of website use was given. Access to the website was given for four months.
Yaphe et al [39] (2000)	United Kingdom	309	DIPEx	NR

[a] Not reported.

[b] DIPEx: Database of Individual Patients' Experiences.

Sample sizes of the studies ranged from 23 to 2125 (samples of the following substudies were combined: Betsch et al [14], Sullivan et al [36], and Winterbottom et al [37]) with a median of 302 per study. Most of the included studies focus on the effects on (potential) patients. Only one study that met our inclusion criteria focused on (future) health care professionals (medical students) [35]. The measures of the studies were (1) general perceptions of patient narratives, including patients' expectations and learning experiences [18,30], self‐help and use of patients’ stories [39], and empowering processes and outcomes experienced by website participants [26]; (2) effects of narratives on patients’ and health care professionals’ behavior, including health care participation [38]; information search, treatment intentions, and decision satisfaction [32,33]; self-efficacy coping with cancer and patient competence [29]; physical activity [27,34]; health care utilization and medication overuse [27]; and performance in examinations [35]; and (3) decision making about dialysis modality [37], tobacco cessation treatment [28], vaccination [14,15], reflux and blood pressure drugs [36], and early-stage breast cancer [31,33].

The degree of exposure to the narrative was reported by 11 out of 17 studies. The reporting included measures such as self-reporting regarding frequency of website visits [26,34], mean visiting times of the websites [27] or mean times spent on the corresponding websites [15,29-31], the length of the narrative videos or expected study length [33,35], the number of participants exposed to the narratives [36], and the timespan for which participants had access to the corresponding websites [38].

The mean score of the quality assessment was 84.5 (range 61-100). The main issues were concerning appropriate sampling strategies [14,30,37], ethical issues regarding the relationship between researchers and participants [14,15,32,38,39], and the transferability of the study findings to a wider population [14,30,32,33,37,39] (Table 3). Among all the experimental studies, allocation concealment and study blinding were not adequately reported.

Table 3. Quality assessment of included studies.

Authors (year)	Abstract and title[a]	Introduction and aims[a]	Method and data[a]	Sampling[a]	Data analysis[a]	Ethics and bias[a]	Results[a]	Transferability or generalizability[a]	Implications and usefulness[a]	Scaled overall score[b]
Aardoom et al [26] (2014)	8	8	8	6	8	8	8	7	6	90.6
Allam et al [27] (2015)	8	8	8	8	8	8	8	8	8	100
Betsch et al [14] (2011)	8	8	8	4	7	2	7	5	6	68.4
Betsch et al [15] (2013)	8	8	8	8	7	2	7	8	8	85.2
Brunette et al [28] (2015)	8	8	8	8	8	8	8	8	8	100
Engler et al [18] (2016)	8	6	7	8	5	8	8	6	6	81.5
Giesler et al [29] (2017)	8	8	8	8	7	8	8	8	8	98
Newman et al [30] (2009)	8	8	4	4	4	8	6	4	6	63
Shaffer et al [31] (2013)	8	8	8	6	8	8	8	8	6	92.6
Shaffer et al [32] (2013b)	8	8	8	5	6	2	6	5	8	70.4
Shaffer et al [33] (2014)	8	7	7	6	5	6	6	4	8	72.1
Schweier et al [34] (2014)	8	8	8	8	8	8	8	8	8	100
Snow et al [35] (2016)	8	8	8	8	8	7	8	8	8	98
Sullivan et al [36] (2018)	8	8	8	8	8	8	8	8	8	100
Winterbottom et al [37] (2012)	7	8	8	4	4	8	8	5	6	74.1
Wise et al [38] (2008)	8	8	8	7	7	2	8	6	8	81.5
Yaphe et al [39] (2000)	8	7	6	7	4	2	6	4	7	61

[a] Sum score ranging from 2 to 8.

[b] Scaled overall score ranging from 0 to 100.

Description of Study Methodologies

The study design varied among the included studies (Table 4): nine used experimental designs, including four RCTs [29-35], and seven used factorial designs [14,15,32,33,36,37], two were descriptive cross-sectional survey studies [26,39], two were mixed method studies [18,30], one was a pre-post pilot study [28], and one was a secondary analysis [38]. Only one study [34] used an intention-to-treat analysis. Furthermore, 6 studies were informed by a theoretical framework, including the social learning theory [34,38], empowerment construct [26], social support features and gamification elements [27], theory of planned behavior [28], and a self-developed taxonomy of patient stories that provides a framework [31].

Table 4. Description of study methodologies.

Authors (year)	Study design	Measures (attitudes, psychometric scales, preferences, behavior, etc)	Type of participants
Aardoom et al [26] (2014)	Cross-sectional (descriptive online survey)	Eating psychopathology, general empowerment, symptom duration, treatment status, and user activity	Website visitors who indicated having eating problems
Allam et al [27] (2015)	5-arm parallel randomized controlled trial	Physical activity, health care utilization, medication overuse, empowerment, and rheumatoid arthritis knowledge	Individuals diagnosed with rheumatoid arthritis
Betsch et al [14] (2011)	Factorial between-subjects design	Perceived risk of side effects and vaccination intentions	Students
Betsch et al [15] (2013)	Factorial between-subjects design	Perceived risk, vaccination intention, and subjective numeracy	General population
Brunette et al [28] (2015)	Pre-post pilot study	Use of cessation treatment	Individuals smoking ≥4 cigarettes
Engler et al [18] (2016)	Mixed method approach including log file analyses, descriptive survey data analyses, and thematic analysis of focus group discussions (only quantitative results are extracted).	Attitudes toward health-related websites in general, perception of krankheitserfahrungen.de in particular	Individuals diagnosed with colorectal, breast, or prostate cancer
Giesler et al [29] (2017)	Randomized two-group between-subjects design with repeated measures.	Coping self-efficacy and patient competencies	Individuals diagnosed with colorectal cancer
Newman et al [30] (2009)	Mixed method. The study involved three stages: (1) focus groups guided the development of a descriptive questionnaire, (2) the questionnaire was modified, and (3) a sample of outpatients was asked to complete the questionnaire. (Only quantitative results are extracted.)	Attitudes toward the website	Individuals diagnosed with an inflammatory rheumatologic condition
Shaffer et al [31] (2013)	Factorial design	Information search, treatment intentions, and decision satisfaction	Women from the general population who were not pregnant and without a breast cancer history
Shaffer et al [32] (2013b)	Factorial design	Treatment preference	Women from the general population who were not pregnant and without a breast cancer history
Shaffer et al [33] (2014)	Factorial design	Treatment preference	Women from the general population who were not pregnant and without a breast cancer history
Schweier et al [34] (2014)	Sequential controlled trial	Diagnosis, BMI, baseline behavior for physical activity, eating routine, exercise frequency and attention paid to healthy diet, and improvements in physical activity and eating behavior	Individuals diagnosed with coronary heart disease
Snow et al [35] (2016)	Exploratory randomized controlled trial	Knowledge demonstration and clinical examination with a simulated patient	Medical students
Sullivan et al [36] (2018)	Factorial design	Risk perception	Individuals with self-reported acid reflux or high blood pressure
Winterbottom et al [37] (2012)	Factorial design	Hypothetical treatment choice	Students
Wise et al [38] (2008)	Secondary analysis	Health care participation and online information use	Women diagnosed with breast cancer
Yaphe et al [39] (2000)	Cross-sectional (descriptive survey study)	Whether and how patients' stories are collected and used	Self-help groups

Outcomes of Studies

Table 5 describes the effect directions of the outcomes of the included studies. The outcomes are grouped along the taxonomy from Shaffer and Zikmund-Fisher [3]. Most studies reported more than one outcome. Therefore, the number of outcomes is higher than the number of included studies.

Table 5. Effects of narratives on outcomes taxonomy.

Taxonomy, outcome, authors (year)	Effect direction		
	Risk	No effect	Benefit
Inform			
Competence			
Giesler et al [29] (2017)	N/A[a]	X[b]	N/A
Snow et al [35] (2016)	N/A	N/A	X
Knowledge			
Allam et al [27] (2015)	N/A	X	N/A
Engler et al [18] (2016)	N/A	N/A	X
Engage			
Empowerment			
Aardoom et al [26] (2014)	N/A	N/A	X
Allam et al [27] (2015)	N/A	N/A	X
Length of information search			
Shaffer et al [31] (2013)	N/A	N/A	X
Shaffer et al [32] (2013b)	N/A	N/A	X
Sharing experiences			
Engler et al [18] (2016)	N/A	N/A	X
Newman et al [30] (2009)	N/A	N/A	X
Yaphe et al [39] (2000)	N/A	N/A	X
Model behavior			
Eating behavior			
Schweier et al [34] (2014)	N/A	X	N/A
Health care utilization			
Allam et al [27] (2015)	N/A	N/A	X
Wise et al [38] (2008)	N/A	N/A	X
Medication overuse			
Allam et al [27] (2015)	N/A	N/A	X
Physical activity			
Allam et al [27] (2015)	N/A	N/A	X
Schweier et al [34] (2014)	N/A	X	N/A
Persuade			
Risk judgments			
Betsch et al [14] (2011)	X	N/A	N/A
Betsch et al [15] (2013)	X	N/A	N/A
Sullivan et al [40] (2018)	X	N/A	N/A
Treatment decisions			
Betsch et al [14] (2011)	X	N/A	N/A
Betsch et al [15] (2013)	X	N/A	N/A
Brunette et al [28] (2015)	N/A	N/A	X
Shaffer et al [33] (2014)	N/A	X	N/A
Winterbottom et al [37] (2012)	X	N/A	N/A
Comfort			

Taxonomy, outcome, authors (year)	Effect direction		
	Risk	No effect	Benefit
Confidence			
Shaffer et al [31] (2013)	N/A	N/A	X
Snow et al [35] (2016)	N/A	N/A	X
Self-efficacy			
Giesler et al [29] (2017)	N/A	X	N/A

[a] Not applicable.

[b] Each X represents an individual study reporting statistically significant risks, no significant effects or statistically significant benefits.

Provide Information

Giesler et al [29] investigated patient competence, including self-regulation, effective coping with emotional distress, explicit dealing with cancer threat, and low avoidance, as a secondary outcome in their study. They reported no significant differences between the intervention and control groups. Snow et al [35] examined the effect of patient narratives describing their colposcopy on fifth-year medical students' proficiency in standard examinations. They reported a significantly better performance in the examination compared with the control group that viewed a clinician describing the procedure (odds ratio [OR] 2.7, 95% CI 1.2-6.1; $P=.02$).

Allam et al [27] reported no significant improvements in the knowledge of rheumatoid arthritis. It should be noted that the initial level of the control group was significant. A study among cancer patients testing narrative cancer modules on the website *krankheitserfahrungen* found that 72% (40/56) agreed or strongly agreed that the internet is supportive to understanding what physicians tell them [18].

Engage

A study by Aardoom et al [26] reported that the exchange of information, finding recognition, sharing experiences with others, and feeling better informed were the most often reported empowering processes and outcomes. The authors concluded that online sources where individuals can share their experiences are promising strategies for successful electronic health (eHealth) initiatives such as *Proud2Bme*. A 5-arm parallel RCT found that levels of empowerment changed over time in study groups having access to online social support (beta=2.59; $P=.03$) or gamified experiences of a website (beta=2.29; $P=.05$) [27].

Participants viewing narratives relating how a patient makes her decision were found to spend more time searching for information regarding breast cancer (narrative condition, mean 5.38 min, SD 2.37, vs no narrative condition, mean 4.92 min SD 2.03 [31]; narrative condition, mean 39.88 min, SD 15.62, vs no narrative condition, mean 35.08 min, SD 16.09 [32]). Furthermore, Shaffer et al [31] reported that participants who viewed narratives containing experiences regarding diagnosis, treatment, or complications with early breast cancer treatments showed greater abilities to imagine how it might be to experience these treatments (imagine a mastectomy in the no narrative condition, mean 4.46, SD 1.21, vs imagine a mastectomy in the narrative condition, mean 4.69, SD 1.02, $t=1.72$; $P=.04$; imagine a lumpectomy with radiation in the no narrative condition, mean 4.44, SD 1.19, vs imagine a lumpectomy with radiation in the narrative condition, mean 4.72, SD 0.94, $t=2.19$; $P=.01$; measured on a 9-point Likert scale).

Findings showed that learning about other peoples' health-related experiences is relevant and helpful [18,30]. Furthermore, patients' stories collected by DIPEx are frequently included in interviews or articles for group newsletters, newspaper articles, or media broadcasts by voluntary organizations [39]. Engler et al [18], eg, reported that 76% (43/56) of their participants agreed that it can be helpful to witness the health-related experiences of others. However, some of the younger participants in the study by Newman et al [30] reported that the site did not cover experiences of younger patients. The participants highlighted the importance of incorporating current and accurate information. Some participants were concerned that the site might be depressing to patients with a new diagnosis [30].

Model Target Behaviors

A statistically significant positive effect on physical activity was reported by Allam et al [27]. In contrast, Schweier et al [34] did not find significant effects on physical activity and eating behavior changes. Health care utilization and medication overuse decreased according to the findings of one study [27]. Furthermore, one study investigated the effects of Web-based narratives and didactic information on health care participation [38]. This study reported positive effects of an eHealth program with narratives (audiovisual and text; beta=.123; $P<.01$) and didactic information (text only; beta=.104; $P<.05$) on health care participation. Health care participation was measured on a 7-item, 5-point response scale. These effects were reported to be significantly greater for African Americans.

Persuade

A total of six studies investigated the effects of narratives on risk judgments [14,36] and treatment decisions, including hypothetical treatment choices between a lumpectomy with radiation or a mastectomy [33], vaccination intentions [14,15], hypothetical dialysis modalities [37], and cessation treatment [28].

Furthermore, two studies [14,15] focused on the effects of statistical and/or narrative information on vaccination decisions. Betsch et al [14] showed that the perceived risk of vaccination increases the more the narratives report adverse events ($F_{2,58}=3.852$; $P<.05$; $\eta^2=0.12$), and if adverse events are reported in a highly emotional manner (mean 15.33, SD 9.27 vs mean 17.52, SD 11.00; $F_{1,297}$=4.197; P<.05; η^2=0.01). Furthermore, they showed that the intention for vaccination decreases when the number of narratives increases ($F_{2,58}$=5.729; P<.01; η^2=0.17), which is partially mediated by an increased perception of risk [14]. Two years later, the same research group published results from a similar setting, which point in the same direction [15]. Sullivan et al [36] investigated the influence of videos on consumers' knowledge, perceptions, and behavioral intentions. Participants were randomly assigned to 1 of 10 fictitious prescription drug websites. The video type (patient testimonial, informational video describing the mechanism of action, or none) and whether the video included drug risks was manipulated on each website. They found that participants who were exposed to any of the videos were less likely to recognize drug risks that were presented only on the website text. Videos that included risk information overall led to increased risk recognition. However, in some risk recognition measures such as risk of fracture, risk of special liver tests, or risk of angioedema, risk recognition decreased for risks that were not presented in the videos but risk information was always present in the website text. Furthermore, the study found no significant effects of risk prominence and type of video condition on physician interaction and search intentions on the internet.

In addition, one pilot study investigated the effect of a website that aimed to engage smokers in a cessation treatment [28]. Among 38 participants who used the website, 18 participants (47%) became abstinent for at least one day, 7 (18.4%) became abstinent for 7 or more days, and 4 (11%) became continuously abstinent. Winterbottom et al [37] demonstrated that hypothetical dialysis treatment choices presented as a patient narrative were more likely to be chosen by the participants than presented by a doctor (both using actors). Another study [33] found no differences in preferences for surgical treatments between women who watched videos that included narratives compared with those who watched a control video.

Provide Comfort

Snow et al [35] reported that students in the narrative condition reported significantly more confidence in comfort with patients' emotions (OR 6.4, 95% CI 2.7-14.9; P<.001). The study by Shaffer et al [31] compared participants who viewed experience narratives with those not viewing experience narratives. They demonstrated increased confidence in the experience narratives condition regarding the ability to make an informed choice (mean 3.77, SD 0.90 vs mean 4.01, SD 0.84; t=2.33; P<.01), to be more thorough in considering relevant factors (mean 4.07, SD 0.73 vs mean 4.21, SD 0.64; t=1.72; P<.04), to be more confident in the awareness of relevant factors (mean 3.29, SD 0.95 vs mean 3.53, SD 0.90; t=2.21; P<.01), and to be more satisfied with their decision-making process (mean 3.76, SD 0.81 vs mean 3.95, SD 0.77; t=2.08; P<.02).

Giesler et al [29] evaluated the colorectal cancer module of the German DIPEx website with regard to coping self-efficacy as the primary outcome and patient competencies as the secondary outcome. The study results did not support the authors' hypothesis that the website increases self-efficacy for coping with cancer or patient competencies such as self-regulation or managing emotional distress at 2 and 6 weeks after baseline.

Discussion

Principal Findings

There is an increasing number of Web-based sources containing research-based, systematically generated accounts of patient illness and health experiences. Although the evidence on the persuasiveness of narrative information on individuals' decision making was reviewed over a decade ago [4], we present, to our knowledge, the first systematic review about the effects of Web-based patient narratives on patients, relatives, or health care professionals.

Our review revealed several beneficial effects for patients and health care professionals. Web-based narratives are an effective way of teaching to improve knowledge and confidence for students as well as for patients [18,35]. Furthermore, research indicates that patients perceive other patients' health experiences as relevant and helpful [18,30]. This finding points to the importance of the quality of health-related information [41]. Compared with the health-related information and experiences on general social media sites, academic research–based patient narratives might be less susceptible to challenges for the quality of health-related information through, eg, spamming, intentional misspelling, or actuality of information [41]. Several quality measures to evaluate the quality of Web-based health information are available [42,43].

Another identified benefit is that participants viewing narratives that contain information on how patients make decisions result in longer search times for information [31,32]. This effect can be a resource to increase, eg, patients' health literacy. However, Shaffer et al [32] also reported that transcripts of the patient videos caused the opposite effect. Participants confronted with text-based narratives spent approximately five fewer minutes for information search. Researchers and health care professionals using patient narrative databases should be aware that the format of patient stories might be similarly important as the content in determining their effect on medical decision making [32].

On the basis of the findings of this review, it remains unclear whether patient narratives can influence patients' target behavior. The results regarding physical activity are equivocal [27,34]. Narratives led to an increase of health care participation and decreased unnecessary health care utilization as well as medication overuse [27,38].

Even though we identified several benefits of patient narratives on the different purposes of narratives, overall, there is little evidence for the effects of Web-based patient narratives in a positive or negative way. The total number of studies we included in the review is small, which is especially challenging in the light of the heterogeneity regarding the sampled population, the study aims, and the heterogeneity of the narratives in itself. Furthermore, the purposes of the narratives presented on the different websites vary considerably from each other. However, patient narratives are not homogenous and have to be evaluated in their context with regard to content, purpose, and patients' evaluative expression, such as expressions of (dis)satisfaction with processes of health care decision making [3]. We concur with the position of Shaffer and Zikmund-Fisher [3] that the role of narratives can only be fully understood if operational definitions of narratives are sufficient. Furthermore, there is a need for more theoretical conceptions about the impact of narratives on specific outcomes. We found that only 6 out of 17 studies were guided by a theoretical framework [31,32,34,37-39]. The lack of theoretical frameworks might partially be explained by the fact that research on narratives on certain patient-relevant outcomes is a relatively new field with a range of potentially relevant outcomes. Giesler et al [29], eg, found no significant differences at 2 weeks between an intervention group that had immediate access to the colorectal cancer module of the DIPEx website and a control group with regard to self-efficacy for coping with cancer and patient competence. Study participants in the intervention group visited the website on average for 42.21 min with 3.31 mean number of sessions. Such findings do not necessarily indicate that there is no effect of patient narratives. It rather highlights that the specific outcomes that were investigated in this study and in the study-specific setting were not significant. Other psychological measures on self-efficacy for coping or on patient competence may have led to different results. Another indication for the overall little evidence is that the effects of several studies reporting significant outcomes are rather small or are only significant under specific experimental conditions. For example, the difference between the mean search time for information regarding breast cancer reported by Shaffer et al [32] is 0.46 min, with an average search time of 5.38 min in the intervention group and 4.92 min in the control group.

Almost one-third of the included studies used study samples that cast doubts whether the results can be generalized to broader patient populations [14,30,32,37], Clear definitions of the basic population and appropriate sampling strategies would be desirable in future studies. Schlesinger et al [44] demonstrated that a rigorous collection of patient narratives can also be incorporated into large patient experience surveys.

At the same time, narratives can bear potential risks in preference-sensitive decisions [24]. There is a growing body of evidence on the effect of narrative bias [4,14,15], where narrative information can override risk judgments. This effect can even occur when base rate information is presented in addition [15]. Narratives are widely used in patient decision aids [45]. Furthermore, it is likely that narratives are used by other patients as decision support tools, although they are not explicitly declared as decision aids. Decision aids are evidence-based tools with an aim to support patients in a value-sensitive way to make specific health care choices [46]. Narratives may reduce the effectiveness of decision aids by presenting unbalanced information or by overriding decision-relevant information through characteristics of the narrator [4,45]. For example, a study conducted by Khangura et al [45] indicates that patient narratives in decision aids were more likely to portray patients that were satisfied with the outcome of their treatment decision. This points to the importance of including disclaimers that highlight the potential for biases in patient narratives [15]. Furthermore, this highlights the need for a careful selection of the presented stories on patient narrative databases by the corresponding research teams in charge for the databases to present a balanced picture of the whole spectrum of health experiences [47]. This might be especially important in narratives about health conditions where public opinions are mixed and biases might be suspected.

Qualitative studies focusing on how individuals use and value personal health-related experiences [10], decision making regarding prostate cancer [11], or information needs of patients with cancer and their views of internet-based health information [12] indicate improvements in decision making [10,11] and in meeting information needs [12]. These findings are not completely in line with our review of quantitative studies. How can this difference be explained? Both approaches study different phenomena. The foundations of the qualitative paradigm are interpretivism and constructivism, where multiple socially constructed realities are investigated [48,49]. On the contrary, the quantitative approach is based on positivism, which assumes that phenomena can be represented by empirical indicators that represent the one and only truth [49]. It can be speculated that the qualitative findings rather represent the *lived experiences* of patients' decision making, whereas the quantitative results represent quantitative measures of the decision-making processes.

Limitations

Our study has several limitations. First, we searched only for papers published in journals, and only in English or German. Papers that were published in books or reports are often not indexed in the databases we have chosen for our search strategy and are therefore not included. Therefore, we may have missed some studies published in languages or places outside our scope. Second, we reviewed only published studies regarding patient narratives. Therefore, we may potentially be confronted with a publication bias in such a way that, eg, negative study results were not published. Third, we decided to include only studies that focused on Web-based narratives and that were generated through a research methodology. Although we have done so to ensure comparability among the studies, we also acknowledge that this decision has led to an exclusion of several studies that investigated the effects of non–Web-based narratives or generated in an unstructured, non–research-based way, eg, in chatrooms or fora. Narratives are valuable resources for the narrators themselves, for other patients and their relatives, and for health care professionals and researchers. Despite the limitations, our findings might be helpful for health care professionals and researchers to understand the possible effects of narratives in health care settings.

Conclusions

In total, we found 17 studies on the effects of Web-based patient narratives. The effects of narratives were classified by purpose—inform, engage, model behavior, persuade, and comfort—using the taxonomy provided by Shaffer and Zikmund-Fisher [3]. Overall, patient narratives seem to be a promising means to improve knowledge of health care professionals and patients. Learning about other patients' experiences is perceived as supportive and relevant. Furthermore, they can positively influence patient empowerment. There is some evidence of beneficial effects on some outcomes, such as information search and the modeling of target behavior such as physical activity, health care participation, and medication overuse. The narratives used in the studies are characterized by considerable heterogeneity, and the investigated outcomes are hardly comparable among each other, which makes an overall judgment difficult.

Acknowledgments

The authors would like to thank Kevin Selby, Erik von Elm, and Nina Streeck for helpful comments on the setting up and drafting of this review.

Conflicts of Interest

All the authors declare that their institution is in charge of DIPEx Switzerland.

Multimedia Appendix 1
Search terms.
[DOCX File , 16 KB - jmir_v22i3e15772_app1.docx]

Multimedia Appendix 2
Investigated databases and descriptions.
[DOCX File , 14 KB - jmir_v22i3e15772_app2.docx]

References

1. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):e1196-e1252 [FREE Full text] [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]
2. World Health Organization. European Health Report 2018: More Than Numbers - Evidence for All. Copenhagen, Denmark: World Health Organization. Regional Office for Europe; 2018.
3. Shaffer VA, Zikmund-Fisher BJ. All stories are not alike: a purpose-, content-, and valence-based taxonomy of patient narratives in decision aids. Med Decis Making 2013 Jan;33(1):4-13. [doi: 10.1177/0272989X12463266] [Medline: 23065418]
4. Winterbottom A, Bekker HL, Conner M, Mooney A. Does narrative information bias individual's decision making? A systematic review. Soc Sci Med 2008 Dec;67(12):2079-2088. [doi: 10.1016/j.socscimed.2008.09.037] [Medline: 18951673]
5. Ziebland S, Lavie-Ajayi M, Lucius-Hoene G. The role of the internet for people with chronic pain: examples from the DIPEx International Project. Br J Pain 2015 Feb;9(1):62-64 [FREE Full text] [doi: 10.1177/2049463714555438] [Medline: 26516557]
6. Ziebland S, Herxheimer A. How patients' experiences contribute to decision making: illustrations from DIPEx (personal experiences of health and illness). J Nurs Manag 2008 May;16(4):433-439. [doi: 10.1111/j.1365-2834.2008.00863.x] [Medline: 18405260]
7. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. Br Med J 2004 Mar 6;328(7439):564 [FREE Full text] [doi: 10.1136/bmj.328.7439.564] [Medline: 15001506]
8. Greenhalgh T. Cultural Contexts of Health: The Use of Narrative Research in the Health Sector. Copenhagen: World Health Organization Regional Office for Europe; 2016.
9. Burkow TM, Vognild LK, Østengen G, Johnsen E, Risberg MJ, Bratvold A, et al. Internet-enabled pulmonary rehabilitation and diabetes education in group settings at home: a preliminary study of patient acceptability. BMC Med Inform Decis Mak 2013 Mar 5;13:33 [FREE Full text] [doi: 10.1186/1472-6947-13-33] [Medline: 23496829]
10. Entwistle VA, France EF, Wyke S, Jepson R, Hunt K, Ziebland S, et al. How information about other people's personal experiences can help with healthcare decision-making: a qualitative study. Patient Educ Couns 2011 Dec;85(3):e291-e298. [doi: 10.1016/j.pec.2011.05.014] [Medline: 21652162]
11. Chapple A, Ziebland S, Herxheimer A, McPherson A, Shepperd S, Miller R. Is 'watchful waiting' a real choice for men with prostate cancer? A qualitative study. BJU Int 2002 Aug;90(3):257-264 [FREE Full text] [doi: 10.1046/j.1464-410x.2002.02846.x] [Medline: 12133062]
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 Educ Couns 2004 Apr;53(1):57-64. [doi: 10.1016/S0738-3991(03)00116-2] [Medline: 15062905]
13. Evans DG, Barwell J, Eccles DM, Collins A, Izatt L, Jacobs C, FH02 Study Group, RGC teams, et al. The Angelina Jolie effect: how high celebrity profile can have a major impact on provision of cancer related services. Breast Cancer Res 2014 Sep 19;16(5):442. [doi: 10.1186/s13058-014-0442-6] [Medline: 25510853]

14. Betsch C, Ulshöfer C, Renkewitz F, Betsch T. The influence of narrative v statistical information on perceiving vaccination risks. Med Decis Making 2011;31(5):742-753. [doi: 10.1177/0272989X11400419] [Medline: 21447730]
15. Betsch C, Renkewitz F, Haase N. Effect of narrative reports about vaccine adverse events and bias-awareness disclaimers on vaccine decisions: a simulation of an online patient social network. Med Decis Making 2013 Jan;33(1):14-25. [doi: 10.1177/0272989X12452342] [Medline: 22875721]
16. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
17. Schaffer R, Kuczynski K, Skinner D. Producing genetic knowledge and citizenship through the internet: mothers, pediatric genetics, and cybermedicine. Sociol Health Illn 2008 Jan;30(1):145-159 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01042.x] [Medline: 18254838]
18. Engler J, Adami S, Adam Y, Keller B, Repke T, Fügemann H, et al. Using others' experiences. Cancer patients' expectations and navigation of a website providing narratives on prostate, breast and colorectal cancer. Patient Educ Couns 2016 Aug;99(8):1325-1332. [doi: 10.1016/j.pec.2016.03.015] [Medline: 27067064]
19. Breuning M, Lucius-Hoene G, Burbaum C, Himmel W, Bengel J. [Patient experiences and patient centeredness : The website project DIPEx Germany]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2017 Apr;60(4):453-461. [doi: 10.1007/s00103-017-2524-y] [Medline: 28251275]
20. Greenhalgh T, Wengraf T. Collecting stories: is it research? Is it good research? Preliminary guidance based on a Delphi study. Med Educ 2008 Mar;42(3):242-247. [doi: 10.1111/j.1365-2923.2007.02956.x] [Medline: 18275411]
21. 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]
22. Hawker S, Payne S, Kerr C, Hardey M, Powell J. Appraising the evidence: reviewing disparate data systematically. Qual Health Res 2002 Nov;12(9):1284-1299. [doi: 10.1177/1049732302238251] [Medline: 12448672]
23. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. Can Med Assoc J 2010 Dec 14;182(18):E839-E842 [FREE Full text] [doi: 10.1503/cmaj.090449] [Medline: 20603348]
24. Wennberg JE, Fisher ES, Skinner JS. Geography and the debate over Medicare reform. Health Aff (Millwood) 2002;21(Suppl1):W96-114. [doi: 10.1377/hlthaff.w2.96] [Medline: 12703563]
25. O'Connor AM, Légaré F, Stacey D. Risk communication in practice: the contribution of decision aids. Br Med J 2003 Sep 27;327(7417):736-740 [FREE Full text] [doi: 10.1136/bmj.327.7417.736] [Medline: 14512487]
26. Aardoom JJ, Dingemans AE, Boogaard LH, van Furth EF. Internet and patient empowerment in individuals with symptoms of an eating disorder: a cross-sectional investigation of a pro-recovery focused e-community. Eat Behav 2014 Aug;15(3):350-356. [doi: 10.1016/j.eatbeh.2014.04.003] [Medline: 25064280]
27. Allam A, Kostova Z, Nakamoto K, Schulz PJ. The effect of social support features and gamification on a Web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015 Jan 9;17(1):e14 [FREE Full text] [doi: 10.2196/jmir.3510] [Medline: 25574939]
28. Brunette MF, Gunn W, Alvarez H, Finn PC, Geiger P, Ferron JC, et al. A pre-post pilot study of a brief, web-based intervention to engage disadvantaged smokers into cessation treatment. Addict Sci Clin Pract 2015 Feb 1;10:3 [FREE Full text] [doi: 10.1186/s13722-015-0026-5] [Medline: 25638283]
29. Giesler JM, Keller B, Repke T, Leonhart R, Weis J, Muckelbauer R, et al. Effect of a website that presents patients' experiences on self-efficacy and patient competence of colorectal cancer patients: web-based randomized controlled trial. J Med Internet Res 2017 Oct 13;19(10):e334 [FREE Full text] [doi: 10.2196/jmir.7639] [Medline: 29030329]
30. Newman MA, Ziebland S, Barker KL. Patients' views of a multimedia resource featuring experiences of rheumatoid arthritis: pilot evaluation of www.healthtalkonline.org. Health Informatics J 2009 Jun;15(2):147-159. [doi: 10.1177/1460458209102974] [Medline: 19474226]
31. Shaffer VA, Hulsey L, Zikmund-Fisher BJ. The effects of process-focused versus experience-focused narratives in a breast cancer treatment decision task. Patient Educ Couns 2013 Nov;93(2):255-264. [doi: 10.1016/j.pec.2013.07.013] [Medline: 23962673]
32. Shaffer VA, Owens J, Zikmund-Fisher BJ. The effect of patient narratives on information search in a web-based breast cancer decision aid: an eye-tracking study. J Med Internet Res 2013 Dec 17;15(12):e273 [FREE Full text] [doi: 10.2196/jmir.2784] [Medline: 24345424]
33. Shaffer VA, Tomek S, Hulsey L. The effect of narrative information in a publicly available patient decision aid for early-stage breast cancer. Health Commun 2014;29(1):64-73. [doi: 10.1080/10410236.2012.717341] [Medline: 23384155]
34. Schweier R, Romppel M, Richter C, Hoberg E, Hahmann H, Scherwinski I, et al. A web-based peer-modeling intervention aimed at lifestyle changes in patients with coronary heart disease and chronic back pain: sequential controlled trial. J Med Internet Res 2014 Jul 23;16(7):e177 [FREE Full text] [doi: 10.2196/jmir.3434] [Medline: 25057119]
35. Snow R, Crocker J, Talbot K, Moore J, Salisbury H. Does hearing the patient perspective improve consultation skills in examinations? An exploratory randomized controlled trial in medical undergraduate education. Med Teach 2016 Dec;38(12):1229-1235. [doi: 10.1080/0142159X.2016.1210109] [Medline: 27573531]

36. Sullivan HW, O'Donoghue AC, Gard Read J, Amoozegar JB, Aikin KJ, Rupert DK. Testimonials and informational videos on branded prescription drug websites: experimental study to assess influence on consumer knowledge and perceptions. J Med Internet Res 2018 Jan 23;20(1):e13 [FREE Full text] [doi: 10.2196/jmir.7959] [Medline: 29362205]
37. Winterbottom AE, Bekker HL, Conner M, Mooney AF. Patient stories about their dialysis experience biases others' choices regardless of doctor's advice: an experimental study. Nephrol Dial Transplant 2012 Jan;27(1):325-331. [doi: 10.1093/ndt/gfr266] [Medline: 21642512]
38. Wise M, Han JY, Shaw B, McTavish F, Gustafson DH. Effects of using online narrative and didactic information on healthcare participation for breast cancer patients. Patient Educ Couns 2008 Mar;70(3):348-356 [FREE Full text] [doi: 10.1016/j.pec.2007.11.009] [Medline: 18201859]
39. Yaphe J, Rigge M, Herxheimer A, McPherson A, Miller R, Shepperd S, et al. The use of patients' stories by self-help groups: a survey of voluntary organizations in the UK on the register of the College of Health. Health Expect 2000 Sep;3(3):176-181 [FREE Full text] [doi: 10.1046/j.1369-6513.2000.00092.x] [Medline: 11281927]
40. Sullivan J, Gillam L, Monagle P. Parents and end-of-life decision-making for their child: roles and responsibilities. BMJ Support Palliat Care 2015 Sep;5(3):240-248. [doi: 10.1136/bmjspcare-2013-000558] [Medline: 24644205]
41. De Martino I, D'Apolito R, McLawhorn AS, Fehring KA, Sculco PK, Gasparini G. Social media for patients: benefits and drawbacks. Curr Rev Musculoskelet Med 2017 Mar;10(1):141-145 [FREE Full text] [doi: 10.1007/s12178-017-9394-7] [Medline: 28110391]
42. 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] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
43. Boyer C, Selby M, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for medical and health Websites. Comput Biol Med 1998 Sep;28(5):603-610. [doi: 10.1016/s0010-4825(98)00037-7] [Medline: 9861515]
44. Schlesinger M, Grob R, Shaller D, Martino SC, Parker AM, Rybowski L, et al. A rigorous approach to large-scale elicitation and analysis of patient narratives. Med Care Res Rev 2018 Oct 6:1077558718803859. [doi: 10.1177/1077558718803859] [Medline: 30293521]
45. Khangura S, Bennett C, Stacey D, O'Connor AM. Personal stories in publicly available patient decision aids. Patient Educ Couns 2008 Dec;73(3):456-464. [doi: 10.1016/j.pec.2008.07.035] [Medline: 18774668]
46. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
47. Grob R, Schlesinger M, Parker AM, Shaller D, Barre LR, Martino SC, et al. Breaking narrative ground: innovative methods for rigorously eliciting and assessing patient narratives. Health Serv Res 2016 Jun;51(Suppl 2):1248-1272 [FREE Full text] [doi: 10.1111/1475-6773.12503] [Medline: 27126144]
48. Berger PL, Luckmann T. The Social Construction of Reality. Frankfurt am Main: Fischer; 2004.
49. Sale JE, Lohfeld LH, Brazil K. Revisiting the quantitative-qualitative debate: implications for mixed-methods research. Qual Quant 2002 Feb;36(1):43-53 [FREE Full text] [doi: 10.1023/A:1014301607592] [Medline: 26523073]

Abbreviations

DIPEx: Database of Individual Patients’ Experiences
eHealth: electronic health
OR: odds ratio
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 05.08.19; peer-reviewed by C Holmberg, H Sullivan; comments to author 14.09.19; revised version received 01.11.19; accepted 16.12.19; published 26.03.20.

Please cite as:
Drewniak D, Glässel A, Hodel M, Biller-Andorno N
Risks and Benefits of Web-Based Patient Narratives: Systematic Review
J Med Internet Res 2020;22(3):e15772
URL: http://www.jmir.org/2020/3/e15772/
doi:10.2196/15772
PMID:32213468

Original Paper

The Impact of Portal Satisfaction on Portal Use and Health-Seeking Behavior: Structural Equation Analysis

Reginald A Silver[1*], DrPH; Chandrasekar Subramaniam[1*], PhD; Antonis Stylianou[1*], PhD

University of North Carolina at Charlotte, Charlotte, NC, United States

[*] all authors contributed equally

Corresponding Author:
Reginald A Silver, DrPH
University of North Carolina at Charlotte
9201 University City Blvd.
Charlotte, NC,
United States
Phone: 1 7046876181
Email: rsilver5@uncc.edu

Abstract

Background: Our study addresses a gap in the modern information systems (IS) use literature by investigating factors that explain patient portal satisfaction (SWP) and perceptions about health-seeking behavior (HSB). A novel feature of our study is the incorporation of actual portal use data rather than the perceptions of use intention, which prevails in the modern IS literature.

Objective: This study aimed to empirically validate factors that influence SWP as an influencing agent on portal use and HSB. Our population segment was comprised of college students with active patient portal accounts.

Methods: Using web-based survey data from a population of portal users (n=1142) in a university health center, we proposed a theoretical model that adapts constructs from the Technology Acceptance Model by Davis, the revised Technology Adoption Model by Venkatesh, the Unified Theory of the Acceptance and Use of Technology 2, and the Health Belief Model by Rosenstock et al. We validated our model using structural equation modeling techniques.

Results: Our model explained nearly 65% of the variance in SWP (R^2=0.6499), nearly 33% of the variance in portal use (R^2=0.3250), and 29% of the variance in HSB (R^2=0.2900). Statistically significant antecedents of SWP included social influence (beta=.160, t_{499}=6.145), habit (beta=.114, t_{499}=4.89), facilitating conditions (beta=.062, t_{499}=2.401), effort expectancy (beta=.311, t_{499}=11.149), and performance expectancy (beta=.359, t_{499}=11.588). SWP influenced HSB (beta=.505, t_{499}=19.705) and portal use (beta=.050, t_{499}=2.031). We did not find a statistically significant association between portal use and HSB (beta=.015, t_{499}=0.513). Perceived severity significantly influenced HSB (beta=.129, t_{499}=4.675) but not portal use (beta=.012, t_{499}=.488).

Conclusions: Understanding the importance of SWP and the role it plays in influencing HSB may point to future technology design considerations for information technology developers and health care providers. We extend current Expectancy Confirmation Theory research by finding a positive association between SWP and portal use.

(J Med Internet Res 2020;22(3):e16260) doi:10.2196/16260

KEYWORDS

patient portals; health care IT; technology usage; health-seeking behavior

Introduction

Background

Patient portals have become instrumental in the engagement of patients in their own care. A number of factors have contributed to the proliferation and evolution of portal systems. Patient portals evolved out of the expansion of electronic health record (EHR) systems that were incentivized by the American Recovery and Reinvestment Act. The act required health care providers to implement EHRs as a mechanism to ensure patient safety, improve the coordination of patient care, and better engage patients and their caregivers, while maintaining privacy and security of personal health care information. Health care providers were instructed to ensure that at least 5% of their patients could access their health information over the Web and to ensure that they could also exchange secure messages with their health care provider [1]. Patient portals became an important tool in promoting the kind of patient engagement prescribed by the American Recovery and Reinvestment Act. Patient portals promote this engagement through functionality that includes the ability to exchange secure communication with health care providers, schedule appointments, review lab results, and renew prescriptions for medications. Even with all of these factors that would seemingly promote high levels of patient portal usage, patient portal usage has been relatively low since the inception of these systems. Peacock et al [2] estimate patient portal usage to be only about 28% for portal users within the first year of activating a portal account. Current literature has provided a basic understanding of factors that influence portal use and how this use is thought to influence outcomes. With this research study, we validate and extend this knowledge base.

Our research effort is intended to fill a gap in current health care information technology (IT) research concerning portal user perceptions about portals and how those perceptions are related to perceived health-seeking behavior (HSB) and satisfaction with the portal. We investigate how portal satisfaction (SWP) may be associated with perceived HSB and with portal use. Furthermore, our study adds to the literature by using objective portal use data as opposed to intention to use, which is prevalent in the current literature. Specific research questions (RQs) include the following:

- *RQ1*: What factors are associated with a patient’s portal satisfaction?
- *RQ2*: Does patient portal satisfaction influence health-seeking behavior?
- *RQ3*: Does patient portal use influence health-seeking behavior?
- *RQ4*: Does patient portal satisfaction influence portal use?
- *RQ5*: Does perceived severity of health condition influence health-seeking behavior?
- *RQ6*: Does perceived severity of health condition influence portal use?

We next review the current literature on patient portals to identify broad findings about and research gaps in portal use. Following this, we describe our research model and hypotheses. We then present our research methods and results. We conclude our paper with a discussion of our findings, limitations, and conclusions.

Literature Review

Existing research has shown that patient portal usage varies by demographic factors such as age, sex, and morbidity level. Portal usage has been demonstrated to be higher among middle-aged female patients with more severe conditions [3]. Repeated portal usage has been demonstrated to be higher among patients with chronic medical conditions [4]. A patient is more likely to have an interest in using the portal to communicate with their health care provider when the patient is dissatisfied with the health care provider’s responsiveness to traditional forms of communication [5]. Other factors that have been demonstrated to influence portal usage include the endorsement of the portal by the health care provider, the utility of portal features, the usability of portal features, and the health literacy of the patient [6]. Conversely, portal usage has been hindered by low health literacy among end users [7].

Knowledge about patient portal usage comes primarily from studies that evaluate usage in the outpatient setting. In the inpatient setting, patient portal usage seems likely to have the same positive effects regarding self-involvement in one’s own care. Inpatient usage, however, has also been low at approximately 23% [8]. Similar patterns of usage can be seen across both the inpatient and outpatient settings. Patients with more severe conditions (eg, surgical patients), as demonstrated by longer lengths of stay, tend to use the portal more than patients with less severe conditions [9]. Portal usage across both inpatient and outpatient settings is impacted by demographic factors such as age, sex, education level, and race. However, a missing link is understanding how patient satisfaction impacts portal use.

Existing studies have been focused largely on describing portal usage through the analysis of system data, but understanding the factors that influence HSB still represents a gap in the existing literature [10]. Even fewer studies have been grounded in some theory-driven framework. Marton and Choo [11] identified only four such theory-driven studies in their 2011 paper on theoretical models for Web-based health information seeking. The theoretical frameworks utilized by those studies comprised expectancy value models, the Theory of Planned Behavior, the Technology Acceptance Model (TAM), and a behavioral model of Web-based information seeking. This gap in contemporary research on patient portal usage may be filled by the expanded use of theoretical frameworks such as the Health Belief Model (HBM), TAM, and the Unified Theory of the Acceptance and Use of Technology 2 (UTAUT2) [12-15].

Research on patient portal usage predicated upon theoretical frameworks is emerging. Tavares and Oliveira [16] used a derivative of Venkatesh’s UTAUT2 theoretical framework to investigate EHR adoption with an added construct of the patient’s self-perception. Unique to their study, Tavares and Oliveira integrated into their theoretical model the construct of patient self-perception that measured perceived vs real severity of a portal user’s chief medical complaint. Tavares and Oliveira found that statistically significant motivators for the behavioral intention to use a patient portal were performance expectancy, effort expectancy, habit, and self-perception. Similar to Tavares and Oliveira, we extend Venkatesh’s UTAUT2 model to include constructs and relationships for perceived HSB and perceived satisfaction with the portal.

Another stream of literature that has emerged focuses on understanding the relationship between user satisfaction with technology (eg, portals) and the actual use of the technology. Deng et al [17] studied the use of mobile internet services and found that user satisfaction positively impacted the intention to continue using the services. Ghobakhloo et al [18] reviewed different theoretical models of technology acceptance and formulated an integrated model of acceptance and satisfaction. Their model proposes that technology satisfaction impacts technology usage, which, in turn, impacts usage behavior [18].

With this research effort, we hope to inform the research community about portal user perceptions and how these perceptions are related to perceived HSB and satisfaction with the portal. We investigate how SWP may be associated with perceived HSB.

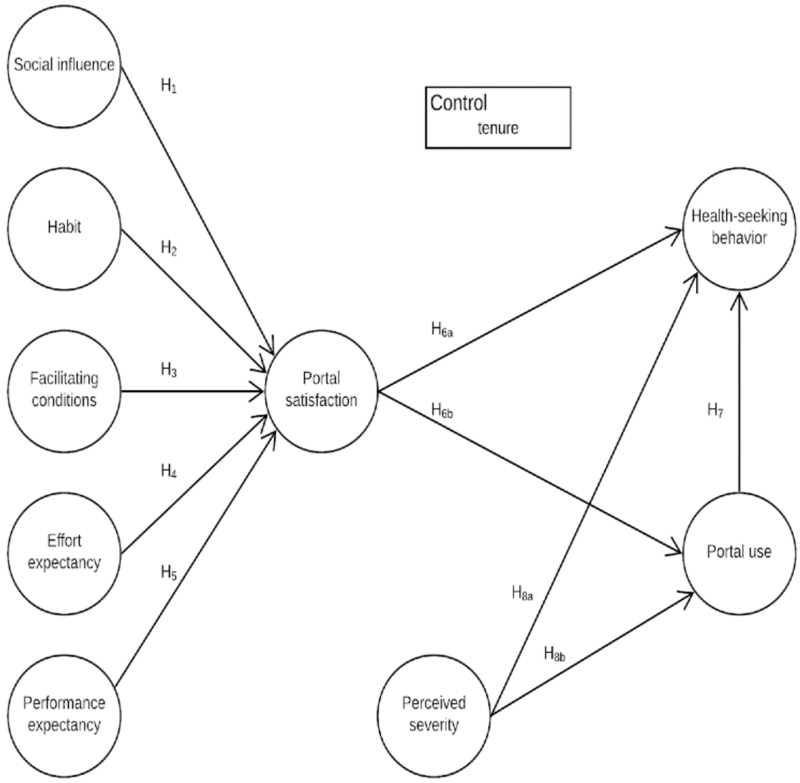
Research Model and Hypotheses

Our theoretical model incorporates social influence, habit, facilitating conditions, effort expectancy, and performance expectancy. Each of these constructs are traditionally used exogenous variables in management information systems (IS) research. We explore how these variables affect satisfaction with the portal and use. Drawing from HBM, we propose a relationship between perceived severity and HSB. Similarly, we propose a relationship between perceived severity and use. Our final hypothesis tests the relationship between portal use and HSB. Our complete research model is shown in Figure 1.

Patients who interact with health care technology, such as patient portals, are usually motivated to use the technology because it might be the main source of contact with their health care provider. Studies that have explored TAMs have consistently identified social influence as an important factor that influences end user behavior toward technology. Social influence is based on social contagion theory, which postulates that an individual's behaviors are subject to those of the people who are important to that individual [19]. These influencers include care givers, health care providers, and close relatives. An example of social influence is when family members discuss the benefits of portals with the patient or are themselves users of portals [20].

Bhattacherjee [21] identified confirmation as an antecedent to technology satisfaction while proposing the use of the Expectation-Confirmation Model in an IS context. Bhattacherjee defined confirmation as the comparison between end users' pre- and post-technology usage expectations. Similar to how Bhattacherjee positioned confirmation as an antecedent of technology satisfaction, we posit that it is feasible that a portal user's social circle can serve as a proxy for confirmation expectation. Portal users with more positive input from their close social circle about portals are likely to be more satisfied themselves with the portal. Although there is little empirical research in this space, we test this influence on satisfaction in our model.

Figure 1. Proposed research model.



H1: Perceived Social Influence is Positively Associated With Portal Satisfaction

Venkatesh et al [22] considered the effect that habit has on technology use. Habit has been demonstrated to predict end user intention to use a technology. Contemporary views of habit have also relied on the definition provided by Limayem et al [23] who defined habit as the extent to which people tend to automatically execute certain behaviors such as technology use because of learning. Amoroso and Lim [24] found a positive correlation between habit and consumer satisfaction. As Amoroso and Lim did find a correlation between habit and satisfaction, we found utility in testing this correlation between habit and satisfaction in our theoretical model. Drawing from these previous uses of habit in IS research, we hypothesize that habit is positively associated with SWP.

H2: Habit is Positively Associated With Portal Satisfaction

Facilitating conditions is a UTAUT construct that describes a person's perception of the resources and any other supporting elements at their disposal that assist them in performing a task (eg, technology use) [22]. Maillet et al [25] tested and found a statistically significant link between facilitating conditions and nurse satisfaction with an EHR. Conditions facilitating the access to and use of the system lead to a positive user experience, resulting in higher user satisfaction. Hence, we posit that facilitating conditions are positively associated with SWP.

H3: Facilitating Conditions are Positively Associated With Portal Satisfaction

The technological characteristics of the patient portal are important in motivating a patient to use the portal. Two important elements of portal technology are effort expectancy and performance expectancy. Effort expectancy is the degree to which the portal is easy to navigate and use, thus requiring less effort by the user [26]. An end user who finds a system intuitive and easy to use will have less frustration in completing tasks while using that system. In addition to finding the previously cited link between facilitating conditions and EHR satisfaction among nurses, Maillet et al [25] tested and found a statistically significant link between effort expectancy and satisfaction. Building on this finding by Maillet et al [25], we explore whether effort expectancy is associated with SWP.

H4: Effort Expectancy is Positively Associated With Portal Satisfaction

Performance expectancy is the degree to which the user perceives that the technology will help in carrying out functions important to the user [22]. For a patient portal user, performance expectancy reflects the patient's perception of how well it will help with better management of the user's health care. Some of the expected benefits of using a portal are viewing lab results on a smart device, easily scheduling or rescheduling appointments, and securely messaging physicians or nurses [27]. If these postadoption expectations are met, then portal users are likely to be satisfied with their portal. We find support for this idea through existing IS research that has used the Expectation-Confirmation Theory (ECT) to demonstrate a positive and statistically significant link between confirmation and satisfaction [28,29]. As technology users confirm that their expectations for the technology are being met, they exhibit satisfaction with that technology. We hypothesize that performance expectancy is positively associated with SWP.

H5: Performance Expectancy is Positively Associated With Portal Satisfaction

Evidence is accumulating in the medical literature that portals are associated with more favorable health outcomes for patients who use portals [30,31]. We suspect that this emerging focus on improving health outcomes through the use of portals is, in some way, connected to the end user's satisfaction with the portal as a tool that supports their HSB. We note that this is an understudied phenomenon that is not well covered in existing research. In our paper, we focus on the patient's satisfaction with the portal and the patient's HSB as the outcome measure.

H6a: Portal Satisfaction is Positively Associated With Health-Seeking Behavior

Wixom and Todd [32] partially explained technology use intention through the influence of attitude, which was influenced by ease of use, which was, in turn, influenced by system satisfaction. Drawing from their contribution, we hypothesize that SWP will result in higher instances of portal use. This connection between satisfaction and use is also supported in the current ECT literature [21,28,32]. From the ECT literature, we note four important contributions: (1) Bhattacherjee [21] established satisfaction as an antecedent of IS continuance (repeat usage after adoption); (2) Thong et al [28] suggested end users are more likely to continue the use of an IT, if they are satisfied with that IT; (3) Oghuma et al [29] demonstrated how perceived service quality and perceived usability can impact satisfaction, which, in turn, impacts use continuance; and (4) Wixom and Todd [32] showed how system satisfaction can indirectly serve as an attitudinal influence on an end user's intention to use a technology.

H6b: Portal Satisfaction is Positively Associated With Portal Use

The connection between SWP and HSB is an emerging concept. To date, there has been very little research into a direct connection between satisfaction with the portal (or health care IT of any kind) and HSB. Kim and Park [33] suggest an indirect link between health belief and the motivation to take actions toward health management. Tustin [34] explored the role that patient satisfaction plays in influencing Web-based health information seeking. Tustin suggests that patients who are dissatisfied with their health care provider are more likely to seek and trust information from sources other than their provider [34]. Although Kim and Park and Tustin attempt to explain relationships between individual factors such as health belief, level of provider satisfaction, technology, and HSB, the shared scope of HSB between these studies has been limited to Web-based health information seeking. We attempt to build upon what previous research has contributed by exploring the relationship between SWP and HSB within the context of an end user's perceived change in health status being based on their having used a portal.

H7: Portal Use is Positively Associated With Health-Seeking Behavior

Some studies on understanding a patient's motivation for connecting with Web sources, such as health websites and patient portals, are rooted in HBM. HBM suggests that a patient's perceived health risk predicts the likelihood of that patient being more engaged in their health care. They do this by seeking more information about their health and adapt their behavior toward better health [35]. We lean on this idea that perceived severity is a determinant of perceived HSB.

H8a: Perceived Severity is Positively Associated With Health-Seeking Behavior

We also expect that perceived severity will be associated with increased portal use. We hypothesize that patients who perceive that they suffer from severe health conditions will demonstrate more frequent portal use.

H8b: Perceived Severity is Positively Associated With Portal Use

Our hypotheses are listed in Table 1. In testing these relationships, we hope to confirm existing research findings such as the work done by Hsu and Lin [36] that demonstrates the strength of social influence on a person's motivation to use a type of IT; in their case, the focus was on the association of social influence with motivation to use blogs.

We consider the addition of the possible link between portal use and HSB to be a novel contribution that has not been heavily investigated in other research. Understanding whether this relationship exists would be informative in determining whether portal use influences HSB. This knowledge could have implications for encouraging HSB.

Table 1. Hypotheses tested.

Hypothesis	Hypothesis description
H_1	Perceived social influence is positively associated with portal satisfaction.
H_2	Habit is positively associated with portal satisfaction.
H_3	Facilitating conditions are positively associated with portal satisfaction.
H_4	Effort expectancy is positively associated with portal satisfaction.
H_5	Performance expectancy is positively associated portal satisfaction.
H_{6a}	Portal satisfaction is positively associated with health-seeking behavior.
H_{6b}	Portal satisfaction is positively associated with portal use.
H_7	Portal use is positively associated with health-seeking behavior.
H_{8a}	Perceived severity is positively associated with health-seeking behavior.
H_{8b}	Perceived severity is positively associated with portal use.

Control Variable: Tenure

For the purposes of our research study, we define tenure as the number of days between a portal user registering for a portal account and the date on which the system report was generated to produce the dataset used in our analysis. We set a cutoff point for tenure of 60 days. This allowed us to eliminate the effect of early portal learners who could have potentially skewed the results of our study. We surmise that end users who have used the portal longer will be more satisfied, borrowing from the ECT concept of satisfaction being correlated with continued use. We also surmise that longer periods of use are more likely to be associated with perceived changes in HSB. By holding tenure as a control variable, our model allows for the testing of any direct associations between tenure and portal use as well as tenure and HSB.

Methods

Data Collection

Web-based survey responses were collected from a total of 1142 respondents who were active portal users within a university student health system. For the purposes of this research study, we define *active* portal users as registered portal users who logged into their portal account during the 2-year data collection period. Portal users responded to the survey via a direct link once they signed into their patient portal account. The delivery method for the survey ensured that we, at no time, were able to personally identify any patient or access specific patient details that could be linked to an individual.

We differentiate our research study from previous portal and technology use studies in that we were able to incorporate an objective measure of use. Portal use was measured as the count of portal visits for each respondent. This information was obtained from the Web logs of the patient portal system. Capturing the count of portal visits in this way facilitated the matching of portal visits to survey responses by way of masked identification numbers that were generated by the health center staff. We were able to match survey responses to portal use while simultaneously maintaining the anonymity of the portal users.

The survey items were used to obtain patient perceptions of the factors that influence their usage and their perceptions about the impact that portal usage may have on their HSB and SWP. Patients responded by providing Likert Scale-type responses (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree).

We also captured high-level demographic information about our survey respondents.

Data Analysis

The mean and standard deviation for the survey responses were calculated in R Studio. Subsequently, the survey data were analyzed using partial least squares structural equation modeling (PLS-SEM) in SmartPLS. PLS-SEM is an appropriate technique for testing relationships within the proposed theoretical model because of the use of latent variables [37]. Given that all of the indicators in our model are reflective, we evaluated the reliability of each construct through the use of Cronbach alpha and composite reliability. To assess convergent validity, we evaluated the average variance extracted (AVE). Discriminant validity was assessed to determine significant interconstruct differences [38]. We evaluated the heterotrait-monotrait ratio of correlations in an effort to determine if discriminant validity exists between the reflective constructs in our proposed model.

Results

Demographic Data

Of the 1142 survey respondents, 705 (61.73%) were female and 437 (38.27%) were male (Table 2).

Ethnicity was self-reported as African American, Native American, Asian or Pacific islander, white, any two or more races, international or unspecified. People that identified as white represented the highest number of respondents, 52.36% (598/1142). People that identified as Native American were among the fewest survey respondents. Only 3 people identified as Native American. Academic standing was recorded as freshman, sophomore, junior, senior, fifth year, or graduate students. Freshmen comprised the largest number of survey respondents, 47.46% (542/1142) and fifth year students comprised the smallest number of respondents, 0.79% (9/1142).

Table 2. Survey respondent demographic data (n=1142).

Characteristics	Value, n
Gender	
Male	437
Female	705
Ethnicity	
African American	172
Native American	3
Any two or more races	82
Asian or Pacific islander	93
White	598
International	68
Not specified	126
Academic standing	
Freshman	542
Sophomore	206
Junior	184
Senior	81
Fifth year	9
Graduate	108
Early college	12

Reliability

Mean response scores and reliability analysis are presented in Table 3.

Social influence, effort expectancy, performance expectancy, HSB, and SWP demonstrated Cronbach alphas that were greater than .80. Habit and facilitating conditions both had weaker Cronbach alpha of .726 and .710, respectively. Perceived severity had the lowest Cronbach alpha of .611, and it actually fell below the widely accepted threshold of .70 [39]. Owing to the exploratory nature of our research study, we also reviewed acceptable thresholds for composite reliability. We noted that the composite reliability for perceived severity exceeded 0.60, which has been viewed as acceptable in an exploratory research context [40]. We therefore retained perceived severity in our model analysis.

We found that all of the values for AVE in our analysis (Table 3) were higher than the Fornell and Locker [41] suggested threshold of 0.50 for AVE. This leads us to conclude that there is sufficient convergent validity for each of our model constructs.

Table 3. Mean response scores and reliability analysis.

Theoretical construct and survey item	Value, mean (SD)	Outer loadings	Cronbach alpha	Composite reliability	Average variance extracted
S[a]			.846	0.907	0.766
S1—People who care for me want me to use the portal.	3.91 (0.872)	0.896			
S2—People who influence me want me to use the portal.	3.76 (0.871)	0.916			
S3—My nurse or physician has encouraged me to use the portal.	3.40 (0.910)	0.811			
H[b]			.726	0.879	0.784
H1—Portal usage has become a habit for me.	3.19 (0.916)	0.896			
H2—I must use the portal on a regular basis to improve my health.	3.21 (0.937)	0.875			
FC[c]			.710	0.873	0.775
FC1—I know how to access the portal.	4.26 (0.821)	0.887			
FC2—The portal works with other technology that I use.	4.16 (0.784)	0.874			
EE[d]			.894	0.950	0.904
EE1—Learning to use the portal was easy for me.	4.02 (0.834)	0.948			
EE2—The portal was easy to navigate and use.	4.04 (0.807)	0.953			
PE[e]			.870	0.939	0.885
PE1—Using the portal will support the care I receive.	3.91 (0.769)	0.943			
PE2—Using the portal allows me to be more involved in my own care.	3.95 (0.769)	0.939			
PS[f]			.611	0.835	0.717
PS1—I believe I am vulnerable to illnesses.	2.60 (1.100)	0.886			
PS2—I believe that my current health conditions are serious.	2.07 (1.039)	0.806			
HSB[g]			.910	0.957	0.917
HSB1—Portal usage has influenced me to adopt healthier behaviors.	3.18 (0.840)	0.957			
HSB2—Portal usage has influenced me to exercise more.	3.06 (0.861)	0.959			
SWP[h]			.824	0.919	0.850
SWP1—I am satisfied with the patient portal.	3.92 (0.734)	0.916			
SWP2—I would recommend using the portal to my friends and family.	3.73 (0.795)	0.928			

[a] S: social influence.

[b] H: habit.

[c] FC: facilitating conditions.

[d] EE: effort expectancy.

[e] PE: performance expectancy.

[f] PS: perceived severity.

[g] HSB: health-seeking behavior.

[h] SWP: portal satisfaction.

Discriminant Validity

Discriminant validity analysis is presented in Table 4.

None of the correlation values in Table 4 exceeds the widely accepted threshold of 0.90 for heterotrait-monotrait values [42,43]. Owing to this observation, we assert that the independent variables in our model demonstrate sufficient discriminant validity.

Table 4. Discriminant validity.

Model constructs	Effort expectancy	Facilitating conditions	Habit	Health-seeking behavior	Perceived severity	Performance expectancy	Satisfaction with portal	Social influence	Tenure
Effort expectancy	N/A[a]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Facilitating conditions	0.821	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Habit	0.419	0.443	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health-seeking behavior	0.362	0.305	0.712	N/A	N/A	N/A	N/A	N/A	N/A
Perceived severity	0.049	0.085	0.239	0.202	N/A	N/A	N/A	N/A	N/A
Performance expectancy	0.685	0.672	0.627	0.535	0.155	N/A	N/A	N/A	N/A
Satisfaction with portal	0.775	0.717	0.665	0.592	0.094	0.857	N/A	N/A	N/A
Social influence	0.402	0.452	0.793	0.564	0.167	0.646	0.67	N/A	N/A
Tenure	0.020	0.024	0.090	0.095	0.156	0.018	0.017	0.106	N/A
Use	0.035	0.021	0.141	0.015	0.106	0.019	0.058	0.06	0.568

[a] Not applicable.

Path Analysis

To test our research hypotheses, we used the results from PLS-SEM. In our initial tests, we conducted the model analysis with a bootstrapping sample of 500. In an effort to assess the robustness of our findings, we subsequently ran the model with a bootstrapping sample of 1000. The overall results of our analysis did not materially change between the two model runs with different bootstrapping sample sizes. The significance of established associations remained consistent between the two iterations of our analysis.

Our model explains 29% of the variance observed in HSB (R^2=0.2900) and 65% of the variance observed in SWP (R^2=0.6499). The results from our PLS-SEM (Figure 2) suggest significant associations among social influence, habit, facilitating conditions, effort expectancy, performance expectancy, and SWP. Perceived severity is significantly associated with HSB, but it is not significantly associated with portal use. Portal use, contrary to our initial supposition, is not significantly associated with HSB.

We found support for 8 out of 10 of our hypotheses (Table 5). The strongest association observed in our model appears to be between SWP and HSB, demonstrated by a path coefficient of .505 (P<.001). We also found that tenure is significantly associated with both HSB and portal use. As tenure was treated as a control variable, we did not draw any major conclusions about level of significance.

Figure 2. Empirical model with path coefficients.

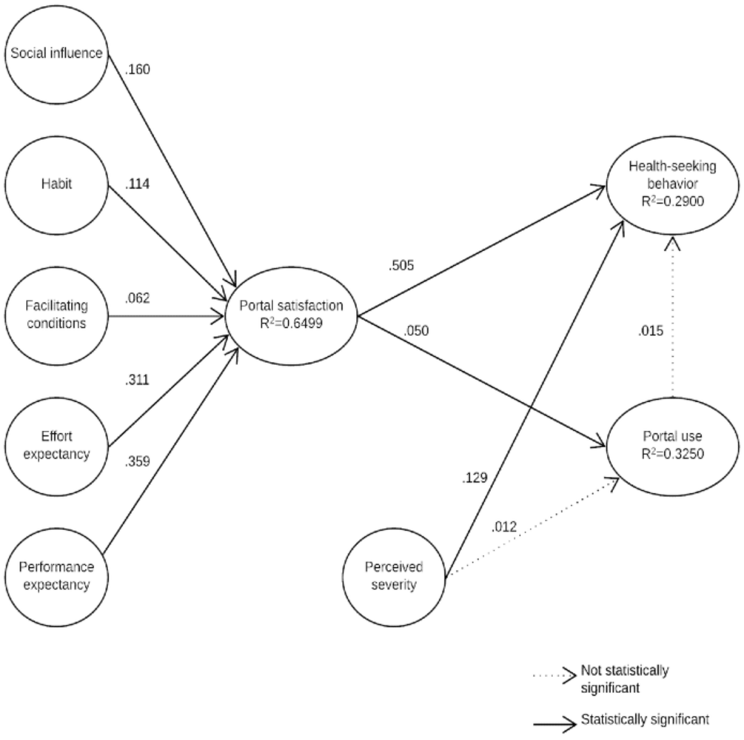


Table 5. Path analysis results: full model.

Dependent and independent variables	Path coefficient (beta)	Standard error	*t* statistic (*df*=499)[a]	Hypothesis	Supported (yes/no)	R^2
Portal satisfaction						0.6499
Social influence	.160	0.025	6.145[b]	H_1	Yes	
Habit	.114	0.023	4.89[b]	H_2	Yes	
Facilitating conditions	.062	0.026	2.401[c]	H_3	Yes	
Effort expectancy	.311	0.028	11.149[b]	H_4	Yes	
Performance expectancy	.359	0.031	11.588[b]	H_5	Yes	
Health-seeking behavior						0.2900
Portal satisfaction	.505	0.026	19.705[b]	H_{6a}	Yes	
Use	.015	0.028	.513	H_7	No	
Perceived severity	.129	0.028	4.675[b]	H_{8a}	Yes	
Tenure	−.116	0.030	3.925[b]	Control	N/A[d]	
Portal use						0.3250
Portal satisfaction	.050	0.025	2.031[c]	H_{6b}	Yes	
Perceived severity	.012	0.025	.488	H_{8b}	No	
Tenure	.566	0.039	14.409[b]	Control	N/A	

[a] 2-tailed *t* test.

[b] *P*<.1.

[c] *P*<.05.

[d] Not applicable.

Discussion

Principal Findings

RQ1. What Factors Are Associated With a Patient's Portal Satisfaction?

We find that SWP is driven most strongly by performance expectancy. If we rely on Venkatesh et al's [14] definition of performance expectancy, wherein performance expectancy is understood to represent the degree to which users expect systems to help them attain a specific performance, we note that portal user satisfaction appears to be driven mostly by what portal users think they will get out of using the system. Our research supports social influence, habit, facilitating conditions, and effort expectancy as factors that are associated with SWP.

RQ2. Does Patient Portal Satisfaction Influence Health-Seeking Behavior?

Our analysis suggests that HSB is significantly associated with SWP. The association between HSB and SWP was the highest association that our model demonstrated between any of the model constructs with a path coefficient of .505.

RQ3. Does Patient Portal Use Influence Health-Seeking Behavior?

Our empirical analysis does not support a relationship between portal use and HSB. We had expected to find a positive influence of portal usage on HSB. Intuitively, as patients more frequently use portals, one would expect to see a strong, positive correlation between this portal use and HSB. Our analysis, however, demonstrates no significant link between portal use and HSB.

RQ4. Does Patient Portal Satisfaction Influence Portal Use?

Patient SWP does appear to influence portal use. We observed a significant positive correlation between SWP and portal use. Although this correlation was not as strong as we had anticipated (beta=.050), the correlation between the two constructs was statistically significant (*P*=.04). This result does add to the existing ECT literature in that it supports the idea that use (or continued use in ECT parlance) can be influenced by end user satisfaction. Our study supports the work of Thong et al [28] in that the relationship between satisfaction and continued use suggests a need for IT developers to focus on developing easy-to-use technology features as a method for ensuring satisfaction and, thereby, driving continued technology use.

RQ5. Does Perceived Severity of Health Condition Influence Health-Seeking Behavior?

Perceived severity of health condition has a weak, positive association with HSB. Intuitively, one would expect to find that the more severe a patient's condition, the more frequently the patient would exhibit HSB. Although we found only a weak, positive association between perceived severity and HSB, the association between the two is statistically significant. This association would have been missed had we eliminated it from our study by solely relying on the Cronbach alpha of .611 (Table 3).

RQ6. Does Perceived Severity of Health Condition Influence Portal Use?

Perceived severity of health condition does not have a significant association with portal use. This finding was counterintuitive in that we expected perceived severity to play a strong role in driving portal use. We expected to find that survey respondents who perceived themselves to have more severe health conditions would report higher levels of portal use. This, however, was not the case.

Limitations

The authors acknowledge geographic limitations to this study. Data were collected and analyzed from a single academic institution. It is possible that there are unique characteristics of this college student population that may prevent the generalizability of our findings to other populations of portal users. These characteristics might include age, education level, technology literacy, and easy access to technology. Future research should include data from the general population for the purposes of comparing, contrasting, and possibly strengthening the generalizability of the findings. The data used in this study were collected using a snapshot of observations within a 2-year time frame. Hence, the study suffers from the general limitations of using snapshot data [44]. Future longitudinal studies can address this limitation and provide greater insight into portal user perceptions over time. Longitudinal studies generally provide more statistically powerful tests [45].

Conclusions

Our research benefited from a sample size of 1142 portal user survey responses. Also, unlike many previous studies, our research benefited from objective data on actual portal use. We tested the overall reliability of the responses, and we found support for existing theories of how concepts from TAMs can be used to explain associations between those widely accepted concepts, and SWP might serve as a proxy for technology satisfaction in future research. We also found utility in these widely accepted technology acceptance constructs as a means of explaining some aspects of HSB.

Our findings, interestingly, showed a minor significant role for perceived severity as it relates to perceived HSB. Although the effects were small, they were significant. This relationship may spark future research interests into the role that perceived severity may play in understanding perceived HSB.

We also determined that there was a link between SWP and HSB. This link may also spark future research prompted by the desire to better understand whether SWP is likely to be a determinant of HSB or health information seeking.

Acknowledgments

The authors would like to extend their gratitude to the staff of the university health system for their support in collecting the survey data.

Conflicts of Interest

None declared.

References

1. Neuner J, Fedders M, Caravella M, Bradford L, Schapira M. Meaningful use and the patient portal: patient enrollment, use, and satisfaction with patient portals at a later-adopting center. Am J Med Qual 2015;30(2):105-113 [FREE Full text] [doi: 10.1177/1062860614523488] [Medline: 24563085]
2. Peacock S, Reddy A, Leveille SG, Walker J, Payne TH, Oster NV, et al. Patient portals and personal health information online: perception, access, and use by US adults. J Am Med Inform Assoc 2017 Apr 1;24(e1):e173-e177. [doi: 10.1093/jamia/ocw095] [Medline: 27413120]
3. Carrell D, Ralston JD. Variation in adoption rates of a patient web portal with a shared medical record by age, gender, and morbidity level. AMIA Annu Symp Proc 2006;2006:871 [FREE Full text] [Medline: 17238491]
4. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
5. Zickmund SL, Hess R, Bryce CL, McTigue K, Olshansky E, Fitzgerald K, et al. Interest in the use of computerized patient portals: role of the provider-patient relationship. J Gen Intern Med 2008 Jan;23(Suppl 1):20-26 [FREE Full text] [doi: 10.1007/s11606-007-0273-6] [Medline: 18095039]
6. Irizarry T, Dabbs AD, 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]
7. Coughlin SS, Stewart JL, Young L, Heboyan V, de Leo G. Health literacy and patient web portals. Int J Med Inform 2018 May;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]
8. Davis SE, Osborn CY, Kripalani S, Goggins KM, Jackson GP. Health literacy, education levels, and patient portal usage during hospitalizations. AMIA Annu Symp Proc 2015;2015:1871-1880 [FREE Full text] [Medline: 26958286]
9. Robinson JR, Davis SE, Cronin RM, Jackson GP. Use of a patient portal during hospital admissions to surgical services. AMIA Annu Symp Proc 2016;2016:1967-1976 [FREE Full text] [Medline: 28269956]
10. Fraccaro P, Vigo M, Balatsoukas P, Buchan IE, Peek N, van der Veer SN. Patient portal adoption rates: a systematic literature review and meta-analysis. Stud Health Technol Inform 2017;245:79-83. [Medline: 29295056]
11. Marton C, Choo CW. A review of theoretical models of health information seeking on the web. J Documentation 2012;68(3):330-352. [doi: 10.1108/00220411211225575]
12. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]
13. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):169-332. [doi: 10.1287/mnsc.46.2.186.11926]
14. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
15. Kirch W. Health belief model. In: Encyclopedia of Public Health. Netherlands: Springer; 2008.
16. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 2;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
17. Deng L, Turner DE, Gehling R, Prince B. User experience, satisfaction, and continual usage intention of IT. Eur J Inf Syst 2010;19(1):60-75. [doi: 10.1057/ejis.2009.50]
18. Ghobakhloo M, Zulkifli NB, Aziz FA. The interactive model of user information technology acceptance and satisfaction in small and medium-sized enterprises. Eur J Econ Finance Adm Sci 2010(19):7-25 [FREE Full text]
19. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. Stat Med 2013 Feb 20;32(4):556-577 [FREE Full text] [doi: 10.1002/sim.5408] [Medline: 22711416]
20. Heath M, Porter TH. Patient health records: an exploratory study of patient satisfaction. Health Policy Technol 2017;6(4):401-409. [doi: 10.1016/j.hlpt.2017.10.002]
21. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001;25(3):351-370. [doi: 10.2307/3250921]
22. Venkatesh V, Thong J, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
23. Limayem M, Hirt S, Cheung C. How habit limits the predictive power of intention: the case of information systems continuance. MIS Q 2007;31(4):705-737. [doi: 10.2307/25148817]

24. Amoroso D, Lim R. The mediating effects of habit on continuance intention. Int J Inf Manag 2017;37(6):693-702. [doi: 10.1016/j.ijinfomgt.2017.05.003]
25. Maillet E, Mathieu L, Sicotte C. Modeling factors explaining the acceptance, actual use and satisfaction of nurses using an Electronic Patient Record in acute care settings: an extension of the UTAUT. Int J Med Inform 2015 Jan;84(1):36-47. [doi: 10.1016/j.ijmedinf.2014.09.004] [Medline: 25288192]
26. 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]
27. Arsand E, Demiris G. User-centered methods for designing patient-centric self-help tools. Inform Health Soc Care 2008 Sep;33(3):158-169. [doi: 10.1080/17538150802457562] [Medline: 18850399]
28. Thong JY, Hong S, Tam KY. The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. Int J Hum-Comput Stud 2006;64(9):799-810. [doi: 10.1016/j.ijhcs.2006.05.001]
29. Oghuma AP, Libaque-Saenz CF, Wong SF, Chang Y. An expectation-confirmation model of continuance intention to use mobile instant messaging. Telemat Inform 2016;33(1):34-47. [doi: 10.1016/j.tele.2015.05.006]
30. 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]
31. Shaw RJ, Ferranti J. Patient-provider internet portals--patient outcomes and use. Comput Inform Nurs 2011 Dec;29(12):714-8; quiz 719. [doi: 10.1097/NCN.0b013e318224b597] [Medline: 21697705]
32. Wixom BH, Todd PA. A theoretical integration of user satisfaction and technology acceptance. Inf Syst Res 2005;16(1):85-102. [doi: 10.1287/isre.1050.0042]
33. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 1;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
34. Tustin N. The role of patient satisfaction in online health information seeking. J Health Commun 2010 Jan;15(1):3-17. [doi: 10.1080/10810730903465491] [Medline: 20390974]
35. Ahadzadeh AS, Sharif SP, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
36. Hsu CL, Lin JC. Acceptance of blog usage: the roles of technology acceptance, social influence and knowledge sharing motivation. Inf Manag 2008;45(1):65-74. [doi: 10.1016/j.im.2007.11.001]
37. Ringle CM, Sarstedt M, Straub D. Editor's comments: a critical look at the use of PLS-SEM in 'MIS Quarterly'. MIS Q 2012;36(1):iii-xiv. [doi: 10.2307/41410402]
38. Chin WW. The partial least squares approach for structural equation modeling. In: Modern Methods for Business Research. New York: Lawrence Erlbaum Associates Publishers; 1998:295-336.
39. Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw-Hill; 1978.
40. Revelle W. Hierarchical cluster analysis and the internal structure of tests. Multivariate Behav Res 1979 Jan 1;14(1):57-74. [doi: 10.1207/s15327906mbr1401_4] [Medline: 26766619]
41. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981;18(1):39-50. [doi: 10.1177/002224378101800104]
42. Hair J, Hult TM, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Second Edition. Thousand Oaks, CA: SAGE Publications; 2016.
43. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2015;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
44. Kaplan EH. Snapshot samples. Socioecon Plann Sci 1997;31(4):281-291. [doi: 10.1016/s0038-0121(97)00011-6]
45. Hedeker D, Gibbons RD. Longitudinal Data Analysis. Hoboken: Wiley & Sons; 2006.

Abbreviations

AVE: average variance extracted
ECT: Expectation-Confirmation Theory
EHR: electronic health record
HBM: Health Belief Model
HSB: health-seeking behavior
IS: information system
IT: information technology
PLS-SEM: partial least squares structural equation modeling
RQ: research question
SWP: portal satisfaction
TAM: Technology Acceptance Model
UTAUT2: Unified Theory of the Acceptance and Use of Technology 2

Edited by G Eysenbach; submitted 13.09.19; peer-reviewed by J Tavares, F Seidl, T Irizarry; comments to author 06.11.19; revised version received 15.11.19; accepted 16.12.19; published 27.03.20.

Please cite as:
Silver RA, Subramaniam C, Stylianou A
The Impact of Portal Satisfaction on Portal Use and Health-Seeking Behavior: Structural Equation Analysis
J Med Internet Res 2020;22(3):e16260
URL: http://www.jmir.org/2020/3/e16260/
doi:10.2196/16260
PMID:32217505

Original Paper

The Ecofisio Mobile App for Assessment and Diagnosis Using Ultrasound Imaging for Undergraduate Health Science Students: Multicenter Randomized Controlled Trial

Mario Lozano-Lozano[1,2,3], PhD; Noelia Galiano-Castillo[1,2,3], PhD; Carolina Fernández-Lao[1,2,3], PhD; Paula Postigo-Martin[1,2], MSc; Francisco Álvarez-Salvago[1], PhD; Manuel Arroyo-Morales[1,2,3], MD, PhD; Irene Cantarero-Villanueva[1,2,3], PhD

[1]Department of Physical Therapy, University of Granada, Granada, Spain
[2]Sport and Health Joint University Institute, Granada, Spain
[3]Biohealth Research Institute in Granada, Granada, Spain

Corresponding Author:
Noelia Galiano-Castillo, PhD
Department of Physical Therapy
University of Granada
Avda de la Ilustración 60
Granada
Spain
Phone: 34 958248769
Email: noeliagaliano@ugr.es

Abstract

Background: Generation Z is starting to reach college age. They have adopted technology from an early age and have a deep dependence on it; therefore, they have become more drawn to the virtual world. M-learning has experienced huge growth in recent years, both in the medical context and in medical and health sciences education. Ultrasound imaging is an important diagnosis technique in physiotherapy, especially in sports pathology. M-learning systems could be useful tools for improving the comprehension of ultrasound concepts and the acquisition of professional competencies.

Objective: The purpose of this study was to evaluate the efficacy and use of an interactive platform accessible through mobile devices—Ecofisio—using ultrasound imaging for the development of professional competencies in the evaluation and diagnosis of sports pathologies.

Methods: Participants included 110 undergraduate students who were placed into one of two groups of a randomized controlled multicenter study: control group (ie, traditional learning) and experimental group (ie, Ecofisio mobile app). Participants' theoretical knowledge was assessed using a multiple-choice questionnaire (MCQ); students were also assessed by means of the Objective Structured Clinical Examination (OSCE). Moreover, a satisfaction survey was completed by the students.

Results: The statistical analyses revealed that Ecofisio was effective in most of the processes evaluated when compared with the traditional learning method: all OSCE stations, *P*<.001; MCQ, 43 versus 15 students passed in the Ecofisio and control groups, respectively, *P*<.001. Moreover, the results revealed that the students found the app to be attractive and useful.

Conclusions: The Ecofisio mobile app may be an effective way for physiotherapy students to obtain adequate professional competencies regarding evaluation and diagnosis of sports pathologies.

Trial Registration: ClinicalTrials.gov NCT04138511; https://clinicaltrials.gov/ct2/show/NCT04138511

(J Med Internet Res 2020;22(3):e16258) doi:10.2196/16258

KEYWORDS

undergraduate; OSCE; mHealth; teaching and learning strategies

Introduction

Generational evolution is nothing new; constant social change poses a challenge for education. During the last several years, concepts referring to subsequent generational learning styles, such as those of Generation X, Generation Y (ie, millennials), or even Generation Z, have been emerging as new ways of understanding new generations [1]. Although most current undergraduate students are within the so-called millennial generation, the next wave of students (ie, Generation Z) is beginning to reach college age [2]. Both groups share many similarities, but the latter has unique characteristics that require new teaching strategies [3]. *Generation Z* is the term used to refer to those born between the mid-90s and the 2000s [4]. This generation has adopted technology from a very early age, creating a deep dependence on it and becoming more drawn to the virtual world [5].

Some studies have shown how these generations absorb and process information in very different ways compared to previous generations [6-8]. These students are characterized as being digital natives and active learners who constantly search for innovation and technology, and who prefer teams, groups, and multitasking [6]. Therefore, they have grown up with several tools that make their learning brain structures different [9]. In this sense, the use of traditional teaching tools is decontextualized [10,11].

In this new teaching-and-learning era, the adaptation of methodologies to new technologies is completely necessary to improve social learning. Educational activity has moved from distance learning to e-learning and, finally, to mobile learning (m-learning), as knowledge has increased exponentially and the demand has escalated [12]. According to Colazzo et al [13], “an educational process can be considered as any learning and teaching activity that is possible through mobile tools or in settings where mobile equipment is available.” M-learning can be understood the same way.

M-learning is surpassing its predecessor (ie, e-learning), since it takes advantage of all its characteristics and overcomes all its disadvantages, making the content much more accessible in time and space [14]. It has become a reality, as proven by many validation studies conducted all over the world [15,16]; it even affects cognitive processes, such as memory. We are becoming symbiotic with our electronic tools, living in an increasingly interconnected society [17], where mobile apps attempt to overcome the problems currently posed by the learning system: the lack of personalized content and appeal to different learning styles, the inability of teachers to apply their true strengths, and the lack of effective reforms at a reasonable cost [18].

There has been an incredible expansion of mobile phone apps in both the medical context [19] and medical and health sciences education. Briz-Ponce et al suggest that more than 37% of health sciences students have used a medical app to improve their learning skills [20]. However, despite the increasing use of m-learning, Po Lau et al highlight the lack of knowledge of m-learning services by public institutions [4]. Even though mobile technology is very well received and health sciences students are well disposed toward its use, its implementation among students has been unsuccessful; this is due to a lack of compliance by the educational systems because m-learning has not been integrated within the standard teaching process [21]. Consequently, as these authors indicate, it is necessary to conduct larger studies to raise awareness of these programs.

There is an increasing number of studies on the use of m-learning in health sciences, such as in surgical skills [15], medicine [21], and nursing [22]. A systematic review in physiotherapy specifically evaluated the effectiveness and user perceptions of online technology for physiotherapy teaching and learning and highlighted its benefits. However, this review did not include any studies that used m-learning as the educational process [23]. Therefore, it is of crucial importance to include this technology as a complement in physical therapy studies.

Our research team has previously conducted some e-learning [24,25] and m-learning studies with physiotherapy students. Our most recent study has shown that palpation skills and ultrasound techniques could be better acquired with a mobile app than by means of traditional learning [26]. In this sense, due to the importance of ultrasound as a diagnostic technique in physiotherapy, especially in sports pathology, m-learning seems to be a useful tool for its study, since it provides the necessary tools for the comprehension of concepts. To our knowledge, there are no studies that propose an m-learning environment to improve the acquisition of professional competencies of sports pathology assessment. Thus, the aim of this study was to evaluate the efficacy and use of an interactive platform accessible through mobile devices using ultrasound imaging for the development of professional competencies in the evaluation and diagnosis of sports pathologies.

Methods

Recruitment and Participants

A total of 110 undergraduate physical therapy students took part in this study; students were enrolled at the Health Science Faculty at the University of Granada, Spain, or at the Campus of Tudela, University of Navarra, Spain. These students were in their first or second year of Physical Therapy Fundamentals and Physical Therapy Assessments courses or in their second year of the Physiology of Effort, Physical Exercise, and Health course during the 2014-2015 academic year. Participants’ theoretical knowledge was assessed and they were also assessed by means of the Objective Structured Clinical Examination (OSCE). The participants were recruited via a public announcement at both universities. All students signed an informed consent form to participate in this study.

Eligibility Criteria

To be eligible for this trial, students needed to meet the following criteria: not having received any previous training in ultrasound imaging and management, having anatomy and biomechanics knowledge, being enrolled in any required subjects, having the basic ability to use a mobile app, and having a mobile phone with Internet access running on the Android operating system.

Study Design

We conducted a multicenter study, which consisted of a double-blinded, randomized controlled study involving volunteer students earning degrees in physical therapy from two public Spanish universities. The study was registered at ClinicalTrials.gov (NCT04138511). We used a computer program to randomly assign a number to each student in both universities. These numbers were provided in numbered, opaque envelopes by an external study member to ensure that the study member responsible for outcome assessments was blinded to intervention and control group assignment. The envelopes were opened after a baseline assessment. The assessment evaluator was blinded to both the assessment and the outcomes. There were 55 participants in the experimental group and 55 in the control group. The students were also blinded to the group to which they were assigned.

After students received theoretical and practical lessons about ultrasound skills in sports pathology areas, the role that the Ecofisio mobile app played was assessed in a traditional way and by means of the OSCE. The OSCE allows us to measure ultrasound competence as expressed in specific observable behaviors. This study was approved by the *Unidad de Calidad, Innovación y Prospectiva*—a quality and innovation body—from the University of Granada, Spain (PID 14-56), and was conducted in accordance with the Declaration of Helsinki [27,28].

The theoretical and practical lessons were conducted by three professors and three teaching fellows; the teacher-to-student ratio was 1:6-8. The same lessons were taught in both universities by the same professors.

There were a total of six contact sessions on site; this was followed by self-study time focused on the theoretical and practical learning of ultrasound imaging procedures, on different areas of the body, regarding sports pathologies. Each study group—Ecofisio intervention group and control group—attended 2 hours of theoretical lessons and 4 hours of practical lessons. The ultrasound imaging sessions were developed following a previously reported methodology [29] and the diagnosis of sports injuries. All participants used the same ultrasound device model—MyLab 25 (Esaote Medical Systems)—with a 12 MHz linear probe. Finally, there was a self-study period, where the Ecofisio group used the mobile app and the control group used traditional study models (eg, books and journal papers).

Both groups had 2 weeks to study after the on-site lessons. To prevent the control group from accessing information related to the mobile app, they were assessed before the Ecofisio group. The Ecofisio mobile app is focused on sports pathologies and contains relevant written and digital information about ultrasound imaging and management, as well as diagnosis of sport injuries. For each anatomical structure, there exists a theoretical description, a drawing with the anatomical description, an image with the specific placement of the ultrasound probe, an ultrasound slice, a diagram of the ultrasound image, and a video of the manual palpation procedure.

Outcome Measures

All variables were measured among both study groups after their self-study periods.

Objective Structured Clinical Examination

An OSCE [30,31] was used to measure participants' hands-on skills in terms of ultrasound management. It is a competency-based evaluation [32] consisting of one station with five specific components: positioning of the patient, positioning of the ultrasound probe, orientation (ie, angle) of the ultrasound probe, handling of the ultrasound probe, and image adjustment. Before starting each component, the participants received accurate information about the task to perform. The examiner used a 5-point Likert scale, ranging from 0 to 4, to grade the specific efficiency of each component. The examiner also registered the required time (seconds) to identify the lesion. During each component, no comments were allowed between the students, the examiners, and the patients. The blinded examiner was an expert in ultrasound management with at least 5 years of experience, and the patient suffered a lesion that students had to locate. There were different clinical cases about different sports injuries.

Multiple-Choice Questionnaire

An evaluation of students' theoretical knowledge was conducted using a multiple-choice questionnaire of 20 questions, where a maximum score of 10 points could be obtained.

Satisfaction Survey

Satisfaction scores related to the interventions were assessed among both groups through a specific 5-point Likert questionnaire, ranging from 1 (*disagree*) to 5 (*strongly agree*), which had been used in similar contexts previously [24-26]. Furthermore, the Ecofisio group completed another satisfaction questionnaire about the usage of the mobile app, ranging from 0 (*totally unsatisfied*) to 10 (*totally satisfied*).

Sample Size Calculation

It would require a power of 90% to detect a significant mean difference of 3.5 (SD 3) points in the palpation assessment via the OSCE, assuming a type 1 error (alpha) of 5% and a type 2 error (beta) of 10%. Considering a dropout rate of 20%, we decided to enroll 55 subjects per group. Before the on-campus sessions, an independent researcher randomly assigned each participant to a group using Epidat 3.1 software (Xeral de Saúde Pública).

Statistical Analysis

Descriptive analyses were obtained for all randomized participants. The Kolmogorov-Smirnov test was applied to test the hypothesis of normality for all variables. We calculated Student *t* tests for variables with normal distribution and Mann-Whitney U tests for the rest as nonparametric tests. The statistical significance was determined by $P<.05$, and all analyses were conducted using IBM SPSS Statistics for Windows, version 22.0 (IBM Corp).

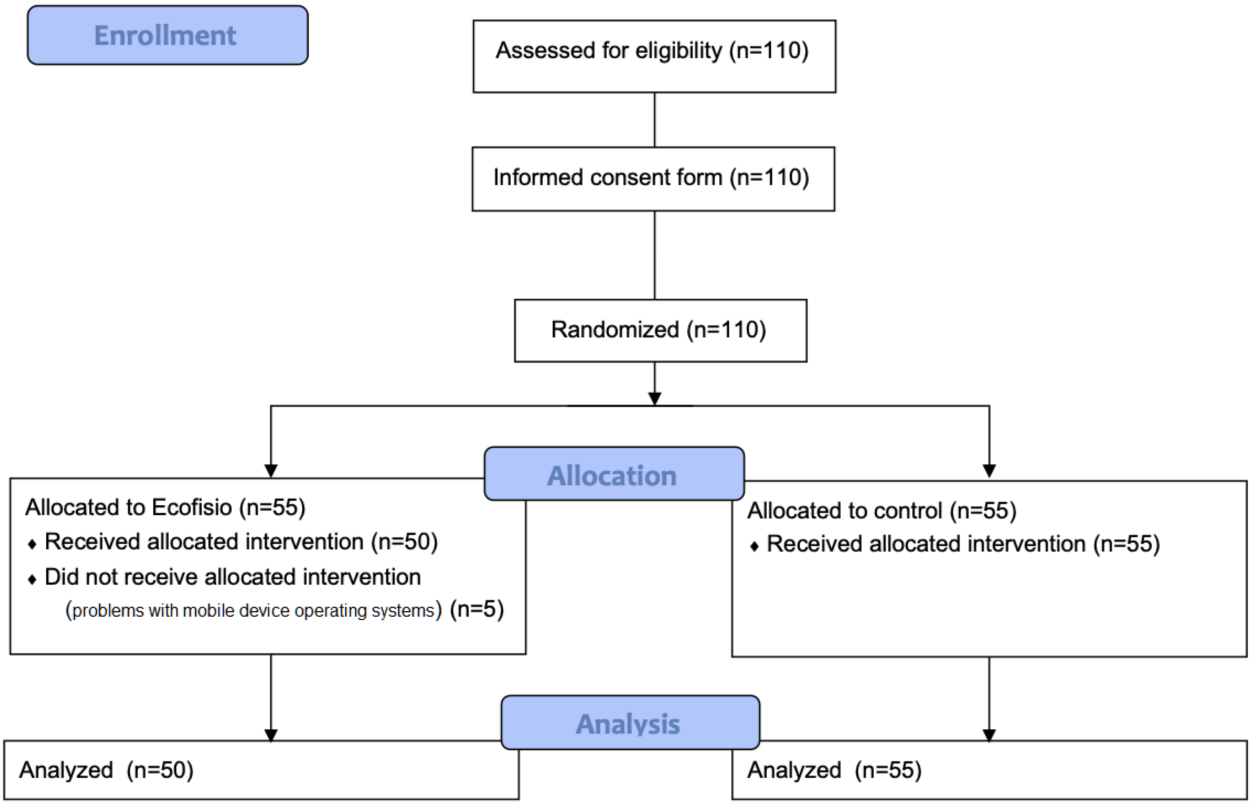
Results

The final sample was comprised of 50 students in the Ecofisio group and 55 students in the control group, with mean ages of 19.8 years (SD 2.1) and 19.7 years (SD 6.0), respectively. At the beginning of the study, there were no group differences in study variables such as gender or previous experience (see Table 1). There were 5 dropouts (5/55, 9%) in the Ecofisio group due to problems with the operating system versions of their mobile devices (see Figure 1).

Table 1. Characteristics of groups.

Characteristic	Ecofisio group (N=50)	Control group (N=55)
Age (years), mean (SD)	19.8 (2.1)	19.7 (6.0)
Gender, n (%)		
Female	30 (60)	32 (58)
Male	20 (40)	23 (42)
Experience with ultrasound, n (%)		
Yes	0 (0)	0 (0)
No	50 (100)	55 (100)
Experience with sports injuries diagnosis, n (%)		
Yes	0 (0)	0 (0)
No	50 (100)	55 (100)
Experience with sports injuries diagnosis with ultrasound, n (%)		
Yes	1 (2)	4 (7)
No	49 (98)	51 (93)

Figure 1. Consolidated Standards Of Reporting Trials (CONSORT) flow diagram of the study.



We found a significant difference in the theoretical exam score (*P*<.001) in favor of the Ecofisio group: mean 7.3 (SD 1.5) points. The results also indicated significant differences in all components assessed during the OSCE stations (*P*<.001 for all) (see Table 2). The Ecofisio group significantly improved in the positioning of the patient (mean score 3.9 [SD 0.3]) and ultrasound probe (mean score 3.9 [SD 0.3]) compared with the control group (mean score 2.3 [SD 1.4] and 1.6 [SD 1.5], respectively). The Ecofisio group also reported higher marks than the control group in the following tasks: orientation of the ultrasound probe (mean score 3.8 [SD 0.6] vs 1.8 [SD 1.7]), handling of the ultrasound probe (mean score 3.3 [SD 0.8] vs 0.7 [SD 0.9]), and image adjustment (mean score 3.0 [SD 1.4] vs 1.6 [SD 1.5]). Finally, the Ecofisio group took significantly more time (*P*<.001) to identify the lesion (mean score 60.2 [SD 25.9]) than did the control group (mean score 45.3 [SD 18.3]). We found significant differences between groups in terms of passing the exam in favor of the Ecofisio group (43/50, 86%, vs 15/55, 27%, *P*<.001) (see Table 2).

Regarding the satisfaction results, we only found significant differences between groups for two items. The Ecofisio group showed a higher satisfaction level than the control group for the following items: *The teacher was competent* (mean score 4.8 [SD 0.4] vs 4.5 [SD 0.5], *P*=.009) and *I believe that the training is applicable* (mean score 3.3 [SD 0.8] vs 3.0 [SD 0.9], *P*=.04) (see Table 3). For the rest of the items, there were no significant differences between groups: *The subject was interesting* (*P*=.08) and *I am satisfied with the training* (*P*=.98) (see Table 4).

Table 2. Differences between groups in data obtained from the Objective Structured Clinical Examination (OSCE) stations 1 and 2.

Station, specific components	Ecofisio group (N=50)	Control group (N=55)	*P* value
1, mean (SD)			
Total theoretical knowledge (maximum 10 points)	7.3 (1.5)	6.3 (1.5)	<.001[a]
2, mean (SD)			
Positioning of patient (maximum 4 points)	3.9 (0.3)	2.3 (1.4)	<.001[a]
Positioning of ultrasound probe (maximum 4 points)	3.9 (0.3)	1.6 (1.5)	<.001[a]
Orientation (ie, angle) of ultrasound probe (maximum 4 points)	3.8 (0.6)	1.8 (1.7)	<.001[a]
Handling of ultrasound probe (maximum 4 points)	3.3 (0.8)	0.7 (0.9)	<.001[a]
Image adjustment (maximum 4 points)	3.0 (1.4)	1.6 (1.5)	<.001[a]
Time to identify the lesion (seconds)	60.2 (25.9)	45.3 (18.3)	<.001[b]
2, total, n (%)			
Suitable	43 (86)	15 (27)	<.001[c]
Not suitable	12 (14)	35 (73)	<.001[c]

[a] Mann-Whitney U test.

[b] One-way analysis of variance (ANOVA).

[c] Chi-square test.

Table 3. Differences between groups regarding training satisfaction.

Item	Ecofisio group (N=50), mean (SD)	Control group (N=55), mean (SD)	*P* value
The teacher was competent (maximum 5 points)	4.8 (0.4)	4.5 (0.5)	.009[a]
The subject was interesting (maximum 5 points)	4.3 (0.5)	4.4 (0.6)	.08[a]
I am satisfied with the training (maximum 5 points)	3.6 (0.7)	3.5 (0.7)	.98
I believe that the training is applicable (maximum 5 points)	3.3 (0.8)	3.0 (0.9)	.04[a]

[a] Mann-Whitney U test.

Table 4. Descriptive data related to the Ecofisio group's satisfaction.

Item	Ecofisio group (N=50), n (%)
Understandable information	
Very clear	40 (80)
Not very clear	10 (20)
Not clear	N/A[a]
Scientific quality of information (good quality)	
Yes	46 (92)
Not quite	4 (8)
No	N/A
Clear presentation of contents	
Yes	37 (74)
In some cases	13 (26)
No	N/A
Overall impression of the app	
Very good	7 (14)
Good	43 (86)
Not very good	N/A
Bad	N/A
Evaluation of the app from 0 to 10	
10	1 (2)
9	11 (22)
8	28 (56)
7	10 (20)
6	N/A
5	N/A
4	N/A
3	N/A
2	N/A
1	N/A
0	N/A

[a] Not applicable.

Discussion

Principal Findings

This work presents the results obtained on the use and efficacy of an interactive platform accessible through mobile devices using ultrasound imaging for the development of professional competencies in the evaluation and diagnosis of sports injuries. The mobile app has demonstrated its effectiveness in the majority of the processes evaluated when compared with the traditional learning method. Moreover, the results revealed that the app is attractive and useful from the point of view of the students.

The findings presented are in line with previous work developed by our research group, where different technological tools were used to acquire professional competencies among physiotherapy students [24-26]. This study improves and adds importance to the ultrasound imaging competencies for the assessment and follow-up in sports injuries, which has great impact on the development of the professional activity of physiotherapists. In the case of the use of a mobile app, this study resulted in an improvement in all the components of the ultrasound assessment, being broader than those presented previously by Fernández-Lao et al [26]; that study, which also used a mobile app, showed no differences between groups in the items *orientation of the ultrasound probe* and *image adjustment*. This may be due to the fact that students had more motivation to correctly diagnose the sports injuries than to improve on other settings presented in the previous work. On the other hand, in our opinion, more experienced teachers could have also had a positive influence on the way information was provided to students.

A surprising result was that there were significant differences between groups in the time required to identify the lesion. The identification time was almost 20 seconds lower for the control group when compared with the Ecofisio group. This finding was also described in our previous work [26]. We hypothesized that the students in the m-learning group took longer to reach the final result because they dedicated more time to finding a clearer image of the lesion. It is also possible that the control group had finished before the m-learning group because they did not find a clear image of the lesion [33]. Nevertheless, decreasing the time to identify the lesion was not an objective of our study and, therefore, the students were not encouraged to complete the assessment in a short time.

According to the results of the postprogram survey on training satisfaction, nearly all the scores were better in the Ecofisio group than in the control group. We think that the students in the group with access to the app felt a greater motivation for the project in general, so they felt more satisfied with the experience; this fact was also stated in all our previously published works [24-26]. This is in accordance with results from Hill et al, where the students preferred new technologies over traditional learning processes [6]. Another previous study developed in the area of physiotherapy, which used a 3D mobile app, also showed a very good level of acceptance among students of anatomy and manual therapy [34]. In our study, the only item in which both groups presented very similar results was *I am satisfied with the training*. It is possible that all the students who were enrolled in the project were motivated because the project was conducted during the course and because of their very active participation in any activity that was proposed. Furthermore, their regular teacher taught the lessons, in which students were encouraged to consider the mobile app as a study complement.

At the end of the project, the students in the Ecofisio group were asked about their level of satisfaction with the app; the results were quite good for all of the items evaluated. A total of 80% of the students gave scores of 8 or more on a 0-10-point rating scale about their overall satisfaction level with the app, similar to our previous studies [24-26]. Fernández-Lao et al [26] reported an average score of 8.2 on the global satisfaction level with their mobile app as a learning complement of palpation and ultrasound imaging skills of the shoulder. This fact demonstrates that, regardless of the subject studied with mobile apps, these tools are useful for students and their reported acceptance levels are very good. This fact highlights the interest and familiarization of *Generation Z* students with new technologies [35,36].

Limitations

Finally, we have to recognize some limitations of this research. Our study results with this mobile app should be supported in different countries and with different languages in order to extrapolate the results to other populations. On the other hand, we are aware of the fact that conducting the project during the course sessions may have prompted extra motivation among the students and that ultrasound imaging was an attractive tool for them to use; therefore, the results were quite good for both study groups. In this sense, it would be interesting to develop an independent study with the mobile app in other settings. Finally, there was no follow-up period with which to evaluate the sustainability of the benefits of this interactive platform. Nevertheless, this study also presents a number of strengths. This project was carried out in two different universities in Spain with an important sample of students; therefore, the results could be generalized to physiotherapy students. To our knowledge, this is one of the few studies that has been developed in the area of physiotherapy, showing the use of new technologies to complement traditional learning. In fact, we believe this is the first m-learning system designed to improve the acquisition of professional competencies in the assessment of sports pathologies. In addition, this m-learning system could be implemented in different areas of knowledge, such as sports medicine or sports sciences.

Conclusions

Ecofisio is an interactive platform accessible through mobile devices, which uses ultrasound imaging. It is an effective way to develop professional competencies in the evaluation and diagnosis of sports pathologies in physiotherapy students, in a way that students find satisfying. Similar experiences could be implemented in different areas of knowledge with good results.

Acknowledgments

We are grateful to Ms Carmen Sainz-Quinn for her assistance with the English language. The authors are also grateful for the collaboration of the students who participated in the study. This study was funded by a grant from the Educational Innovation Unit of the University of Granada, Spain (PID 14-56). This study was conducted thanks to the additional funding from 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
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 348 KB - jmir_v22i3e16258_app1.pdf]

References

1. Jones VR. Essentials for engaged 21st-century students. Tech 2012 Oct;87(7):16-19 ISSN: ISSN-1527-1803.
2. Eckleberry-Hunt J, Lick D, Hunt R. Is medical education ready for Generation Z? J Grad Med Educ 2018 Aug 02;10(4):378-381. [doi: 10.4300/jgme-d-18-00466.1]
3. Shatto B, Erwin K. Teaching Millennials and Generation Z: Bridging the generational divide. Creat Nurs 2017 Feb 01;23(1):24-28. [doi: 10.1891/1078-4535.23.1.24] [Medline: 28196564]
4. Lau K, Chiu D, Ho K, Lo P, See-To EW. Educational usage of mobile devices: Differences between postgraduate and undergraduate students. J Acad Librariansh 2017 May;43(3):201-208. [doi: 10.1016/j.acalib.2017.03.004]
5. Pérez-Escoda A, Castro-Zubizarreta A, Fandos-Igado M. Digital skills in the Z Generation: Key questions for a curricular introduction in primary school. Comunicar: Revista Científica de Comunicación y Educación 2016 Oct 01;24(49):71-79. [doi: 10.3916/c49-2016-07]
6. Hill RL, Swanson A, Kirwan J. EI and learning styles: The effect on learning of generationals. In: Proceedings of the International Conference on Interactive Collaborative Learning. 2017 Presented at: International Conference on Interactive Collaborative Learning; September 21-23, 2016; Belfast, UK p. 101-115 URL: https://link.springer.com/chapter/10.1007/978-3-319-50340-0_8 [doi: 10.1109/mic.2015.58]
7. Hopkins L, Hampton BS, Abbott JF, Buery-Joyner SD, Craig LB, Dalrymple JL, et al. To the point: Medical education, technology, and the millennial learner. Am J Obstet Gynecol 2018 Feb;218(2):188-192. [doi: 10.1016/j.ajog.2017.06.001] [Medline: 28599897]
8. Hampton D, Welsh D, Wiggins A. Learning preferences and engagement level of Generation Z nursing students. Nurse Educ 2019 Jun 14:1. [doi: 10.1097/NNE.0000000000000710] [Medline: 31219957]
9. Desy JR, Reed DA, Wolanskyj AP. Milestones and Millennials: A perfect pairing-competency-based medical education and the learning preferences of Generation Y. Mayo Clin Proc 2017 Feb;92(2):243-250. [doi: 10.1016/j.mayocp.2016.10.026] [Medline: 28160874]
10. Grossman P, McDonald M. Back to the future: Directions for research in teaching and teacher education. Am Educ Res J 2008 Mar;45(1):184-205. [doi: 10.3102/0002831207312906]
11. Hora M, Smolarek B, Martin K, Scrivener L. Exploring the Situated and Cultural Aspects of Communication in the Professions: Implications for Teaching, Student Employability, and Equity in Higher Education. Am Educ Res J 2019 Apr 09;56(6):2221-2261. [doi: 10.3102/0002831219840333]
12. Muttappallymyalil J, Mendis S, John LJ, Shanthakumari N, Sreedharan J, Shaikh RB. Evolution of technology in teaching: Blackboard and beyond in medical education. Nepal J Epidemiol 2016 Oct;6(3):588-592 [FREE Full text] [doi: 10.3126/nje.v6i3.15870] [Medline: 27822404]
13. Colazzo L, Molinari A, Ronchetti M, Trifonova A. Towards a multi-vendor mobile learning management system. In: Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn). 2003 Presented at: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn); November 7-11, 2003; Phoenix, AZ p. 2097-2100 URL: https://pdfs.semanticscholar.org/eabb/bd1313e9c5dec868d2736233533b7149f8e9.pdf
14. Farley H, Murphy A, Johnson C, Carter B, Lane M, Midgley W, et al. How do students use their mobile devices to support learning? A case study from an Australian regional university. J Interact Media Educ 2015;1(14):13. [doi: 10.5334/jime.ar]
15. Al-Jundi W, Kayssi A, Papia G, Dueck A. Smart(phone) learning experience among vascular trainees using a response system application. J Surg Educ 2017;74(4):638-643. [doi: 10.1016/j.jsurg.2016.12.006] [Medline: 28130100]
16. Lazarus L, Sookrajh R, Satyapal KS. Tablet technology in medical education in South Africa: A mixed methods study. BMJ Open 2017 Jul 31;7(7):e013871 [FREE Full text] [doi: 10.1136/bmjopen-2016-013871] [Medline: 28760784]
17. Sparrow B, Liu J, Wegner D. Google effects on memory: Cognitive consequences of having information at our fingertips. Science 2011 Aug 05;333(6043):776-778. [doi: 10.1126/science.1207745] [Medline: 21764755]
18. Newstead B, Wright CM, Colby SJ. The Bridgespan Group. 2010 Jun 28. Next generation learning: Can we crack four problems to unleash quality education for all? URL: https://tinyurl.com/vrmlk5x [accessed 2020-01-19]
19. Woodill G, Udell C. Mobile Learning in Medicine and Healthcare: Professional Education Applications. Morton, IL: Float Mobile Learning; 2011. URL: http://info.floatlearning.com/hs-fs/hub/241955/file-285652672-pdf/docs/mobile-health/mobile-learning-medicine-healthcare-professional-education-applications.pdf [accessed 2020-01-19]
20. Briz-Ponce L, Juanes-Méndez J, García-Peñalvo F. Exploring mobile learning apps for medical students and health care professionals. In: Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'16). 2016 Presented at: Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'16); November 2-4, 2016; Salamanca, Spain p. 461-465 URL: https://tinyurl.com/vufgvzh [doi: 10.1145/3012430.3012557]
21. Patil RN, Almale BD, Patil M, Gujrathi A, Dhakne-Palwe S, Patil AR, et al. Attitudes and perceptions of medical undergraduates towards mobile learning (m-learning). J Clin Diagn Res 2016 Oct;10(10):JC06-JC10 [FREE Full text] [doi: 10.7860/JCDR/2016/20214.8682] [Medline: 27891356]

22. Mather CA, Gale F, Cummings EA. Governing mobile technology use for continuing professional development in the Australian nursing profession. BMC Nurs 2017;16:17 [FREE Full text] [doi: 10.1186/s12912-017-0212-8] [Medline: 28428731]
23. Macznik AK, Ribeiro DC, Baxter GD. Online technology use in physiotherapy teaching and learning: A systematic review of effectiveness and users' perceptions. BMC Med Educ 2015 Sep 28;15:160 [FREE Full text] [doi: 10.1186/s12909-015-0429-8] [Medline: 26415677]
24. 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]
25. Cantarero-Villanueva I, Fernández-Lao C, Galiano-Castillo N, Castro-Martín E, Díaz-Rodríguez L, Arroyo-Morales M. Evaluation of e-learning as an adjunctive method for the acquisition of skills in bony landmark palpation and muscular ultrasound examination in the lumbopelvic region: A controlled study. J Manipulative Physiol Ther 2012;35(9):727-734. [doi: 10.1016/j.jmpt.2012.10.007] [Medline: 23206968]
26. 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]
27. Manzini JL. Declaración de Helsinki: Principios éticos para la investigación médica sobre sujetos humanos. Acta Bioeth 2000;6(2):321-334 ISSN 1726-569X [FREE Full text] [doi: 10.4067/s1726-569x2000000200010]
28. World Medical Association General Assembly. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. J Int Bioethique 2004 Mar;15(1):124-129. [Medline: 15835069]
29. Martinoli C. Musculoskeletal ultrasound: Technical guidelines. Insights Imaging 2010 Jul;1(3):99-141 [FREE Full text] [doi: 10.1007/s13244-010-0032-9] [Medline: 23100193]
30. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J 1975 Feb 22;1(5955):447-451 [FREE Full text] [doi: 10.1136/bmj.1.5955.447] [Medline: 1115966]
31. Davies R, Ellerton C, Evans C. Reaching consensus on measuring professional behaviour in physical therapy objective structured clinical examinations. Physiotherapy Can 2017 Feb;69(1):65-72. [doi: 10.3138/ptc.2015-45e]
32. Carraccio C, Englander R. The objective structured clinical examination: A step in the direction of competency-based evaluation. Arch Pediatr Adolesc Med 2000 Jul;154(7):736-741. [doi: 10.1001/archpedi.154.7.736] [Medline: 10891028]
33. Payne E, Brown G. Communication and practice with examination criteria: Does this influence performance in examinations? Assess Eval High Educ 2011 Oct;36(6):619-626. [doi: 10.1080/02602931003632373]
34. Noguera JM, Jiménez JJ, Osuna-Pérez MC. Development and evaluation of a 3D mobile application for learning manual therapy in the physiotherapy laboratory. Comput Educ 2013 Nov;69:96-108. [doi: 10.1016/j.compedu.2013.07.007]
35. Whipple EC, Allgood KL, Larue EM. Third-year medical students' knowledge of privacy and security issues concerning mobile devices. Med Teach 2012;34(8):e532-e548. [doi: 10.3109/0142159X.2012.670319] [Medline: 22489969]
36. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: A systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]

Abbreviations

m-learning: mobile learning
OSCE: Objective Structured Clinical Examination

Edited by G Eysenbach; submitted 15.09.19; peer-reviewed by AJ Martín-Pérez, F Artacho-Cordon; comments to author 16.10.19; revised version received 24.10.19; accepted 13.11.19; published 10.03.20.

Please cite as:
Lozano-Lozano M, Galiano-Castillo N, Fernández-Lao C, Postigo-Martin P, Álvarez-Salvago F, Arroyo-Morales M, Cantarero-Villanueva I
The Ecofisio Mobile App for Assessment and Diagnosis Using Ultrasound Imaging for Undergraduate Health Science Students: Multicenter Randomized Controlled Trial
J Med Internet Res 2020;22(3):e16258
URL: https://www.jmir.org/2020/3/e16258
doi:10.2196/16258
PMID:32154784

Original Paper

Development and Assessment of a Gastroscopy Electronic Learning System for Primary Learners: Randomized Controlled Trial

Shuang Li[1], MMSc; Guoqing Li[1], MMSc; Ying Liu[1], MMSc; Wanying Xu[1], MMSc; Ningning Yang[1], MMSc; Haoyuan Chen[1], MMSc; Ning Li[1], MMSc; Kunpeng Luo[1], BM; Shizhu Jin[1], MD, PhD

Department of Gastroenterology and Hepatology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

Corresponding Author:
Shizhu Jin, MD, PhD
Department of Gastroenterology and Hepatology
The Second Affiliated Hospital of Harbin Medical University
246 Xuefu Road, Nangang District
Harbin, 150086
China
Phone: 86 13654699988
Email: drshizhujin@hrbmu.edu.cn

Abstract

Background: Endoscopic examination is a popular and routine procedure for the diagnosis and treatment of gastrointestinal (GI) diseases. Skilled endoscopists are in great demand in clinical practice, but the training process for beginners to become endoscopy specialists is fairly long. Convenience and a self-paced, learner-centered approach make electronic learning (e-learning) an excellent instructional prospect.

Objective: This study aimed to develop and apply an e-learning system in gastroscopy teaching and learning and to evaluate its effectiveness and user satisfaction.

Methods: The e-learning software *Gastroscope Roaming System* was developed for primary learners. The system simulates the real structure of the upper gastrointestinal (UGI) tract to teach the main characteristics of gastroscopy under both normal conditions and conditions of common UGI tract diseases. A randomized controlled trial was conducted. Participants were randomly allocated to an e-learning group (EG)or a non–e-learning control group after a pretest. On completing the training, participants undertook a posttest and gastroscopy examination. In addition, the EG completed a satisfaction questionnaire.

Results: Of the 44 volunteers, 41 (93%) completed the gastroscopy learning and testing components. No significant pretest differences were found between the intervention and control groups (mean 50.86, SD 6.12 vs mean 50.76, SD 6.88; *P*=.96). After 1 month of learning, the EG's posttest scores were higher (mean 83.70, SD 5.99 vs mean 78.76, SD 7.58; *P*=.03) and improved more (*P*=.01) than those of the control group, with better performance in the gastroscopy examination (mean 91.05, SD 4.58 vs mean 84.38, SD 5.19; *P*<.001). Overall, 85% (17/20) of the participants were satisfied with the e-learning system, and 95% (19/20) of the participants considered it successful.

Conclusions: E-learning is an effective educational strategy for primary learners to acquire skills in gastroscopy examination and endoscopic imaging of the GI tract.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IOR-17013091; http://www.chictr.org.cn/showproj.aspx?proj=22142

(J Med Internet Res 2020;22(3):e16233) doi:10.2196/16233

KEYWORDS

e-learning; gastroscopy; endoscopy; randomized controlled trial

Introduction

Background

With the development of digestive endoscopy technology, gastrointestinal (GI) diseases are increasingly treatable by digestive endoscopy. Skilled and experienced endoscopists are needed in clinical practice worldwide, yet the training process for primary learners is difficult and time consuming. Hence, educating endoscopists has become a global challenge in medical education.

Traditionally, experienced endoscopists have played an important role in training primary learners, and primary learners have worked with their tutors to practice endoscopy on patients [1]. It is common for primary learners to feel nervous and anxious and to be unable to obtain satisfactory cooperation from their patients. In addition, patients may refuse to permit primary learners to practice GI endoscopy on them. Traditional clinical training may fail to provide adequate information on multiple needs and cognitive deficits. Endoscopic simulators, including ex vivo animal tissue models, live animal models, mechanical models, and virtual reality computer simulators [2,3], have been widely used in academic practice. However, problems remain with regard to the high cost of teaching, limited resource space, and small audience, and primary learners can only study the structure of the upper gastrointestinal (UGI) tract and cannot further study the endoscopic manifestations of UGI diseases. Thus, improving the instructional programs and models for primary learners is a key issue in endoscopy training.

Many institutes are pioneering electronic learning (e-learning) methods as a cost-effective alternative to traditional methods [4,5]. E-learning is a well-established approach to learning via electronic-/computer-based, mostly Web-based, programs [6]. The technology is based on the use of the internet to deliver a broad array of educational materials and training procedures to enhance knowledge and performance [7,8]. Numerous researchers have described the use of e-learning in medical education [9-18]. Compared with traditional teaching environments, e-learning has several advantages, such as asynchrony, cost savings, individualized learning, greater accessibility, greater ease of distribution, and up-to-date content, which may overcome the difficulties and dilemmas experienced during the early phases of endoscopic training [19-22]. Learners may be attracted to e-learning because it centers them in the learning procedure, in contrast to their role as passive recipients in traditional training methods [23]. Learners enrolled in an e-learning program can choose the content that they access, the sequence in which it is studied, and the space used for learning based on their individual experience and personal learning objectives without time or space limitations [24,25].

Aim of This Study

In this context, we designed and developed the Gastroscope Roaming System (GRS) to create, assess, and implement an integrated e-learning gastroscopic education package for primary learners. The system focuses on teaching the principal characteristics of gastroscopy examination of the human UGI under normal conditions and under conditions of routine UGI diseases. We aimed to (1) develop a gastroscopy e-learning system for primary learners, (2) evaluate the efficacy and effectiveness of the system, and (3) assess the satisfaction of primary learners.

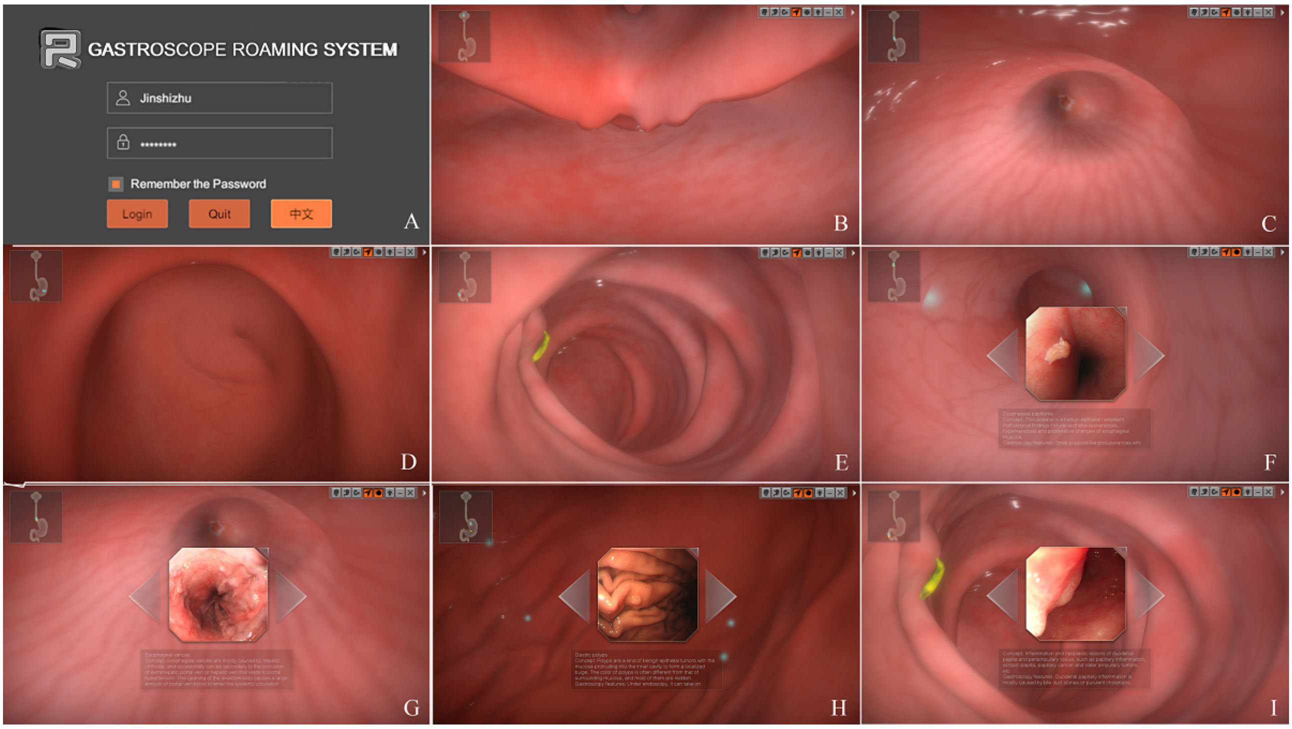
Methods

Development and Description of the Electronic Learning System

We collaborated with the Harbin University of Science and Technology to develop the GRS. The system was built to be bilingual in Chinese and English; the language can be freely switched on the log-in interface, and the system can be used by primary learners of gastroscopy in China and worldwide. First, the user interface of the system was designed and created. A 3-dimensional model of the UGI tract, including the esophagus, stomach, and duodenum, was obtained after a review of various data sources (mainly pictures and videos of gastroscopy examinations from the Digestive Endoscopy Center of the Second Affiliated Hospital of Harbin Medical University [HMU]) and was modeled by Unity 3D, Autodesk Maya, Zbrush, and Substance Painter software. Then, the 3-dimensional model was transformed into an operation scene, and the physiological function was simulated and rendered (Adobe Photoshop software for material design and production and Maya software for animation). The C++ programming language was used to design the roaming, interactive, and cognitive functions of the digestive organ structure and to realize the camera placement, path switching, and perspective switching. In addition, embedded links and demonstrations of lesion cases were created. Currently, the program requires a specified account number and password for a user to log in.

The two GRS modules were designed for primary learners to gain knowledge of the UGI anatomic structure, relevant endoscopy images, and protocol requirements and to improve their confidence. Module A contains normal human anatomic structures, with anatomic color illustrations from the pharyngeal portion to the descending part of the duodenum of the human UGI. Module B contains UGI images showing both normal conditions and conditions of routine UGI diseases. The software can also simulate UGI endoscopy and the route of the gastroscopy lens. The incidences of typical diseases along the endoscopic examination path are labeled based on clinical images of these diseases. Dozens of common UGI diseases are included for learning purposes. There are corresponding bilingual explanations under the images, including disease characteristics and endoscopic manifestations. Screenshots are presented in Figure 1.

Figure 1. Gastroscope Roaming System's software screen captures. (A) Software system log-in interface. (B) Normal esophageal entrance screenshot. (C) Normal cardia entrance screenshot. (D) Normal pyloric entrance screenshot. (E) Normal duodenal papilla screenshot. (F-I) Screenshots of common diseases in different parts of the disease learning state, and the disease description and gastroscopic characteristics are given below each image.



Jinshizhu

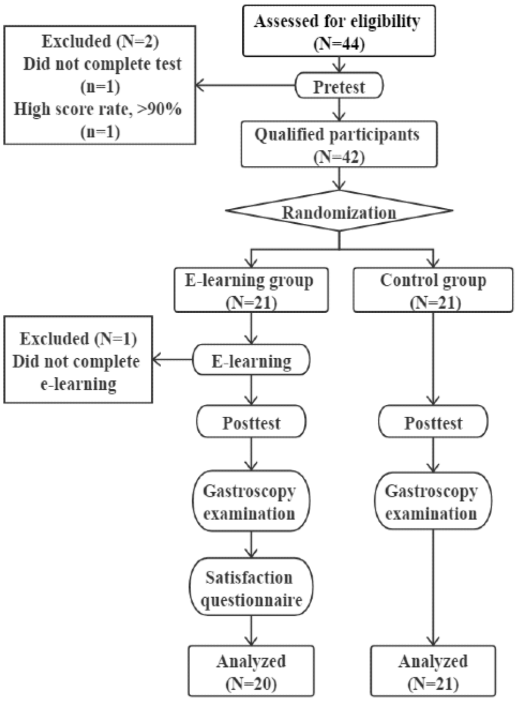
Participant Recruitment and Implementation Strategy

We issued an experimental information notice to recruit participants in the hospital. The 44 volunteers were all master degree students majoring in gastroenterology and hepatology at the Second Affiliated Hospital of HMU who had signed a consent form to participate in this trial. The teachers at the endoscopy center verified the basic information of each volunteer and played a key supervisory role in the entire process of learning and evaluation. All participants had acquired basic GI theoretical knowledge in their previous years of study, but because the GRS was designed for primary learners, the participants could not perform a gastroscopy and confirmed that they had not yet received systematic professional education in gastroscopy. In addition, they needed access to a procedure on their own computers to operate the sample version of the e-learning system. All participants completed a pretest questionnaire; those who did not complete the pretest or did too well in the pretest (correct rating >90% or higher) were excluded because the e-learning was designed for beginners who had not previously mastered sufficient endoscopy knowledge [26].

The experiment was conducted from November 2018 to December 2018. The flow of the experiment is shown in Figure 2. The 44 eligible volunteers were asked to complete the pretest in 30 min, and the unqualified participants were eliminated based on the results. The randomization sequence was computer generated using IBM SPSS version 24. Then, the qualified participants were randomly divided into two groups: the e-learning group (EG) and the non–e-learning control group (CG). The enrolled students spent 1 month in theoretical and practical training for gastroscopy examinations in our on-campus endoscopy center (Endoscopy Center, the Second Affiliated Hospital of HMU). A specific username and password were assigned to each EG participant to log in to the GRS during the trial. The EG had access to traditional learning combined with e-learning at any time and location. During the same period, the CG did not have access to the e-learning tool and learned only through the traditional mode of reading textbooks and other written materials in addition to practical training. All participants received routine gastroscopy training at the endoscopy center during the trial.

After the 1-month gastroscopy study, all participants were required to complete the posttest and gastroscopy examination assessment. The EG also completed an additional questionnaire to determine the level of satisfaction with the GRS.

Figure 2. Flow diagram showing the details of the participant enrollment, randomization, and tests. e-learning: electronic learning.



Learning Outcome Evaluation Strategy

The potential and advantages of e-learning may not always be perceived as leading to significant improvements in educational outcomes. Therefore, it is necessary to demonstrate the rationality of e-learning through the assessments of effectiveness and efficacy. Kirkpatrick built a famous framework to evaluate the learning effect in the 1950s [27], which can be used to assess GRS effectiveness [24,28,29]. The framework describes four assessment levels: reaction, learning, behavior, and results [19,30]. Most studies that evaluate e-learning processes rely only on users' learning and reaction rather than behavior and results [21,30-33]. GRS was designed to help improve the gastroscopic practice skills of primary learners. On this basis, e-learning in this study was further evaluated at the *behavior* level to reveal the efficiency and reliability of e-learning in medical education.

Pre- and Postintervention Tests

For the *learning* level assessment, the most conventional and reasonable test method is the comparison of pre- and postintervention test scores. In our experiment, the participants completed a pretest before gastroscopy learning and a posttest after 1 month of training. The pretest and posttest contained the same set of questions, but the order of the questions differed to ensure that the difficulty and reference standards of the two tests were at the same level [23,34]. The tests consisted of 50 multiple-choice questions, each of which had 4 to 5 response options and a single correct answer. Each question had a value of two points, with a total possible score of 100 points. The main topics of the multiple-choice questions were the elements of the basic operation of gastroscopy, cognition of the anatomical structure of the UGI, mastery of the characteristics of the structure of the UGI under gastroscopy, and the identification of and distinction between images of various common diseases of the UGI under gastroscopy. Most endoscopic images used in the questions were taken from the linked pictures used in the GRS system, and the rest were from our endoscopy center. The test was compiled by the authors of this paper and the designer of the GRS software. In total, 2 endoscopy experts from the Department of Gastroenterology, the Second Affiliated Hospital of HMU, reviewed the test questions to ensure that they were suitable for this experiment. The test paper (pretest version) is provided as Multimedia Appendix 1.

To ensure the fairness of the test and the reliability of the scores, we placed the participants in a classroom with a signal shield and conducted a closed-volume test with 2 invigilators and timers. The examination time was set at 30 min. When the time ended, the participants were asked to stop, and the papers were collected immediately. After the pretest, the participants could not see their own answers or the correct answers to ensure the validity of the posttest.

Gastroscopy Examination Test

As our experiment was based on the combination of e-learning and traditional endoscopy teaching (ie, blended learning), we tested the participants' actual gastroscopy examination performance to evaluate the *behavior* level. After 1 month of training, the students undertook a gastroscopy examination, with the examination order based on drawing lots, and made real-time video recordings. All the patients participating in the gastroscopy test had taken an appointment for a general gastroscopy at the outpatient department and had agreed to a gastroscopy performed by primary endoscopists. As many unexpected problems appeared in the actual operation, such as patients with a special physiological structure or patients with lesions who were difficult to identify, teachers were present during the procedure. Each participant had three opportunities to select the performance that he or she thought was the most satisfactory for archiving. A total of 5 expert endoscopists (from the Affiliated Hospital of HMU) who were not aware of the participants' grouping scored the students' gastroscopy examination videos based on their experience in gastroscopy. The evaluation standard applied in the examination was the scale developed jointly by experts, including forward operation of gastroscopy, withdrawal observation, patients' comfort and satisfaction, overall gastroscopy examination time and fluency, clarity of collected images, and diagnostic accuracy. We removed the highest score and the lowest score and obtained the average final score. The grading table is provided as Multimedia Appendix 2.

Gastroscope Roaming System Satisfaction Questionnaire

The *reaction* level was assessed by a satisfaction questionnaire. We used the questionnaire developed by Wang [35,36], which has been widely cited in evaluations of e-learning satisfaction [23]. This questionnaire has been shown to reach a reliability (Cronbach alpha) of .95 [35]. As the current version of the GRS is asynchronous, we did not use the *learning community* quality of Wang's questionnaire. With the exception of the last two questions about overall satisfaction and success, we divided the remaining 22 questions into four modules: content, interface, testing, and personalization. We translated the questionnaire into Chinese and used a 5-point Likert scale to estimate each question anchored with *strongly disagree* to *strongly agree* and then counted the total points [37]. All the students in the EG completed the satisfaction questionnaire after completing their study. The questionnaire is provided as Multimedia Appendix 3.

Statistical Analysis

IBM SPSS version 24 was used for data analysis. The analysis included simple frequencies and descriptive analyses (means and standard deviations). First, the interaction of variables was verified by a general linear model. A paired sample *t* test was used to compare the test scores before and after the intervention in each group, and an independent sample *t* test was used to compare the pretest, posttest, and gastroscopy examination scores between the EG and the CG. The differences in the sex ratio and grade composition of the participants were analyzed by the chi-square test. For all the statistical analyses, we considered the significance level to be .05. Data analysis was used to determine whether there was a significant difference between the two experimental groups.

Results

Sample Description

Of the 44 masters degree students from the Second Affiliated Hospital of HMU who were assessed for eligibility, 42 (95%) who met the inclusion criteria completed the pretest and were included in the study. Two participants were excluded because one failed to complete the pretest and the other had a pretest score that was too high (>90%). Of the remaining participants, 21 were assigned to the EG and 21 to the CG following the principles of randomized controlled trials. During the learning period, only 1 student in the EG dropped out of the course. Therefore, data for 20 students in the EG and 21 students in the CG were analyzed. All 20 of the remaining students in the EG completed the satisfaction questionnaire. The baseline characteristics were similar in both the groups. Table 1 summarizes the characteristics of the participants.

Table 1. Demographic and other characteristics of participants.

Baseline characteristics	Electronic learning group (n=20)	Control group (n=21)	*P* value
Age (years)			.77
Range	23-27	23-26	
Mean	24.05	24.14	
Gender, n (%)			.39
Male	6 (30)	9 (43)	
Female	14 (70)	12 (57)	
Current year of residency training, n (%)			.44
Postgraduate year 1	12 (60)	15 (71)	
Postgraduate year 2	8 (40)	6 (29)	
Pretest score			.96
Range	40-60	38-62	
Mean	50.86	50.76	

Statistics and Analysis of Examination Scores

The mean pretest score (SD) in the EG was 50.86 (SD 6.11), and it improved to 83.70 (SD 5.99) in the posttest (*P*<.001). In contrast, the mean pretest score (SD) in the CG was 50.76 (SD 6.88), and it increased to 78.76 (SD 7.58) in the posttest (*P*<.001). After using the general linear model to analyze the difference between the EG and CG in the pretest and posttest, no interaction was found between the between-group variable and the within-group variable (*P*=.11). Therefore, a *t* test was used for further analysis. The results showed that there were no significant differences between the EG and CG (*P*=.96) in the pretest. After 1 month of learning, the posttest results of the EG were better than those of the CG (*P*=.03), and the test scores of the EG improved more than those of the CG (*P*=.01).

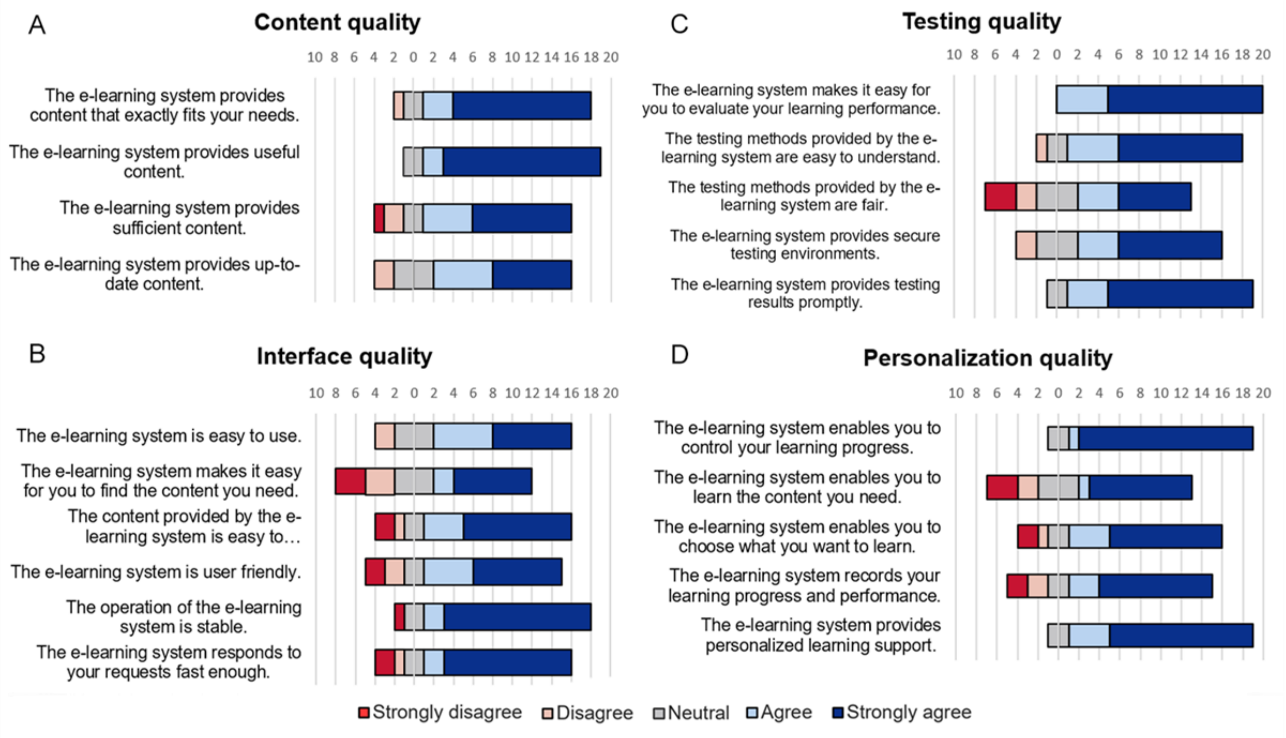
Regarding the gastroscopy examination scores, the mean score (SD) was 91.05 (SD 4.58) in the EG and 84.38 (SD 5.19) in the CG (*P*<.001).

User Satisfaction With the Gastroscope Roaming System

The overall satisfaction rate with the e-learning course was 77.0% (308/400). Only 11.5% (46/400) of the students were biased against e-learning, and the remaining 11.5% (46/400) maintained a neutral attitude. The last two overall satisfaction problems showed that 85% (17/20) of the participants were satisfied with the GRS, and 95% (19/20) of the participants thought it was successful. The satisfaction level for each problem is detailed in Figure 3.

Regarding the content quality (Question 1-Question 4), 80% (64/80) of the learners thought that the course fit their needs and was sufficient, useful, and up-to-date. Regarding the interface quality (Question 5- Question 10), 73.3% (88/120) of the participants found the e-learning course to be stable, user friendly, and fast and agreed that it was easy to find the content needed. Regarding the testing quality (Question 11-Question 15), 80% (80/100) of the learners assessed the testing of the course as fair, secure, prompt, easy to understand, and easy to evaluate. Regarding the personalization quality (Question 16-Question 20), 76% (76/100) of the students thought that the GRS enabled them to learn the content needed, choose what to learn, control their learning progress, record their learning progress, and provide personalized learning support. Hence, there were more oppositional responses to the interface and personalization qualities than to the other qualities.

Figure 3. Satisfaction analysis of the system by Gastroscope Roaming System participants in four modules: (A) content quality, (B) interface quality, (C) testing quality, and (D) personalization quality. Each module lists the problems it contains. e-learning: electronic learning.



Discussion

Principal Findings

The first main outcome of this study was the development of an e-learning software named *Gastroscope Roaming System* for primary learners. To date, the conventional gastroscopy teaching mode has been face-to-face tutorials. The number of learners is, therefore, very small; the instruction is lengthy; and it is not convenient to practice only on patients. Textbooks and handouts appear to be the extensive way of disseminating knowledge in education [38]. However, for students in the field of medical operations, especially endoscopy teaching, traditional lectures and spectrograms are not conducive to learning how to operate and gain experience effectively. In addition, gastroscopy simulators have problems of high cost, limited teaching resources, and low popularity. As constructed in this experiment, an internet-based e-learning system offers tremendous advantages over the traditional teaching methods because there are no limits on the number of learners or learning time and place. Students can study gastroscopy with their laptop at any time and place instead of being confined to an observational study in the narrow gastroscopy room, thus enabling them to more reasonably allocate and use their study time. Gastroscopy learning has three aspects: knowledge, skills, and experience. Through conventional lectures or hands-on seminars, endoscopists can acquire only the relevant knowledge and techniques [39]. Therefore, we collected many endoscopic images of common UGI diseases and inserted them into the software for primary learners to accumulate experience. Given the creation of the new gastroscope learning method, we can expect to apply this method to more endoscopic learning and training.

The second principal result was the comparison of the learning results of the GRS-based EG with those of the traditional learning group, showing that e-learning can improve the learning efficiency of primary gastroscopists. Notably, the EG in this experiment used a combination of e-learning and traditional learning, that is, blended learning, which seemed to be more effective than single e-learning with respect to knowledge acquisition. Although e-learning may not be an alternative to traditional learning in some cases, it can provide supplements to traditional education and is a useful complement to traditional education [40-42].

Before gastroscopy learning, there was no significant difference between the two groups in the predicted test, indicating that the level of the participants was similar. After 1 month of study, the theoretical test scores of all the students significantly improved, and the score of the EG increased slightly more than that of the CG. However, the gastroscopy examination score of the EG was notably higher than that of the CG after 1 month of learning, indicating that e-learning is more effective than traditional learning, especially in improving students' gastroscopic handling ability. The theoretical-level test verified the participants' basic knowledge of gastroscopy and their ability to diagnose disease, whereas the gastroscopy examination verified their practical ability. In other words, gastroscopy ability=cognitive ability+practical ability. The technical level is most important for beginners learning gastroscopy. Therefore, we believe that the GRS is essential for helping students master gastroscopy skills in a short time. The GRS provides learners with a virtual UGI space, which enables them to explore the structure of the UGI freely and then, through repeated learning, to master and familiarize themselves with the structure of the UGI. In the process of a real gastroscopy procedure, they can operate skillfully and confidently, accelerate their learning speed, and avoid the harm caused to patients by confusion and fear. Moreover, by repeatedly studying the endoscopic pictures and characteristics of common diseases shown in the GRS, students can learn to accurately identify lesions in actual gastroscopy.

The third main result was that the satisfaction questionnaire showed that the overall satisfaction rate of the students in the EG was very high. Using Wang's questionnaire, we determined how satisfied the students were with each part. We found that the students were more satisfied with the content and testing qualities and less satisfied with the interface and personalization qualities. In terms of the interface quality, students found that the operational interface of the software was somewhat complex, difficult to understand, and not smooth. This response may be because of the software design using the *W/A/S/D* keys on keyboard with mouse movement. The students had different proficiency levels in computer operation and different computer configurations.

Limitations

Our study also has limitations. For example, the small sample size in a single institution affects the generalizability of the results. The main reason was that our GRS software was in the early stage of construction. We will continue to improve, upgrade, and expand its content in the future. Therefore, the software was applied only on a small scale in the hospital with which the author is affiliated to assess the efficiency and effectiveness of the e-learning model. Regarding the assessment of the real gastroscopy procedure, because of differences in the physiological structure of the UGI and the pain and tolerance levels of each patient, it was not guaranteed that each student would face the same difficulty in the assessment. Even with three opportunities, the participants might not have been able to perform at the real level of gastroscopy, although the probability of this problem is low. Finally, we should carefully consider the validity of the satisfaction assessment. We used an unproven Chinese version of the user satisfaction questionnaire because we could not find a Chinese-validated translation for e-learning systems in the literature.

Conclusions

The results of our study suggest that e-learning is an effective educational strategy for primary learners to acquire skills in gastroscopy examination and master the characteristics of endoscopic images of the UGI. The GRS may serve as a meaningful supplementary approach to on-campus endoscopy education. The novel computer assistant module in the field of endoscopic teaching will strengthen the training of skilled endoscopists. We will continue to upgrade and improve the GRS system to enhance the authenticity of the simulation and to enrich the variety of diseases. In summary, e-learning can be widely used as an effective assistant to supplement routine teaching and learning in medical education.

Acknowledgments

This study was funded by the China Medical Board (CMB; project number 15-238). The authors are grateful to HMU for training the authors and to the leaders, teachers, and graduate students who helped with this study. The authors sincerely thank the CMB for its support and the animation faculty, computer center, and students of Harbin University of Science and Technology for their technical cooperation.

Authors' Contributions

SJ and SL contributed to study concept and design. SL, GL, YL, and WX contributed to trial implementation and supervision. SL, NY, HC, NL, and KL contributed to acquisition of data and statistical analysis. SJ and SL drafted the manuscript. SJ contributed to study supervision. All authors contributed to interpretation of the results, critical revision of the manuscript for important intellectual content, and final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Multiple-choice question test paper for evaluating Gastroscope Roaming System learning efficiency.
[PDF File (Adobe PDF File), 629 KB - jmir_v22i3e16233_app1.pdf]

Multimedia Appendix 2
Grading table for gastroscopy examination.
[PDF File (Adobe PDF File), 65 KB - jmir_v22i3e16233_app2.pdf]

Multimedia Appendix 3
Gastroscope Roaming System satisfaction questionnaire.
[PDF File (Adobe PDF File), 81 KB - jmir_v22i3e16233_app3.pdf]

Multimedia Appendix 4
CONSORT-EHEALTH (V 1.6.1) -Submission/Publication Form.
[PDF File (Adobe PDF File), 3494 KB - jmir_v22i3e16233_app4.pdf]

References

1. Ekkelenkamp VE, Koch AD, de Man RA, Kuipers EJ. Training and competence assessment in GI endoscopy: a systematic review. Gut 2016 Apr;65(4):607-615. [doi: 10.1136/gutjnl-2014-307173] [Medline: 25636697]
2. ASGE Technology Committee, Goodman AJ, Melson J, Aslanian HR, Bhutani MS, Krishnan K, et al. Endoscopic simulators. Gastrointest Endosc 2019 Jul;90(1):1-12. [doi: 10.1016/j.gie.2018.10.037] [Medline: 31122746]
3. Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM. Virtual reality simulation training in endoscopy: a Cochrane review and meta-analysis. Endoscopy 2019 Jul;51(7):653-664. [doi: 10.1055/a-0894-4400] [Medline: 31071757]
4. Thorne CJ, Lockey AS, Bullock I, Hampshire S, Begum-Ali S, Perkins GD, Advanced Life Support Subcommittee of the Resuscitation Council (UK). E-learning in advanced life support--an evaluation by the Resuscitation Council (UK). Resuscitation 2015 May;90:79-84. [doi: 10.1016/j.resuscitation.2015.02.026] [Medline: 25766092]
5. Taveira-Gomes T, Saffarzadeh A, Severo M, Guimarães MJ, Ferreira MA. A novel collaborative e-learning platform for medical students - ALERT STUDENT. BMC Med Educ 2014 Jul 14;14:143 [FREE Full text] [doi: 10.1186/1472-6920-14-143] [Medline: 25017028]
6. Holzinger A, Kickmeier-Rust MD, Wassertheurer S, Hessinger M. Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. Comput Educ 2008;52(2):292-301. [doi: 10.1016/j.compedu.2008.08.008]
7. Obdeijn MC, Bavinck N, Mathoulin C, van der Horst CM, Schijven MP, Tuijthof GJ. Education in wrist arthroscopy: past, present and future. Knee Surg Sports Traumatol Arthrosc 2015 May;23(5):1337-1345. [doi: 10.1007/s00167-013-2592-y] [Medline: 23835770]
8. Kim K, Han J, Park IB, Kee C. Medical education in Korea: the e-learning consortium. Med Teach 2009 Sep;31(9):e397-e401. [doi: 10.1080/01421590902744902] [Medline: 19811175]
9. Sajeva M. E-learning: web-based education. Curr Opin Anaesthesiol 2006 Dec;19(6):645-649. [doi: 10.1097/ACO.0b013e328010bec8] [Medline: 17093369]
10. Lahti M, Hätönen H, Välimäki M. Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: a systematic review and meta-analysis. Int J Nurs Stud 2014 Jan;51(1):136-149. [doi: 10.1016/j.ijnurstu.2012.12.017] [Medline: 23384695]
11. Feng JY, Chang H, Chang HY, Erdley WS, Lin CH, Chang YJ. Systematic review of effectiveness of situated e-learning on medical and nursing education. Worldviews Evid Based Nurs 2013 Aug;10(3):174-183. [doi: 10.1111/wvn.12005] [Medline: 23510119]
12. Perkins GD, Fullerton JN, Davis-Gomez N, Davies RP, Baldock C, Stevens H, et al. The effect of pre-course e-learning prior to advanced life support training: a randomised controlled trial. Resuscitation 2010 Jul;81(7):877-881. [doi: 10.1016/j.resuscitation.2010.03.019] [Medline: 20409627]
13. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med 2006 Mar;81(3):207-212. [doi: 10.1097/00001888-200603000-00002] [Medline: 16501260]
14. Beux PL, Fieschi M. Virtual biomedical universities and e-learning. Int J Med Inform 2007;76(5-6):331-335. [doi: 10.1016/S1386-5056(07)00060-3] [Medline: 17407747]

15. Wutoh R, Boren SA, Balas EA. eLearning: a review of internet-based continuing medical education. J Contin Educ Health Prof 2004;24(1):20-30. [doi: 10.1002/chp.1340240105] [Medline: 15069909]
16. Wong G, Greenhalgh T, Pawson R. Internet-based medical education: a realist review of what works, for whom and in what circumstances. BMC Med Educ 2010 Feb 2;10:12 [FREE Full text] [doi: 10.1186/1472-6920-10-12] [Medline: 20122253]
17. Chumley-Jones HS, Dobbie A, Alford CL. Web-based learning: sound educational method or hype? A review of the evaluation literature. Acad Med 2002 Oct;77(10 Suppl):S86-S93. [doi: 10.1097/00001888-200210001-00028] [Medline: 12377715]
18. Cook DA, Garside S, Levinson AJ, Dupras DM, Montori VM. What do we mean by web-based learning? A systematic review of the variability of interventions. Med Educ 2010 Aug;44(8):765-774. [doi: 10.1111/j.1365-2923.2010.03723.x] [Medline: 20633216]
19. Abdelhai R, Yassin S, Ahmad MF, Fors UG. An e-learning reproductive health module to support improved student learning and interaction: a prospective interventional study at a medical school in Egypt. BMC Med Educ 2012 Mar 20;12:11 [FREE Full text] [doi: 10.1186/1472-6920-12-11] [Medline: 22433670]
20. Moreira IC, Ventura SR, Ramos I, Rodrigues PP. Development and assessment of an e-learning course on breast imaging for radiographers: a stratified randomized controlled trial. J Med Internet Res 2015 Jan 5;17(1):e3 [FREE Full text] [doi: 10.2196/jmir.3344] [Medline: 25560547]
21. Maloney S, Chamberlain M, Morrison S, Kotsanas G, Keating JL, Ilic D. Health professional learner attitudes and use of digital learning resources. J Med Internet Res 2013 Jan 16;15(1):e7 [FREE Full text] [doi: 10.2196/jmir.2094] [Medline: 23324800]
22. Pinto A, Brunese L, Pinto F, Acampora C, Romano L. E-learning and education in radiology. Eur J Radiol 2011 Jun;78(3):368-371. [doi: 10.1016/j.ejrad.2010.12.029] [Medline: 21255951]
23. Oliveira AC, Mattos S, Coimbra M. Development and assessment of an E-learning course on pediatric cardiology basics. JMIR Med Educ 2017 May 10;3(1):e10 [FREE Full text] [doi: 10.2196/mededu.5434] [Medline: 28490416]
24. Ward JP, Gordon J, Field MJ, Lehmann HP. Communication and information technology in medical education. Lancet 2001 Mar 10;357(9258):792-796. [doi: 10.1016/S0140-6736(00)04173-8] [Medline: 11253986]
25. Dhir SK, Verma D, Batta M, Mishra D. E-learning in medical education in India. Indian Pediatr 2017 Oct 15;54(10):871-877 [FREE Full text] [doi: 10.1007/s13312-017-1152-9] [Medline: 29120336]
26. Nakanishi H, Doyama H, Ishikawa H, Uedo N, Gotoda T, Kato M, et al. Evaluation of an e-learning system for diagnosis of gastric lesions using magnifying narrow-band imaging: a multicenter randomized controlled study. Endoscopy 2017 Oct;49(10):957-967. [doi: 10.1055/s-0043-111888] [Medline: 28637065]
27. Alliger GM, Janak EA. Kirkpatrick's levels of training criteria: thirty years later. Pers Psychol 1989 Jun;42(2):331-342. [doi: 10.1111/j.1744-6570.1989.tb00661.x]
28. Kirkpatrick D. Great ideas revisited. Techniques for evaluating training programs. Revisiting Kirkpatrick's Four-Level Model. Train Dev 1996;50(1):54-59 [FREE Full text]
29. Galloway DL. Evaluating distance delivery and e-learning is Kirkpatrick's model relevant? Perform Improv 2005 Apr;44(4):21-27. [doi: 10.1002/pfi.4140440407]
30. Santos G, Leite A, Figueiredo P, Pimentel N, Flores-Mir C, de Melo NS, et al. Effectiveness of E-learning in oral radiology education: a systematic review. J Dent Educ 2016 Sep;80(9):1126-1139 [FREE Full text] [Medline: 27587580]
31. Carriero A, Bonomo L, Calliada F, Campioni P, Colosimo C, Cotroneo A, et al. E-learning in radiology: an Italian multicentre experience. Eur J Radiol 2012 Dec;81(12):3936-3941. [doi: 10.1016/j.ejrad.2012.07.007] [Medline: 22902406]
32. Yao K, Uedo N, Muto M, Ishikawa H, Cardona HJ, Filho EC, et al. Development of an E-learning system for the endoscopic diagnosis of early gastric cancer: an international multicenter randomized controlled trial. EBioMedicine 2016 Jul;9:140-147 [FREE Full text] [doi: 10.1016/j.ebiom.2016.05.016] [Medline: 27333048]
33. Stark CM, Graham-Kiefer ML, Devine CM, Dollahite JS, Olson CM. Online course increases nutrition professionals' knowledge, skills, and self-efficacy in using an ecological approach to prevent childhood obesity. J Nutr Educ Behav 2011;43(5):316-322. [doi: 10.1016/j.jneb.2011.01.010] [Medline: 21906545]
34. Hashemikamangar SS, Yazdanpanah F, Mirzaii M, Yazdani R, Karazifard MJ, Yasini E. Efficacy of E-learning via the website of Tehran University of Medical Sciences for diagnosing tooth discolorations and treatment planning by senior dental students. Acta Med Iran 2016 Aug;54(8):536-541 [FREE Full text] [Medline: 27701725]
35. Wang YS. Assessment of learner satisfaction with asynchronous electronic learning systems. Inf Manag 2003 Oct;41(1):75-86 [FREE Full text] [doi: 10.1016/s0378-7206(03)00028-4]
36. AL-Smadi M, Wesiak G, Guetl C, Holzinger A. Assessment for/as Learning: Integrated Automatic Assessment in Complex Learning Resources for Self-Directed Learning. In: Proceedings of the 2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems. 2012 Presented at: CISIS'12; July 4-6, 2012; Palermo, Italy p. 929-934. [doi: 10.1109/CISIS.2012.210]
37. Hersey P, McAleer S. Developing an e-learning resource for nurse airway assistants in the emergency department. Br J Nurs 2017 Feb 23;26(4):217-221. [doi: 10.12968/bjon.2017.26.4.217] [Medline: 28230439]

38. Trelease RB. From chalkboard, slides, and paper to e-learning: How computing technologies have transformed anatomical sciences education. Anat Sci Educ 2016 Nov;9(6):583-602. [doi: 10.1002/ase.1620] [Medline: 27163170]
39. Baumgart DC, Wende I, Grittner U. Tablet computer-based multimedia enhanced medical training improves performance in gastroenterology and endoscopy board style exam compared with traditional medical education. Gut 2016 Mar;65(3):535-536. [doi: 10.1136/gutjnl-2015-309895] [Medline: 26123026]
40. Zhan X, Zhang Z, Sun F, Liu Q, Peng W, Zhang H, et al. Effects of improving primary health care workers' knowledge about public health services in rural China: a comparative study of blended learning and pure E-learning. J Med Internet Res 2017 May 1;19(5):e116 [FREE Full text] [doi: 10.2196/jmir.6453] [Medline: 28461286]
41. Bollinger R, Chang H, Jafari R, O'Callaghan T, Ngatia P, Settle D, et al. Leveraging information technology to bridge the health workforce gap. Bull World Health Organ 2013 Nov 1;91(11):890-892 [FREE Full text] [doi: 10.2471/BLT.13.118737] [Medline: 24347719]
42. 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 4;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]

Abbreviations

CG: non–e-learning control group
CMB: China Medical Board
e-learning: electronic learning
EG: e-learning group
GI: gastrointestinal
GRS: Gastroscope Roaming System
HMU: Harbin Medical University
UGI: upper gastrointestinal

Edited by G Eysenbach; submitted 18.09.19; peer-reviewed by R Khan, S Sarbadhikari, A Holzinger; comments to author 09.10.19; revised version received 12.02.20; accepted 22.02.20; published 23.03.20.

Please cite as:
Li S, Li G, Liu Y, Xu W, Yang N, Chen H, Li N, Luo K, Jin S
Development and Assessment of a Gastroscopy Electronic Learning System for Primary Learners: Randomized Controlled Trial
J Med Internet Res 2020;22(3):e16233
URL: http://www.jmir.org/2020/3/e16233/
doi:10.2196/16233
PMID:32202507

Original Paper

Serious Gaming During Multidisciplinary Rehabilitation for Patients With Chronic Pain or Fatigue Symptoms: Mixed Methods Design of a Realist Process Evaluation

Miel AP Vugts[1], PhD; Aglaia MEE Zedlitz[2], PhD; Margot CW Joosen[1,3], PhD; Hubertus JM Vrijhoef[4,5], PhD

[1]Tranzo Scientific Centre for Care and Wellbeing, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands
[2]Leiden Institute for Brain and Cognition, Department of Health, Medical and Neuropsychology, Leiden University, Leiden, Netherlands
[3]Department of Human Resource Studies, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands
[4]Department of Patient & Care, Maastricht University Medical Center, Maastricht, Netherlands
[5]Panaxea, Amsterdam, Netherlands

Corresponding Author:
Miel AP Vugts, PhD
Tranzo Scientific Centre for Care and Wellbeing
Tilburg School of Social and Behavioral Sciences
Tilburg University
PO Box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 134664294
Email: m.a.p.vugts@uvt.nl

Abstract

Background: Serious gaming could support patients in learning to cope with chronic pain or functional somatic syndromes and reduce symptom burdens.

Objective: To realize this potential, insight is needed into how, why, for whom, and when it works in actual treatment circumstances.

Methods: Following a realist approach, process evaluations were performed before, during, and after a two-armed, natural quasi-experiment (n=275). A group of patients with interfering chronic pain or fatigue symptoms received a short additional blended mindfulness-based serious gaming intervention during a multidisciplinary rehabilitation program. A control group only received the regular rehabilitation program. During two sessions before and one session after the experiment, expectations about serious gaming processes were discussed in focus groups with local care providers, implementers, and experts. Patients participated in a survey (n=114) and in semistructured interviews (n=10). The qualitative data were used to develop tentative expectations about aspects of serious gaming that, in certain patients and circumstances, trigger mechanisms of learning and health outcome change. Hypotheses about indicative quantitative data patterns for tentative expectations were formulated before inspecting, describing, and analyzing—with regression models—routinely collected clinical outcome data. An updated program theory was formulated after mixing the qualitative and quantitative results.

Results: Qualitative data showed that a subset of patients perceived improvement of their self-awareness in moments of daily social interactions. These results were explained by patients, who played the serious game LAKA, as a “confrontation with yourself,” which reflected self-discrepancies. Important characteristics of serious gaming in the study’s context included innovation factors of relative advantage with experiential learning opportunity, compatibility with the treatment approach, and the limited flexibility in regard to patient preferences. Perceived patient factors included age and style of coping with stress or pain. Learning perceptions could also depend on care provider role-taking and the planning and facilitating (ie, local organization) of serious gaming introduction and feedback sessions in small groups of patients. Quantitative data showed very small average differences between the study groups in self-reported depression, pain, and fatigue changes (-.07<beta<-.17, all 95% CI upper bounds <0), which were mediated by small group differences in mindfulness (beta=.26, 95% CI .02-.51). Mindfulness changes were positively associated with patient involvement in serious gaming (n=114, beta=.36, *P*=.001). Acceptance of serious gaming was lower in older patients. Average health outcome changes went up to a medium size in patients that reported lower active coping with stress and lower pain coping before serious gaming. Mindfulness changes and gaming acceptance perceptions covaried with group structure and immediate feedback sessions after serious gaming.

Conclusions: This study developed transferable insight into how and why serious gaming can facilitate additional learning about coping in order to reduce burdens of chronic pain or fatigue symptoms in certain patients and in actual treatment circumstances. Future studies are needed to continue the development of this fallible theory. Such research will further support decisions about using, designing, allocating, and tailoring serious gaming to optimize important patient health benefits.

Trial Registration: Netherlands Trial Register NTR6020; https://www.trialregister.nl/trial/5754

(J Med Internet Res 2020;22(3):e14766) doi:10.2196/14766

KEYWORDS

serious gaming; eHealth; chronic pain; medically unexplained symptoms; implementation; realist evaluation

Introduction

Background

Serious games could someday have substantial health impact by facilitating patients in learning to cope with chronic pain (CP) or functional somatic syndrome (FSS). A serious game is a kind of computer game that not only aims to provide *fun*, but also to inform, instruct, and modify [1]. In many patients with CP or FSS, functioning can improve with the moderation of psychosocial consequences or perpetuating factors of their symptoms [2,3]. Increasing evidence suggests that serious gaming can facilitate processes of change in health behaviors and their antecedents, as well as in mental health in patients of any age [1,4]. Randomized controlled trials with *stand-alone* serious gaming interventions have been reported [5,6]. The findings suggested clinically relevant mental health effects compared to waiting-list conditions, or an equivalent effect compared to *traditional* face-to-face delivery modes. However, complementary evaluation studies of strong internal and external validity were suggested [1,4]; they are needed to inform and legitimize the implementation of serious games on a large scale, which, assuming that development costs are high and variable costs low, is a plausible requirement to realize value for societies.

This study aims to develop theory with which future developers and implementers of serious games can improve their design and integration in regular health care settings. To achieve this, realist process evaluations were embedded into a quasi-experiment with the serious game LAKA.

Burdens of Chronic Pain and Functional Somatic Syndromes

CP is pain, with or without a specific organic cause, that persists longer than a usual 3- to 6-months of organic recovery [3]. An FSS is characterized by a persistent pattern of bodily symptoms that cannot be sufficiently pathologically explained after adequate physical examination [2]. Close to 1 in 5 adults has intense pain (ie, rating 5/10 or more, for at least 6 months) [7-9]. FSSs characterized by irritable bowels (11%), chronic fatigue (1%), tension headache (2%), or tinnitus (10%-15%) are prevalent as well [10-12]. The severity and chronicity of lower back pain, which is the most common symptom of both CP and FSS, may be associated with more functional disturbance during people’s lifetime than any other disease [13].

CP and FSS often come with psychological or social burdens. Comorbid mental disorders were found in 35% of patients with CP [14]. Similarly, substantial minorities of patients with FSS also have another FSS or a psychological disorder [15,16]. Compared with individuals without pain, those who experience more than 3 months of disabling pain (7.4%) are about six times more likely to be absent from work and two times more likely to visit a doctor [17]. In addition, economic costs of sick leave and early retirement exceed the total amount of medical expenditures associated with both CP and FSS [18-20].

Biopsychosocial Management

For managing burdens of CP and FSS, biopsychosocial approaches are considered effective [2,3]. Meta-analyses show positive effects for physical and emotional functioning of various treatment options (ie, medication, complementary medicine, psychological therapy, or multidisciplinary rehabilitation) that are small or medium sized, at most [21-28]. Recommended treatment includes steps of conservative medication, psychotherapy, and physiotherapy, and a multi- or interdisciplinary rehabilitation program when previous treatments do not suffice [2]. Multidisciplinary rehabilitation programs vary in content but commonly include medical, psychological, physical, and occupational interventions [21,26]. New knowledge, with which treatment effectiveness for different CP and FSS patients may be improved, comes from research into the following: (1) the biological mechanisms of specific symptom patterns [29] and (2) how *adequately* (eg, feasibly or cost-effectively) already understood biopsychosocial mechanisms are targeted by certain treatment approaches, strategies, and modes of delivery [30,31]. Regarding the latter, current interest goes to the potential of facilitating intervention with computer technology [32,33].

How and Why Biopsychosocial Approaches Work, for Whom, and When

CP and FSS conditions are diverse, but a common, partial, explanation is given by the sensitization of the central nervous system [34]. Biopsychosocial approaches target the following: (1) peripheral systems with pharmacotherapy (ie, *bottom-up* approaches) as well as (2) cortical brain systems with neuroscience education, cognitive behavioral therapy, and exercise (ie, *top-down* approaches). Changes in coping with stress and pain, avoidance beliefs, rumination, acceptance, or catastrophizing mediate intervention effects on experienced physical and mental functioning [35,36]. An approach for changing coping responses is to promote individual mindfulness (ie, self-awareness, self-regulation, and self-transcendence) in response to adverse experiences (eg, negative emotions or physical sensations) [37]. Another approach is to restructure particular cognitive antecedents of pain-related experiences and behaviors [38-41]. Whether and how these approaches differ with respect to underlying change mechanisms and outcomes has been debated [38,39].

Literature about patient factors for effectiveness provide limited evidence to inform practice. Some studies claimed that neither demographic nor psychological differences between patients predict clinically important variation in treatment effects [42,43]. Other studies, however, stressed that fear-avoidance beliefs, pain acceptance, or depressive symptoms can predict treatment gains to some extent and should therefore be targeted early or additionally in treatment [44-47].

More insight is also needed about the characteristics of interventions and treatment contexts that are responsible for varying intervention effects. Meta-analytic results showed that computer-based intervention (eg, delivery over the Internet) can provide modest chances for symptom reduction, similarly to face-to-face group therapy of similar content [30]. It also showed uncertainty about the transferability from results in self-selecting participants to wider populations and settings. Researchers have been repeatedly suggesting that better explanations of varying effectiveness levels between studies requires *contextualization*; for instance, consideration of (1) factors such as fidelity of implementation, program compositions, and comparisons and (2) interdependencies and interactions between such factors [31,44,47-49].

How and Why Serious Gaming May Work, for Whom, and When

For serious games, there is a similar need for theories of what works how, for whom, and when [50,51]. Games or virtual reality have motivational qualities that may support the extinction of phobias, distraction from pain, repetitive physical or cognitive training, or learning about cognitive antecedents (eg, self-efficacy) of health behavior [52-54]. Gaming features, including stories and interactivity, may strengthen behavioral change processes with engagement, intrinsic motivation, positive affect, and sense of presence when processing educational content [52,55]. Various studies showed that effects of serious games can vary with intervention factors (eg, participatory design and duration) and users (eg, gender, age, intelligence, and gaming experience) [1,51,56]. Individual experiences of serious gaming may be difficult to predict and may be context dependent [57,58]. Debriefing is a way of dealing with such variation by leveraging recipient experiences after serious gaming for transferrable individual learning results as intended [59].

Effectiveness of Serious Gaming During Treatment for Adults With Chronic Pain or Functional Somatic Syndrome

To the best of the authors' knowledge, the first outcome evaluation of serious gaming for reducing disease burden in patients with CP or FSS was reported [30,60]. The effect of serious gaming was studied in a quasi-experiment with patients with mostly chronic musculoskeletal pain and psychosocial problems. A comparison was made of symptom changes during a regular multidisciplinary rehabilitation program (100 hours) between (1) an intervention group that received an additional *blended* intervention with the game LAKA (4 extra hours) and (2) a control group that received no serious gaming. A very small acceleration of physical and emotional symptom reduction was attributed to serious gaming: a standardized regression coefficient for the group × time effect of 0.12 [60]. This effect size estimate corresponds with a comparable estimate from an earlier meta-analytic subgroup analysis (ie, a 0.13 standardized average difference in short-term behavioral or mental health outcomes between groups following multi-component interventions *with* and *without* serious gaming) [1]. This suggests that a couple of hours of serious gaming as part of treatment does not provide a general clinically relevant benefit, per se, but can add to the effectiveness of the treatment as a whole. Plausibly, serious games and other computer applications for CP or FSS patients reach stronger, average effects when delivered as stand-alone interventions in other circumstances (eg, when compared with no active treatment) [1,30].

Existing Theories to Inform Intervention Theory

Existing theories on related topics provide a starting point for building transferrable insights into how and why features of a serious gaming could work in certain contexts for patients with CP or FSS (see Textbox 1). Changes of rehabilitation outcomes due to serious gaming are potentially explained with theories about health behavior, including mindfulness [37], relational framing [61], psychological well-being [62], or self-discrepancy theory (SDT) [63]. Why serious gaming features could be a distinctive trigger for learning has been explained with theories about motivational and affective responses to virtual or computer gaming environments [52,64]. Other potentially applicable theories described a comprehensive range of potential context factors for implementing health care innovations [65,66]. More research upon which theory building rested is summarized in Textbox 1 [67-71].

Textbox 1. Candidate theories for explaining effects of serious gaming: potentially applicable, formal explanatory frameworks.

Rehabilitation mechanisms:

- One theoretical model aligned particularly well with initial developer ideas about the intervention mechanisms [37]. Mindfulness covers several mental training practices and their consequences for cognitive functioning. These consequences are described by self-awareness, self-regulation (ie, the ability to change one’s own behavior), and self-transcendence (S-ART) (ie, prosocial characteristics). Three subtypes of practices are distinguished. Two of these are explicitly instructed and encouraged in LAKA (ie, focused attention and open monitoring), and one (ie, ethical enhancement) was integrated into immersive simulation tasks (ie, virtual social interactions or encounters). The S-ART model describes how these three types of practices influence the brain networks that support intention and motivation, attention regulation, emotion regulation, extinction and reconsolidation, prosociality, nonattachment, and decentering. From this theory, it is plausible to suggest that the features of a game such as LAKA can support an introduction to mindfulness practice for novices, rather than a tool for long-term exercise.
- Relational frame theory [61] underlies third-wave cognitive behavioral change approaches, such as acceptance and commitment therapy. It explains the harmful outcomes of cognitive fusion (ie, fusing thoughts with reality) and experiential avoidance (ie, attempts to avoid thoughts, feelings, memories, physical sensations, and other internal experiences) in functional contexts.
- Eudaemonist approaches to psychological well-being [62]. The rehabilitation approach, and the intended role of serious gaming in it, is to direct attention toward participation in social roles and psychological well-being rather than to necessarily change particular symptom-related cognitions or behaviors. With tasks in LAKA, players are challenged—or enabled to experiment with—vicarious prosocial behavior. Intrinsically or autonomously motivated prosocial behavior is known to promote psychological well-being [67].
- Self-discrepancy theory (SDT) can also be a useful explanatory framework for serious gaming effects. This was also hinted at in previous pilot results [68]. SDT predicts particular affective states (eg, dissatisfaction, disappointment, and agitation) from the degree and nature of discrepancy between a person’s actual self-state and a self-script (ie, ideal or ought to states from own or others' perspective) [63]. Self-discrepancies also partially explain comorbidity of depression and chronic pain (CP) [69,70]: when pain or fatigue symptoms become chronic, self-scripts conditional on the absence of those symptoms may be a source of emotional disturbance and maladaptive activity patterns. SDT was used to predict functional improvement in CP patients by motivating pursuit of possible selves or perceptions, such as interest, approval, or acceptance, that are not actually conditional to physical symptoms [71].

Mechanisms of computer gaming:

- SDT could also play a role in understanding responses to computer gaming. Previous research showed that perceived opportunity for realizing ideal selves drives engagement in computer games, and especially in individuals with larger actual-ideal self-discrepancies [64].
- In addition, various processes of motivation, affect, and immersion during serious gaming were considered to potentially strengthen learning or behavioral change effects [52].

Study Objectives

This study primarily aims to provide in-depth information to recipients, developers, evaluators, and implementers on contexts in which certain patients with CP or FSS can use features of serious gaming in ways that result in clinically important health benefits. The overall research question is as follows: (1) *How*, (2) *why*, (3) *for whom*, and (4) *when* does serious gaming lead to learning and health outcome change? Accordingly, the objectives are enhanced propositions about the following:

1. How do blended and mindfulness-based serious gaming interventions lead to additional learning and health outcome improvement?
2. Why do serious gaming interventions lead to additional learning and health outcome improvement? What generic, formal theory supports the explanation?
3. For which patients with CP or FSS are serious gaming interventions feasible and effective with respect to additional learning and health outcome improvement?
4. Under what circumstances of implementation in a regular multidisciplinary rehabilitation treatment are serious gaming interventions feasible and effective with respect to additional learning and health outcome improvement?

Methods

Process Evaluation Approach

The process evaluation uses a realist evaluation approach [72]. Empirical process evaluations are the preferred method to gain a thorough understanding of which intervention features contribute to clinically relevant benefit, how, why, for whom, and under which *complex* circumstances in real health care contexts [73,74]. Such insights can benefit future choices of development, evaluation, and implementation (eg, selective allocation, tailored design, and rollout in other settings). Realist principles can guide theory-based and contextual, sensitive, process evaluations of complex *programs* or *interventions*—terms are used interchangeably [73,75].

A *program theory* is developed on the basis of empirical evidence that indicates whether initial ideas are supported or must be refuted, extended, or refined. Program theories are built up from configurations (C) of intervention characteristics (I) in context (C) triggering a mechanism (M) that leads to certain outcomes (O) (ICMO configurations or ICMO-Cs). An intervention is something new to, or extracted from, a pre-existing situation, certain aspects of which interact with the intervention to elicit causal effects (ie, context). A *mechanism* is an underlying (ie, invisible) causal force of empirical events, that is, not a description of successive events under counterfactual situations [76]. Mechanisms have often been framed as the reasoning with which recipients respond to the resources of interventions [77], but ideas about what they are can vary with the topic of interest. The ICMO-C is also applied to conceptualize on a *middle range* of abstraction for "dealing with different spheres of behaviors and structures to transcend sheer description" [76]. Realist approaches use induction, to discover regularities in empirical phenomena, in alternation with deduction, to formulate expected observable consequences of abstract formal theories, for inferring the *best* explanation of observable outcome patterns (ie, *abduction*) [78].

Initial Program Theory

An initial ICMO-C was formulated by taking assumptions from several of the introduced theories on related topics (see Textbox 2 as well as Multimedia Appendix 1 for a comprehensive overview). In addition, it was expected that serious gaming would be more generally accepted and better adhered to after improving the delivery modes of serious gaming in the treatment context, as suggested by previous pilot study results [67].

Textbox 2. Initial intervention-context-mechanism-outcome (ICMO) configuration.

Intervention in context:

Mindfulness-based serious gaming delivered in blended form (ie, a combination of computer gaming and face-to-face guidance) as a standard component during multidisciplinary rehabilitation (ie, intervention in context) is accepted and adhered to by patients (ie, feasibility mechanisms) with a complex chronic pain or functional somatic syndrome condition (ie, patient in context).

Mechanism:

Serious gaming, as such, can provide complementary features, such as sounds, visuals, stories, or covert learning strategies, that trigger distinctive experiential, affective, or motivational qualities. This can include the degree or valence of affect or sense of presence (ie, gaming mechanism). These experiences strengthen learning results with respect to *mindfulness*, coping flexibility, or psychological well-being (ie, rehabilitation mechanism or intermediate outcome).

Outcome:

The learning results subsequently contribute to reductions in physical and emotional symptoms of pain intensity, fatigue, and depression (ie, rehabilitation outcome).

Mixed Methods Design

The process evaluation is designed as an embedded, two-armed, natural, quasi-experimental, mixed methods study [66]. It was carried out by an integrated team of researchers, who are trained in various quantitative and qualitative methods. Priority was given to quantitative methods for investigating patterns in routine clinical patient outcome assessments. Patients were recruited who were following a standardized 16-week multidisciplinary rehabilitation program, of 100 hours on average, in one of four treatment sites of a single Dutch rehabilitation center. At two sites, an additional 4-hour *blended* intervention with the mindfulness game LAKA was provided during the second half (weeks 9-12) of the rehabilitation program (ie, intervention condition). In the other two sites, gaming was not offered (ie, control condition). Measurements were routinely taken at baseline (t0), at an intermediate time point after 8 weeks of treatment (t1), and posttreatment after 16 weeks (t2). Nonintrusive qualitative data collection took place before, during, and after the experiment. Analyses of qualitative data for formulating hypotheses preceded the inspection and analyses of the quantitative outcome data (see addendum to the registration in the Netherlands Trial Register, NTR6020). All steps from the qualitative and quantitative research are presented in Table 1.

The alternative hypotheses formulated in step 2 specify certain patterns of linear relationships between variables that operationalize outcomes and characteristics of the intervention in context. Such statements were listed, a priori, to strengthen support, or refutations, of tentative ICMO elements. An ICMO-C gains support, or refutation, with quantitative results that should or should not be considered very unlikely under the assumption that the ICMO element is superfluous, as follows:

1. The effect of serious gaming on patient outcomes (ie, depressive mood is mediated by change in learning results, such as mindfulness).
2. Sense of presence and positive affect during serious gaming are positively related with changes in learning results and health outcomes.
3. Learning results, subsequent health outcome change, or game acceptance perceptions vary with patient-level factors of age, coping (ie, an active style and existing use of various alternate ways of pain coping), and room for health improvement (ie, psychological symptoms).
4. Learning results, subsequent health outcome change, or game acceptance perceptions vary with differences in the organization of sessions (ie, timing, presence of other patients, care providers, and changes of intervention in context over time).

The Adjudicating Formal Theory and Formulating Hypotheses section explains the ICMO elements to which the hypotheses relate.

Table 1. Steps of recruitment, data collection (steps A-G), and analysis (steps 1-5).

Protocol steps	Research activity
A (data collection)	Recruit stakeholders and perform focus group interviews (two sessions)
B (data collection)	Recruit patients
C (data collection)	Collect post-serious gaming feedback from professional and patient users
D (data collection)	Purposively select patients for semistructured interviews
E (data collection)	Perform semistructured interviews with patients
F (data collection)	Perform stakeholder focus group interview (third session)
1 (analysis)	In iteration with steps E and F: Code intervention (I), context (C), mechanism (M), and outcome (O) elements in all the qualitative information Describe ICMO relationships per individual patient interview Compare patient-level findings with focus group data collected from other stakeholders
2 (analysis)	Interpret mechanisms on the basis of formal theory (ie, adjudication) Formulate quantitatively testable hypotheses before outcome inspection
G (data collection)	Retrieve quantitative data from patient records
3 (analysis)	Describe quantitative data as a means to triangulate the qualitative results
4 (analysis)	Test hypotheses with statistical models
5 (analysis)	Mix the results of different methods: summarize how they are interpreted to support, refute, refine, or extend initial expectations Construct summary ICMO configurations (*middle-range* theory) based on findings of this study Propose a transferable program theory after comparison of findings from this study with those of previous studies

Setting and Participants

The setting and participants were fully described in the outcome evaluation report [59]. Treatment and control sites had similar protocols, sizes, histories, and absence of disruptive activities during the study period. Patients with a physicians' indication of eligibility for multidisciplinary rehabilitation received informed consent from a familiar care provider soon after the second part of rehabilitation treatment started. Participation included additional data collection for the study and permission to process codified, routinely collected, clinical patient data. A total of 275—156 intervention group and 119 control group participants—out of the 329 eligible patients participated (83.6%). Patients in the sample were, on average, 44 years of age (SD 11.3, range 18-67). The proportion of females was 69.8% (192/275). Almost half of the patients (134/275, 48.9%) reported a symptom duration of over 2 years. Patients (N=275) mostly had musculoskeletal pain and concomitant psychosocial problems. The findings of the outcome evaluation suggested no confounding influences by treatment site or patient baseline variables that differed, statistically significantly, in mean levels between the study groups.

Other participants included care providers (ie, three psychologists and one physiotherapist), local managers who were responsible for providing the serious gaming intervention on site, and stakeholders with relevant expertise in information and communication technology (ICT), serious gaming, rehabilitation medicine, health psychology, and spiritual counseling. All were familiar with the setting and provided informed consent on the participation in focus group sessions.

Interventions

Features of the multidisciplinary treatment and additional serious gaming intervention offered to the intervention group were previously reported in detail [66]. A short description of LAKA is given in Textbox 3. The clinic facilitated tablet computers, suitable rooms with Wi-Fi connections, and the automated planning of four 1-hour sessions, for 1-6 patients simultaneously, in connection to regular therapy hours. Three psychologists and a physiotherapist (two per site) provided support during the first (ie, introduction) and fourth (ie, debriefing) sessions. Topics of debriefing were technology acceptance, play experiences, and *learning transfer* to daily life. Access to LAKA during the second and third sessions was provided on site by local staff members. In addition, patients could play LAKA at home on a tablet computer with an Internet connection. This connection made it possible to download LAKA via an app store and to safely exchange data with the rehabilitation center for access to the app, storage of progress, and feedback on performance. All patients were expected to attend a debriefing session after completing the game at least once: this took 2.5 hours on average. This intended *dose* was based on what had been a natural amount to patients during a feasibility study [68].

Textbox 3. Short description of the serious game LAKA.

LAKA challenges patients to take the role of a virtual character (ie, avatar). The gameplay includes prompts for monitoring and evaluating *satisfaction* about selected *responses* in virtual social encounters. These optional responses are descriptions of implementation intentions for acts, phrases, and postures in social interaction scenarios that players can select for their avatar. This avatar represents themselves on a virtual trip around the world. Optional responses—five at each moment—vary in the degree of correspondence with the principles of *generosity*, *moral discipline*, *patience or forbearance*, and *enthusiastic perseverance*. Each selected response has salient (eg, emotional expressions) and realistic (ie, not predictable or moralistic) consequences for the avatar. These consequences are intended to evoke reflection. Neutral to positive indirect performance feedback reinforces the monitoring task. Moreover, scenarios prompt instructions for 3-minute focused attention or open monitoring (ie, meditation) exercises. Encouragement stimulates users to repeat the exercises at any convenient moments in daily life.

Qualitative Data

Interviews and feedback surveys contained open-ended questions, which were intentionally free from theoretical preconceptions. Such open-ended questions were posed first to give space for explanations of initial reactions, including unexpected ones. Topics and additional follow-up questions for interviews were based on the existing theoretical frameworks on related topics (see Introduction section). The prioritization of topics varied with the expected knowledge areas of the participants [79]. Participating patients were considered to be most aware of outcomes and characteristics of the intervention in context, while care providers knew most about rehabilitation mechanisms.

Focus groups were first held with stakeholders who represented various roles and areas of expertise. Participants prepared for the focus groups by playing LAKA and reading pilot study results and adapted intervention specifications. At the beginning of the session, the research questions were presented for collecting local experts' initial expectations. In *part two*, care providers and implementers built upon shared expectations for specifying local implementation procedures in further detail. The second and third focus group interviews were held with two care providers who worked at the same intervention site. At the beginning of the second focus group, initial expectations about intervention effects on rehabilitation mechanisms were explained. During the natural experiment, care providers also shared feedback informally with MAPV. The third focus group with care providers first generally addressed their postexperimental experiences. These were about how recipients responded to the intervention in the context, including how those responses were shaped by care providers' own reactions. The discussion was enriched by sharing tentative qualitative results (ie, parts of tentative program theory).

The first source of qualitative patient information consisted of the given responses on two open-text requests after finishing playing LAKA: “Please describe in your own words what you experienced when you played LAKA” and “Please explain how the LAKA sessions, according to you, will contribute to your daily life.” Furthermore, semistructured interviews were held with patients. These interviews were held with (1) intervention group patients who were purposively selected by differences in outcome expectations and (2) control group participants with similar baseline characteristics as interviewed intervention group patients. Interviews took place when patients had completed their rehabilitation program. Main topics of each interview were *what worked and how*, in order to change courses of symptoms and learning in the specific situations of the patient throughout rehabilitation (eg, “Could you explain how your health status developed throughout the rehabilitation program? What aspects of the rehabilitation program really made a difference for you? What was the role of serious gaming, if any, assuming differences occurred? What aspect of the game contributed like that?“). A data saturation point was reached when no new information emerged by asking patients about three *good* and three *bad* things about serious gaming. The more initial responses from recipients were already collected and analyzed, the more often MAPV shared an interpretation of a fragment in order to question its accuracy. If potentially relevant topics for the development of program theory were not addressed at the initiative of patients, they were sometimes introduced with questions that referred more directly to specific theory-based expectations (eg, ”Were you concentrating while playing LAKA?“).

Quantitative Data

The initial program theory informed the collection of quantitative data sources. Table 2 lists the times of assessment and the internal consistency information of all quantitative measurements that were processed for hypotheses testing. Health outcomes were measured with validated questionnaires on depressive mood, pain intensity, and fatigue as part of the routine clinical data collection [68,80-82]. *Learning result* was measured with a mindfulness questionnaire, of which unpublished psychometric results gave support for being a valid measure (see Multimedia Appendix 2, which contains information in support of a good internal consistency, low social desirability, and patterns of association with other constructs that are similar to those of other validated mindfulness questionnaires) [83,84].

For additional indications about potential mechanisms, a short *postgaming* feedback questionnaire with complementary scales was added to the routine clinical measurement system. This postgaming feedback questionnaire included Likert-scale items on sense of presence (ie, involvement and realism subscales) [85], positive and negative affect [86], and patient acceptance perceptions with regard to the serious gaming intervention [87,88].

Indicators of patient factors included clinical data about demographic, health status, and coping [89,90] variables as measured at the latest available time point before serious gaming. Classification norms from manuals of the questionnaires were used to divide patients over subgroups. Log data about the circumstances of the intervention in context were disaggregated to the patient level (see Table 2). This includes dummy variables for indicating each intervention site, the care provider who provided debriefing sessions, and the social *structure* of debriefing sessions: a care provider with a patient alone, a dyad, or a group. Moreover, differences between dates indicated coincidental variation by central planning and general changes of serious gaming sessions in time.

Table 2. Overview of quantitative data.

Classification as intervention, context, mechanism, or outcome, and theoretical construct	Variables (operationalization)	Time of measure (Cronbach alpha)
Rehabilitation outcome (health outcomes)[a]		
Depressive mood	Symptom Checklist-90 (SCL-90) depression subscale [80]	t0[b], t1[c], t2[d] (.91, .91, .91)
Pain intensity	Numerical rating scale (current) 0-100 [81]	t0[b], t1[c], t2[d] (N/A[e])
Fatigue	Checklist of individual strength [82]	t0[b], t1[c], t2[d] (.86, .95, .96)
Rehabilitation mechanism, serious gaming outcome, learning result		
Mindfulness	Sums of scores of three subscales for *mental stability*, *forbearance*, and *enthusiastic perseverance* behavior (eg, “Also in a turbulent environment, I can concentrate well” and “I remain patient until I see the solution”)	t0, t1, t2 (.94, .95, .95)
Gaming performance	Standardized values for responses (ordinal scales 1-5)[f]	Log data (N/A)
Serious gaming feasibility outcome, implementation fidelity		
Adherence	Progress: number of encounters completed	Log data (N/A)
	Attendance of serious gaming sessions	Log data (N/A)
Serious gaming mechanisms		
Involvement	Igroup Sense of Presence Questionnaire [85]	Postgaming (.76)
Realism	Igroup Sense of Presence Questionnaire [85]	Postgaming (.69[g])
Experiential qualities		
Positive affect	Positive affect scale: Positive and Negative Affect Schedule (PANAS)-short form [86] (1-5)	Postgaming (.86)
Negative affect	Negative affect scale: Positive and Negative Affect Schedule (PANAS)-short form [86] (1-5)	Postgaming (.80)
Serious gaming feasibility mechanism, intervention		
Game acceptance perceptions	The average score was taken from the following Likert-scale items[h]: From the UTAUT2[i] questionnaire [87]: • Perceived usefulness (“By following the LAKA module, I could achieve my health goals more quickly”); • Perceived ease of use (“Learning how to use LAKA was easy for me”); • Perceived trust (“The LAKA module offers services in my best interest”); and • Perceived enjoyment (“Following the LAKA module was enjoyable”). From the EgameFlow questionnaire [88]: • Clear goals (“Overall goals of LAKA were presented clearly”); • Challenge (“LAKA provides different levels of challenges that tailor to different players”); and • Perceived learning (“The LAKA module increased my knowledge”).	Postgaming (.83)

Classification as intervention, context, mechanism, or outcome, and theoretical construct	Variables (operationalization)	Time of measure (Cronbach alpha)
Outcome expectations	"Use the following slider (0-10) to indicate to what extent you expect that the LAKA sessions contribute to your daily life"	Postgaming (N/A)
Patient factors		
Demographics[j]	Age in years (>45=high)	t0 (N/A)
	Female or male	t0 (N/A)
	Socioeconomic status by neighborhood	t0 (N/A)
Coping with stress and pain	Utrecht Coping List [89]. Active coping scale (>20=high)	t1 (.81)
	Pain Coping and Cognitions List [90]. Pain coping scale (low=a standardized residual score <-1 after regression on relevant factors: gender with male as reference [beta=.14, *P*=.03] and socioeconomic status [beta=.16, *P*=.02])	t1 (.85)
Psychological distress	SCL-90. Total score (>132 is high or very high)	t1 (.97)
Intervention in context		
Setting (treatment site)	Four dummy variables: • Site 1; • Site 2; • Site 3; and • Site 4.	t0 (N/A)
Intervention	Identifiers of intervention versus control group (ie, dummy); intervention (site 1-2) versus control (site 3-4)	t0 (N/A)
Inner setting (provider)	Five dummy variables for the care provider from whom debriefing was received: • Care provider 1 vs other; • Care provider 2 vs other; • Care provider 3 vs other; • Care provider 4 vs other; and • Stand-in care provider.	Log data (N/A)
Implementation (development of quality over time)	Introduction session dates	Log data (N/A)

Classification as intervention, context, mechanism, or outcome, and theoretical construct	Variables (operationalization)	Time of measure (Cronbach alpha)
Planning	Differences in days between the following: • Intermediate outcome assessment and introduction session; • Introduction and debriefing sessions; • Debriefing and posttreatment assessment; and • Last time playing LAKA and debriefing (short <3 days).	Log data (N/A)
Social *structure* (of serious gaming sessions)	Three dummy variables on the numbers of patients who were simultaneously attending debriefing sessions (as determined by planning): • The patient was in a group (2 or more other patients were present); • The patient was alone (ie, one-on-one debriefing by a care provider); and • There was 1 other patient (dyad).	Log data (N/A)

[a] This report omits a primary outcome measure for pain catastrophizing, as previous outcome evaluations found no indication that variance in this outcome was attributable to serious gaming.

[b] t0: at baseline.

[c] t1: at intermediate time point after 8 weeks of treatment.

[d] t2: after 16 weeks of treatment.

[e] N/A: not applicable.

[f] Log data included log ins, selections made by patients for their avatar (ie, gaming performance), and last completed encounter. These were automatically collected through a secure Internet connection for saving progress and providing performance feedback. Elucidating whether game scores are valid indicators of learning is a technical challenge warranting focused research attention.

[g] Internal consistency of this scale was considered insufficient; therefore, sensitivity analyses were performed in which scores were replaced by the average of three scores, excluding one item, for which Cronbach alpha=.73. There were no relevant changes of the results.

[h] These items were selected because they had the strongest factor loadings within the scales to which they belong as established in a pilot study [68]. Together, the items formed an internally consistent scale.

[i] UTAUT2: second generation of the Unified Theory of Acceptance and Use of Technology.

[j] More information was retrieved for the evaluation project. Information on demographics that are only processed for the outcome evaluation and not for the present process evaluation, specifically, were excluded.

Qualitative Analyses (Steps 1 and 2)

All interview texts (ie, verbatim transcripts) were coded by MAPV in ATLAS.ti, version 7.5.16 (Scientific Software Development GmbH), using sensitizing concepts for intervention, context, mechanism, and outcome unless new codes were needed to cover *unexpected* statements. The first four patient interviews were also independently coded by MCWJ (two interviews) and AMEEZ (two interviews). Unresolved differences were discussed with HJMV. It was counted how often codes for open-feedback text fragments recurred to get a sense of the relative importance of various themes and coverage of those by the interview data.

In step 1, ICMO relationships were determined per patient interview when explicit statements were found about relationships between positive outcome expectations, mechanisms, and the intervention in context (see Multimedia Appendix 3 for illustrations). In addition, intervention or context factors were coded as *barriers* when they were explicitly related to suboptimal, weak, or absent activation mechanisms or outcomes. Tentative ICMOs, documented down in theoretical notes, summarized regularities that were found when comparing individual patient results with each other and with the qualitative data that were obtained from other stakeholders. Theoretical notes formed the basis for regular bilateral or team discussions that continued during the next step.

In step 2, decisions were made about which formal theoretical perspectives were most suitable for explaining the qualitative data about intervention mechanisms. This was a starting point for adding, actually resuming, a deductive process to specify indicative patterns of quantitative data (ie, observable effects) of tentative ICMOs. Hypotheses were formulated that were testable with the available (ie, uninspected) quantitative data. Formulations of hypotheses had to take into account the limitations of the sample size, avoiding too many or too complicated hypotheses, and a lack of possibilities to add new quantitative measurement instruments at that time. The research protocol was ethically assessed before data collection began.

Quantitative Analyses (Steps 3 and 4)

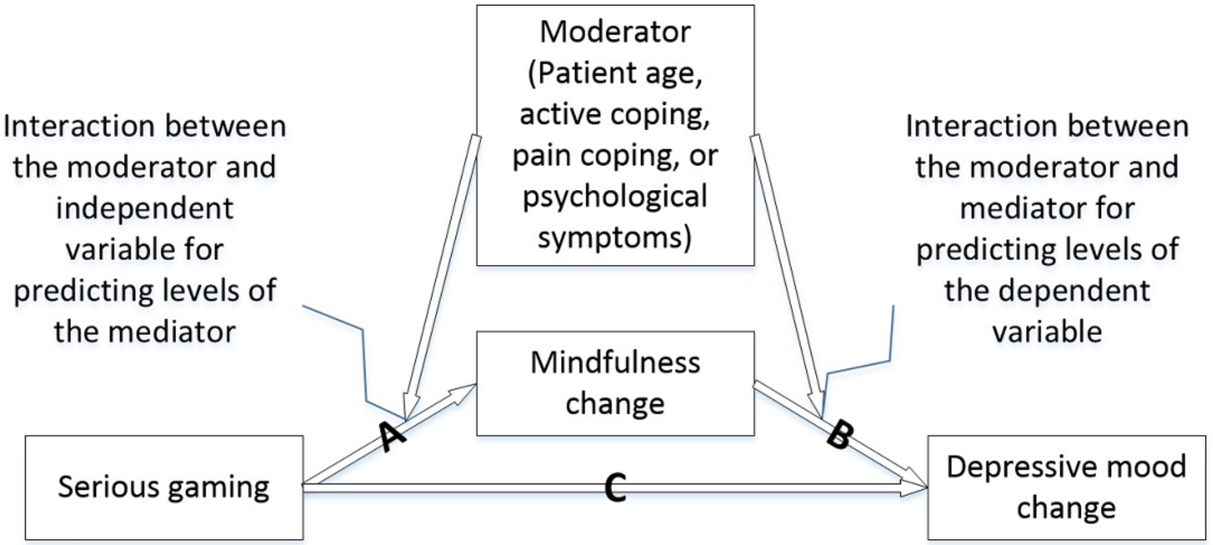
The ggplot2 package from R, version 1.1.463 (The R Foundation), was used to visualize levels of change in health outcomes, change of mindfulness, and game acceptance levels between subgroups of patients in context. Provisional classifications of *favorable* or *not favorable* were made with combinations of variables about the following: (1) whether a patient received the serious game intervention and how the implementation went, individually, and (2) his or her pre-existing characteristics.

All prediction models were executed in SPSS, version 24.0 (IBM Corp), using applications of the ordinary-least-squares estimation algorithm. All models were based on complete cases. All variables, except for dummy variables and dates, were centered or standardized before inclusion in statistical models. Outcome change indicators are standardized residual scores after regression of postintervention (t2) scores on preintervention (t1) scores. Decisions on the null hypotheses are informed by two-sided *P* values and 95% CIs. Various sensitivity analyses were performed to indicate the robustness of the methodological choices, including missing data handling. These analyses are discussed among the strengths and limitations.

The first type of procedure applied Hayes' PROCESS macro, version 2.16, for SPSS to investigate hypotheses 1 and 3 [91]. These procedures calculate CIs using a nonparametric bootstrap approach that, by default, generates 5000 samples. Hypothesis 1 was investigated by fitting models of *type 4*. These models specified estimations for effects of study group on changes of the three health outcomes as mediated by mindfulness changes. Mediation models were extended to moderated mediation models of *type 58* for testing hypothesis 3 (see Figure 1). These models calculated the following: (1) parameters for the interaction between the moderating patient factors and study group for predicting levels of the mediator (ie, mindfulness change), (2) parameters for the interaction between the moderating patient factors and the mediator for predicting levels of the health outcomes, and (3) an index of moderated mediation assessing differences in the total indirect effects between two patient subgroups.

Secondly, multivariable linear regression models were used. A model with mindfulness as the dependent variable, and the subscales of sense of presence and affect as the independent variables, provided relevant test statistics to decide on hypothesis 2. A similar regression model, with game acceptance perceptions as the dependent variable, and patient factors (ie, continuous) as the independent variables, informed, in part, hypothesis 3. Three sets of regression models were used to test whether the variables for *intervention* or *context* factors, as the independent variables, affected the variables of (1) game acceptance, (2) mindfulness change, or (3) depression change. To prevent too many factors from being added to a single model, *intervention in context* variables were added and removed one by one after controlling for patient age and active coping style. Only the planning variables (ie, date differences) were added and removed simultaneously to isolate unique influences, taking interdependency into account.

Figure 1. Illustration of the moderated mediation models.



Mixing the Results of Different Methods (Step 5)

The final middle-range ICMO-Cs were constructed by assuming elements of the initial ICMO-Cs, refinements, and extensions based on the results from steps 1 and 2 that were not refuted based on the quantitative results from steps 3 and 4. The final conclusions are based on comparisons of the results, with respect to the elements of the ICMO-C, between this study and previous studies.

Results

Participants

Postgaming feedback data were collected from 114 out of 156 (73.1%) patients of the intervention group. Nonresponders (mean 40.6 years old, SD 11.7, *P*=.02) were relatively younger than responders (mean 45.5 years old, SD 11.2) and completed, on average, 1.5 more encounters (15.1 versus 13.6) in the game LAKA (*P*=.02). No other statistically significant differences were observed between these groups. Two care providers participated in the first focus group, together with a rehabilitation physician and serious gaming expert, a spiritual counselor, an executive and ICT expert, and two managers. The four care providers—three psychologists and one physiotherapist—who regularly facilitated introduction and debriefing sessions participated in the second and third focus groups. The 8 interviewed patients from the intervention group reported differences in the conditions in which they received serious gaming, their demographics, intervention experiences, and outcome levels (see Table 3). All patients reported having at least completed a secondary education. Recruitment of control patients was stopped after two, short, 10-15-minute interviews, because the interviews did not provide useful data.

Table 3. Characteristics collected from interviews with patient participants from the intervention group.

Characteristic	Patient number							
	1	2	3	4	5	6	7	8
Age in years	51	34	46	45	55	47	56	55
Gender (female=1; male=0)	1	0	0	1	0	1	0	1
Highest education level[a]	MV[b]	2	3	3	2	3	3	MV
Encounters completed (range 0-28)	14	15	24	28	12	16	16	1
Site ID	1	2	2	2	1	1	2	2
Symptom Checklist-90 (SCL-90) depression score (range 16-80): t1 to t2[c] decrease	2	4	17	7	17	32	12	1
Mindfulness (range 49-245): t1 to t2 increase	6	10	24	11	27	35	28	43
Perceived outcome (range 0-10)	9	8	7	6	5	5	1	0
Negative affect (range 0-20)	0	3	2	5	10	1	2	20
Positive affect (range 0-20)	17	18	11	19	6	12	3	0
Involvement (range 0-6)	6	4.5	4.5	4	3	4.75	2.25	3
Realism (range 0-6)	6	4	2.25	3.25	3.5	2.75	3	0
Group size debriefing (range 1-6)	1	5	2	1	6	5	2	3

[a] Highest education level: secondary (ie, high school level) as highest=2; tertiary (ie, college or university level) as highest=3.

[b] MV: missing value.

[c] t1: intermediate time point after 8 weeks of treatment (ie, before serious gaming); t2: after 16 weeks of treatment.

Results of Coding Qualitative Materials (Step 1)

Learning Results

In all interviews except for one, patients mentioned barriers—intervention or context—for serious gaming outcomes. ICMO-Cs emerged from the six out of eight interviews with patients that had moderate-to-high outcome expectations: the scores were 5-9 out of 10. Among all 114 survey respondents, 61 (53.5%) had at least moderate outcome expectations. Statements about outcomes were coded using sensitizing concepts about mindfulness (ie, a heightened self-awareness about certain dysfunctional mental states when reacting to situations of daily life, including loss of attention, rumination, rigidity, moodiness [sad, anxious, and irritable], automaticity, and prejudice). Also, heightened self-awareness for positive changes was noted (eg, self-regulation toward calm, alert, self-accepting, prosocial, or assertive reactions).

> *With the whole happening of LAKA ... [patient explains] ... which makes you approach and do things less rigid and or short-sighted.* [Patient #5]

Serious Gaming Mechanisms That Lead to Learning Results (Mechanism and Outcome)

Intervention group patients who elucidated outcomes also spontaneously recognized active intervention elements (ie, resources). They spoke about *encounters*, *reflections*, mental training instructions, and debriefing. These were mostly connected to a form of *reasoning* that was commonly described as being *confronted with oneself*. Explanations of this common experience included awareness of alternate response options, emotional consequences of those responses, nonautomaticity, effort to maintain focus of attention, and transfers of gaming experiences into situations of daily life, for example:

> *You are confronted actually ... that is where awareness begins. In everyday life you often have those situations in which you do not even realize that you can go left or right. And yes, with LAKA you really get that choice and then you really have to start thinking.* [Patient #2]

Particular qualities of experience during gaming, including positive affect and *involvement*, were more often described by patients who also described positive learning results. However, none of the patients attributed their learning results to these gaming experiences explicitly and spontaneously:

> *I was really into it, the journey across the world ... and you can completely forget the world around you ... Where does that help for? Maybe that when you're busy with something ... you're just really focused on doing it, and not being distracted ... that concentration.* [Patient #1]

Patients' reasoning seemed to correspond with care providers' expectations, before the experiment (ie, second focus group), that LAKA might stimulate self-reflection and behavioral adaptation by showing opportunity thereof. After the experiment, care providers (ie, third focus group) emphasized the necessity of debriefing for many patients, for transferring their experiences into desired learning results. Debriefing group discussions provided the opportunity for patients to express their experiences *in* LAKA or *about* LAKA. These are vicarious experiences, mediated by the avatar, versus the nonmediated ones, such as declarations about a lack of identification with the avatar or the liking or disliking of entertainment-oriented features. Such nonmediated experiences were expected in the first focus group, when participants critically and jokingly wondered why one would not use something else to trigger experiences as a basis for discussion (eg, offering pastry). After the experiment, care providers agreed that the best results are gained by discussing both the mediated and nonmediated experiences in debriefing, albeit depending on situational and patient needs:

> *Those mini-games in between, which ought to be less valuable: for reflection, you can get more out of that than from the encounters, because: I never go to Istanbul, and I do not like temples. So, I just clicked something. That mini-game is really stupid, I had to start all over again! You can reflect nicely on that ... Why were you not interested in Istanbul? What does that say about you, your daily life, and your symptoms?* [Health care provider]

Characteristics of Intervention in Context for Triggering Mechanisms

Patients made both positive and negative remarks about serious gaming as part of the treatment program. They agreed that learning mechanisms were compatible to other educational and psychotherapeutic approaches in the rehabilitation program. Two patients explained that the opportunity for experiential learning supported the learning transfer. This offers a relative advantage to regular, more text-based modalities. However, outcomes were regarded as suboptimal with regard to the ambiguous and noncompelling feedback. Other generally experienced barriers were the limited duration, design quality aspects, and personalization or inflexibility of LAKA, as follows: (1) response options are prescripted, (2) the virtual world and the activities in it are neither exhaustive nor adjusted to with respect to what individual patients find important or valuable in their lives, (3) too much time could be spent on *nonactive* elements, and (4) a third-person, instead of first-person, perspective was used. Some example quotes follow:

> *It [reflection] should be, as far as I am concerned, more in the game and immediately after the choices you have made. Let consequences being "lived through" and then get back on; if this, then what?* [Patient #7]

> *The fact that all the answers that are given [options for responding in encounters] did not apply to me; I found that very difficult.* [Patient #5]

> *You are traveling the world, but it is actually about life. In that world you encounter something that you do not meet at home on the couch. In that journey you can put all kinds of aspects of life ... Do you bring your partner? Are you going alone? Those are actually very essential choices. Why does that man [avatar] have to travel alone?* [Expert]

The central planning of sessions by the clinic elicited varying views about the timing and presence of other patients in serious gaming sessions. One expert (first focus group) expected it to be better when sessions took place within a short period of time (ie, a maximum of 2 weeks). Patients clearly preferred that debriefings took place shortly after playing. This could matter for *confrontations* because recall of relevant serious gaming experiences gets more problematic over time. Some patients argued that when serious gaming was provided earlier, it would have helped more in combination with previous program elements. Others appreciated that gaming was not introduced earlier because its rationale would be more difficult to understand, or use of the computer-based modality would be considered more burdensome. One patient liked to have a debriefing with a care provider one-on-one, while other patients found it interesting to experience serious gaming (eg, performing meditation) in a room with other patients and to hear about the experiences of others.

> *In one scenario or one session, quite a lot happens: all the choices that people make. You could have concluded each session with an evaluation and reflect on what they just did. In the last session, it was not always clear what had happened and what they had experienced.* [Health care provider #2]

Before the experiment, care providers could only speculate about patient factors. They considered that processes for restoring social roles and pain acceptance throughout rehabilitation are too heterogeneous for making general predictive statements. After the experiment, both patients and care providers expected that a younger age, habit, or self-efficacy regarding technology usage contribute to positive patient expectations and experiences. Patients who expected benefit from serious gaming explained that they became aware of dysfunctional cognitions, moods, or behaviors. Respondents with low outcome expectations considered that they did not have emotional problems, or could control them, and focused on other means for handling their pain and other symptoms (eg, physical exercise) at that time. Care providers came to consider that patients who experience more difficulties and show a greater need for guidance when using LAKA (ie, to get in control and transfer experiences toward learning results) could be those who may benefit the most from it. This seemed consistent with what patients admitted about themselves, albeit in less explicit terms, for example:

> *When I first came there, and received such a tablet computer, games and such on computers really were not my thing. So, to me it was all "abracadabra" what happened. I have been fighting with myself for the first 50 minutes; what am I supposed to do with this? And then you try something.* [Patient #5]

To care providers, patient expressions of difficulties encountered when using LAKA were also useful input for debriefings. When guiding patients, care providers had prioritized reflection over convenience (eg, by offering practical information and assisting with forgotten passwords, computer and Internet settings, and game controls). A care provider explained how role performance in guidance could be influenced by an ambiguous attitude toward the compatibility of computer-mediated learning and personal work styles:

> *I think our role is to say at the beginning: "I am not going to tell you much. You get started, then I give information about how the device works, and eventually we will evaluate it" ... I must be self-critical: I was motivated to provide those sessions but skeptical, because games are not my thing. That is not good, because that influences your performance. Giving people little information has helped me not to be influenced.* [Health care provider #1]

Adjudicating Formal Theory and Formulating Hypotheses (Step 2)

With regard to why serious gaming can enhance learning outcomes, *confrontation with oneself* was identified as a necessary form of reasoning. Based on this finding, SDT was chosen as an appropriate general theoretical framework for understanding underlying motivational and affective processes of learning from serious gaming by patients.

Various hypotheses were formulated and tested with available quantitative data for additional indications about the validity of ICMO elements. Hypothesis 1 focused on establishing a quantitative data pattern indicating the initially expected and qualitatively consistent premise that *mindfulness* is a rehabilitation mechanism, or interim learning outcome, that can be triggered by a serious gaming intervention. Hypothesis 2 was aimed at determining unique covariation between measures of learning results and potential gaming mechanisms, including valence of affect and sense of presence. Although it was initially considered possible to indicate mechanisms with a simple linear relationship between learning and valence of affect, there was no clear basis for this in qualitative results or in the SDT. A positive relationship between the subjective experience of involvement during play and mindfulness was hinted at in the qualitative data, but without explanation of why such a pattern would be generated. Hypothesis 3 focused on determining patient subgroups for whom particularly *good* or *bad* outcomes are attributable to serious gaming. Such an empirical pattern was seen as an indication that the self-discrepancies that some of the patients become aware of through serious gaming, that give reason to change behavior, may be absent in other patients. The latter being patients who find themselves free of unsolvable emotional stress, or assume that they can cope with their pain in different ways (ie, no current demand for mindfulness intervention). At the same time, individual inclinations to experience a lack of control could have a negative influence on the use of serious gaming and its results. Hypothesis 4 focused on remarkable covariation between perceived outcomes and factors of serious gaming intervention in context.

Quantitative Analyses (Steps 3 and 4)

Overview

Figures 2 and 3 visualize scores for mindfulness change, depression change, and postgaming acceptance perceptions, separately, between contexts that were tentatively considered as *favorable* or *not favorable*. Relatively steep changes in mindfulness and depression scores in the *right* directions (see Figure 2) and more stable acceptance levels (see Figure 3) were visible for subgroups of patients of the intervention group with *favorable conditions* of the patient—at most, 45 years of age; *high* psychological symptoms; and not a high level of active coping or pain coping—or the intervention in context—debriefing is received in a group within 4 weeks after introduction and, at most, 2 days after completing at least 50% of a complete playthrough in LAKA.

Figure 2. Outcome changes by conditions of the patient and intervention in context. Full intervention-in-context (IC) fidelity: debriefing is received in a group within 4 weeks after introduction and, at most, 2 days after completing at least 50% of a complete playthrough in LAKA. Favorable patient condition before gaming: at most, 45 years of age, high psychological symptoms, and active or pain coping levels are not high.

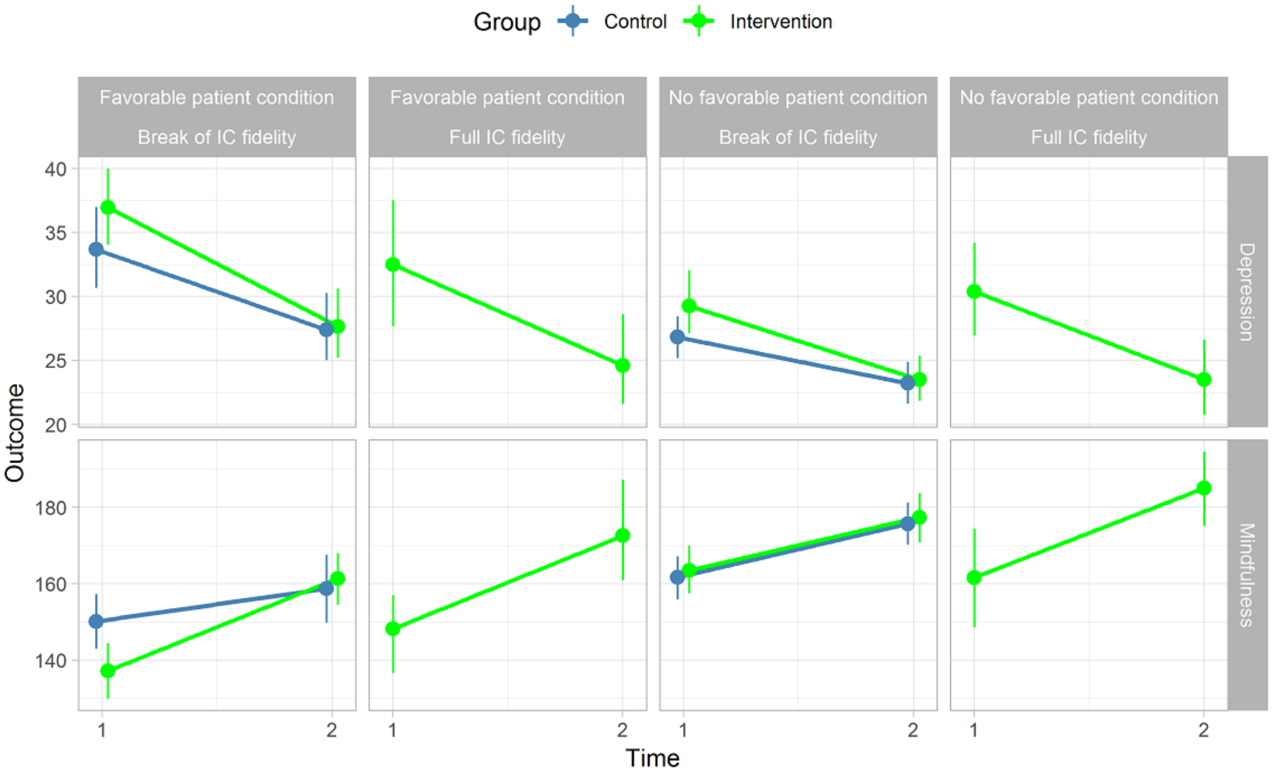
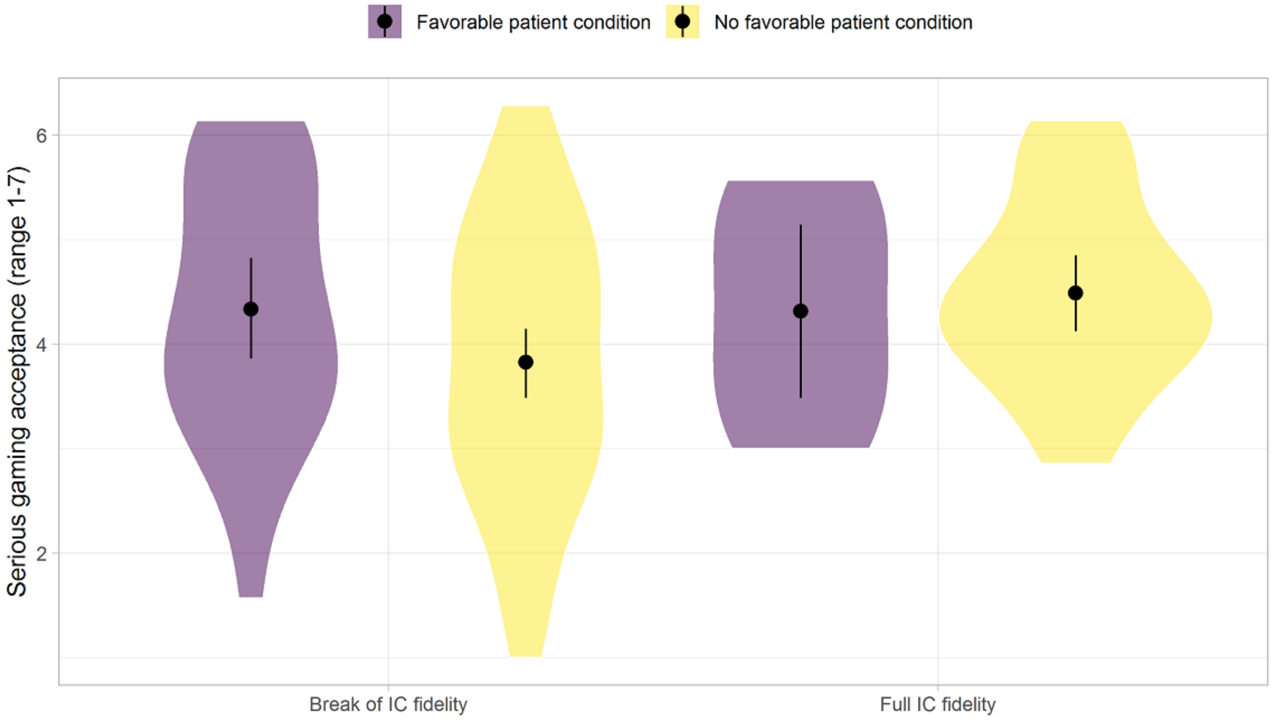


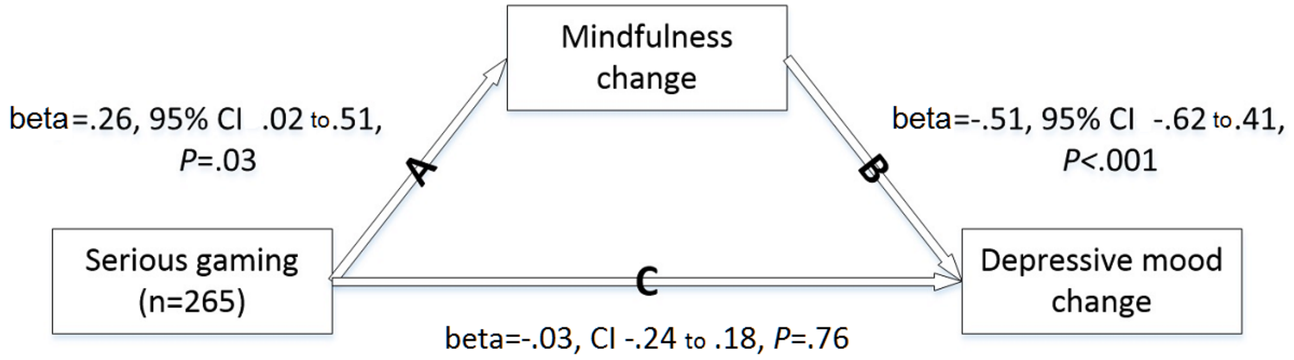
Figure 3. Gaming acceptance perceptions by conditions of the patient and intervention in context. Full intervention-in-context (IC) fidelity: debriefing is received in a group within 4 weeks after introduction and, at most, 2 days after completing at least 50% of a complete playthrough in LAKA. Favorable patient condition before gaming: at most, 45 years of age, high psychological symptoms, and active or pain coping levels are not high. The width of the "violins" represents the number of observations.



The Effect of Group on Health Outcomes Mediated by Mindfulness (Hypothesis 1)

Figure 4 displays regression coefficients for the effects of study group on mindfulness change (path A), effects of mindfulness change on depressive mood change (path B), and effects of study group on depressive mood change after controlling for mindfulness change (path C). Together, the results show that mindfulness changes mediated the relationship between study group (ie, independent variable) and depression changes (ie, dependent variable). The indirect effect was very small (beta=-.14, 95% CI -.27 to -.02). Similar results were obtained with fatigue (beta=-.15, 95% CI -.29 to -.01) and current pain (0-100 numeric rating scale; beta=-.09, 95% CI -.19 to -.01) as the dependent variable.

Figure 4. The mediated effect of serious gaming on depressive mood changes through mindfulness changes.



The Relationships Between Serious Gaming Experiences, Learning, and Health Change (Hypothesis 2)

Mean levels for affect valence after serious gaming were 3.00 (SD 0.91, range 1-5) for positive affect and 1.53 (SD 0.63, range 1-5) for negative affect. Mean scores for the subscales of sense of presence were 2.59 (SD 1.29) for involvement and 2.40 (SD 1.05) for realism. Intercorrelations between positive affect, sense of presence, and game acceptance were of moderate sizes (.57<ρ<.59, *P*<.001). Regression of *mindfulness* change on positive affect, negative affect, involvement, realism, and outcome expectations (ie, control variable) only showed statistical significance for involvement (N=113, R^2=.10, beta=.36, *P*=.001).

Differences in Intervention Effects on Outcomes by Patient Factor (Hypothesis 3)

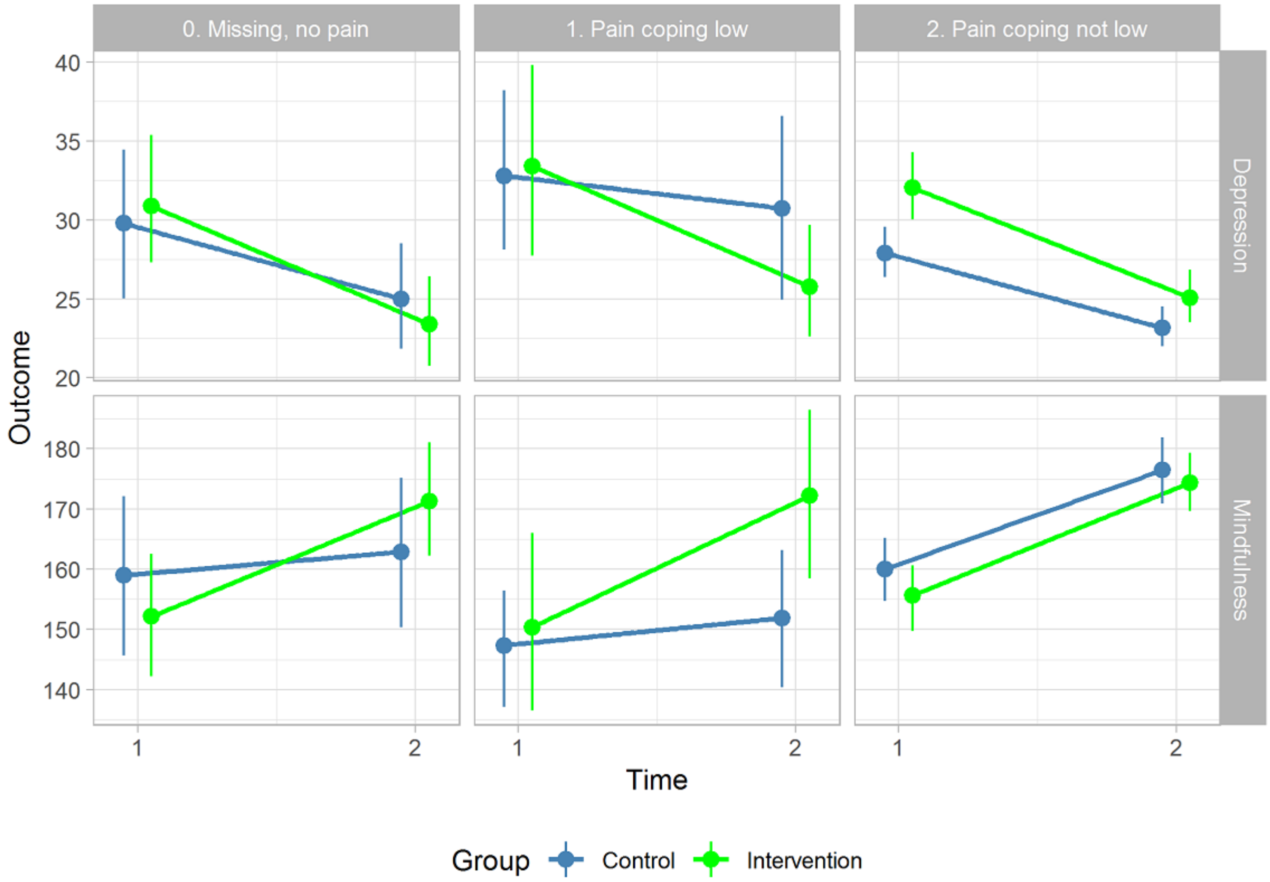
Age appeared to be the only patient factor with a notable association with game acceptance scores (b=-.025, *P*=.04). Moderated mediation analysis did not show statistically significant differences in the indirect intervention effect on depression change through mindfulness change between the two age groups.

A stronger indirect intervention effect on health outcome change was seen in the subgroup of participants without *high* active coping scores. For absence (score<21) versus presence of high active coping (213/265, 80.4%), the moderated mediation index amounted to -.26 (95% CI -.06 to -.52).

Within the subgroup of 33 out of 214 patients (15.4%) with low pain coping scores (18/121, 14.9%, from the intervention group and 15/93, 16%, from the control group), the indirect effect estimate (beta) was -.79 (95% CI -1.50 to -.31) stronger than within the 181 out of 214 (84.6%) other patients. Figure 5 shows that changes of mindfulness (ie, positive) and depressive mood (ie, negative) were particularly weak for control group patients with low pain coping.

Finally, no difference in the overall indirect effect was seen between subgroups determined by high or very high psychological distress: index of moderated mediation was -.19 (95% CI -.40 to -.02).

Figure 5. Group differences in average depression and mindfulness changes by pain coping.



Differences in Outcomes by Intervention in Context (Hypothesis 4)

Multilinear regression models showed that none of the objective indicators for intervention in context had a notable direct association with health outcome changes. For mindfulness change, model prediction improved by the addition of the dummy variable for *group* as a debriefing session structure (R^2_{change}=.03, b=.36, *P*=.04). Generally, addition of the planning variables did not improve the mindfulness change model (R^2_{change} =.05, *P*=.073). The parameter estimates were as follows: b=-.001, SE=.006, *P*=.84 for the date differences between preintervention assessment (t1) and the introduction session; b=-.02, SE=.01, *P*=.06 for the date differences between introduction and debriefing; and b=-.02, SE=.01, *P*=.03 for date differences between debriefing and postintervention assessment (t2). Bootstrapping or removing outlying mindfulness change values (Z>3 or Z<-3) generally accentuated the directions in which these findings pointed. Regressing game acceptance perceptions on a dummy variable for debriefing *shortly* after playing LAKA for the last time (ie, within less than 3 days) also resulted in model improvement (R^2_{change}=.08, b=.72, SE=.23, *P*=.002).

Mixing the Qualitative and Quantitative Results (Step 5)

Two original ICMO-Cs were constructed by updating initial formulations on the basis of the qualitative and quantitative results (see Textbox 4).

Textbox 4. Final configurations of contexts, mechanisms, and outcomes.

Intervention-context-mechanism-outcome configuration (ICMO-C) 1. Serious gaming acceptance perceptions are a feasibility mechanism and a context for learning and health outcome change that get hindered when:

- design qualities with respect to clear feedback and tailoring to recipient preferences are limited (I in C);
- players attribute negative perceptions to their older age (C); and
- implementation processes (ie, central planning) do not facilitate immediate debriefing after play for sharing available memories (I in C).

ICMO-C 2: A stronger self-awareness in daily life (ie, rehabilitation mechanism or serious gaming outcome) and subsequent changes in health effects (ie, rehabilitation outcomes) are triggered by involvement in serious gaming tasks, wherein both the actual self and discrepant self-scripts (ie, serious gaming mechanisms) are being processed. This occurs when:

- the innovation of treatment delivery modes is relatively advantageous (ie, adds experiential learning) and compatible with the rationale of a larger treatment program (I in C);
- the patient has limited inclination to manage with stress or pain in active or variegated ways (patient in C); and
- organizational implementation processes are characterized by adequate planning of patient guidance; sessions can address an actual need or discrepancy in a timely manner and debriefing is followed in a small group with other patients (I in C).

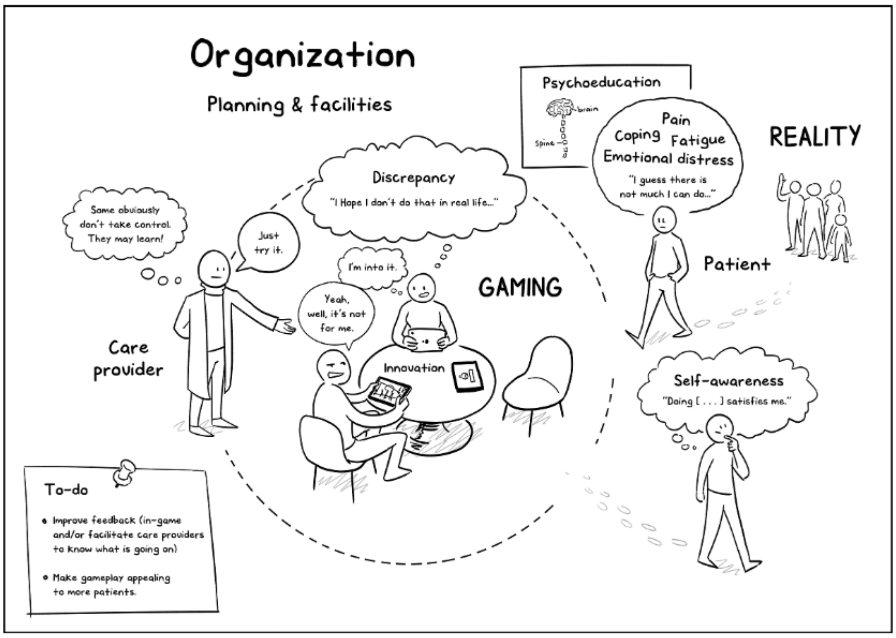
Discussion

Summary of Findings and Comparison With Existing Literature

Overview

This study informs future recipients, developers, evaluators, implementers, and policy makers about using, developing, or implementing serious gaming in a health care context. The question was how, why, for whom, and when facilitating additional serious gaming during multidisciplinary rehabilitation is effective for patients with CP or FSS. Qualitative and quantitative research was used to update an initial theory about mechanisms through which elements of a serious gaming intervention—with the game LAKA—lead to relevant positive effects of learning and physical and emotional symptom reduction. Moreover, insights have been developed as to which patients and circumstances of intervention such effects do and do not occur. Figure 6 presents a visual representation of the identified processes (ie, corresponding to Textbox 3). In the following sections, the contributions of this study to the further development of initial program theory are summarized (see Textbox 2).

Figure 6. A visual summary of program theory about serious gaming during the multidisciplinary rehabilitation of patients with complex chronic pain or functional somatic syndrome.



How Serious Gaming Facilitates Outcome Change During Multidisciplinary Rehabilitation

The *how* question was addressed by (1) asking patients how serious gaming during multidisciplinary rehabilitation influenced health outcomes, if perceived to be the case, and by (2) estimating how much additional change in learning about mindfulness and health outcomes was attributable to the serious gaming intervention and subjective experiences thereof. First, the findings supported the premise that serious gaming can be a relevant assistive tool to activate a rehabilitation mechanism of open and nonjudgmental *self-awareness*. Second, initial ideas about how serious gaming interventions can facilitate learning in this respect can be refined based on our findings. Prompts that encourage patients to reflect on discrepancies between their current behavior and their goals constitute a *resource* within the LAKA game. It evoked a kind of *reasoning* that patients referred to as a *confrontation with yourself*. In addition, the initial beliefs that debriefings facilitate learning transfers between participants and contexts were supported [52,59]. An unexpected finding was that a lack of perceived behavioral control in patients during serious gaming sessions, according to care providers, signaled a need for support as well as a potential to benefit from serious gaming. Third, this study refuted a general and simple positive linear pattern of stronger learning outcomes as patients experience more positive affect or virtual presence while playing a serious game. It was found, however, that a stronger subjective involvement (ie, attention devoted during serious gaming) had a positive association with mindfulness and symptom changes. The explanation for this remained implicit.

Why Serious Gaming Can Facilitate Learning and Subsequent Health Changes

The *why* question was addressed by adjudicating theory that can serve as a general *underlying* explanation for the *being confronted* experiences of patients. The SDT was chosen as a good starting point for predicting how mindfulness-based serious gaming can support learning and related affective and motivational processes. SDT is considered complementary to a model of mindfulness for explaining mechanisms [37]. Mindfulness can be seen as a way of processing self-discrepancies through which one is capable of moderating their affective and behavioral consequences [37,92]. This is illustrated by the common instructions of focused-attention meditation exercises that were also instructed in LAKA: *gently* return focus of attention to a chosen mental or sensory object (ie, *ideal*) when the present object of attention (ie, *actual*) is discrepant from that object. From this, it follows that *self-awareness* as a learning result from serious gaming concerns both the actual behavior of patients in daily social life as well as new or existing *noncontingent* scripts (ie, behavioral ideals and norms).

For Whom Does Serious Gaming Facilitate Outcome Change

Initially there were hardly any starting points for developing propositions regarding which patients with CP or FSS serious game intervention would be more or less effective. Findings from this study suggested theoretical extension with patient factors for feasibility and effectiveness. First, it is now suggested that perceptions of serious gaming acceptance can be weaker in patients with a relatively older age. This does not say that age, per se, is an explanation. Second, chances of a relevant effect of serious gaming on depressive symptoms may be greater in patients who experience less control over the consequences of stressors or pain in their daily lives. In an identifiable subgroup of patients with lower scores on active coping and pain coping scales, the average intervention effects were *small* or *medium* instead of *very small*, as found for all patients together [30]. This finding is remarkable as more active coping predicted better health outcome changes during regular rehabilitation. Moreover, pilot study and qualitative data suggested that a more active coping style facilitates usage of serious gaming. On the other hand, the effectiveness of serious gaming can now be called into question for patients who do have a tendency toward active and variegated ways of coping with stress or pain, and who already improve when receiving other means of treatment.

When Serious Gaming Affects Outcome Change

Circumstances determining the generation of mechanisms that lead, or do not lead, to serious gaming outcomes were examined qualitatively by asking patients for strengths and weaknesses of a serious gaming intervention during multidisciplinary rehabilitation. Responses were classified by factors on innovation, care provider, and organizational levels. Where available, quantitative data were used to identify patterns of covariation between indicators of relevant circumstances of serious game intervention in context and acceptance perceptions, learning outcomes, or health outcomes.

Based on the findings, it is proposed that blended serious gaming interventions trigger the mechanisms of effectiveness (1) when adding a relatively advantageous feature to the treatment (ie, experiential learning in addition to text-based psychoeducation), (2) when these features are compatible with the rationale of existing treatment (ie, adaptation to a CP or FSS condition by shifting attention away from self-scripts that actually depend on a known or indefinite underlying pathology toward scripts that are not), and (3) when organizational processes of planning and facilitation of instruction and feedback sessions are being implemented at high fidelity (ie, central planning and facilitating of debriefing in small patient groups). Moreover, serious gaming interventions may be perceived more positively by patients (1) when debriefing is always provided immediately after play or automatic feedback functions are adequate and (2) when software design adapts well to recipient needs and preferences (eg, regarding game environments and freedom of choice).

Strengths and Limitations

Paradigm

A major strength of this study is that the realist evaluation approach led to transferable findings on a theoretical level. A new example is added to a few previous ones about how quantitative methods can be used within a realistic evaluation paradigm [93-95]. Limitations related to the use of *realistic* principles are due to the late adoption during evaluation. This could have influenced initial program theory formulations (eg, conflation of intervention characteristics with mechanisms in the protocol stage) and, therefore, possibly also the data collection [96]. Perhaps more elaborate initial theoretical development could have enabled the study to (1) pose more or better follow-up questions to participants during interviews, (2) narrow the sets of sensitizing concepts and topics, or (3) determine more specific criteria for data saturation. Also, it could have led to different procedural choices. For example, interview data that resulted from the procedure of matching control group with intervention group cases failed to result in similar cases, with the exception of exposure to serious gaming and outcomes thereof. This was considered a reminder that a counterfactual logic of analysis does not suit a realist approach of discovering generative causal intervention effects.

Design and Procedures

General strengths and weaknesses of the mixed methods design were previously reported [60,96], but additional points on execution are noted here. Embedded experimental mixed methods ideally suited the questions and realistic approach of this study [97]. To realists, various research methods are commensurable and complementary and none is generally preferable. Controlled experiments can provide precise estimates of outcome pattern attributions to experimental conditions, but different methods are needed when well-defined ideas are lacking about how and why intervention effects are generated [98].

The validity of our findings varies across the study objectives and corresponding program theory elements. This is because not all qualitative findings were also triangulated with quantitative indications. As the study protocol had to be ethically approved before data collection started, the selection of quantitative measures could not be based on tentative ICMO-Cs after qualitative investigation. In realist evaluation, like in any other methodological approach, data collection is ideally informed by the most recent theoretical insights. Moreover, intervention and context factors had limited variation because only a single setting and a single serious game were studied. However, strong representativeness was achieved in the sampling of patients.

Methods

Finally, the findings of this study are to be interpreted in the light of theory development. Future studies are needed to proceed in the development of realistic propositions with newly collected data and complementary methodological strengths. With regard to the qualitative methods of this study, trustworthiness is supported by some techniques (ie, part of the qualitative data was independently coded). Regarding our quantitative analyses, limiting the number of variables by formulating hypotheses before quantitative data inspection, using qualitative findings, reduced the risk of capitalizing on chance for finding statistically significant results. Moreover, calculations of key parameters were repeated after particular methodological changes in order to check for the robustness of our findings. More detailed considerations about sensitivity analyses, data quantity, data quality, and missing data are presented in Textbox 5. Still, many decisions related to the calculations of variables and model specifications were not fully detailed in the study protocol or trial register.

Textbox 5. Additional details on the quantitative analyses.

- Sample size calculation did not specifically anticipate the testing of complicated, configurational propositions. Still, the 265 out of 275 cases with complete outcome data provided sufficient statistical power for moderated mediation modeling unless true direct effects were very small [98].
- Potentially relevant influences on the results due to nonresponse to the postgaming survey were explored by running chi-square or Student t tests on the differences in patient factors and gameplay behaviors between those who did and did not complete this survey.
- We did not apply methods to correct for multiple hypotheses testing.
- The psychometric qualities of the mindfulness measurements available for this study were assessed during an audit by an independent knowledge institute, but not during the peer-review process of an international scholarly journal.
- The results of sensitivity analyses showed that none of the presented estimates changed in an important manner after methodological changes. First, an intention-to-treat analysis was performed that yields unbiased estimates of the group × time effect on the intermediate outcome (ie, mindfulness), assuming that the outcome data of some respondents (ie, <4%) were missing at random. The moderated mediation models were refitted after replacing health outcome and learning outcome indicators with different residual posttreatment scores (t2): that is, the residual scores after regression on both the pretreatment score (t1) and the possible confounding variables. The latter refers to the variables for which a potential difference was found between the study group averages [60]. Finally, sensitivity of the coefficients produced with multivariable linear regression models was checked by the performance of bootstrapping with 1000 samples and/or removal of outlying cases on the dependent variable: standardized scores >3 or <-3.

Future Directions

This study illustrates that relevant insights for optimizing biopsychosocial management of CP or FSS (ie, including eHealth) can be developed with an explanatory and context-sensitive evaluative approach. Ideally, such research precedes and goes along with design and assessments of cost-effectiveness and quality improvement with suitable experimental methods (eg, randomized controlled or pragmatic, single- or multicenter, and individual allocation or stepped-wedge methods) [73]. Previously, researchers made similar suggestions for researching effectiveness of treatments for patients with CP, FSS, or mental health problems [31,45,99-101]. After all, consensus building about appropriate allocation and tailoring of treatment to patients has been relying on limited evidence [102]. As realist approaches focus on such issues, there is a large but yet unproven promise for methodological progress to support new insights for intervention quality improvement, treatment allocation decisions, personalization, and setting patient expectations. Specific areas for future theory-oriented research on serious games are implied in Table 4. This table gives an overview of considerations about how the findings of this study relate to those of other studies in various disciplines [1,3,30,31,52,53,59,68,85,87,96,103-116].

Table 4. Previous research findings related to the results of this study.

Question	Comparable research findings
How or why (rehabilitation mechanism)	The mediation analysis in this study (see Figure 4) showed similar results to a previous meta-analytic mediation analysis, which was mostly based on experimental comparisons of mental health outcome data with intervention groups that received comprehensive mindfulness programs and *passive* control groups (ie, standard treatment or waiting-list controls) [103]. Both studies found a *very small* indirect average effect of mindfulness-based intervention on mental health outcomes via *small* effects on mindfulness changes. The meta-analysis, however, found a significant direct intervention effect that was not seen in this study. A plausible explanation for this difference is that other ways in which mindfulness interventions work are nonspecific and had already been elicited by the other face-to-face-delivered techniques (eg, neuroscience education and cognitive restructuring) that the control group patients in our study received. In line with previous studies that suggested a limited sustainability of behavioral change effects, this study showed a decrease in the learning results of participants in the intervention group as the amount of time between debriefing and posttreatment assessment increased [1].
How or why (serious gaming mechanism)	Drawing attention to discrepancies between a person's current behavior and the person's previously set outcome goals, behavioral goals, or action plans constitutes an acknowledged behavioral change technique [104]. This technique apparently was the strongest, or most distinctive, trigger among other techniques that were integrated in the game LAKA [96]. LAKA does not contain explicit educational elements to change specific cognitive antecedents of health behaviors, as has been common in other health behavior games [52]. A crucial property for influencing self- or emotional regulation processes with games is user identification with avatars in storylines [105]. Herein, avatars have heroic or antiheroic qualities and challenges (eg, growing and defeating an antagonist). In LAKA, the avatar starts his or her journey in a state of discrepancy with respect to a relevant goal or ideal, and is committed to change. Users are then given control over avatar behaviors that determine the goal state. Distraction could be an alternative reason for a positive association between involvement in serious gaming and changes in pain intensity [53]. Finally, findings of this study correspond with a previous study that discovered positive behavioral change effects by eliciting self-discrepancy in videogames when users are primed to self-regulate with a prevention instead of a promotion focus [106]. Compared to other virtual environments and games, the average level of involvement in LAKA had a similar, relatively high, level [85]. Previous studies suggested that behavioral change outcomes are positively moderated by involvement in immersive virtual environments, as compared to, for example, text-based webpages [52,107-109].
For whom	Associations between age and technology acceptance have been inconsistent across contexts of use [110]. Previous pilot results hinted more specifically at an explanation for a pattern of lower acceptance with higher age. There is a positive moderation effect of age on the relationship between anxiety about technology use and acceptance [68]. Explanations may lie in general beliefs that games require young people's skills, or that older people are usually late adopters of technology. Previous work supports the finding that both *active* coping and pain coping, using many different strategies, are often, but not always, adaptive for patients with chronic pain as they are often used in the flexible coping by patients who remain hopeful and optimistic about being able to achieve important life goals [3]. Improvement of mindfulness, through acceptance of aversive experience, emotional regulation, and perceived support, can enhance adaptation to chronic pain (eg, moderate depressive symptoms) independently of pain coping [111,112]. Furthermore, psychological approaches may generally be more effective in patients who have more to gain with respect to emotional functioning [30,31].
When	The importance of adopting a valid rationale in developing games for health was emphasized in existing guidelines [113]. The indications from this study that immediate face-to-face interactions are important correspond to the notion that debriefings are necessary for generally satisfying learning experiences after serious gaming [59]. The role of group dynamics for effective debriefings were previously touched upon, but not well understood [59]. A previous implementation study of Internet-based cognitive behavioral therapy identified intervention adaptability as an important innovation-level barrier for usage [114]. Care providers' initial skepticism about serious gaming corresponds to previous findings on care provider perspectives on digitalized interventions, including serious gaming, for mental health improvement [115,116]. The proportions of patients who logged in to use LAKA (n=155), among those who were part of the intervention group (n=156) or had access (n=171), were much higher under the circumstances of this evaluation study than in the previous feasibility pilot (ie, n=71, n=116, and n=410, respectively). During the pilot, serious gaming was available only for one computer platform. It was also not an embedded treatment component and did not include care provider support [68]. Under such *noncommittal* circumstances, usage may be less well explained by a process of *going with the flow*, and more so by individual planning processes [87].

Conclusions and Recommendations

Following a realistic approach, this study started with the selection of potentially relevant theories and ended with an integrated fallible model [98]. Serious gaming (ie, a 4-hour intervention during an intensive multidisciplinary rehabilitation program) can offer relevant additional opportunities for enhancing self-awareness and reducing emotional symptoms by triggering increased attention and self-discrepancies. This can be achieved in CP or FSS patients who experience relatively less control over their current stress or pain. This also depends on the fidelity of implementation processes (ie, planning and facilitating) in such a way that gaming can be immediately followed up by exchanging experiences between patients under professional guidance, toward goals compatible with the overall treatment. At the same time, there is no sign as of yet of potential clinically relevant effects of serious game interventions in other treatment circumstances and among patients with CP or FSS.

This already offers some transferable insights for patients, care providers, health care organizations, insurers, policy makers, and businesses who will make future decisions on the future implementation of serious gaming. It may not be wise to offer this form of serious gaming, with similar late timing during rehabilitation, to all patients, including those claiming that their stress and pain are already under control.

Realizing the promise of serious gaming depends on ongoing efforts to develop transferable understandings of how they work, for whom, when, and why. In doing so, focus can be informed by perspectives of important stakeholders. The theoretical scope could be extended to the outcome domains of quality of life of patients, adoption and implementation by organizations, and cost-effectiveness. Understanding about the mechanisms of learning from serious games could be improved by applying SDT when formulating predictions about cognitive and emotional consequences of certain changes in the design and context of use. Finally, future studies could zoom in and out on micro- or macro-level conditions for serious gaming with impact. For example, patients, software developers, and data scientists could collaborate to optimize software (eg, in-game feedback) for adaptation to recipient characteristics for learning and health improvement. Care providers can also be involved in optimizing feedback provision (eg, enabling them to interpret log data) and in optimizing the fit with work routines. Lastly, on an organizational or societal level, policy makers could participate in theorizing on preconditions for implementation, such as financial arrangements and dissemination platforms.

Acknowledgments

Bruno Marchal provided constructive feedback on the study during two occasions. Imke Evers assisted in conducting the first focus group. Essential contributions to the development of LAKA are acknowledged in the evaluation protocol.

Conflicts of Interest

During data collection, MAPV was provided with time and occasion by his employer Ciran to conduct independent doctoral research, by way of agreement, at Tranzo, Scientific Centre for Care and Wellbeing. The terms of this arrangement have been reviewed and approved by Tranzo in accordance with its policy on objectivity in research. Ciran paid personal fees to HJMV outside of the submitted work. MJ and AZ have nothing to declare.

Multimedia Appendix 1
Initial developer assumptions and evaluators’ interpretations on the basis of relevant existing scientific theories.
[PDF File (Adobe PDF File), 805 KB - jmir_v22i3e14766_app1.pdf]

Multimedia Appendix 2
Additional information about the mindfulness measure.
[PDF File (Adobe PDF File), 765 KB - jmir_v22i3e14766_app2.pdf]

Multimedia Appendix 3
Coding illustrations for individual-level intervention-context-mechanism-outcome configurations and barriers.
[PDF File (Adobe PDF File), 1471 KB - jmir_v22i3e14766_app3.pdf]

References

1. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
2. Henningsen P, Zipfel S, Herzog W. Management of functional somatic syndromes. Lancet 2007 Mar 17;369(9565):946-955. [doi: 10.1016/S0140-6736(07)60159-7] [Medline: 17368156]
3. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: Scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
4. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: A systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
5. Li J, Theng Y, Foo S. Game-based digital interventions for depression therapy: A systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):519-527 [FREE Full text] [doi: 10.1089/cyber.2013.0481] [Medline: 24810933]

6. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: Randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
7. Johannes CB, Le TK, Zhou X, Johnston JA, Dworkin RH. The prevalence of chronic pain in United States adults: Results of an Internet-based survey. J Pain 2010 Nov;11(11):1230-1239. [doi: 10.1016/j.jpain.2010.07.002] [Medline: 20797916]
8. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
9. Jackson T, Thomas S, Stabile V, Han X, Shotwell M, McQueen K. Prevalence of chronic pain in low-income and middle-income countries: A systematic review and meta-analysis. Lancet 2015 Apr 27;385 Suppl 2:S10. [doi: 10.1016/S0140-6736(15)60805-4] [Medline: 26313056]
10. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: A meta-analysis. Clin Gastroenterol Hepatol 2012 Jul;10(7):712-721.e4. [doi: 10.1016/j.cgh.2012.02.029] [Medline: 22426087]
11. Langguth B, Kreuzer P, Kleinjung T, De Ridder D. Tinnitus: Causes and clinical management. Lancet Neurol 2013 Sep;12(9):920-930 [FREE Full text] [doi: 10.1016/s1474-4422(13)70160-1]
12. Johnston S, Brenu EW, Staines D, Marshall-Gradisnik S. The prevalence of chronic fatigue syndrome/myalgic encephalomyelitis: A meta-analysis. Clin Epidemiol 2013;5:105-110 [FREE Full text] [doi: 10.2147/CLEP.S39876] [Medline: 23576883]
13. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: Estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014 Jun;73(6):968-974. [doi: 10.1136/annrheumdis-2013-204428] [Medline: 24665116]
14. Von Korff M, Crane P, Lane M, Miglioretti DL, Simon G, Saunders K, et al. Chronic spinal pain and physical-mental comorbidity in the United States: Results from the national comorbidity survey replication. Pain 2005 Feb;113(3):331-339. [doi: 10.1016/j.pain.2004.11.010] [Medline: 15661441]
15. Janssens KA, Zijlema WL, Joustra ML, Rosmalen JG. Mood and anxiety disorders in chronic fatigue syndrome, fibromyalgia, and irritable bowel syndrome: Results from the LifeLines Cohort Study. Psychosom Med 2015 May;77(4):449-457. [doi: 10.1097/PSY.0000000000000161] [Medline: 25768845]
16. Zöger S, Svedlund J, Holgers K. Relationship between tinnitus severity and psychiatric disorders. Psychosomatics 2006;47(4):282-288. [doi: 10.1176/appi.psy.47.4.282] [Medline: 16844885]
17. Häuser W, Wolfe F, Henningsen P, Schmutzer G, Brähler E, Hinz A. Untying chronic pain: Prevalence and societal burden of chronic pain stages in the general population. A cross-sectional survey. BMC Public Health 2014 Apr 13;14:352 [FREE Full text] [doi: 10.1186/1471-2458-14-352] [Medline: 24725286]
18. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724. [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
19. Gustavsson A, Bjorkman J, Ljungcrantz C, Rhodin A, Rivano-Fischer M, Sjolund K, et al. Socio-economic burden of patients with a diagnosis related to chronic pain: Register data of 840,000 Swedish patients. Eur J Pain 2012 Feb;16(2):289-299 [FREE Full text] [doi: 10.1016/j.ejpain.2011.07.006] [Medline: 22323381]
20. Konnopka A, Kaufmann C, König HH, Heider D, Wild B, Szecsenyi J, et al. Association of costs with somatic symptom severity in patients with medically unexplained symptoms. J Psychosom Res 2013 Oct;75(4):370-375. [doi: 10.1016/j.jpsychores.2013.08.011] [Medline: 24119945]
21. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. BMJ 2015 Feb 18;350:h444 [FREE Full text] [doi: 10.1136/bmj.h444] [Medline: 25694111]
22. Vos-Vromans DC, Smeets RJ, Huijnen IP, Köke AJ, Hitters WM, Rijnders LJ, et al. Multidisciplinary rehabilitation treatment versus cognitive behavioural therapy for patients with chronic fatigue syndrome: A randomized controlled trial. J Intern Med 2016 Mar;279(3):268-282 [FREE Full text] [doi: 10.1111/joim.12402] [Medline: 26306716]
23. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: An overview of Cochrane Reviews. Cochrane Database Syst Rev 2017 Apr 24;4:CD011279 [FREE Full text] [doi: 10.1002/14651858.CD011279.pub3] [Medline: 28436583]
24. Enthoven WT, Roelofs PD, Deyo RA, van Tulder MW, Koes BW. Non-steroidal anti-inflammatory drugs for chronic low back pain. Cochrane Database Syst Rev 2016 Feb 10;2:CD012087. [doi: 10.1002/14651858.CD012087] [Medline: 26863524]
25. Vickers AJ, Vertosick EA, Lewith G, MacPherson H, Foster NE, Sherman KJ, Acupuncture Trialists' Collaboration. Acupuncture for chronic pain: Update of an individual patient data meta-analysis. J Pain 2018 May;19(5):455-474 [FREE Full text] [doi: 10.1016/j.jpain.2017.11.005] [Medline: 29198932]
26. Flor H, Fydrich T, Turk DC. Efficacy of multidisciplinary pain treatment centers: A meta-analytic review. Pain 1992 May;49(2):221-230. [doi: 10.1016/0304-3959(92)90145-2] [Medline: 1535122]
27. Zech N, Hansen E, Bernardy K, Häuser W. Efficacy, acceptability and safety of guided imagery/hypnosis in fibromyalgia: A systematic review and meta-analysis of randomized controlled trials. Eur J Pain 2017 Feb;21(2):217-227. [doi: 10.1002/ejp.933] [Medline: 27896907]

28. Williams A, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012 Nov 14;11:CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
29. Kuner R, Flor H. Structural plasticity and reorganisation in chronic pain. Nat Rev Neurosci 2017 Jan 20;18(2):113. [doi: 10.1038/nrn.2017.5] [Medline: 28704354]
30. Vugts MAP, Joosen MCW, van der Geer JE, Zedlitz AMEE, Vrijhoef HJM. The effectiveness of various computer-based interventions for patients with chronic pain or functional somatic syndromes: A systematic review and meta-analysis. PLoS One 2018;13(5):e0196467 [FREE Full text] [doi: 10.1371/journal.pone.0196467] [Medline: 29768436]
31. Gilpin HR, Keyes A, Stahl DR, Greig R, McCracken LM. Predictors of treatment outcome in contextual cognitive and behavioral therapies for chronic pain: A systematic review. J Pain 2017 Oct;18(10):1153-1164. [doi: 10.1016/j.jpain.2017.04.003] [Medline: 28455249]
32. Macea DD, Gajos K, Daglia Calil YA, Fregni F. The efficacy of Web-based cognitive behavioral interventions for chronic pain: A systematic review and meta-analysis. J Pain 2010 Oct;11(10):917-929. [doi: 10.1016/j.jpain.2010.06.005] [Medline: 20650691]
33. Gatchel RJ, McGeary DD, McGeary CA, Lippe B. Interdisciplinary chronic pain management: Past, present, and future. Am Psychol 2014;69(2):119-130. [doi: 10.1037/a0035514] [Medline: 24547798]
34. Nijs J, Malfliet A, Ickmans K, Baert I, Meeus M. Treatment of central sensitization in patients with 'unexplained' chronic pain: An update. Expert Opin Pharmacother 2014 Aug;15(12):1671-1683. [doi: 10.1517/14656566.2014.925446] [Medline: 24930805]
35. Vowles KE, Witkiewitz K, Sowden G, Ashworth J. Acceptance and commitment therapy for chronic pain: Evidence of mediation and clinically significant change following an abbreviated interdisciplinary program of rehabilitation. J Pain 2014 Jan;15(1):101-113. [doi: 10.1016/j.jpain.2013.10.002] [Medline: 24373572]
36. Wessels T, van Tulder M, Sigl T, Ewert T, Limm H, Stucki G. What predicts outcome in non-operative treatments of chronic low back pain? A systematic review. Eur Spine J 2006 Nov;15(11):1633-1644. [doi: 10.1007/s00586-006-0073-4] [Medline: 16575598]
37. Vago DR, Silbersweig DA. Self-awareness, self-regulation, and self-transcendence (S-ART): A framework for understanding the neurobiological mechanisms of mindfulness. Front Hum Neurosci 2012;6:296 [FREE Full text] [doi: 10.3389/fnhum.2012.00296] [Medline: 23112770]
38. Turner JA, Anderson ML, Balderson BH, Cook AJ, Sherman KJ, Cherkin DC. Mindfulness-based stress reduction and cognitive behavioral therapy for chronic low back pain: Similar effects on mindfulness, catastrophizing, self-efficacy, and acceptance in a randomized controlled trial. Pain 2016 Nov;157(11):2434-2444 [FREE Full text] [doi: 10.1097/j.pain.0000000000000635] [Medline: 27257859]
39. Hofmann SG, Asmundson GJG. Acceptance and mindfulness-based therapy: New wave or old hat? Clin Psychol Rev 2008 Jan;28(1):1-16. [doi: 10.1016/j.cpr.2007.09.003] [Medline: 17904260]
40. Veehof MM, Trompetter HR, Bohlmeijer ET, Schreurs KM. Acceptance- and mindfulness-based interventions for the treatment of chronic pain: A meta-analytic review. Cogn Behav Ther 2016;45(1):5-31. [doi: 10.1080/16506073.2015.1098724] [Medline: 26818413]
41. Pots WT, Trompetter HR, Schreurs KM, Bohlmeijer ET. How and for whom does Web-based acceptance and commitment therapy work? Mediation and moderation analyses of Web-based ACT for depressive symptoms. BMC Psychiatry 2016 May 23;16:158 [FREE Full text] [doi: 10.1186/s12888-016-0841-6] [Medline: 27215733]
42. Turner JA, Holtzman S, Mancl L. Mediators, moderators, and predictors of therapeutic change in cognitive-behavioral therapy for chronic pain. Pain 2007 Feb;127(3):276-286. [doi: 10.1016/j.pain.2006.09.005] [Medline: 17071000]
43. Smeets RJ, Maher CG, Nicholas MK, Refshauge KM, Herbert RD. Do psychological characteristics predict response to exercise and advice for subacute low back pain? Arthritis Rheum 2009 Sep 15;61(9):1202-1209 [FREE Full text] [doi: 10.1002/art.24731] [Medline: 19714601]
44. van der Hulst M, Vollenbroek-Hutten MM, Groothuis-Oudshoorn KG, Hermens HJ. Multidisciplinary rehabilitation treatment of patients with chronic low back pain: A prognostic model for its outcome. Clin J Pain 2008 Jun;24(5):421-430. [doi: 10.1097/AJP.0b013e31816719f5] [Medline: 18496307]
45. Miles CL, Pincus T, Carnes D, Homer KE, Taylor SJ, Bremner SA, et al. Can we identify how programmes aimed at promoting self-management in musculoskeletal pain work and who benefits? A systematic review of sub-group analysis within RCTs. Eur J Pain 2011 Sep;15(8):775.e1-775.e11. [doi: 10.1016/j.ejpain.2011.01.016] [Medline: 21354838]
46. Samwel HJ, Kraaimaat FW, Crul BJ, van Dongen RD, Evers AW. Multidisciplinary allocation of chronic pain treatment: Effects and cognitive-behavioural predictors of outcome. Br J Health Psychol 2009 Sep;14(Pt 3):405-421. [doi: 10.1348/135910708X337760] [Medline: 18718110]
47. Broderick JE, Keefe FJ, Schneider S, Junghaenel DU, Bruckenthal P, Schwartz JE, et al. Cognitive behavioral therapy for chronic pain is effective, but for whom? Pain 2016 Sep;157(9):2115-2123. [doi: 10.1097/j.pain.0000000000000626] [Medline: 27227692]
48. Pincus T, McCracken LM. Psychological factors and treatment opportunities in low back pain. Best Pract Res Clin Rheumatol 2013 Oct;27(5):625-635. [doi: 10.1016/j.berh.2013.09.010] [Medline: 24315144]

49. Turk DC. The potential of treatment matching for subgroups of patients with chronic pain: Lumping versus splitting. Clin J Pain 2005;21(1):44-55; discussion 69. [doi: 10.1097/00002508-200501000-00006] [Medline: 15599131]
50. Baranowski T, Baranowski J, Thompson D, Buday R. Behavioral science in video games for children's diet and physical activity change: Key research needs. J Diabetes Sci Technol 2011 Mar 01;5(2):229-233 [FREE Full text] [doi: 10.1177/193229681100500204] [Medline: 21527086]
51. Bul KCM, Doove LL, Franken IHA, Oord SV, Kato PM, Maras A. A serious game for children with attention deficit hyperactivity disorder: Who benefits the most? PLoS One 2018;13(3):e0193681 [FREE Full text] [doi: 10.1371/journal.pone.0193681] [Medline: 29543891]
52. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: Video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
53. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: A systematic review. Clin Psychol Rev 2010 Dec;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]
54. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
55. Cerasoli CP, Nicklin JM, Ford MT. Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. Psychol Bull 2014 Jul;140(4):980-1008. [doi: 10.1037/a0035661] [Medline: 24491020]
56. DeSmet A, Thompson D, Baranowski T, Palmeira A, Verloigne M, De Bourdeaudhuij I. Is participatory design associated with the effectiveness of serious digital games for healthy lifestyle promotion? A meta-analysis. J Med Internet Res 2016 Apr 29;18(4):e94 [FREE Full text] [doi: 10.2196/jmir.4444] [Medline: 27129447]
57. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. J Educ Psychol 2013;105(2):249-265. [doi: 10.1037/a0031311]
58. Kim B, Park H, Baek Y. Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning. Comput Educ 2009 May;52(4):800-810 [FREE Full text] [doi: 10.1016/j.compedu.2008.12.004]
59. Crookall D. Serious games, debriefing, and simulation/gaming as a discipline. Simul Gaming 2011 Jan 06;41(6):898-920 [FREE Full text] [doi: 10.1177/1046878110390784]
60. Vugts MA, Joosen MC, Mert A, Zedlitz AM, Vrijhoef HJ. Effectiveness of serious gaming during the multidisciplinary rehabilitation of patients with complex chronic pain or fatigue: Natural quasi-experiment. J Med Internet Res 2018 Aug 15;20(8):e250 [FREE Full text] [doi: 10.2196/jmir.9739] [Medline: 30111527]
61. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies: Republished article. Behav Ther 2016 Nov;47(6):869-885. [doi: 10.1016/j.beth.2016.11.006] [Medline: 27993338]
62. Ryff CD, Singer BH. Know thyself and become what you are: A eudaimonic approach to psychological well-being. J Happiness Stud 2006 Oct 5;9(1):13-39. [doi: 10.1007/s10902-006-9019-0]
63. Higgins ET. Self-discrepancy: A theory relating self and affect. Psychol Rev 1987 Jul;94(3):319-340. [Medline: 3615707]
64. Przybylski AK, Weinstein N, Murayama K, Lynch MF, Ryan RM. The ideal self at play: The appeal of video games that let you be all you can be. Psychol Sci 2012 Jan 01;23(1):69-76. [doi: 10.1177/0956797611418676] [Medline: 22173739]
65. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 2004 Mar 15;180(S6):S57-S60. [Medline: 15012583]
66. 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]
67. Weinstein N, Ryan RM. When helping helps: Autonomous motivation for prosocial behavior and its influence on well-being for the helper and recipient. J Pers Soc Psychol 2010 Feb;98(2):222-244. [doi: 10.1037/a0016984] [Medline: 20085397]
68. Vugts MA, Joosen MC, van Bergen AH, Vrijhoef HJ. Feasibility of applied gaming during interdisciplinary rehabilitation for patients with complex chronic pain and fatigue complaints: A mixed-methods study. JMIR Serious Games 2016 Apr 01;4(1):e2 [FREE Full text] [doi: 10.2196/games.5088] [Medline: 27036094]
69. Waters SJ, Keefe FJ, Strauman TJ. Self-discrepancy in chronic low back pain: Relation to pain, depression, and psychological distress. J Pain Symptom Manage 2004 Mar;27(3):251-259. [doi: 10.1016/j.jpainsymman.2003.07.001] [Medline: 15010103]
70. Kindermans HP, Huijnen IP, Goossens ME, Roelofs J, Verbunt JA, Vlaeyen JW. "Being" in pain: The role of self-discrepancies in the emotional experience and activity patterns of patients with chronic low back pain. Pain 2011 Feb;152(2):403-409. [doi: 10.1016/j.pain.2010.11.009] [Medline: 21216100]
71. Sutherland R, Morley S. Self-pain enmeshment: Future possible selves, sociotropy, autonomy and adjustment to chronic pain. Pain 2008 Jul 15;137(2):366-377. [doi: 10.1016/j.pain.2007.09.023] [Medline: 17977661]
72. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. BMC Med 2016 Jun 24;14(1):96 [FREE Full text] [doi: 10.1186/s12916-016-0643-1] [Medline: 27342217]
73. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]

74. Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ: John Wiley & Sons; 2008.
75. Fletcher A, Jamal F, Moore G, Evans RE, Murphy S, Bonell C. Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions. Evaluation (Lond) 2016 Jul;22(3):286-303 [FREE Full text] [doi: 10.1177/1356389016652743] [Medline: 27478401]
76. Pawson R, Tilley N. Realistic Evaluation. London, UK: SAGE Publications; 1997.
77. Dalkin SM, Greenhalgh J, Jones D, Cunningham B, Lhussier M. What's in a mechanism? Development of a key concept in realist evaluation. Implement Sci 2015 Apr 16;10:49 [FREE Full text] [doi: 10.1186/s13012-015-0237-x] [Medline: 25885787]
78. Eastwood JG, Kemp LA, Jalaludin BB. Realist theory construction for a mixed method multilevel study of neighbourhood context and postnatal depression. Springerplus 2016;5(1):1081 [FREE Full text] [doi: 10.1186/s40064-016-2729-9] [Medline: 27468381]
79. Manzano A. The craft of interviewing in realist evaluation. Evaluation 2016 Apr 05;22(3):342-360 [FREE Full text] [doi: 10.1177/1356389016638615]
80. Ettema JHM, Arrindell WA. SCL-90. Handleiding bij een multidimensionele psychopathologie-indicator. Lisse, the Netherlands: Swets & Zeitlinger; 2003.
81. Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain 1994 Feb;56(2):217-226. [doi: 10.1016/0304-3959(94)90097-3] [Medline: 8008411]
82. Dittner AJ, Wessely SC, Brown RG. The assessment of fatigue: A practical guide for clinicians and researchers. J Psychosom Res 2004 Feb;56(2):157-170. [doi: 10.1016/S0022-3999(03)00371-4] [Medline: 15016573]
83. Goldberg SB, Wielgosz J, Dahl C, Schuyler B, MacCoon DS, Rosenkranz M, et al. Does the Five Facet Mindfulness Questionnaire measure what we think it does? Construct validity evidence from an active controlled randomized clinical trial. Psychol Assess 2016 Aug;28(8):1009-1014 [FREE Full text] [doi: 10.1037/pas0000233] [Medline: 26460893]
84. Spinhoven P, Huijbers MJ, Ormel J, Speckens AEM. Improvement of mindfulness skills during mindfulness-based cognitive therapy predicts long-term reductions of neuroticism in persons with recurrent depression in remission. J Affect Disord 2017 Apr 15;213:112-117. [doi: 10.1016/j.jad.2017.02.011] [Medline: 28213122]
85. Lessiter J, Freeman J, Keogh E, Davidoff J. A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. Teleoperators Virtual Env 2001 Jun;10(3):282-297 [FREE Full text] [doi: 10.1162/105474601300343612]
86. Thompson E. Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). J Cross Cult Psychol 2016 Jul 26;38(2):227-242 [FREE Full text] [doi: 10.1177/0022022106297301]
87. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
88. Fu F, Su R, Yu S. EGameFlow: A scale to measure learners’ enjoyment of e-learning games. Comput Educ 2009 Jan;52(1):101-112 [FREE Full text] [doi: 10.1016/j.compedu.2008.07.004]
89. Schreurs P, Van Willige G, Brosschot J, Tellegen B, Graus G. De Utrechtse Coping lijst: Omgaan met problemen en gebeurtenissen (herziene druk). Lisse, the Netherlands: Swets & Zeitlinger; 1993.
90. Spinhoven P, Ter Kuile M, Kole-Snijders AM, Hutten Mansfeld M, Den Ouden DJ, Vlaeyen JW. Catastrophizing and internal pain control as mediators of outcome in the multidisciplinary treatment of chronic low back pain. Eur J Pain 2004 Jun;8(3):211-219. [doi: 10.1016/j.ejpain.2003.08.003] [Medline: 15109971]
91. Hayes AF. PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling [White paper]. 2012. URL: https://pdfs.semanticscholar.org/e9bb/7b23993113a73ee1ff6cde5ff9a4164f946e.pdf?_ga=2.155998271.523161736.1579062123-536124310.1579062123 [accessed 2020-01-14]
92. Crane C, Barnhofer T, Duggan D, Hepburn S, Fennell M, Williams J. Mindfulness-based cognitive therapy and self-discrepancy in recovered depressed patients with a history of depression and suicidality. Cognit Ther Res 2008 Jun 7;32(6):775-787 [FREE Full text] [doi: 10.1007/s10608-008-9193-y]
93. Ford JA, Jones A, Wong G, Clark A, Porter T, Steel N. Access to primary care for socio-economically disadvantaged older people in rural areas: Exploring realist theory using structural equation modelling in a linked dataset. BMC Med Res Methodol 2018 Jun 19;18(1):57 [FREE Full text] [doi: 10.1186/s12874-018-0514-x] [Medline: 29914411]
94. Eastwood JG, Jalaludin BB, Kemp LA, Phung HN. Bayesian hierarchical spatial regression of maternal depressive symptoms in South Western Sydney, Australia. Springerplus 2014;3:55 [FREE Full text] [doi: 10.1186/2193-1801-3-55] [Medline: 24555171]
95. Eastwood JG, Jalaludin BB, Kemp LA, Phung HN. Realist identification of group-level latent variables for perinatal social epidemiology theory building. Int J Health Serv 2014;44(3):407-433. [doi: 10.2190/HS.44.3.a] [Medline: 25618983]
96. Vugts MAP, Joosen MCW, Mert A, Zedlitz A, Vrijhoef HJM. Serious gaming during multidisciplinary rehabilitation for patients with complex chronic pain or fatigue complaints: Study protocol for a controlled trial and process evaluation. BMJ Open 2017 Jun 08;7(6):e016394 [FREE Full text] [doi: 10.1136/bmjopen-2017-016394] [Medline: 28600377]
97. Maxwell JA, Mittapalli K. Realism as a stance for mixed methods research. In: Tashakkori AM, Teddlie CB, editors. SAGE Handbook of Mixed Methods in Social & Behavioral Research. 2nd edition. Thousand Oaks, CA: SAGE Publications; 2010:145-168.

98. Van Belle S, Wong G, Westhorp G, Pearson M, Emmel N, Manzano A, et al. Can "realist" randomised controlled trials be genuinely realist? Trials 2016 Jul 07;17(1):313 [FREE Full text] [doi: 10.1186/s13063-016-1407-0] [Medline: 27387202]
99. Duncan C, Weich S, Fenton S, Twigg L, Moon G, Madan J, et al. A realist approach to the evaluation of complex mental health interventions. Br J Psychiatry 2018 Aug;213(2):451-453. [doi: 10.1192/bjp.2018.96] [Medline: 30027875]
100. McIntosh G, Steenstra I, Hogg-Johnson S, Carter T, Hall H. Lack of prognostic model validation in low back pain prediction studies: A systematic review. Clin J Pain 2018 Aug;34(8):748-754. [doi: 10.1097/AJP.0000000000000591] [Medline: 29406366]
101. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of eHealth interventions in somatic diseases: A systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
102. van Hooff ML, van Loon J, van Limbeek J, de Kleuver M. The Nijmegen decision tool for chronic low back pain. Development of a clinical decision tool for secondary or tertiary spine care specialists. PLoS One 2014;9(8):e104226 [FREE Full text] [doi: 10.1371/journal.pone.0104226] [Medline: 25133645]
103. Gu J, Strauss C, Bond R, Cavanagh K. How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. Clin Psychol Rev 2015 Apr;37:1-12. [doi: 10.1016/j.cpr.2015.01.006] [Medline: 25689576]
104. 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]
105. Villani D, Carissoli C, Triberti S, Marchetti A, Gilli G, Riva G. Videogames for emotion regulation: A systematic review. Games Health J 2018 Apr;7(2):85-99. [doi: 10.1089/g4h.2017.0108] [Medline: 29424555]
106. Jin SA. Self-discrepancy and regulatory fit in avatar-based exergames. Psychol Rep 2012 Dec;111(3):697-710. [doi: 10.2466/06.07.21.PR0.111.6.697-710] [Medline: 23402039]
107. Diemer J, Alpers GW, Peperkorn HM, Shiban Y, Mühlberger A. The impact of perception and presence on emotional reactions: A review of research in virtual reality. Front Psychol 2015;6:26 [FREE Full text] [doi: 10.3389/fpsyg.2015.00026] [Medline: 25688218]
108. Price M, Mehta N, Tone EB, Anderson PL. Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. J Anxiety Disord 2011 Aug;25(6):763-770 [FREE Full text] [doi: 10.1016/j.janxdis.2011.03.004] [Medline: 21515027]
109. Kelders SM, Sommers-Spijkerman M, Goldberg J. Investigating the direct impact of a gamified versus nongamified well-being intervention: An exploratory experiment. J Med Internet Res 2018 Jul 26;20(7):e247 [FREE Full text] [doi: 10.2196/jmir.9923] [Medline: 30049669]
110. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a Web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
111. Agar-Wilson M, Jackson T. Are emotion regulation skills related to adjustment among people with chronic pain, independent of pain coping? Eur J Pain 2012 Jan;16(1):105-114. [doi: 10.1016/j.ejpain.2011.05.011] [Medline: 21669543]
112. McCracken LM, MacKichan F, Eccleston C. Contextual cognitive-behavioral therapy for severely disabled chronic pain sufferers: Effectiveness and clinically significant change. Eur J Pain 2007 Apr;11(3):314-322. [doi: 10.1016/j.ejpain.2006.05.004] [Medline: 16762571]
113. Graafland M, Dankbaar M, Mert A, Lagro J, De Wit-Zuurendonk L, Schuit S, et al. How to systematically assess serious games applied to health care. JMIR Serious Games 2014 Nov 11;2(2):e11 [FREE Full text] [doi: 10.2196/games.3825] [Medline: 25654163]
114. Hadjistavropoulos HD, Nugent MM, Dirkse D, Pugh N. Implementation of Internet-delivered cognitive behavior therapy within community mental health clinics: A process evaluation using the consolidated framework for implementation research. BMC Psychiatry 2017 Sep 12;17(1):331 [FREE Full text] [doi: 10.1186/s12888-017-1496-7] [Medline: 28899365]
115. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of eMental health by psychologists: The construction of the Levels of Adoption of eMental Health Model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
116. Eichenberg C, Grabmayer G, Green N. Acceptance of serious games in psychotherapy: An inquiry into the stance of therapists and patients. Telemed J E Health 2016 Nov;22(11):945-951. [doi: 10.1089/tmj.2016.0001] [Medline: 27045452]

Abbreviations

CP: chronic pain
FSS: functional somatic syndrome
ICMO: intervention-context-mechanism-outcome
ICMO-C: intervention-context-mechanism-outcome configuration
ICT: information and communication technology

SDT: self-discrepancy theory
t0: baseline
t1: intermediate time point after 8 weeks of treatment
t2: posttreatment after 16 weeks

Edited by G Eysenbach; submitted 20.05.19; peer-reviewed by A van der Horst, A Schäfer, C Yeager; comments to author 09.07.19; revised version received 31.10.19; accepted 15.12.19; published 09.03.20.

Please cite as:
Vugts MAP, Zedlitz AMEE, Joosen MCW, Vrijhoef HJM
Serious Gaming During Multidisciplinary Rehabilitation for Patients With Chronic Pain or Fatigue Symptoms: Mixed Methods Design of a Realist Process Evaluation
J Med Internet Res 2020;22(3):e14766
URL: https://www.jmir.org/2020/3/e14766
doi:10.2196/14766
PMID:32149720

Original Paper

Precourse Preparation Using a Serious Smartphone Game on Advanced Life Support Knowledge and Skills: Randomized Controlled Trial

Pariwat Phungoen[1], MD; Songwoot Promto[1], MD; Sivit Chanthawatthanarak[1], MD; Sawitree Maneepong[2], BSN; Korakot Apiratwarakul[1], MD; Praew Kotruchin[1], MD, PhD; Thapanawong Mitsungnern[1], MD

[1]Department of Emergency Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

[2]CPR Training Unit, Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

Corresponding Author:
Thapanawong Mitsungnern, MD
Department of Emergency Medicine
Faculty of Medicine
Khon Kaen University
123 Mitraparb Rd Muang
Faculty of Medicine, Khon Kaen University
Khon Kaen, 40002
Thailand
Phone: 66 43366869
Email: thapanawong@hotmail.com

Abstract

Background: In the past several years, gamified learning has been growing in popularity in various medical educational contexts including cardiopulmonary resuscitation (CPR) training. Furthermore, prior work in Basic Life Support (BLS) training has demonstrated the benefits of serious games as a method for pretraining among medical students. However, there is little evidence to support these benefits with regard to Advanced Life Support (ALS) training.

Objective: We compare the effects of a brief precourse ALS preparation using a serious smartphone game on student knowledge, skills, and perceptions in this area with those of conventional ALS training alone.

Methods: A serious game (Resus Days) was developed by a Thai physician based on global ALS clinical practice guidelines. Fifth-year medical students were enrolled and randomized to either the game group or the control group. Participants in both groups attended a traditional ALS lecture, but the game group was assigned to play Resus Days for 1 hour before attending the lecture and were allowed to play as much as they wished during the training course. All students underwent conventional ALS training, and their abilities were evaluated using multiple-choice questions and with hands-on practice on a mannequin. Subject attitudes and perceptions about the game were evaluated using a questionnaire.

Results: A total of 105 students participated in the study and were randomly assigned to either the game group (n=52) or the control group (n=53). Students in the game group performed better on the ALS algorithm knowledge posttest than those in the control group (17.22 [SD 1.93] vs 16.60 [SD 1.97], *P*=.01; adjusted mean difference [AMD] 0.93; 95% CI 0.21-1.66). The game group’s pass rate on the skill test was also higher but not to a statistically significant extent (79% vs 66%, *P*=.09; adjusted odds ratio [AOR] 2.22; 95% CI 0.89-5.51). Students indicated high satisfaction with the game (9.02 [SD 1.11] out of 10).

Conclusions: Engaging in game-based preparation prior to an ALS training course resulted in better algorithm knowledge scores for medical students than attending the course alone.

Trial Registration: Thai Clinical Trials Registry HE611533; https://tinyurl.com/wmbp3q7

(J Med Internet Res 2020;22(3):e16987) doi:10.2196/16987

KEYWORDS

CPR training; gamified learning; medical education; serious game learning

Introduction

Advanced Life Support (ALS), widely accepted as the gold standard of care in patients with cardiopulmonary arrest, requires specialized skills and knowledge [1,2]. Training in ALS is thus recommended for health care providers who are likely to encounter either in-hospital or out-of-hospital cardiac arrest (OHCA) [2,3]. The ALS training course covers the knowledge, skills, and attitudes required for managing both cardiac arrest and periarrest problems. It includes lecture-based teaching, workshops, skill stations, simulation-based training with high-fidelity mannequins, and debriefing [4-6]. In order to maximize learning outcomes, the current American Heart Association guidelines for cardiopulmonary resuscitation (CPR) recommend students prepare in advance for the ALS training course [7]. This can be addressed in a variety of ways including the provision of precourse reading (CD-based, e-learning–based, etc), online and precourse testing, and opportunities to practice pertinent technical skills [6]. However, there has been no evidence demonstrating the benefits of precourse training when taken in conjunction with a conventional ALS training program [7-10].

Over the past several years, gamified learning has been growing in popularity in various medical educational contexts including CPR training [11-20]. The benefits of this learning method are that it improves learning outcomes by creating a high level of engagement among participants, facilitating learners' holistic understanding of scientific concepts, and providing flexible learning methods and real-time feedback [18,21-23]. Recent studies have found that gamified learning can improve medical knowledge, skills, attitudes, and satisfaction when compared with traditional education methods [12,20,24]. Furthermore, prior work in Basic Life Support (BLS) training has demonstrated the benefits of serious games as a method for pretraining among medical students [18]. However, there is little evidence to support these benefits with regard to ALS training, with only one study reporting that game-based training augments retention of acquired skills and knowledge [13].

Resus Days is a serious smartphone game developed by a Thai physician using standard ALS guidelines and is available online (Figure 1) [25]. The aim of the game is to familiarize students with the ALS algorithms for cardiac arrest and periarrest scenarios, leading to shorter knowledge acquisition time. The objective of this study was to evaluate the effects of precourse preparation using this game on students' ALS knowledge, skills, and perceptions. We hypothesized that adding this game to the traditional precourse training would lead to higher ALS knowledge scores.

Figure 1. Screenshot from Resus Days.

Methods

Study Design

This was a prospective, randomized controlled parallel group trial in a tertiary university hospital in Thailand. It was approved by the Khon Kaen University institutional committee for ethics in human research and registered with the Thai Clinical Trials Registry [HE611533]. Written informed consent was obtained from each participant prior to enrollment.

Participants and Randomization

Participants were fifth-year medical students at the Khon Kaen University Faculty of Medicine who attended the ALS training course offered by the faculty’s CPR unit in April 2019. Eligibility criteria included not having participated in any type of ALS course in the past and never having played Resus Days. After enrollment, participants were randomly assigned to one of two groups (control group or game group) at a 1:1 allocation ratio based on a computer-generated random number list [26].

Sample Size Calculation

The sample size for analysis of covariance of two levels and two covariates was determined based on our pilot data using power analysis. We hypothesized that the difference in knowledge between the two groups was approximately 5%, and that there would be moderate consistency between pretest and posttest scores. The power analysis was determined using an alpha of .05 and power of 0.80. This resulted in an estimated desired effect sample size of at least 22 subjects in each group.

Study Interventions

Prior to any intervention, we tested students’ baseline ALS knowledge using a multiple-choice pretest. Students in the game group then played Resus Days using their own smartphone for 1 hour in normal mode. After this, students in both groups attended an ALS training course taught by CPR instructors. Students in the game group were allowed to play Resus Days as much as they wished during the 2-day ALS training course.

Game Description

Resus Days was developed by Rath Panyowat. It is a mobile simulation game that allows health care professionals to practice CPR through repetitive playing. Players must resuscitate a simulated patient in various scenarios by choosing treatment methods based on the ALS algorithm until the patient’s cardiac rhythm becomes normal. There are two game modes: normal and physician. In normal mode, a player receives hints regarding diagnosis and treatment and gains points if the prescribed treatment is correct. In physician mode, there are no hints and points are subtracted if the prescribed treatment is incorrect. There are a total of 7 scenarios including cardiac arrest, bradycardia, tachycardia, and several simulated megacode (mixed algorithm) scenarios. To win in each scenario, the player must strictly follow the ALS algorithm within the time allotted (Figure 1).

Advanced Life Support Training Course

The ALS training course was held over 2 days. The first day of the program consisted of approximately 6 hours of lectures, and the second day consisted of approximately 8 hours of hands-on workshops, skill station activities, simulation-based training with high-fidelity ALS simulator mannequins with the SimPad system (Laerdal Medical). The student-to-instructor ratio was 1:6. At the end of the course, student ALS knowledge was reassessed using multiple-choice questions in accordance with ALS guidelines, and their practical skills were evaluated individually by examiners blinded to group allocation. We used a questionnaire to assess student attitudes and perceptions regarding both the course and the game.

Assessment of Learning Outcomes, Attitudes, and Perceptions

Knowledge Test

The knowledge test was divided into two parts: one to assess knowledge of the ALS algorithm and another to assess general ALS knowledge. The first part consisted of 20 multiple-choice questions, each with 5 possible answers, and the second consisted of 30 multiple-choice questions. The ALS algorithm test focused on management of life-threatening conditions (using the ALS adult cardiac arrest algorithm) and symptomatic bradycardia and tachycardia. The general ALS knowledge test focused on BLS knowledge, general ALS knowledge, and postresuscitation care (see Multimedia Appendix 1).

Advanced Life Support Skill Tests

Each student was given a series of three different 5-minute megacode (mixed algorithm) simulated ALS scenarios (chosen randomly using the closed envelope method), and their performance was evaluated by certified ALS course providers (see Multimedia Appendix 2 for details regarding each scenario). Students passed the skill tests if they had scores of 80% or greater and no critical errors.

Questionnaire Regarding Participant Attitudes and Perceptions

We asked students to rate their attitudes and perceptions regarding the game on a scale of 1 to 10 (1=disagree completely and 10=agree completely). An open-ended question was used to assess any problems students had while playing the game, opinions about the game, and recommendations for further improvement (see Multimedia Appendix 3).

Game Score

Students received a score from 0 (if the patient died) to 3 (if patient survived and the most appropriate treatments were prescribed) for each scenario. The highest possible overall score was 21 points (see Multimedia Appendix 4).

Statistical Analysis

Categorical variables were expressed as frequencies and percentages. Continuous data were expressed as means and standard deviations. Differences between groups in terms of baseline characteristics were compared using an independent sample *t* test. An analysis of covariance model was used to compare posttest scores between the two groups adjusted for baseline score measurements. The Pearson chi-square test was used to compare the skill-test pass rate and other binary variables. The correlation between game scores before ALS training and knowledge pretest scores and between game scores post-ALS training and knowledge posttest scores were analyzed using Pearson correlation coefficients. All data analyses were performed using Stata version 10 (StataCorp LLC).

Results

A total of 105 students participated in the study and were randomly assigned to either the game group (n=52) or the control group (n=53; Figure 2). All students completed the trial and were included in the data analysis. Demographic characteristics of the two study groups are shown in Table 1. There were no significant differences between the groups in terms of sex, age, grade point average, CPR experience, CPR confidence, or learning outcome scores.

Figure 2. Flow diagram of participant selection.

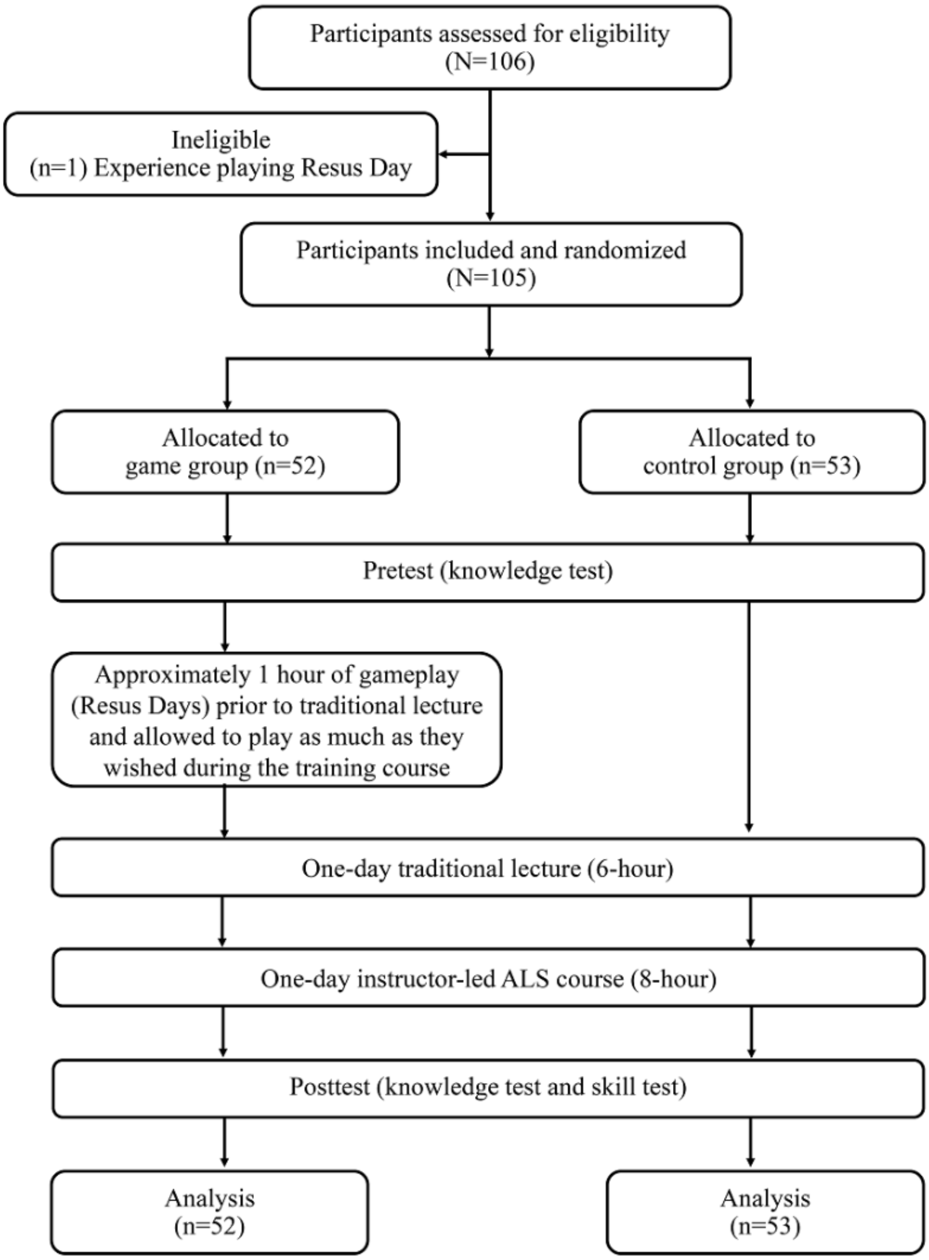


Table 1. Participant demographics.

Characteristic	Game group (n=52)	Control group (n=53)	*P* value
Age in years, mean (SD)	23.06 (1.02)	23.02 (1.07)	.85
Male, n (%)	33 (63)	34 (64)	.94
GPA[a], range 0.00-4.00, mean (SD)	3.33 (0.32)	3.33 (0.38)	.97
CPR[b] member experience, n (%)	38 (73)	42 (79)	.61
Knowledge pretest score, mean (SD)			
Algorithm knowledge test	7.19 (3.02)	8.17 (3.17)	.11
General knowledge test	12.54 (2.90)	12.79 (3.04)	.66
Pretraining CPR confidence (1-10), mean (SD)	3.04 (2.06)	3.17 (2.27)	.86

[a] GPA: grade point average.

[b] CPR: cardiopulmonary respiration.

Students in the game group performed better on the ALS algorithm knowledge posttest than those in the control group (17.22 [SD 1.93] vs 16.60 [SD 1.97], *P*=.01; adjusted mean difference [AMD] 0.93; 95% CI 0.21 to 1.66). They also had a higher pass rate on the skill test (79% vs 66%, *P*=.09; adjusted odds ratio [AOR] 2.22; 95% CI 0.89 to 5.51) and indicated greater confidence in performing CPR (7.87 [SD 1.05] vs 7.85 [SD 2.26], *P*=.51; 95% CI –0.45 to 0.92; Table 2). However, these differences were not statistically significant. Students also indicated high satisfaction with the game (9.02 [SD 1.11] out of 10). The correlations between pre- and posttraining game scores and knowledge scores are shown in Figures 3 and 4. There was little correlation between pretraining game and pretest scores (*P*=.01, *r*=.37; Figure 3).

Student attitudes and perceptions regarding Resus Days are summarized in Table 3. Most students were satisfied with the game and indicated that it helped them memorize the CPR algorithm and drug dosages and increased their confidence in making decisions.

Table 2. Comparison of learning outcomes between the game group and control group.

Characteristic	Game group (n=52)	Control group (n=53)	AMD[a]/AOR[b] (95% CI)	*P* values
Algorithm knowledge score, mean (SD)				
Posttest score[c]	17.33 (1.93)	16.60 (1.97)	0.93 (0.21 to 1.66)[a]	.01
Score improvement (post-pre)	10.13 (3.00)	8.43 (3.10)	—	
General knowledge score				
Posttest score[d], mean (SD)	22.88 (2.49)	23.45 (2.89)	–1.69 (–6.11 to 2.72)[a]	.45
Score improvement (post-pre), mean (SD)	10.35 (3.00)	10.66 (3.63)	—	
ALS[e] skill test pass[f], n (%)	41 (79)	35 (66)	2.22 (0.89 to 5.51)[b]	.09
CPR[g] confidence[d] (postcourse), mean (SD)	7.87 (1.05)	7.85 (2.26)	0.23 (–0.45 to 0.92)[a]	.51

[a] AMD: adjusted mean difference.

[b] AOR: adjusted odds ratio.

[c] Analysis of covariance model adjusted for baseline measurement.

[d] Analysis of covariance model adjusted for baseline measurement and interaction.

[e] ALS: Advanced Life Support.

[f] Binary logistic regression (adjusted for grade point average, sex, and knowledge pretest score).

[g] CPR: cardiopulmonary resuscitation.

Figure 3. Correlation between game score before Advanced Life Support training and knowledge pretest score was low (*P*=.01, r=.37).

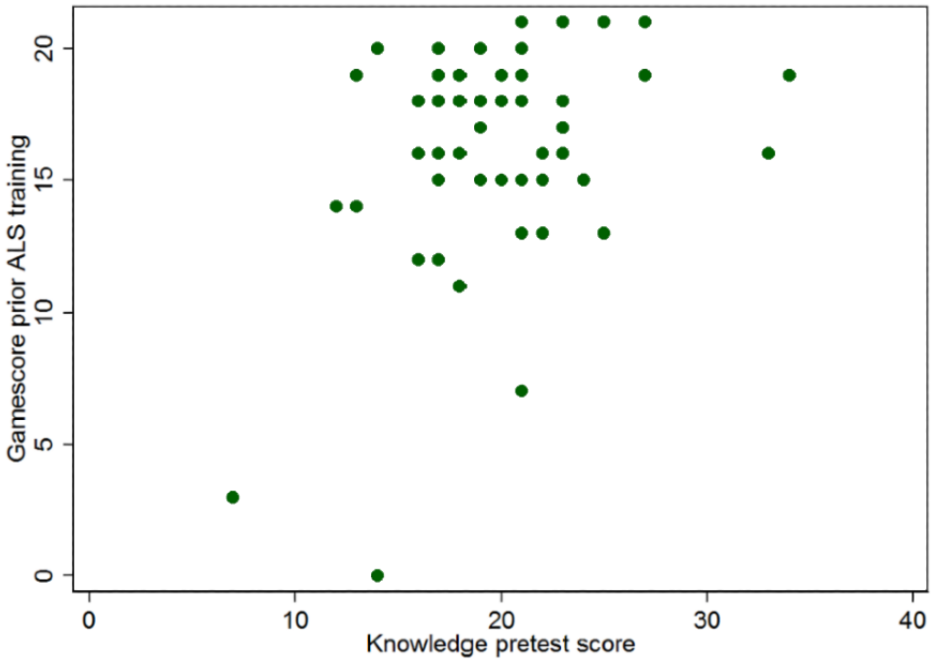


Figure 4. Correlation between game score post-Advanced Life Support training and knowledge posttest score was low (*P*=.07, r=.25).

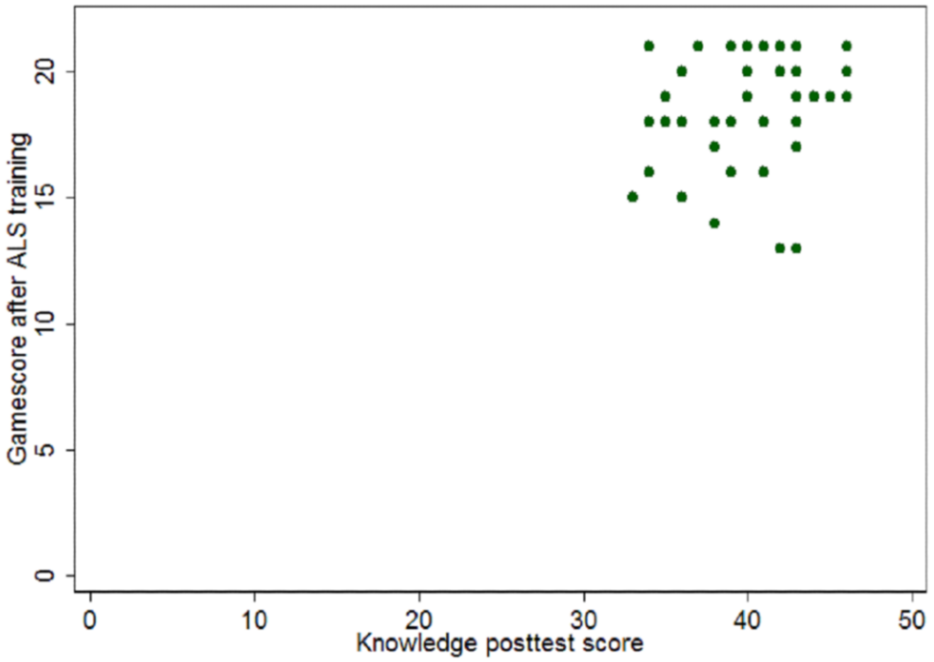


Table 3. Participant attitudes and perceptions regarding Resus Days based on a score of 1=strongly disagree to 10=strongly agree.

Attitudes and perceptions	Participant in game group (n=52)
Overall satisfaction with Resus Days, median (IQR[a])	9 (8-10)
Resus Days helped improve my CPR[a] knowledge, median (IQR)	9 (8-10)
If you think game helped improve your ALS[a] knowledge, please specify a reason (students could select more than one answer; n=62), n(%)	
The game helped me become familiar with cardiac arrest scenarios	15 (24)
The game emphasized critical points in the ALS algorithm	13 (21)
Repeatedly playing the game helped me memorized the CPR algorithm and correct medication dosages	10 (16)
The game improved my decision-making ability and made me more confident	7 (11)
Other	17 (27)
Suggestions for improving the game, n (%)	
Create more scenarios (eg, postcardiac arrest care, myocardial infarction)	22 (42)
Make changes with regard to gameplay	6 (12)
Allow the player to choose the drug doses	5 (10)
Bug fixes	2 (4)
Other	4 (8)
None	13 (25)
Frequency of playing any other online games, n (%)	
1 day per week	14 (27)
Every day	12 (23)
<1 day per week	11 (21)
Every 2 to 3 days	8 (15)
Never	7 (14)

[a] IQR: interquartile range.

[b] CPR: cardiopulmonary resuscitation.

[c] ALS: Advanced Life Support.

Discussion

Principal Findings

ALS certification is mandatory for all medical students entering their clinical years. This study demonstrates the merits of using a serious video game to augment ALS knowledge, especially regarding the ALS algorithm. In addition, the game group had a higher pass rate on the ALS skill test and indicated greater satisfaction with their training.

Previous studies have also examine the effectiveness of games in helping students practice their resuscitation skills. Creutzfeldt et al [18], for example, conducted a study that evaluated the efficacy of game technology as part of a pre-BLS training program for medical students, which led to improved BLS knowledge and performance. These improvements were associated with adherence to the guidelines, which resulted in students performing chest compressions with the appropriate frequency. However, that study had a limited sample size, and its findings required confirmation. Our findings are consistent with those of Creutzfeldt et al [18], but we had a larger sample size and applied game-based learning to training in ALS, which has a more complicated algorithm. The game group exhibited clear improvement with regard to ALS algorithm knowledge. They also demonstrated improvements in terms of skill, but not to a statistically significant extent.

Several studies have reported beneficial effects of game-based training on medical student learning outcomes [18,21-23]. A large number of the medical students in our study had positive attitudes about Resus Days, indicating that the game helped them become more familiar with various cardiac arrest scenarios prior to the ALS training course. These findings are in line with those of a study by Cheng et al [22], which found that game-based learning helped learners develop a holistic understanding of scientific concepts. Furthermore, some students stated that playing the game helped them memorize the CPR algorithm and medications. This may be due to improvements in cognitive function as a result of repetitive playing. Step-by-step game playing, in which scenarios become increasingly complicated, has been shown to stimulate critical thinking and increase engagement [18,21].

Another interesting finding was the low correlation between student game scores and their knowledge scores. This can be explained by the fact that some students may not have been familiar with the game mechanics at first, resulting in them getting low scores despite being knowledgeable about ALS and having good pretest or posttest scores. Because of this, we do not recommend using a player's game score as a substitute for traditional testing. However, regardless of players' scores, their knowledge was significantly improved by playing the game. This is partially because the game provided the correct answers at the end of each scenario even if the player failed.

Several methods have been proposed to prepare trainees before attending ALS training courses. We found that engaging in game-based preparation prior to precourse ALS lectures (consisting of precourse algorithm rehearsal and lectures accompanied by Microsoft PowerPoint slides or lecture notes) was more beneficial than conventional precourse lectures alone. Although it was only an add-on intervention, the observed benefits demonstrated the ability of gamified training to improve a complicated medical course.

Most of the studies that have demonstrated the benefits of game-based learning have been conducted in high-income countries [20]. This study, however, was conducted in a low-middle income country and found that, even in this setting, game-based learning was practical and effective. In the near future, this may be used to help train large numbers of medical students or personnel in remote areas who have little or no access to traditional classroom-based lectures. However, if game-based learning is to replace conventional ALS lectures, a randomized controlled trial using a well-designed game may be necessary.

Strengths and Limitations

One strength of our study is that it is the first to demonstrate the effectiveness of ALS pretraining using a smartphone game and addresses the feasibility of gamified learning in training students to treat highly complicated medical conditions. Second, we used a game that is available in an online store and therefore available to anyone with access to a compatible device. Although previous studies have developed their own games as tools for research [15,17,19], the initial design of such a study is costly and time consuming. Moreover, using a custom game specifically developed for a particular study would limit its generalizability. Third, as our study was a randomized controlled trial, differences in baseline characteristics between students in the game and control groups were limited. Finally, we performed a comparison of gamified learning with a standard ALS training course that was provided by certified CPR providers and that met international standards.

However, our study also had some potential limitations. First, this was a single-center study, which limits the generalizability of the results. Second, we evaluated students based on their knowledge scores, which are imperfect indicators of how someone will perform in real-world scenarios. Third, the sample size was only large enough to evaluate student knowledge scores and not their results on the ALS skill test. Fourth, most of the participants in this study were medical students who were familiar with smartphone games. It is unclear how effective this kind of gamified learning would be in an older population such as senior doctors. Fifth, the students in the game group were assigned to play Resus Days for 1 hour before attending the lecture and were allowed to play as much as they wished during the training course; however, we did not have specific protocol and did not collect any data regarding the frequency of gameplay during the 2-day training. Finally, the content in this game was developed by an individual physician based on standard guidelines, not by a content expert or trusted organization.

Conclusions

Our results demonstrate that incorporating a serious smartphone game (Resus Days) in medical students' ALS training leads to better algorithm knowledge scores than conventional training alone. In addition, students in the game group indicated high satisfaction with their training. Although there was no significant difference between the two groups in terms of their pass rates on the ALS skills test, those who trained using the serious game tended to do better.

Acknowledgments

We would like to thank Dr Rath Panyowat, who developed Resus Days. We would like to express our gratitude to Associate Professor Virat Klungboonkrong, Associate Professor Boonsong Patjanasoontorn, and Associate Professor Pattarapong Makarawate at the CPR training unit at Khon Kaen University Faculty of Medicine's Srinagarind Hospital for their advice and assistance in project coordination. Finally, we thank Dr Dylan Southard, Associate Professor Chulathida Chomchai, and Professor Timothy E O'Brien for English manuscript editing. This study was funded by a grant from the Faculty of Medicine, Khon Kaen University, Thailand (grant number IN62216).

Authors' Contributions

PP conceived the study, designed the trial, obtained research funding, recruited participants, and managed the data. SP conceived the study, recruited participants, managed the data, and drafted the manuscript. SC and SM recruited participants. PK and KA drafted the manuscript. TM takes responsibility for the paper as a whole, and all authors contributed substantially to its revision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Knowledge test.

[PDF File (Adobe PDF File), 459 KB - jmir_v22i3e16987_app1.pdf]

Multimedia Appendix 2
Skill tests.
[PDF File (Adobe PDF File), 364 KB - jmir_v22i3e16987_app2.pdf]

Multimedia Appendix 3
Questionnaire given to intervention group (adapted from original questionnaire in Thai language).
[PDF File (Adobe PDF File), 95 KB - jmir_v22i3e16987_app3.pdf]

Multimedia Appendix 4
Resus Days screenshots.
[PDF File (Adobe PDF File), 906 KB - jmir_v22i3e16987_app4.pdf]

Multimedia Appendix 5
CONSORT-EHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 576 KB - jmir_v22i3e16987_app5.pdf]

References

1. Link MS, Berkow LC, Kudenchuk PJ, Halperin HR, Hess EP, Moitra VK, et al. Part 7: Adult Advanced Cardiovascular Life Support: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2015 Nov 03;132(18 Suppl 2):S444-S464. [doi: 10.1161/CIR.0000000000000261] [Medline: 26472995]
2. Lockey A, Lin Y, Cheng A. Impact of adult advanced cardiac life support course participation on patient outcomes: a systematic review and meta-analysis. Resuscitation 2018 Aug;129:48-54. [doi: 10.1016/j.resuscitation.2018.05.034] [Medline: 29902494]
3. Moretti MA, Cesar LAM, Nusbacher A, Kern KB, Timerman S, Ramires JAF. Advanced cardiac life support training improves long-term survival from in-hospital cardiac arrest. Resuscitation 2007 Mar;72(3):458-465. [doi: 10.1016/j.resuscitation.2006.06.039] [Medline: 17307620]
4. Jensen ML, Lippert F, Hesselfeldt R, Rasmussen MB, Mogensen SS, Jensen MK, et al. The significance of clinical experience on learning outcome from resuscitation training: a randomised controlled study. Resuscitation 2009 Feb;80(2):238-243. [doi: 10.1016/j.resuscitation.2008.10.026] [Medline: 19058890]
5. Baskett PJF, Nolan JP, Handley A, Soar J, Biarent D, Richmond S, European Resuscitation Council. European Resuscitation Council guidelines for resuscitation 2005. Section 9. Principles of training in resuscitation. Resuscitation 2005 Dec;67 Suppl 1:S181-S189. [doi: 10.1016/j.resuscitation.2005.10.006] [Medline: 16321713]
6. Greif R, Lockey AS, Conaghan P, Lippert A, De Vries W, Monsieurs KG, et al. European Resuscitation Council Guidelines for Resuscitation 2015. Resuscitation 2015 Oct;95:288-301. [doi: 10.1016/j.resuscitation.2015.07.032]
7. 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]
8. Perkins GD, Fullerton JN, Davis-Gomez N, Davies RP, Baldock C, Stevens H, et al. The effect of pre-course e-learning prior to advanced life support training: a randomised controlled trial. Resuscitation 2010 Jul;81(7):877-881. [doi: 10.1016/j.resuscitation.2010.03.019] [Medline: 20409627]
9. Perkins GD, Kimani P, Bullock I, Clutton-Brock T, Davies RP, Gale M, Electronic Advanced Life Support Collaborators. Improving the efficiency of advanced life support training: a randomized, controlled trial. Ann Intern Med 2012 Jul 03;157(1):19-28. [doi: 10.7326/0003-4819-157-1-201207030-00005] [Medline: 22751757]
10. Thorne C, Lockey A, Bullock I, Hampshire S, Begum-Ali S, Perkins G. e-Learning in advanced life support: an evaluation by the Resuscitation Council (UK). Resuscitation 2015 May;90:79-84. [doi: 10.1016/j.resuscitation.2015.02.026]
11. Cheng A, Nadkarni VM, Mancini MB, Hunt EA, Sinz EH, Merchant RM, et al. 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):1. [doi: 10.1161/cir.0000000000000583] [Medline: 29930020]
12. Kron FW, Gjerde CL, Sen A, Fetters MD. Medical student attitudes toward video games and related new media technologies in medical education. BMC Med Educ 2010;10:50 [FREE Full text] [doi: 10.1186/1472-6920-10-50] [Medline: 20576125]
13. Buttussi F, Pellis T, Cabas Vidani A, Pausler D, Carchietti E, Chittaro L. Evaluation of a 3D serious game for advanced life support retraining. Int J Med Inform 2013 Sep;82(9):798-809. [doi: 10.1016/j.ijmedinf.2013.05.007] [Medline: 23763908]
14. Chalhoub M, Khazzaka A, Sarkis R, Sleiman Z. The role of smartphone game applications in improving laparoscopic skills. Adv Med Educ Pract 2018;9:541-547 [FREE Full text] [doi: 10.2147/AMEP.S162619] [Medline: 30123018]

15. de Sena DP, Fabrício DD, da Silva VD, Bodanese LC, Franco AR. Comparative evaluation of video-based on-line course versus serious game for training medical students in cardiopulmonary resuscitation: a randomised trial. PLoS One 2019;14(4):e0214722 [FREE Full text] [doi: 10.1371/journal.pone.0214722] [Medline: 30958836]
16. Harrington CM, Chaitanya V, Dicker P, Traynor O, Kavanagh DO. Playing to your skills: a randomised controlled trial evaluating a dedicated video game for minimally invasive surgery. Surg Endosc 2018 Sep;32(9):3813-3821. [doi: 10.1007/s00464-018-6107-2] [Medline: 29445864]
17. Boada I, Rodriguez-Benitez A, Garcia-Gonzalez JM, Olivet J, Carreras V, Sbert M. Using a serious game to complement CPR instruction in a nurse faculty. Comput Methods Programs Biomed 2015 Nov;122(2):282-291. [doi: 10.1016/j.cmpb.2015.08.006] [Medline: 26319184]
18. Creutzfeldt J, Hedman L, Felländer-Tsai L. Effects of pre-training using serious game technology on CPR performance: an exploratory quasi-experimental transfer study. Scand J Trauma Resusc Emerg Med 2012 Dec 06;20:79 [FREE Full text] [doi: 10.1186/1757-7241-20-79] [Medline: 23217084]
19. Creutzfeldt J, Hedman L, Heinrichs L, Youngblood P, Felländer-Tsai L. Cardiopulmonary resuscitation training in high school using avatars in virtual worlds: an international feasibility study. J Med Internet Res 2013 Jan 14;15(1):e9 [FREE Full text] [doi: 10.2196/jmir.1715] [Medline: 23318253]
20. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
21. Zhonggen Y. A meta-analysis of use of serious games in education over a decade. Int J Comput Games Technol 2019 Feb 03;2019:1-8. [doi: 10.1155/2019/4797032]
22. Cheng M, Lin Y, She H, Kuo P. Is immersion of any value? Whether, and to what extent, game immersion experience during serious gaming affects science learning. Br J Educ Technol 2016 Jan 04;48(2):246-263. [doi: 10.1111/bjet.12386]
23. Garneli V, Giannakos M, Chorianopoulos K. Serious games as a malleable learning medium: the effects of narrative, gameplay, and making on students' performance and attitudes. Br J Educ Technol 2016 Apr 29;48(3):842-859. [doi: 10.1111/bjet.12455]
24. Subhash S, Cudney EA. Gamified learning in higher education: a systematic review of the literature. Comput Hum Behav 2018 Oct;87:192-206. [doi: 10.1016/j.chb.2018.05.028]
25. Panyowat R. Resuscitation Code Simulation Game. 2018 May 31. URL: https://resusdays.com/ [accessed 2020-02-21]
26. Reference randomizer. URL: https://www.randomizer.org/ [accessed 2020-02-21]

Abbreviations

ALS: Advanced Life Support
AMD: adjusted mean difference
AOR: adjusted odds ratio
BLS: Basic Life Support
CPR: cardiopulmonary resuscitation
OHCA: out-of-hospital cardiac arrest

Edited by G Eysenbach; submitted 09.11.19; peer-reviewed by J Wilkinson, APS Trakulsrichai; comments to author 17.12.19; revised version received 20.01.20; accepted 09.02.20; published 09.03.20.

Please cite as:
Phungoen P, Promto S, Chanthawatthanarak S, Maneepong S, Apiratwarakul K, Kotruchin P, Mitsungnern T
Precourse Preparation Using a Serious Smartphone Game on Advanced Life Support Knowledge and Skills: Randomized Controlled Trial
J Med Internet Res 2020;22(3):e16987
URL: http://www.jmir.org/2020/3/e16987/
doi:10.2196/16987
PMID:32149711

Original Paper

The Association Between Pain Relief Using Video Games and an Increase in Vagal Tone in Children With Cancer: Analytic Observational Study With a Quasi-Experimental Pre/Posttest Methodology

Mario Alonso Puig[1], MD; Mercedes Alonso-Prieto[2], MD; Jordi Miró[3], PhD; Raquel Torres-Luna[2], MSN; Diego Plaza López de Sabando[4], MD; Francisco Reinoso-Barbero[2,5], PhD

[1]Juegaterapia Foundation, Madrid, Spain

[2]Pediatric Pain Unit, Anesthesiology-Critical Care Service, University La Paz Hospital, Madrid, Spain

[3]Department of Psychology, Unit for the Study and Treatment of Pain-ALGOS, Rovira i Virgili University, Tarragona, Spain

[4]Pediatric Hemato-oncology Service, University La Paz Hospital, Madrid, Spain

[5]Department of Anatomy-Histology and Neuroscience, School of Medicine, Universidad Autónoma de Madrid, Madrid, Spain

Corresponding Author:
Mercedes Alonso-Prieto, MD
Pediatric Pain Unit
Anesthesiology-Critical Care Service
University La Paz Hospital
Paseo de la Castellana, 261
Madrid, 28046
Spain
Phone: 34 676687569 ext 676687569
Email: meme27@hotmail.es

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/7/e19961/

Abstract

Background: Patients with secondary pain due to mucositis after chemotherapy require treatment with morphine. Use of electronic video games (EVGs) has been shown to be an effective method of analgesia in other clinical settings.

Objective: The main objective of this study was to assess the association between the use of EVGs and the intensity of pain caused by chemotherapy-induced mucositis in pediatric patients with cancer. The secondary objective was to assess the association between changes in pain intensity and sympathetic-parasympathetic balance in this sample of pediatric patients.

Methods: Clinical records were compared between the day prior to the use of EVGs and the day after the use of EVGs. The variables were variations in pupil size measured using the AlgiScan video pupilometer (IDMed, Marseille, France), heart rate variability measured using the Analgesia Nociception Index (ANI) monitor (Mdoloris Medical Systems, Loos, France), intensity of pain measured using the Numerical Rating Scale (score 0-10), and self-administered morphine pump parameters.

Results: Twenty patients (11 girls and nine boys; mean age 11.5 years, SD 4.5 years; mean weight 41.5 kg, SD 20.7 kg) who met all the inclusion criteria were recruited. EVGs were played for a mean of 2.3 (SD 1.3) hours per day, resulting in statistically significant changes. After playing EVGs, there was significantly lower daily morphine use (before vs after playing EVGs: 35.9 vs 28.6 µg/kg/day, *P*=.003), lower demand for additional pain relief medication (17 vs 9.6 boluses in 24 hours, *P*=.001), lower scores of incidental pain intensity (7.7 vs 5.4, *P*=.001), lower scores of resting pain (4.8 vs 3.2, *P*=.01), and higher basal parasympathetic tone as measured using the ANI monitor (61.8 vs 71.9, *P*=.009). No variation in pupil size was observed with the use of EVGs.

Conclusions: The use of EVGs in pediatric patients with chemotherapy-induced mucositis has a considerable analgesic effect, which is associated physiologically with an increase in parasympathetic vagal tone despite lower consumption of morphine.

(J Med Internet Res 2020;22(3):e16013) doi:10.2196/16013

KEYWORDS

hematology oncology; pediatric patient; acute pain; patient-controlled analgesia; video pupilometer; analgesia nociception index

Introduction

Pediatric patients with cancer may experience chronic pain associated with the development of the illness itself, which is caused by compression or infiltration of nearby organs or nerves or by pathologic bone fractures. However, in this population, it is very common to observe acute pain specifically associated with the diagnostic or therapeutic procedures necessary for treatment [1]. Chemotherapy-induced mucositis (International Classification of Diseases 10th Revision: K12.30; International Classification of Diseases 11th Revision: DA01.11) is a relatively common complication associated with the use of chemotherapy in the treatment of young patients with cancers, such as leukemia [2]. It is associated with intense pain, and it has a considerable impact on patient quality of life, patient mood, and the current and future evolution of the illness [3].

Acute nociception has been found to be associated with changes in the sympathetic-parasympathetic balance, including the response of pupil dilation and respiratory variations involving heart rate. A positive association has been proven between the magnitude of these vegetative changes and pain intensity [4].

Given the impact of acute pain associated with chemotherapy-induced mucositis, it is very important to administer adequate analgesia. Some of the recommended measures include lidocaine mouthwash and regulated use of analgesics, particularly strong ones such as morphine, with a patient-controlled analgesia (PCA) pump, which allows the patient to administer additional doses of morphine during peaks of intense pain [5]. However, the use of this potent opioid analgesia is not free of likely adverse effects, such as nausea or vomiting, drowsiness, itching, and urinary retention, and very serious or fatal complications, such as respiratory depression [6].

Distraction has been widely used to reduce the intensity of pain and improve function in individuals with pain, especially young individuals [7]. So-called new technologies are increasingly being used to facilitate distraction and increase effectiveness in different groups of patients with pain [8]. Electronic video games (EVGs) are of particular interest in children and adolescents owing to their proven record of attracting the attention of these individuals [9]. Although studies have shown strong positive associations between the use of EVGs and the reduction of pain intensity [10], as well as the functional adaptation of patients while undergoing different potentially painful medical interventions [7], no studies to date have proven the effectiveness of this approach for this type of pain associated with cancer-related mucositis.

Given these considerations, the main objective of this study was to assess the association between the use of EVGs and the intensity of pain (measured using the Numerical Rating Scale [NRS]), which was caused by chemotherapy-induced mucositis, in a sample of pediatric patients with cancer. As a secondary objective, the study assessed the associations of changes in pain intensity with vegetative changes in the sympathetic-parasympathetic balance assessed using a Analgesia Nociception Index (ANI) monitor (Mdoloris Medical Systems, Loos, France) and variations in pupil size assessed using AlgiScan (IDMed, Marseille, France) in this sample of pediatric patients. We hypothesized that EVG use is significantly and negatively associated with pain intensity ratings and morphine consumption. We also hypothesized that EVG use is positively associated with vagal tone.

Methods

Study Design

We designed an analytic observational study with a quasi-experimental pre/posttest methodology.

Participants

Participants in this study were children and adolescents with pain secondary to mucositis, who were receiving treatment in our tertiary hospital in relation to cancer [4]. In order to take part in this study, potential participants had to provide approval. In addition, written informed consent was requested from their parents. We excluded children complaining of pain prior to mucositis. Finally, potential participants could not have any intellectual disability that might interfere with the correct use of video games.

The study included all patients aged between 4 and 17 years, who were hospitalized at the children's oncology ward of the hospital and who received treatment between January 2016 and December 2017 for acute pain caused by grade 3 or 4 mucositis in the context of treatment for cancer [2]. According to the World Health Organization, the grades of mucositis are as follows: grade 1, soreness with or without erythema; grade 2, erythema ulcers (patient can consume a solid diet); grade 3, ulcers and extensive erythema (patient cannot consume a solid diet); and grade 4, mucositis to the extent that alimentation is not possible. Grades 3 and 4 are considered to indicate severe mucositis. Children with an intellectual inability to understand video games and those who refused to play them were excluded from the study.

Procedure

Potential participants were invited when a PCA pump (model Sapphire; Hospira, Illinois, USA) with morphine for pain relief was indicated. At that point, the goals of the study were explained. Patients were included in the treatment protocol for acute pain at the children's pain service, which involved a PCA pump that administered morphine chloride (1 mg morphine/kg body weight up to a maximum of 50 mg, diluted in 100 mL of saline) with the following parameters: continuous perfusion at 0-1 mL/h and a bolus on demand of 1-2 mL every 5 minutes, up to a maximum of 10 mL in 4 hours.

If a patient was under 12 years old, one of the parents or guardians was considered responsible for signing the consent form.

Those interested in taking part were offered the possibility of playing video games on PlayStation Vita (Sony Interactive Entertainment España, Madrid, Spain) on demand, with a range of different alternatives available depending on the patient's preference and age as follows: age-rated puzzles, sports, platforms, and strategy games.

After 24 hours of established analgesia, patients were usually visited by members of the children's pain service at 9 am for 30 minutes to compile information on the variables adopted (ie, pain and anxiety intensities, PCA parameters, and ANI and AlgiScan values). Figure 1 shows one of the participants with a PCA morphine pump while playing an EVG. The information on outcome variables was again collected during a visit at the same time on the following day after having played the EVG.

Figure 1. Child playing an electronic video game while being fitted with a patient-controlled analgesia pump and monitored for respiratory heart rate variability.

The study was approved by the Ethics Committee at La Paz University Hospital on May 21, 2015 (code: PI-1217).

Measures

With regard to demographic and descriptive variables, information on age and sex was collected for descriptive purposes.

Participants were asked to report their pain intensity using the NRS-11 (scores 0-10). Participants were asked to rate their average pain intensity during the past 24 hours by choosing a single whole number between 0 (no pain) and 10 (pain as bad as could be), which best represented the pain during most of the day under rest conditions (basal or resting pain) and especially when patients tried to swallow even their own saliva (incidental pain). The NRS-11 has been found to provide valid and reliable scores when used in young people, including children as young as 6 years of age [11]. We eventually included a 4-year-old girl because she was very co-operative and understood the NRS-11 perfectly (we compared her NRS score with the score of an observational pain scale validated for her age and noted the same score).

In addition to this information, we developed a survey to collect information about the following: (1) PCA use in the previous 24 hours (information about the total morphine dose administered and the required bolus/administered bolus); (2) ANI parameters (where 100 indicates maximum parasympathetic predominance and 0 indicates maximum sympathetic predominance; maximum, minimum, and mean values during the 30-minute visit); and (3) parameters for pupil dilation (in mm) according to the pupillometry index at three time points (beginning of the visit, after 15 minutes, and end of the visit).

Finally, information was collected about the amount of time spent playing EVGs, with children's relatives asked to report the daily total minutes played.

Data Analysis Plan

We first calculated percentages, means, and SDs of the study variables for descriptive purposes. A Wilcoxon test was performed for categorical variables (ie, pain scores), and a Student *t* test was performed for ordinal variables (ie, dose of morphine, number of boluses required, ANI score, and pupil size).

Results

Sample Description

Of 25 patients invited to join the study, 20 (80%) agreed to participate in the study. Of these 20 patients, 12 had acute lymphoblastic leukemia, five had received a bone marrow transplant, and the remaining three had acute myeloid leukemia. One 6-year-old girl showed clinical signs of probable septic shock that required admission to the pediatric intensive care unit, and she was excluded from the study.

Of the remaining 19 patients, 10 were girls and nine were boys, with a mean age of 11.5 (SD 4.5; range 4-17) years. EVGs were played for a mean of 2.3 (SD 1.3) hours per day. Table 1 shows the other descriptive details and the study variables.

Table 1. Parameters of the patient-controlled analgesia pump, Analgesia Nociception Index monitor, and AlgiScan monitor on treatment days.

Variable	Without video games, mean (SD)	With video games, mean (SD)	Student *t* test ($t_{1,18}$)	*P* value
Daily total dose of morphine (μg/kg)	35.9 (27.0)	28.6 (27.1)	3.51	.003
Number of required boluses of morphine	17 (14.9)	9.5 (9.5)	2.87	.01
ANI[a], medial	62.1 (12.4)	71.9 (14.8)	−3.01	.009
ANI, maximum	79.6 (12.6)	86.0 (9.4)	−2.01	.06
ANI, minimum	48.2 (9.6)	55.7 (14.6)	−2.20	.045
Pupil size (mm)	4.9 (0.3)	5.3 (0.5)	−0.37	.37

[a] ANI: Analgesia Nociception Index.

Effects of the Use of Electronic Video Games on Subjective Variables

The results of the completed analyses showed that the mean incidental pain intensity reported in the previous 24 hours reduced significantly after playing EVGs (mean score with vs without EVGs: 5.4, SD 2.7 vs 7.7, SD 2.3; Z=−3.189; *P*=.001) (Figure 2). Similarly, the mean basal pain intensity reported in the previous 24 hours reduced significantly after playing EVGs (mean score with vs without EVGs: 3.2, SD 2.4 vs 4.8, SD 2.8; Z=−2.570; *P*=.01) (Figure 3).

Figure 2. Distribution of hours of playing video games with the maximum variation in incidental pain with and without the use of video games in each patient.

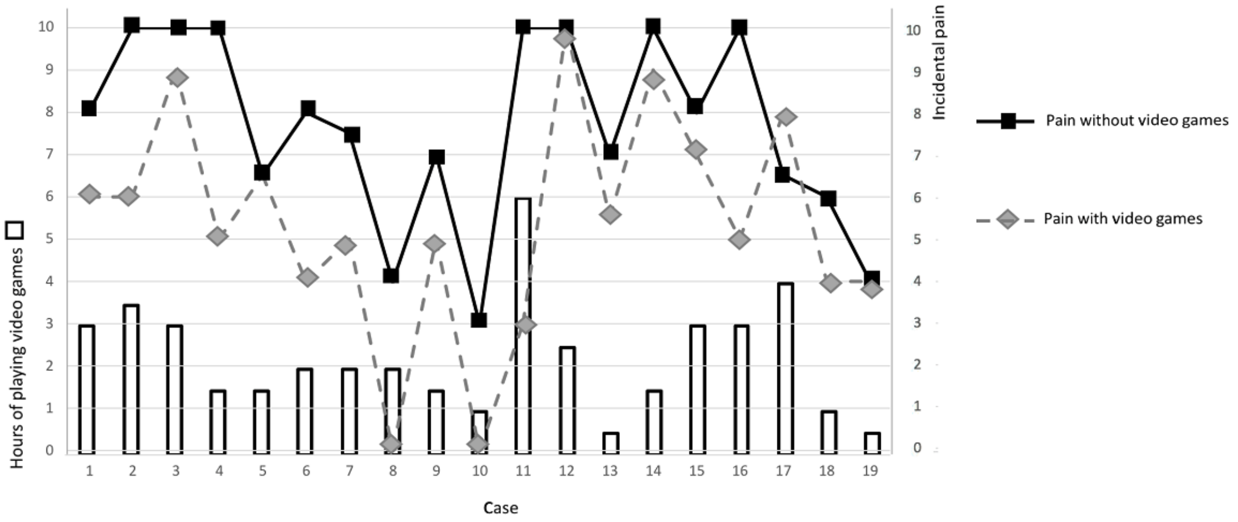
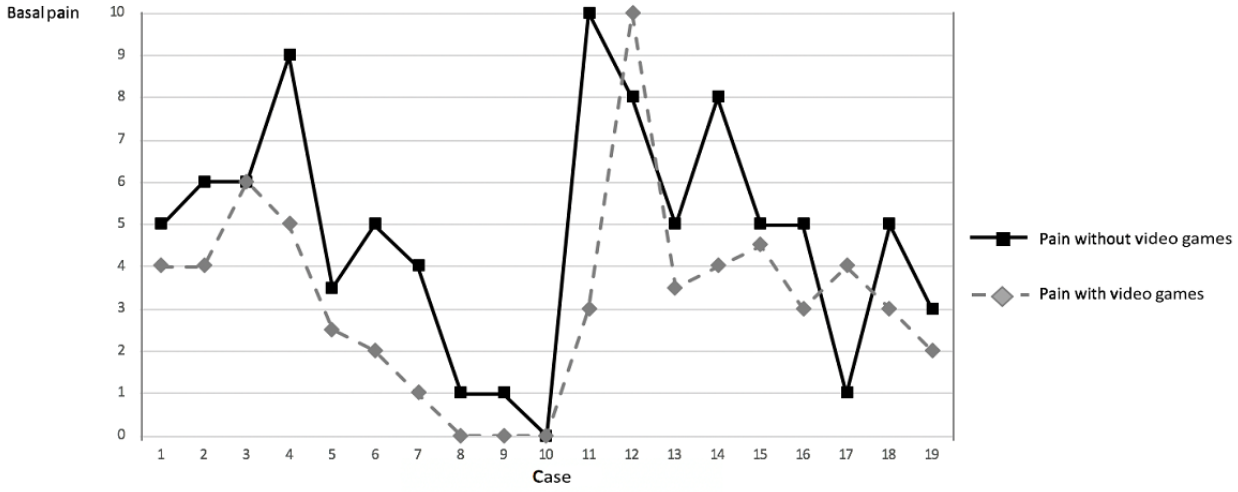


Figure 3. Variations in basal daily pain with and without the use of video games in each patient.



Effects of the Use of Electronic Video Games on Objective Variables

The parameters of the electronic PCA pump showed a 20% decrease in the daily consumption of morphine on the day the video games were played (mean value [µg/kg] with vs without EVGs: 28.6, SD 27.1 vs 35.9, SD 27.0; *P*=.003), with a 44% decrease in the number of on-demand boluses (mean number with vs without EVGs: 9.5, SD 9.5 vs 17, SD 14.9; *P*=.01).

The ANI parameters showed a statistically significant 14% increase in parasympathetic baseline tone on the day after playing the video games (mean value with vs without EVGs: 71.9, SD 14.8 vs 62.1, SD 12.4; *P*=.009) (Table 1). However, pupil size showed no changes in sympathetic tone between the two days (Table 1).

Discussion

To the best of our knowledge, this is the first clinical trial to study the efficacy of playing video games to reduce the pain associated with chemotherapy-induced mucositis related to the treatment of young patients with cancer. The results were as expected. This is to say that as hypothesized, participants reported lower levels of pain intensity after playing video games (significant reduction by 30% regarding incidental pain and 33% regarding basal pain; Figure 3). Furthermore, the results showed that this analgesic effect was accompanied by a vegetative association involving an increase in the parasympathetic tone of the autonomic nervous system. To date, this vegetative influence had not been studied in the management of acute pain (with a duration of a few days) in oncology or other fields.

The EVGs were especially effective in reducing incidental pain that is characteristically associated with mucositis (eg, when patients are not able to even swallow saliva), and we observed a highly remarkable reduction in the peak of pain, along with a 44% decrease in the need for additional morphine boluses to control incidental pain, resulting in a 20% decrease in the daily total dose of morphine administered. This is of particular importance as a reduction in the dose of morphine would logically result in a reduction in common morphine-associated side effects, such as paralytic ileus, nausea or vomiting, and pruritus [6].

Interestingly, the effect was so intense that it influenced the sympathetic-parasympathetic balance of the patients. A sympathetic-parasympathetic imbalance is caused by the presence of a harmful stimulus that tends toward sympathetic activation, while physiological recovery is achieved by the parasympathetic system [12].

The patients included in this study had secondary pain caused by mucositis. Chemotherapy-induced mucositis causes acute pain with a high intensity regarding pain sensation [13], as can be seen from the high initial pain levels reported by our patients, despite the regulated use of intravenous analgesia. In the case of our patients' mucositis, their pain could occasionally exacerbate, leading to very intense pain and great suffering, as well as emotional stress for their relatives [2]. This condition could also have a negative effect on our patients' prognosis, as it could be a source of infectious complications, such as sepsis, which was experienced by one of the patients who had consented to participate but required admission to the intensive care unit.

Our results are consistent with evidence-based medicine, which has shown that EVGs are effective in pediatric patients as both psychological therapy and treatment for boosting physical function [14]. In fact, psychologically, EVGs have been shown to reduce the anxiety associated with hospitalization or cause a state similar to mental relaxation in children [15]. Furthermore, games with an interactive element have been used to encourage physical exercise and thus reduce chronic pain in children with juvenile arthritis [16].

As analgesia, video games have been proven to be useful in reducing pain among children, especially in the case of pain associated with procedures for venipuncture [7], wound cleaning and dressing changes in burn victims [10], and other painful procedures related to other chronic illnesses [17].

The mean duration of video game use in this study was close to 3 hours per day, which is the value established in several studies as the threshold after which video games start to be more harmful than beneficial to health [18]. Thus, the general recommendation to not exceed this time of 3 hours a day of playing video games in hospitalized children should be considered by both parents and responsible physicians.

Video games have already been shown to be effective in pediatric oncology patients, encouraging physical rehabilitation during the patient's recovery phase after disease remission [19] and allowing relief from pain and anxiety associated with the placement of percutaneous central catheters [20]. As of now and according to the findings of this study, it is possible to recommend video games during the acute phase of painful oncological mucositis.

The ANI monitor measures parasympathetic activity through the analysis of respiratory sinus arrhythmia, assessing heart rate variability induced by each respiratory cycle (spontaneous or artificial). In conscious patients, the ANI indicates acute pain and stress levels. Normal values are higher than 50, and higher levels indicate higher parasympathetic-sympathetic balance. This device has been used successfully to measure the level of nociception during different painful situations, including surgery in children [21]. However, in this study, a 14% increase in the ANI was noted during the visit on the day after playing video games, suggesting that the beneficial effects of video games did not have a short duration but extended for several hours. However, no difference in pupil size was found despite the fact that pupil size directly depends on the opiate dose [22]. In this study, there was a 20% reduction in the dose of morphine, but this was not accompanied by an increase in pupil size. This finding might be associated with a high parasympathetic tone (confirmed by the ANI), which promotes pupil myosis and probably compensates the effect of a reduction in the opiate dose.

If the results of this study are confirmed in future work, the clinical implications of the findings would be great, as EVGs could be included as part of the nonpharmacological treatment plan for cancer-related mucositis in pediatric patients.

This study has some limitations that should be considered when interpreting the results. First, the sample size was small. Additional research with a larger sample size would be needed to help determine the reliability of the findings. Second, although the study design allowed for the evaluation of concurrent associations among the scores of the study variables, we were not able to test for causal associations. Thus, clinical trials with control groups are needed to evaluate the causal influence of the use of video games on outcome variables.

Despite the limitations of this study, the findings provide important additional information on the potential value of using video games to reduce pain and suffering, as well as enhance health recovery by increasing vagal tone in a very safe and efficient way. The use of video games for a mean of approximately 2 hours in children with intense mucositis relieved their pain by 30%, with a 14% increase in vagal tone, and at the same time, it reduced the daily dose of morphine by 20%.

The findings are consistent with a biopsychosocial model of pain, supporting the use of medical or physical and psychosocial interventions to prevent or manage pain in young patients [23,24]. Additional research to identify the best methods to provide this integrated help to young patients undergoing painful medical procedures is warranted.

Acknowledgments

We would like to thank Juegaterapia Foundation for raising and promoting this study among children with cancer. JM's work is supported by the Agency for Administration of University and Research Grants (AGAUR; SGR-1321), Obra Social de Caixabank, the Spanish Ministry of Science and Competitiveness (PSI2015-70966-P; PSI2016-82004-REDT), and Fundación Grünenthal. This study was possible thanks to the free transfer of consoles, video games, and other resources by Sony Interactive Entertainment-Spain.

Conflicts of Interest

None declared.

References

1. Bouya S, Balouchi A, Maleknejad A, Koochakzai M, AlKhasawneh E, Abdollahimohammad A. Cancer Pain Management Among Oncology Nurses: Knowledge, Attitude, Related Factors, and Clinical Recommendations: a Systematic Review. J Cancer Educ 2019 Oct;34(5):839-846. [doi: 10.1007/s13187-018-1433-6] [Medline: 30315497]
2. Donohoe C, Bosi JK, Sykes A, Lu Z, Mandrell B. Clinical Characteristics of Children and Adolescents Undergoing Hematopoietic Cell Transplantation Who Develop Oral Mucositis. Oncol Nurs Forum 2018 Jul 02;45(4):457-462 [FREE Full text] [doi: 10.1188/18.ONF.457-462] [Medline: 29947353]
3. Mertens R. [Pain therapy in pediatric oncology: pain experience, drugs and pharmacokinetics]. Anasthesiol Intensivmed Notfallmed Schmerzther 2011 Nov;46(11-12):736-742. [doi: 10.1055/s-0031-1297180] [Medline: 22147611]
4. Subramaniam SD, Doss B, Chanderasekar LD, Madhavan A, Rosary AM. Scope of physiological and behavioural pain assessment techniques in children - a review. Healthc Technol Lett 2018 Aug;5(4):124-129 [FREE Full text] [doi: 10.1049/htl.2017.0108] [Medline: 30155264]

5. Collins JJ, Geake J, Grier HE, Houck CS, Thaler HT, Weinstein HJ, et al. Patient-controlled analgesia for mucositis pain in children: a three-period crossover study comparing morphine and hydromorphone. J Pediatr 1996 Nov;129(5):722-728. [doi: 10.1016/s0022-3476(96)70156-7] [Medline: 8917240]
6. Nelson KL, Yaster M, Kost-Byerly S, Monitto CL. A national survey of American Pediatric Anesthesiologists: patient-controlled analgesia and other intravenous opioid therapies in pediatric acute pain management. Anesth Analg 2010 Mar 01;110(3):754-760. [doi: 10.1213/ANE.0b013e3181ca749c] [Medline: 20185654]
7. Inan G, Inal S. The Impact of 3 Different Distraction Techniques on the Pain and Anxiety Levels of Children During Venipuncture: A Clinical Trial. Clin J Pain 2019 Feb;35(2):140-147. [doi: 10.1097/AJP.0000000000000666] [Medline: 30362982]
8. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
9. Law EF, Dahlquist LM, Sil S, Weiss KE, Herbert LJ, Wohlheiter K, et al. Videogame distraction using virtual reality technology for children experiencing cold pressor pain: the role of cognitive processing. J Pediatr Psychol 2011 Jan;36(1):84-94 [FREE Full text] [doi: 10.1093/jpepsy/jsq063] [Medline: 20656761]
10. Das DA, Grimmer KA, Sparnon AL, McRae SE, Thomas BH. The efficacy of playing a virtual reality game in modulating pain for children with acute burn injuries: a randomized controlled trial [ISRCTN87413556]. BMC Pediatr 2005 Mar 03;5(1):1 [FREE Full text] [doi: 10.1186/1471-2431-5-1] [Medline: 15745448]
11. Sánchez-Rodríguez E, Miró J, Castarlenas E. A comparison of four self-report scales of pain intensity in 6- to 8-year-old children. Pain 2012 Aug;153(8):1715-1719. [doi: 10.1016/j.pain.2012.05.007] [Medline: 22703691]
12. Ledowski T. Objective monitoring of nociception: a review of current commercial solutions. Br J Anaesth 2019 Aug;123(2):e312-e321 [FREE Full text] [doi: 10.1016/j.bja.2019.03.024] [Medline: 31047645]
13. White MC, Hommers C, Parry S, Stoddart PA. Pain management in 100 episodes of severe mucositis in children. Paediatr Anaesth 2011 Apr;21(4):411-416. [doi: 10.1111/j.1460-9592.2010.03515.x] [Medline: 21251145]
14. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
15. Culbert T. Perspectives on Technology-Assisted Relaxation Approaches to Support Mind-Body Skills Practice in Children and Teens: Clinical Experience and Commentary. Children (Basel) 2017 Apr 04;4(4) [FREE Full text] [doi: 10.3390/children4040020] [Medline: 28375179]
16. Arman N, Tarakci E, Tarakci D, Kasapcopur O. Effects of Video Games-Based Task-Oriented Activity Training (Xbox 360 Kinect) on Activity Performance and Participation in Patients With Juvenile Idiopathic Arthritis: A Randomized Clinical Trial. Am J Phys Med Rehabil 2019 Mar;98(3):174-181. [doi: 10.1097/PHM.0000000000001001] [Medline: 30020092]
17. Hua Y, Qiu R, Yao W, Zhang Q, Chen X. The Effect of Virtual Reality Distraction on Pain Relief During Dressing Changes in Children with Chronic Wounds on Lower Limbs. Pain Manag Nurs 2015 Oct;16(5):685-691. [doi: 10.1016/j.pmn.2015.03.001] [Medline: 25972074]
18. Sekiguchi T, Hagiwara Y, Yabe Y, Tsuchiya M, Itaya N, Yoshida S, et al. Playing video games for more than 3 hours a day is associated with shoulder and elbow pain in elite young male baseball players. J Shoulder Elbow Surg 2018 Sep;27(9):1629-1635. [doi: 10.1016/j.jse.2018.06.005] [Medline: 30054242]
19. Kauhanen L, Järvelä L, Lähteenmäki PM, Arola M, Heinonen OJ, Axelin A, et al. Active video games to promote physical activity in children with cancer: a randomized clinical trial with follow-up. BMC Pediatr 2014 Apr 05;14:94 [FREE Full text] [doi: 10.1186/1471-2431-14-94] [Medline: 24708773]
20. Windich-Biermeier A, Sjoberg I, Dale JC, Eshelman D, Guzzetta CE. Effects of distraction on pain, fear, and distress during venous port access and venipuncture in children and adolescents with cancer. J Pediatr Oncol Nurs 2007;24(1):8-19. [doi: 10.1177/1043454206296018] [Medline: 17185397]
21. Julien-Marsollier F, Rachdi K, Caballero M, Ayanmanesh F, Vacher T, Horlin A, et al. Evaluation of the analgesia nociception index for monitoring intraoperative analgesia in children. Br J Anaesth 2018 Aug;121(2):462-468 [FREE Full text] [doi: 10.1016/j.bja.2018.03.034] [Medline: 30032886]
22. Migeon A, Desgranges F, Chassard D, Blaise BJ, De Queiroz M, Stewart A, et al. Pupillary reflex dilatation and analgesia nociception index monitoring to assess the effectiveness of regional anesthesia in children anesthetised with sevoflurane. Paediatr Anaesth 2013 Dec;23(12):1160-1165. [doi: 10.1111/pan.12243] [Medline: 23910160]
23. Kumar A, Das S, Chauhan S, Kiran U, Satapathy S. Perioperative Anxiety and Stress in Children Undergoing Congenital Cardiac Surgery and Their Parents: Effect of Brief Intervention-A Randomized Control Trial. J Cardiothorac Vasc Anesth 2019 May;33(5):1244-1250. [doi: 10.1053/j.jvca.2018.08.187] [Medline: 30243867]
24. Nelson SM, Cunningham NR, Kashikar-Zuck S. A Conceptual Framework for Understanding the Role of Adverse Childhood Experiences in Pediatric Chronic Pain. Clin J Pain 2017 Mar;33(3):264-270 [FREE Full text] [doi: 10.1097/AJP.0000000000000397] [Medline: 27275737]

Abbreviations

ANI: Analgesia Nociception Index
EVG: electronic video game
NRS: Numerical Rating Scale
PCA: patient-controlled analgesia

Edited by G Eysenbach; submitted 01.09.19; peer-reviewed by P Copanitsanou, X Garcia-Eroles, R Lee; comments to author 19.09.19; revised version received 11.11.19; accepted 16.12.19; published 30.03.20.

Please cite as:
Alonso Puig M, Alonso-Prieto M, Miró J, Torres-Luna R, Plaza López de Sabando D, Reinoso-Barbero F
The Association Between Pain Relief Using Video Games and an Increase in Vagal Tone in Children With Cancer: Analytic Observational Study With a Quasi-Experimental Pre/Posttest Methodology
J Med Internet Res 2020;22(3):e16013
URL: http://www.jmir.org/2020/3/e16013/
doi:10.2196/16013
PMID:32224482

Original Paper

Learning Latent Space Representations to Predict Patient Outcomes: Model Development and Validation

Subendhu Rongali[1], BTech; Adam J Rose[2], MSc, MD; David D McManus[3], MSc, MD; Adarsha S Bajracharya[3], MD; Alok Kapoor[3,4], MSc, MD; Edgard Granillo[5], MD; Hong Yu[1,3,6,7], MA, MS, PhD

[1]College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, United States
[2]Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, United States
[3]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States
[4]Meyers Primary Care Institute, Worcester, MA, United States
[5]Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States
[6]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States
[7]Center for Healthcare Organization and Implementation Research, Bedford Veterans Affairs Medical Center, Bedford, MA, United States

Corresponding Author:
Hong Yu, MA, MS, PhD
Department of Computer Science
University of Massachusetts Lowell
1 University Ave
Lowell, MA, 01854
United States
Phone: 1 978 934 3620
Email: Hong_Yu@uml.edu

Abstract

Background: Scalable and accurate health outcome prediction using electronic health record (EHR) data has gained much attention in research recently. Previous machine learning models mostly ignore relations between different types of clinical data (ie, laboratory components, International Classification of Diseases codes, and medications).

Objective: This study aimed to model such relations and build predictive models using the EHR data from intensive care units. We developed innovative neural network models and compared them with the widely used logistic regression model and other state-of-the-art neural network models to predict the patient's mortality using their longitudinal EHR data.

Methods: We built a set of neural network models that we collectively called as long short-term memory (LSTM) outcome prediction using comprehensive feature relations or in short, CLOUT. Our CLOUT models use a correlational neural network model to identify a latent space representation between different types of discrete clinical features during a patient's encounter and integrate the latent representation into an LSTM-based predictive model framework. In addition, we designed an ablation experiment to identify risk factors from our CLOUT models. Using physicians' input as the gold standard, we compared the risk factors identified by both CLOUT and logistic regression models.

Results: Experiments on the Medical Information Mart for Intensive Care-III dataset (selected patient population: 7537) show that CLOUT (area under the receiver operating characteristic curve=0.89) has surpassed logistic regression (0.82) and other baseline NN models (<0.86). In addition, physicians' agreement with the CLOUT-derived risk factor rankings was statistically significantly higher than the agreement with the logistic regression model.

Conclusions: Our results support the applicability of CLOUT for real-world clinical use in identifying patients at high risk of mortality.

(J Med Internet Res 2020;22(3):e16374) doi:10.2196/16374

KEYWORDS

predictive modeling; neural networks; ablation; patient mortality

Introduction

Background

High-precision predictive modeling of clinical outcomes (eg, adverse events such as the onset of disease and death) is a clinically important but computationally challenging task. If physicians can be notified about the risks of adverse events in advance, they may be able to take steps to prevent them. Electronic health records (EHRs) are widely used in US hospitals and are becoming more mature over time [1]. They have been actively researched for predictive modeling [2-8].

Almost 6 million patients are admitted annually to intensive care units (ICUs) in the United States for airway support, for hemodynamic or respiratory monitoring, and to stabilize acute or life-threatening medical problems [9-15]. Patients in ICUs are vulnerable to many acute diseases and often suffer from chronic illness, but the leading causes of death in the ICU are multi-organ failure, sepsis, and cardiovascular disease. Approximately 10% to 30% of adult patients die before hospital discharge in ICUs [16-30]. Regression models have been widely used for predicting mortality for ICU patients [31]. Goal-directed sepsis care represents an example of a successful, evidence-based approach to the care of critically ill patients with sepsis that uses predictive modeling to target patients at high risk for mortality with life-saving upstream therapies [21].

During the past several years, neural network models have shown a great success for many artificial intelligence applications including computer vision, natural language processing, and predictive modeling [4,32-34]. Neural network-based predictive models include the convolutional neural network (CNN) and recurrent neural network (RNN) framework.

Although studies show that CNN models do not necessarily outperform conventional predictive models such as regression models [35], RNNs [36] have been shown to work well with sequential data such as longitudinal EHRs. There have been promising results regarding the use of RNNs in clinical applications such as diagnosis predictions [6,37,38]. Autoencoders [39] are another class of neural networks that extract rich representations using large unlabeled EHR data and have shown state-of-the-art performance in prediction [4].

Although NN-based predictive models have been developed, most models are based on *bag of features*, and few have explicitly modeled the complex relationships between different types of EHR data. Clinical events and diagnoses are not isolated but instead are complex, multifaceted, and often correlated. For example, diagnostic testing leads to a new finding, which may lead to a specific treatment. Therefore, we believe it is important to account for such relationships to improve the predictive power of a model.

Objective

The main objective of this work is to develop innovative prediction models to accurately predict patient mortality using patients' longitudinal EHR data. An important component of our models is a correlational neural network, which is a special neural network model that accounts for correlations between different types of features. We modeled the relationships between different types of clinical features in the EHR through a correlational neural network and integrated them into LSTM-based predictive models for improved performance.

Contributions

Our main contributions include learning of latent features from different clinical data types and integrating the learned latent features for outcome prediction using longitudinal EHR data. Our results show that the integration of latent features yielded the highest results for predicting patient mortality using the ICU data.

In addition to evaluating our CLOUT models using the traditional evaluation metrics such as sensitivity, specificity, and area under the receiver operating characteristic curve, we studied the interpretability of our predictive models. Specifically, we designed a simple ablation experiment [40] to identify important features (or risk factors). Our evaluation results show that physicians were more in agreement with the risk factors ranked by CLOUT than the ones ranked by the commonly used logistic regression model.

In summary, our contributions are twofold: (1) We developed an innovative long short-term memory (LSTM)–based predictive model where a correlational neural network is integrated to identify relationships and latent representations of different clinical features. Our CLOUT model has state-of-the-art performance in mortality prediction, surpassing other competitive NN models and a logistic regression model. (2) We provide a comprehensive evaluation of risk factors identified by our neural network models. Our results show that the risk factors identified by the CLOUT model agree with physicians' assessment, suggesting that CLOUT could be used in real-world clinical settings.

Methods

The Medical Information Mart for Intensive Care-III Dataset

All models are trained and evaluated on the Medical Information Mart for Intensive Care-III (MIMIC-III) dataset; an EHR dataset made publicly available by the Massachusetts Institute of Technology Laboratory for Computational Physiology. MIMIC-III has been widely used for predictive models [41]. The dataset contains 7537 patients with two or more encounters, which is the subset we used to build our CLOUT and baseline models. We call this dataset p-MIMIC. Some demographic information for patients in this dataset is given in Table 1.

Table 1. Patient demographic information (N=7537).

Characteristic	Values
Age (years)	
Mean	74.74
Median	66.00
Sex, n (%)	
Male	4190 (55.59)
Female	3347 (44.41)
Race, n (%)	
White	5644 (74.88)
Black	867 (11.50)
Hispanic	277 (3.68)
Asian	226 (3.00)
Other/unknown	523 (6.94)

We require two or more encounters because we remove the last encounter while making predictions, requiring us to have at least one other encounter with data. We use patient mortality as our outcome label. This label is obtained in the MIMIC dataset from the hospital records and the social security death records. In our dataset of 7537 patients, we have 2825 (37.9%) documented deaths. Further details about MIMIC are covered in Multimedia Appendix 1.

The dataset was further divided into train, validation, and test splits, each containing approximately 69.99% (5275/7537), 9.99% (753/7537), and 20.02% (1509/7537) of the patients, respectively. Once we picked the optimal model hyper-parameters using the validation set, the model was retrained on the combined train-validation set, which contained 79.98% (6028/7537) of the data.

Baselines—Reverse Time Attention Model, Time-Aware Reverse Time Attention Model, Logistic Regression Models

Our first set of baseline models are versions of the *RETAIN* model, which is one of the few publicly available predictive models for EHRs. RETAIN was built on an RNN model, and evaluation has shown that it achieved both state-of-the-art performance and interpretability [6].

RETAIN by itself does not incorporate temporal information beyond the RNN framework; such fine-grained temporal information may be important to patient outcomes. For example, the severity of 2 acute myocardial infarctions separated by different durations could have different clinical implications. On the other hand, there is an option to include the time features to the encounter vector. Therefore, we implemented time-aware RETAIN (TaRETAIN) models as additional baselines by concatenating time information to the input features. We experimented with two different approaches to create the time feature: number of days elapsed since the first encounter and number of days elapsed since the previous encounter. We call these 2 models *TaRETAIN-first* and *TaRETAIN-previous*.

Another baseline model is *logistic regression* as it has been commonly used with EHR data. Although logistic regression is best in interpretability, it is difficult to incorporate temporal information. We therefore combined all the information documented in an encounter to form 1 feature vector for each patient. Our logistic regression model was also augmented with the l2 penalty.

The CLOUT Models

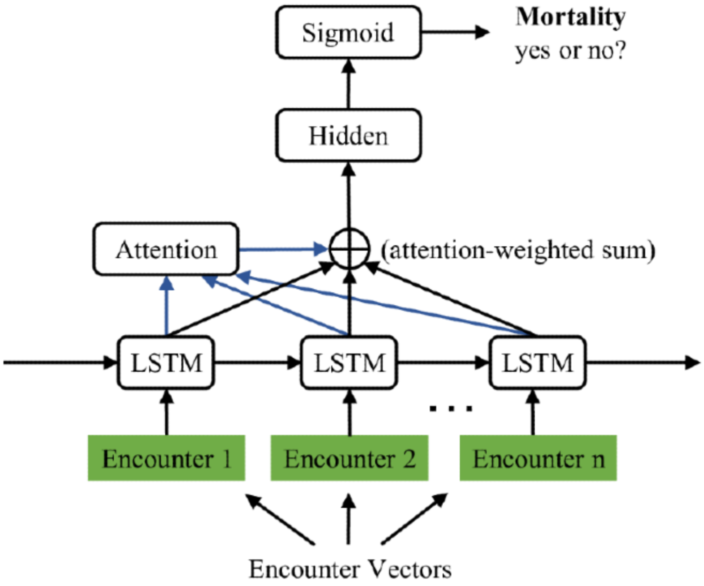
The CLOUT models are built upon the state-of-the-art LSTM framework. We provide a description of relevant concepts or components that are built into our CLOUT models in Multimedia Appendix 2.

Unlike other RNN models, LSTMs can learn dependencies over longer intervals more efficiently [42]. In this study, CLOUT represented all LSTM-based predictive models we built for EHRs. The central architecture, as shown in Figure 1, is an attention-based LSTM model that processed the encounter vectors and made a binary class prediction.

Given a patient with encounters, the encounter vectors derived from a CLOUT model are e_1, e_2, ... e_n. We ran the encounter vectors through the LSTM framework to get the hidden vectors at each time step, h_1, h_2, ... h_n. We then used the attention module to find the weighted sum of these hidden vectors . Formally, $H = a_1 . h_1 + a_2 . h_2 + ... a_n . h_n$. The vector was then sent through a linear layer, and the output was squashed between 0 and 1 using a sigmoid function. This final output represented the probability of a positive class, which in our current application was the probability that the patient died.

Note that our LSTM architecture is commonly used for sequence data. The innovation of this work is the representation of the encounter vector that integrates different types of EHR data, which we will describe below.

Figure 1. Our model architecture. LSTM: long short-term memory.



A Simple Concatenation Model

In this version of CLOUT, the encounter vector was derived by a simple concatenation of different types of features. Every patient encounter had a set of documented International Classification of Diseases (ICD) codes, medications, and laboratory components. We converted these to 3 bit-vectors, , , and , respectively, each of the size of the vocabularies. Bit-vectors are vectors of size equal to the length of the vocabulary with 1 at the index where the feature is documented and 0 everywhere else. We passed these bit-vectors through linear embedding layers to get their dense vector representations. We concatenated these dense vector representations and passed the resultant vector through a nonlinear function such as the rectified linear unit [43] to get the final encounter representation, .

Representation Through Concatenation With Autoencoders

Recent work on word embeddings called ELMo [44] has shown that integrating different levels of representations learned by neural networks improves predictive performance in natural language processing applications, as different layers represent different characteristics of input data. Building on the same concept, we created a CLOUT model that integrates the representations of input features learned from an autoencoder with our inputs before sending them through the prediction layer. The hidden layer representations contain valuable information about the relationships between different input features, and by including this information along with the actual input features, we enable the model to make predictions with more knowledge. We integrate the representations using concatenation.

The Latent Space Representation

ICD codes, medications, and laboratory results are not isolated unrelated clinical information. They are clinically intertwined or correlated. For example, as stated earlier, medications depend upon the diagnoses of the patient in that encounter. To capture the correlations among EHR data, we added a multi-view latent space component, as shown in Figure 2, by adapting a correlational neural network [45] framework.

We used a correlational neural network for 3 views (ICD codes, medications, and laboratory components) to construct the latent representation for our latent space model. This component is graphically shown in Figure 3.

The latent space representation is a measure of the patient condition—a combination of related information from diagnosis codes, medications, and laboratory components. The details of this component are further described in Multimedia Appendix 3.

To integrate latent space representation into the encounter vector, we first projected the encounter into this latent space to get the latent space vector, l. We simultaneously performed all of the operations in the simple concatenation version to find the encounter vector of that version, e_c. The final encounter vector was the concatenation of l and e_c. The model described here is shown in Figure 2. Note that the c-operation stands for concatenation.

To evaluate the effectiveness of the correlational neural network, we also implemented a traditional autoencoder with one hidden layer f and one output layer g with the goal to reconstruct the input using a hidden representation of lower dimensions. We called this model CLOUT-autoencoder**.**

Figure 2. Model for constructing the encounter vector. ReLU: rectified linear unit; ICD: International Classification of Diseases.

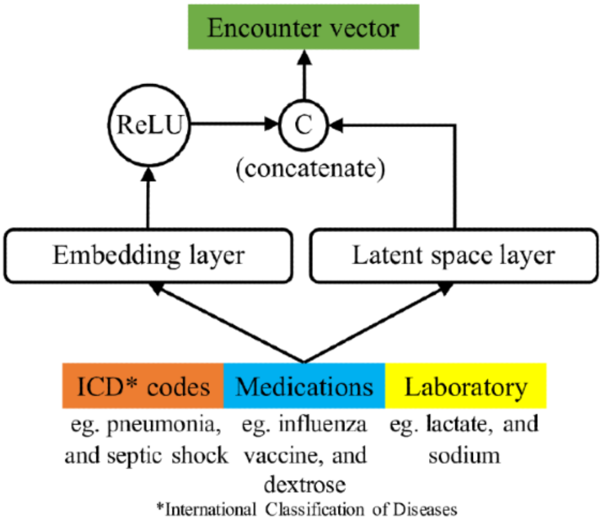
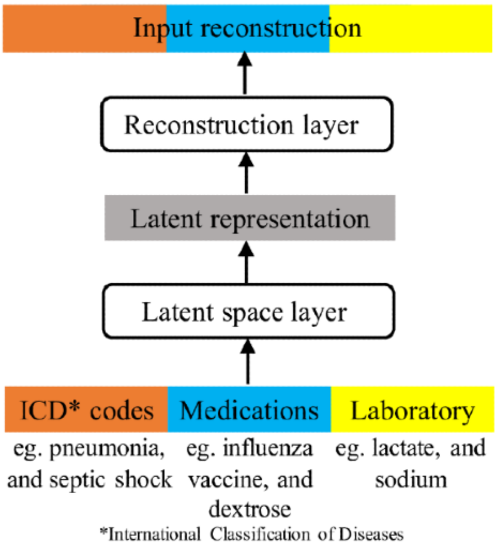


Figure 3. The correlational neural network for our 3 views. ICD: International Classification of Diseases.



Evaluation

We evaluated each of the baseline and CLOUT models on the p-MIMIC dataset. We obtain true-positives (TP), false-positives (FP), true-negatives (TN), and false-negatives (FN). We report area under the receiver operating characteristic curve (AUC-ROC) scores for all models, and precision , recall , and F1-scores for the top performing models.

Risk Factor Experiment With Physicians

Predictive models would be of limited clinical use if the models are not interpretable. To interpret or identify the risk factors in our CLOUT models, we conduct an ablation experiment, which has been widely used for feature engineering. We perturb the patient data to zero out the contribution of a feature and calculate the corresponding difference in output. This classical method shows the contribution of each feature, which may correspond to the risk score.

Recall that each of our CLOUT models outputs a probability score that indicates mortality risk. So, the difference in output would be the reduction in this probability, which we call the attribution weight of the given feature. We calculated the attribution weight for each ICD code, medication, and laboratory component that is documented in the patient's EHRs. These features would then constitute the risk factors associated with the mortality, and the attribution weight represents the strength of the association.

Although ablation experiments have been widely used for feature engineering and interpretation of neural network models in many applications [46], they have not been evaluated for identifying risk factors of patient outcome based on longitudinal EHRs.

Therefore, we designed a comprehensive evaluation of the risk factors ranked by CLOUT and compared them with ones ranked by a logistic regression model. Specifically, we ranked the risk factors at the patient level and population level. At the patient level, each risk factor (ie, feature or variable) is weighted by its contribution to the correct prediction to the patient. We ranked the risk factors at the population level by aggregating and normalizing the attribution weights of features across the patient population.

Experiment Design

Using stratified random sampling, we selected a subset of risk factors from the prediction models CLOUT and logistic regression, respectively, and asked 5 unbiased physicians (4 internists and 1 cardiologist), who were not privy to the reasons for doing the ranking, to independently judge the clinical relevance of those risk factors.

To reduce the total number of features that the physicians need to evaluate, we selected features from CLOUT. Specifically, for each feature, the ablation experiment output a relevance score. We bin the features into 3 groups: (1) top 20 features, (2) 20-50 features, and (3) the remaining features. From each bin, we randomly selected 4 features. We then randomly selected 1 patient and accordingly obtained a total of 12 features for that patient. We also obtained the ranked list of features by population and followed a similar bin strategy to select another 18 features distributed across the different feature sets (we purposely selected those features that differ from the features we selected from the sample patient so that we could maximize our evaluation features). Therefore, we selected a total of 30 features (12 by a patient and 18 by the population).

We randomized those 30 features and asked the 5 physicians who are blinded to the CLOUT rankings to evaluate, for each feature, its clinical relevance. Specifically, we asked each physician to score the feature (1-5, with 1 as the least relevant and 5 the most relevant) based on their clinical knowledge or guidelines.

We calculated the Pearson correlation coefficient between physicians' scores for pairwise agreements between physicians, and between the CLOUT scores and physicians' scores. We also performed a *t* test for statistical significance. We used the same 30 features to evaluate the logistic regression model and, in this case, using the weight assigned by the logistic regression model for the ranking.

Finally, we performed another evaluation where we first averaged the scores of all the physicians to obtain a representative gold standard. We then computed the correlation coefficients between these scores and the scores from our models and the logistic regression baseline.

Results

Model Performance

During our experiments, we found that models using abnormal laboratory components as input (ie, binary coding of normal/abnormal) performed better than those using all the laboratory components. Therefore, the results presented here for the p-MIMIC dataset used only the abnormal labs recorded in patient encounters through an abnormal flag.

As shown in Table 2, the AUC-ROC results for our CLOUT models are significantly better (*P*<.001) than both the RETAIN and the logistic regression models. The AUC-ROC curves for the representative models are presented in Figure 4. Our CLOUT model with concatenated latent space representation (Figure 2) achieved 0.89 AUC-ROC score, which is more than 0.06 absolute increase over the ICD-RETAIN, logistic regression, and simple LSTM models and a 0.02 increase over RETAIN with all codes. To get a better understanding of our results, we also present the precision and recall scores for each class for the top models in Table 3.

Our latent space representation model also slightly outperformed the traditional autoencoder CLOUT model, although it is not statistically significant. An important result here is the integration of different levels of representations (input space and from either an autoencoder or a correlational autoencoder) substantially improves the performance of a model, which outperforms one that uses autoencoder alone. The code for our models and experiments can be found at our CLOUT repository [47].

Table 2. Area under the receiver operating characteristic curve scores for different models.

Method	Area under the receiver operating characteristic curve, mean (SD)
Logistic regression	0.82 (0.0103)
RETAIN[a] (only ICD[b])	0.82 (0.0924)
TaRETAIN[c]-*first* (only ICD)	0.82 (0.0118)
TaRETAIN-*prev* (only ICD)	0.82 (0.0919)
RETAIN (all codes)	0.86 (0.0105)
Long short-term memory with only ICD codes	0.83 (0.0104)
CLOUT[d]—only autoencoder	0.80 (0.0116)
CLOUT—only latent space	0.81 (0.0082)
CLOUT—simple concatenation	0.88 (0.0096)
CLOUT—autoencoder concatenation	0.88 (0.0107)
CLOUT—latent space concatenation	*0.89 (0.0138)* [e]

[a] RETAIN: Reverse Time Attention model.

[b] ICD: International Classification of Diseases.

[c] TaRETAIN: time-aware RETAIN.

[d] CLOUT: L(STM) Outcome prediction using Comprehensive features relations.

[e] Best performing model.

Figure 4. The area under the receiver operating characteristic curves for various models. RETAIN: Reverse Time Attention model; CLOUT: L(STM) Outcome prediction using Comprehensive feature relations.

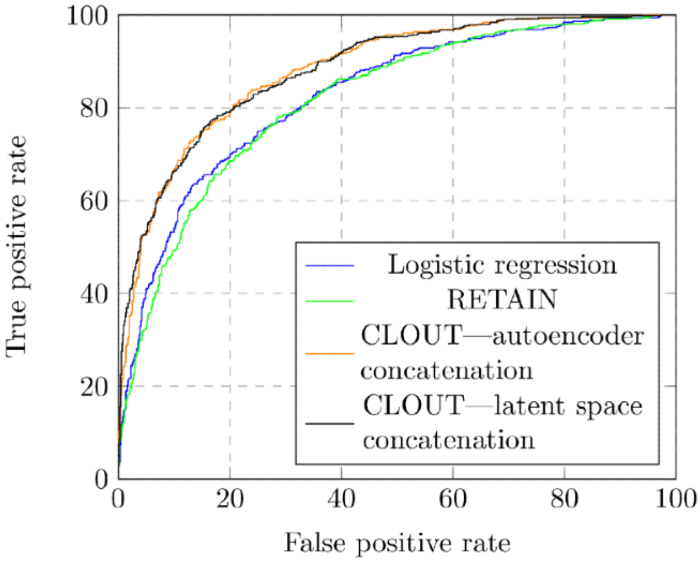


Table 3. Precision, recall, and F-scores for top CLOUT[a] models.

Method and class	Precision	Recall	F-score
CLOUT—Simple concatenation			
0	0.85	0.82	0.83
1	0.71	0.76	0.73
Average	0.80	0.79	0.80
CLOUT—Autoencoder concatenation			
0	0.85	0.85	0.85
1	0.74	0.74	0.74
Average	0.81	0.81	0.81
CLOUT — Latent space concatenation			
0	0.84	0.88	0.86
1	0.78	0.72	0.72
Average	0.82	0.82	0.82

[a] CLOUT: L(STM) Outcome prediction using Comprehensive features relations.

Risk Factors

To measure agreements among physicians, we compute the Pearson correlation coefficient between their scores. For patient-specific features, Table 4 shows the Pearson correlation coefficient between each pair of physicians and also between different models and the physicians. With the physician gold standard ratings computed by averaging, we found that our model had a correlation coefficient of 0.64, which is higher (4.9%) than the correlation coefficient of 0.61 with the logistic regression model.

Table 4. Pearson correlation coefficients for agreement between physicians and models.

Agreement	Physician 1, *r*	Physician 2, *r*	Physician 3, *r*	Physician 4, *r*	Physician 5, *r*	Mean (SD)
Physician-physician agreement						
Physician 1	1.00	0.81	0.56	0.61	0.88	0.72 (0.13)
Physician 2	0.81	1.00	0.87	0.65	0.86	0.80 (0.09)
Physician 3	0.56	0.87	1.00	0.49	0.69	0.65 (0.14)
Physician 4	0.61	0.65	0.49	1.00	0.61	0.59 (0.06)
Physician 5	0.88	0.86	0.69	0.61	1.00	0.76 (0.11)
Physician-model agreement						
Logistic regression	0.60	0.63	0.53	0.32	0.52	0.52 (0.11)
RETAIN[a]	0.65	0.72	0.61	0.30	0.58	0.57 (0.14)
CLOUT[b]—only autoencoder	−0.07	0.13	0.21	*0. 55*[c]	0.17	0.20 (0.20)
CLOUT—only latent space	0.42	0.77	*0. 64*	0.35	0.53	0.54 (0.15)
CLOUT—simple concatenation	0.52	0.64	0.70	0.19	*0. 67*	0.54 (0.19)
CLOUT—autoencoder concatenation	0.54	0.70	0.64	0.14	0.62	0.53 (0.20)
CLOUT—latent space concatenation	*0. 69*	*0. 77*	0.59	0.18	*0. 67*	*0.58 (0.21)*

[a] RETAIN: Reverse Time Attention model.

[b] CLOUT: L(STM) Outcome prediction using Comprehensive features relations.

[c] Italicization signifies highest physician-model agreement in the column.

Discussion

Principal Findings

In this study, we have developed innovative CLOUT models and compared them with other state-of-the-art predictive models with respect to performance on mortality prediction. We found that the performance of almost every CLOUT model surpassed the competitive baseline models (eg, RETAIN). The results support that LSTM is a state-of-the-art framework for EHR-based predictive modeling.

Our results showed that the integration of different levels of latent representations (input space and from either an autoencoder or a correlational autoencoder) substantially improves the performance from 0.80 to 0.88 AUC-ROC. The rich representation may provide extra information to the model, which in turn helps the model make better predictions. The integration of different types of features (ie, ICD codes, laboratories, and medications) however had a mixed result. Specifically, the CLOUT model that incorporated only the abnormal laboratory results slightly surpassed the CLOUT model that incorporated all 3 features. This supported the importance of laboratory results for predicting mortality. Our results also suggested that there may be noisy information in the features. When CLOUT was implemented with the latent vectors included, it had the highest performance, an AUC-ROC score of 0.89 and an F1 score of 0.82. The result supports our approach of using the correlational neural network to identify latent vectors to best represent different but related clinical observations or variables.

On the other hand, when we incorporated temporal information as a feature, we showed little improvement in performance using RETAIN. A possible future direction is to explore time-dependent attentions, which may allow the model to integrate the temporal information in the architecture.

For the risk factors identified by our models, the average correlation coefficient between the physicians is mean 0.71 (SD 0.13), and the average Pearson correlation coefficients between CLOUT and the physicians and between logistic regression and physicians were 0.58 (SD 0.21) and 0.52 (SD 0.11), respectively. These results show a significant difference between the agreement among physicians and the agreement between the logistic regression model and the physicians (*P*=.04). In contrast, the difference in agreement between the CLOUT models and physicians is not statistically significant, strongly supporting the validity of risk factors and their ranking identified by CLOUT.

We also calculated the agreement with RETAIN for reference, and we found that the average was 0.57 (SD 0.14), which is still slightly less than the CLOUT model, with CLOUT losing out a lot with physician 4. Other CLOUT models also have slightly lower scores as reported in Table 4, but it is notable that the latent vector models that use the correlational autoencoder have better correlations (0.58, SD 0.21) with physicians than the ones that use a simple autoencoder (0.53, SD 0.20). The evaluation with our gold standard (the average physician scores) also informs us that CLOUT selects more meaningful features.

Our results show that physician 4 had a low correlation score with other physicians as well as with our CLOUT models. For example, lactulose enema, and encephalopathy not otherwise specified were scored as 2 by physician 4, whereas all the other physicians gave scores of 4 or greater. When we removed physician 4, the correlation between the latent space CLOUT model and the physicians improved from 0.58 to 0.68.

For population-level features, we performed similar evaluation between physician scores and the CLOUT model scores, and the average correlation coefficient values were ICD codes −0.19, medications −0.43, and laboratory components −0.37, which are lower than patient-specific interpretations. This is not surprising as many risk factors (eg, severe diseases) are rare events that are not present for patients in general.

Furthermore, CLOUT models captured important risk factors while making predictions. In general, our CLOUT models show that patients with diagnosis codes representing cranial nerve disorder and cystic liver disease were marked with a high risk of mortality. This is reasonable as those are diseases with a high risk of mortality.

Limitations

Our dataset was constructed from EHR data and is, hence, prone to standard data quality issues that EHRs typically have, as documented in the literature. EHRs are known to have missing diagnoses and medication codes for patients when compared with insurance claims. Furthermore, our analysis of ICU admissions does not account for death because of accidental circumstances such as car crashes. We used all the information exactly as it appears as it is infeasible to comb through all the records to pick patients for the study. Another limitation we would like to report is the absence of vital sign features in our dataset, which we ignore because of the involved preprocessing steps that are required to handle missing numerical values.

The CLOUT models have significant limitations as well. First, similar to most predictive models, the risk factors identified by the CLOUT models include cofounding variables. For example, we found that patients who have a prescription for a scopolamine patch have high-risk scores. This is a medication prescribed to terminally ill patients as part of palliative care regimen to reduce excessive airway secretion. So, in this case, the actual reason for palliative care is a strong risk factor for death, not the medication, which is a confounding factor. Another limitation of our work is that our models are very dependent on the population size. Bias could be introduced when the size is small. However, such limitations exist in most predictive models not reviewed or guided by physician oversight.

Comparison With Prior Work

We surveyed a variety of approaches to compare our models. This includes statistical approaches [48-52], deep learning–based approaches [6,38,53-55], and other phenotyping efforts [4,5]. We also surveyed papers on interpretability. A detailed analysis of all this can be found in Multimedia Appendix 4.

Conclusions

EHRs are widely available and have enormous untapped potential to predict patients' health outcome. EHR-based predictive models are potentially hugely useful for clinical decision support. Our experiments show that incorporating comprehensive clinical information is useful and can improve predictions and that integrating latent space representations learned through a correlational neural network to clinical information led to the best performing CLOUT model. Our risk factor experiment with physicians also suggests that CLOUT models find more clinically relevant risk factors. Our results strongly support that CLOUT may be a useful tool to generate clinical prediction models, especially among hospitalized and critically ill patient populations.

The future directions include new models to incorporate the temporal information and methods to integrate clinical notes for predictive models. We may also explore other models to integrate different views, including the Capsule network model [56].

Acknowledgments

This work was supported in part by the grant R01HL137794 from the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. HY’s time was also supported by grants 5R01HL125089, R01HL135129, R01DA045816, and R01LM012817. DM’s time was also supported by grants R01HL137734, R01HL126911, R01HL13660, and R01HL141434 from the National Heart, Lung, and Blood Institute.

Conflicts of Interest

DM has received research grant support from Apple Computer, Bristol-Myers Squibb, Boehringher-Ingelheim, Pfizer, Samsung, Philips Healthcare, Care Evolution, and Biotronik; has received consultancy fees from Bristol-Myers Squibb, Pfizer, Flexcon, and Boston Biomedical Associates; and has inventor equity in Mobile Sense Technologies, Inc, Connecticut.

Multimedia Appendix 1
The Medical Information Mart for Intensive Care-III, preprocessing, and outcome label.
[DOCX File , 16 KB - jmir_v22i3e16374_app1.docx]

Multimedia Appendix 2
Relevant machine learning components.
[DOCX File , 14 KB - jmir_v22i3e16374_app2.docx]

Multimedia Appendix 3
Correlational neural network.
[DOCX File , 15 KB - jmir_v22i3e16374_app3.docx]

Multimedia Appendix 4
Comparison with prior work.
[DOCX File , 17 KB - jmir_v22i3e16374_app4.docx]

References

1. Kharrazi H, Gonzalez C, Lowe K, Huerta T, Ford E. Forecasting the maturation of electronic health record functions among US hospitals: retrospective analysis and predictive model. J Med Internet Res 2018;20(8):e10458 preprint. [doi: 10.2196/preprints.10458]
2. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of electronic health records in US hospitals. N Engl J Med 2009 Apr 16;360(16):1628-1638. [doi: 10.1056/NEJMsa0900592] [Medline: 19321858]
3. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting clinical events via recurrent neural networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
4. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 May 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
5. Zhou J, Wang F, Hu J, Ye J. From Micro to Macro: Data Driven Phenotyping by Densification of Longitudinal Electronic Medical Records. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.: Association for Computing Machinery, New York, NY, United States; 2014 Presented at: KDD'14; August 24 - 27, 2014; New York, NY, USA p. 135-144. [doi: 10.1145/2623330.2623711]
6. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. In: Proceedings of the 2016 Conference on Neural Information Processing Systems. 2016 Presented at: NIPS'16; December 5-10, 2016; Barcelona, Spain p. 3504-3512.
7. Esteban C, Staeck O, Baier S, Yang Y, Tresp V. Predicting Clinical Events by Combining Static and Dynamic Information Using Recurrent Neural Networks. In: Proceedings of the 2016 IEEE International Conference on Healthcare Informatics.: IEEE; 2016 Presented at: ICHI'16; October 4-7, 2016; Chicago, IL, USA p. 93-101. [doi: 10.1109/ICHI.2016.16]
8. Che Z, Purushotham S, Khemani R, Liu Y. arXiv preprints. 2015. Distilling Knowledge from Deep Networks with Applications to Healthcare Domain URL: https://arxiv.org/abs/1512.03542 [accessed 2020-03-09]
9. Wunsch H, Angus DC, Harrison DA, Linde-Zwirble WT, Rowan KM. Comparison of medical admissions to intensive care units in the United States and United Kingdom. Am J Respir Crit Care Med 2011 Jun 15;183(12):1666-1673. [doi: 10.1164/rccm.201012-1961OC] [Medline: 21471089]

10. Wunsch H, Wagner J, Herlim M, Chong DH, Kramer AA, Halpern SD. ICU occupancy and mechanical ventilator use in the United States. Crit Care Med 2013 Dec;41(12):2712-2719 [FREE Full text] [doi: 10.1097/CCM.0b013e318298a139] [Medline: 23963122]
11. Barrett M, Smith M, Elixhauser A, Honigman L, Pines J. NCBI - NIH. 2014. Utilization of Intensive Care Services, 2011: Statistical Brief #185 URL: https://www.ncbi.nlm.nih.gov/pubmed/25654157 [accessed 2020-03-11]
12. Edwards JD, Houtrow AJ, Vasilevskis EE, Rehm RS, Markovitz BP, Graham RJ, et al. Chronic conditions among children admitted to US pediatric intensive care units: their prevalence and impact on risk for mortality and prolonged length of stay*. Crit Care Med 2012 Jul;40(7):2196-2203 [FREE Full text] [doi: 10.1097/CCM.0b013e31824e68cf] [Medline: 22564961]
13. Krmpotic K, Lobos A. Clinical profile of children requiring early unplanned admission to the PICU. Hosp Pediatr 2013 Jul;3(3):212-218 [FREE Full text] [doi: 10.1542/hpeds.2012-0081] [Medline: 24313089]
14. Harrison W, Goodman D. Epidemiologic trends in neonatal intensive care, 2007-2012. JAMA Pediatr 2015 Sep;169(9):855-862. [doi: 10.1001/jamapediatrics.2015.1305] [Medline: 26214387]
15. Pollack MM, Holubkov R, Funai T, Clark A, Berger JT, Meert K, Eunice Kennedy Shriver National Institute of Child Health and Human Development Collaborative Pediatric Critical Care Research Network. Pediatric intensive care outcomes: development of new morbidities during pediatric critical care. Pediatr Crit Care Med 2014 Nov;15(9):821-827 [FREE Full text] [doi: 10.1097/PCC.0000000000000250] [Medline: 25226501]
16. Dombrovskiy VY, Martin AA, Sunderram J, Paz HL. Rapid increase in hospitalization and mortality rates for severe sepsis in the United States: a trend analysis from 1993 to 2003. Crit Care Med 2007 May;35(5):1244-1250. [doi: 10.1097/01.CCM.0000261890.41311.E9] [Medline: 17414736]
17. Druml W, Lenz K, Laggner AN. Our paper 20 years later: from acute renal failure to acute kidney injury--the metamorphosis of a syndrome. Intensive Care Med 2015 Nov;41(11):1941-1949. [doi: 10.1007/s00134-015-3989-5] [Medline: 26297119]
18. Elias KM, Moromizato T, Gibbons FK, Christopher KB. Derivation and validation of the acute organ failure score to predict outcome in critically ill patients: a cohort study. Crit Care Med 2015 Apr;43(4):856-864. [doi: 10.1097/CCM.0000000000000858] [Medline: 25746746]
19. Levy MM, Dellinger RP, Townsend SR, Linde-Zwirble WT, Marshall JC, Bion J, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. Intensive Care Med 2010 Feb;36(2):222-231 [FREE Full text] [doi: 10.1007/s00134-009-1738-3] [Medline: 20069275]
20. Randolph AG, McCulloh RJ. Pediatric sepsis: important considerations for diagnosing and managing severe infections in infants, children, and adolescents. Virulence 2014 Jan 1;5(1):179-189 [FREE Full text] [doi: 10.4161/viru.27045] [Medline: 24225404]
21. Gupta RG, Hartigan SM, Kashiouris MG, Sessler CN, Bearman GM. Early goal-directed resuscitation of patients with septic shock: current evidence and future directions. Crit Care 2015 Aug 28;19:286 [FREE Full text] [doi: 10.1186/s13054-015-1011-9] [Medline: 26316210]
22. Weiss SL, Fitzgerald JC, Pappachan J, Wheeler D, Jaramillo-Bustamante JC, Salloo A, Sepsis Prevalence, Outcomes,Therapies (SPROUT) Study Investigators and Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Network. Global epidemiology of pediatric severe sepsis: the sepsis prevalence, outcomes, and therapies study. Am J Respir Crit Care Med 2015 May 15;191(10):1147-1157 [FREE Full text] [doi: 10.1164/rccm.201412-2323OC] [Medline: 25734408]
23. Wunsch H, Guerra C, Barnato AE, Angus DC, Li G, Linde-Zwirble WT. Three-year outcomes for Medicare beneficiaries who survive intensive care. J Am Med Assoc 2010 Mar 3;303(9):849-856. [doi: 10.1001/jama.2010.216] [Medline: 20197531]
24. Halpern NA, Pastores SM. Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. Crit Care Med 2010 Jan;38(1):65-71. [doi: 10.1097/CCM.0b013e3181b090d0] [Medline: 19730257]
25. Talmor D, Shapiro N, Greenberg D, Stone PW, Neumann PJ. When is critical care medicine cost-effective? A systematic review of the cost-effectiveness literature. Crit Care Med 2006 Nov;34(11):2738-2747. [doi: 10.1097/01.CCM.0000241159.18620.AB] [Medline: 16957636]
26. Banerjee R, Naessens JM, Seferian EG, Gajic O, Moriarty JP, Johnson MG, et al. Economic implications of nighttime attending intensivist coverage in a medical intensive care unit. Crit Care Med 2011 Jun;39(6):1257-1262 [FREE Full text] [doi: 10.1097/CCM.0b013e31820ee1df] [Medline: 21317642]
27. Pronovost PJ, Needham DM, Waters H, Birkmeyer CM, Calinawan JR, Birkmeyer JD, et al. Intensive care unit physician staffing: financial modeling of the Leapfrog standard. Crit Care Med 2004 Jun;32(6):1247-1253. [doi: 10.1097/01.ccm.0000128609.98470.8b] [Medline: 15187501]
28. Parikh A, Huang SA, Murthy P, Dombrovskiy V, Nolledo M, Lefton R, et al. Quality improvement and cost savings after implementation of the Leapfrog intensive care unit physician staffing standard at a community teaching hospital. Crit Care Med 2012 Oct;40(10):2754-2759. [doi: 10.1097/CCM.0b013e31825b26ef] [Medline: 22824939]
29. Kumar S, Merchant S, Reynolds R. Tele-ICU: efficacy and cost-effectiveness approach of remotely managing the critical care. Open Med Inform J 2013;7:24-29 [FREE Full text] [doi: 10.2174/1874431101307010024] [Medline: 24078857]

30. Young LB, Chan PS, Lu X, Nallamothu BK, Sasson C, Cram PM. Impact of telemedicine intensive care unit coverage on patient outcomes: a systematic review and meta-analysis. Arch Intern Med 2011 Mar 28;171(6):498-506. [doi: 10.1001/archinternmed.2011.61] [Medline: 21444842]
31. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. Lancet Respir Med 2015 Jan;3(1):42-52 [FREE Full text] [doi: 10.1016/S2213-2600(14)70239-5] [Medline: 25466337]
32. Chorowski JK, Bahdanau D, Serdyuk D, Cho K, Bengio Y. Attention-Based Models for Speech Recognition. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. 2015 Presented at: NIPS'15; December 2015; Montreal, Quebec, Canada p. 577-585.
33. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput 1989;1(4):541-551. [doi: 10.1162/neco.1989.1.4.541]
34. Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing.: ACL; 2014 Presented at: EMNLP'14; October 25–29, 2014; Doha, Qatar p. 1724-1734. [doi: 10.3115/v1/d14-1179]
35. Xie J, Wang Q. Benchmark Machine Learning Approaches with Classical Time Series Approaches on the Blood Glucose Level Prediction Challenge. In: Proceedings of the 2018 International Joint Conference on Artificial Intelligence. 2018 Presented at: IJCAI'18; July 13-19, 2018; Stockholm, Sweden.
36. Mikolov T, Karafiat M, Burget L, Cernocky J, Khudanpur S. Recurrent Neural Network Based Language Model. In: 11th Annual Conference of the International Speech Communication Association. 2010 Presented at: ISCA'10; September 26-30, 2010; Makuhari, Chiba, Japan.
37. Lipton ZC, Kale DC, Elkan C, Wetzel R. arXiv preprints. 2015. Learning to Diagnose with LSTM Recurrent Neural Networks URL: https://arxiv.org/abs/1511.03677 [accessed 2020-03-09]
38. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
39. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 2010;11:3371-3408 [FREE Full text]
40. Meyes R, Lu M, de Puiseau CW, Meisen T. arXiv preprints. 2019. Ablation Studies in Artificial Neural Networks URL: https://arxiv.org/abs/1901.08644 [accessed 2020-03-09]
41. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
42. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
43. Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning. 2010 Presented at: ICML'10; June 21-24, 2010; Haifa, Israel p. 807-814.
44. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep Contextualized Word Representations. In: Proceedings of NAACL-HLT 2018. 2018 Presented at: ACL'18; June 1 - 6, 2018; New Orleans, Louisiana, USA p. 2227-2237.
45. Chandar S, Khapra MM, Larochelle H, Ravindran B. Correlational neural networks. Neural Comput 2016 Feb;28(2):257-285. [doi: 10.1162/NECO_a_00801] [Medline: 26654210]
46. Munkhdalai T, Yu H. arXiv preprints. 2016. Neural Semantic Encoders URL: http://arxiv.org/abs/1607.04315 [accessed 2020-03-09]
47. GitHub. CLOUT Repository URL: https://github.com/subendhu19/CLOUT [accessed 2020-03-09]
48. Harrell Jr FE. Hbiostat. 2014. Regression Modeling Strategies URL: http://hbiostat.org/doc/rms.pdf [accessed 2020-03-09]
49. Genders TS, Steyerberg EW, Alkadhi H, Leschka S, Desbiolles L, Nieman K, CAD Consortium. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating, and extension. Eur Heart J 2011 Jun;32(11):1316-1330. [doi: 10.1093/eurheartj/ehr014] [Medline: 21367834]
50. Binenbaum G, Ying GS, Quinn GE, Dreiseitl S, Karp K, Roberts RS, Premature Infants in Need of Transfusion Study Group. A clinical prediction model to stratify retinopathy of prematurity risk using postnatal weight gain. Pediatrics 2011 Mar;127(3):e607-e614 [FREE Full text] [doi: 10.1542/peds.2010-2240] [Medline: 21321036]
51. Tanner L, Schreiber M, Low JG, Ong A, Tolfvenstam T, Lai YL, et al. Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS Negl Trop Dis 2008 Mar 12;2(3):e196 [FREE Full text] [doi: 10.1371/journal.pntd.0000196] [Medline: 18335069]
52. Pinzón-Sánchez C, Cabrera V, Ruegg P. Decision tree analysis of treatment strategies for mild and moderate cases of clinical mastitis occurring in early lactation. J Dairy Sci 2011 Apr;94(4):1873-1892 [FREE Full text] [doi: 10.3168/jds.2010-3930] [Medline: 21426977]
53. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery

and Data Mining. 2017 Presented at: KDD'17; August 13 - 17, 2017; Halifax, NS, Canada p. 1903-1911. [doi: 10.1145/3097983.3098088]

54. Ma T, Xiao C, Wang F. Health-ATM: A Deep Architecture for Multifaceted Patient Health Record Representation and Risk Prediction. In: Health-ATM: A Deep Architecture for Multifaceted Patient Health Record Representation and Risk Prediction. 2018 Presented at: SIAM'18; May 2018; San Diego, California, USA p. 261-269. [doi: 10.1137/1.9781611975321.30]
55. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 1;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
56. Sabour S, Frosst N, Hinton GE. Dynamic Routing Between Capsules. In: Proceedings of the 31st Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17; December 4-9, 2017; Long Beach, CA, USA p. 3586-3866.

Abbreviations

CNN: convolutional neural network
CLOUT: L(STM) Outcome prediction using Comprehensive features relations
EHR: electronic heath record
ICD: International Classification of Diseases
ICU: intensive care unit
LSTM: long short-term memory
MIMIC-III: Medical Information Mart for Intensive Care-III
RETAIN: Reverse Time Attention model
RNN: recurrent neural network
TaRETAIN: time-aware RETAIN

Edited by M Focsa, G Eysenbach; submitted 23.09.19; peer-reviewed by H Kharrazi, A Leichtle; comments to author 18.11.19; revised version received 27.01.20; accepted 14.02.20; published 23.03.20.

Please cite as:
Rongali S, Rose AJ, McManus DD, Bajracharya AS, Kapoor A, Granillo E, Yu H
Learning Latent Space Representations to Predict Patient Outcomes: Model Development and Validation
J Med Internet Res 2020;22(3):e16374
URL: http://www.jmir.org/2020/3/e16374/
doi:10.2196/16374
PMID:32202503

Original Paper

The Mechanisms Responsible for Improved Information Transfer in Avatar-Based Patient Monitoring: Multicenter Comparative Eye-Tracking Study

David Werner Tscholl[1], MD; Julian Rössler[1], MD; Lucas Handschin[1], MD; Burkhardt Seifert[2], PhD; Donat R Spahn[1], MD; Christoph B Nöthiger[1], MD

[1]Institute of Anesthesiology, University Hospital Zurich, Zurich, Switzerland

[2]Department of Biostatistics, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

Corresponding Author:
David Werner Tscholl, MD
Institute of Anesthesiology
University Hospital Zurich
Raemistrasse 100
Zurich, 8091
Switzerland
Phone: 41 78 663 67 87
Email: davidtscholl@me.com

Abstract

Background: Patient monitoring is central to perioperative and intensive care patient safety. Current state-of-the-art monitors display vital signs as numbers and waveforms. Visual Patient technology creates an easy-to-interpret virtual patient avatar model that displays vital sign information as it would look in a real-life patient (eg, avatar changes skin color from healthy to cyanotic depending on oxygen saturation). In previous studies, anesthesia providers using Visual Patient perceived more vital signs during short glances than with conventional monitoring.

Objective: We aimed to study the deeper mechanisms underlying information perception in conventional and avatar-based monitoring.

Methods: In this prospective, multicenter study with a within-subject design, we showed 32 anesthesia providers four 3- and 10-second monitoring scenarios alternatingly as either routine conventional or avatar-based in random sequence. All participants observed the same scenarios with both technologies and reported the vital sign status after each scenario. Using eye-tracking, we evaluated which vital signs the participants had visually fixated (ie, could have potentially read and perceived) during a scenario. We compared the frequencies and durations of participants' visual fixations of vital signs between the two technologies.

Results: Participants visually fixated more vital signs per scenario in avatar-based monitoring (median 10, IQR 9-11 versus median 6, IQR 4-8, *P*<.001; median of differences=3, 95% CI 3-4). In multivariable linear regression, monitoring technology (conventional versus avatar-based monitoring, difference=−3.3, *P*<.001) was an independent predictor of the number of visually fixated vital signs. The difference was less prominent in the longer (10-second) scenarios (difference=−1.5, *P*=.04). Study center, profession, gender, and scenario order did not influence the differences between methods. In all four scenarios, the participants visually fixated 9 of 11 vital signs statistically significantly longer using the avatar (all *P*<.001). Four critical vital signs (pulse rate, blood pressure, oxygen saturation, and respiratory rate) were visible almost the entire time of a scenario with the avatar; these were only visible for fractions of the observations with conventional monitoring. Visual fixation of a certain vital sign was associated with the correct perception of that vital sign in both technologies (avatar: phi coefficient=0.358; conventional monitoring: phi coefficient=0.515, both *P*<.001).

Conclusions: This eye-tracking study uncovered that the way the avatar-based technology integrates the vital sign information into a virtual patient model enabled parallel perception of multiple vital signs and was responsible for the improved information transfer. For example, a single look at the avatar's body can provide information about: pulse rate (pulsation frequency), blood pressure (pulsation intensity), oxygen saturation (skin color), neuromuscular relaxation (extremities limp or stiff), and body temperature (heatwaves or ice crystals). This study adds a new and higher level of empirical evidence about why avatar-based monitoring improves vital sign perception compared with conventional monitoring.

(J Med Internet Res 2020;22(3):e15070) doi:10.2196/15070

KEYWORDS

computers; diagnosis; visual perception; awareness; patient safety

Introduction

The World Health Organization considers continuous patient monitoring to be “extremely important” for the safety of the more than 313 million patients undergoing surgery worldwide each year [1,2]. In operating rooms and intensive care units around the world, monitors help millions of health care providers make critical treatment decisions [3,4]. However, previous research has found that conventional patient monitoring based on numbers and waveforms is not ideally suited for transferring patient status information to health care providers [5-7]. These studies recommend the development of new technologies to improve information transfer, especially from short glances at the monitors, because that is how care providers perform monitoring in real life.

In a previous comparative study with conventional monitoring, we found that anesthesia professionals were able to perceive more vital signs when monitoring with Visual Patient, a technology integrating vital sign information into an easy-to-interpret animated avatar model of the patient’s status, designed according to principles of user-centered design [4,8,9]. When using avatar-based monitoring, participants rated their self-confidence in the correctness of their diagnoses as higher and their subjectively perceived workload as lower.

Although the biocular human visual field encompasses approximately 214 arc degrees horizontally and 150 arc degrees vertically, we can only see sharply in a circular area of approximately 2 arc degrees in the center of our visual field, named the fovea [10]. While reading, we move our eyes to let the light reflected from the words fall through the pupil and the lens directly onto the fovea. At a distance of approximately one arm’s length, the foveal region in which we can see sharp, colorful images is approximately the size of a thumbnail or a circle with a radius of 2 centimeters [11,12].

Human eye movements take place in the form of so-called fixations and saccades. Visual fixations are the periods during which the gaze rests on a position, and information can reach the visual cortex and potentially be interpreted. Saccades are the rapid movements of the eyeballs between fixations [13]. Eye-tracking technology can record both visual fixations and saccades. For this study, we systematically recorded, analyzed, and compared eye-tracking data of participants who watched patient-monitoring scenarios alternatingly as conventional and avatar-based patient monitoring. The rationale of this study was to uncover the underlying functional principles in both monitoring technologies through eye-tracking analysis. These results may be useful for improved understanding of avatar-based monitoring and, across domains, for the future development of user interfaces designed to transfer relevant information as efficiently as possible. We hypothesize that avatar-based monitoring (Visual Patient) facilitates information perception through its compact layout, which enables users to visually fixate on more vital signs.

Methods

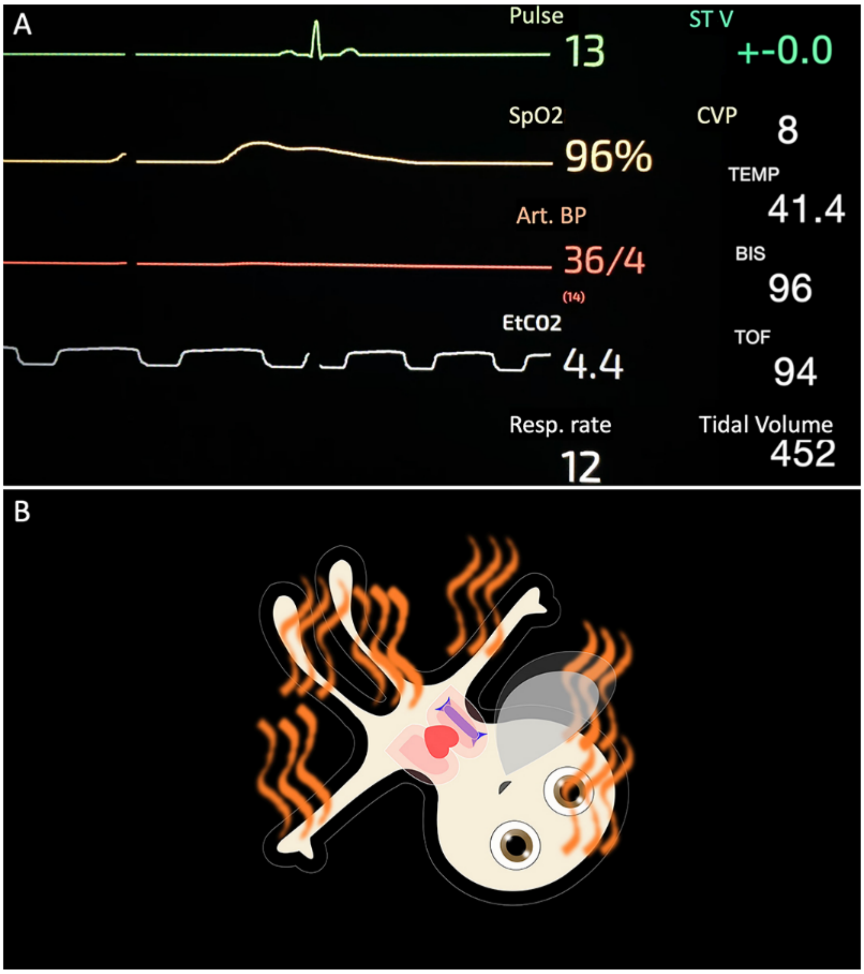
In this paper, we describe the analysis of eye-tracking data that was collected as part of a multimethod laboratory study. The primary objective of that study was to compare the perceptual performance of anesthesia professionals using newly developed avatar-based technology with state-of-the-art number- and waveform-based patient monitoring [8].

The Ethics Committee of the Canton of Zurich, Switzerland, reviewed the study protocol and issued a declaration of nonresponsibility specifying that the research project did not fall into the scope of the Human Research Act (Business Administration System for Ethics Committees, number: 2016-00103). Nevertheless, we obtained written consent for the use of the collected data for scientific purposes from all participants.

Description of Visual Patient Technology

Visual Patient, as used in this study, can display the 11 most frequently monitored vital signs: pulse rate, blood pressure, oxygen saturation, ST segment of the electrocardiogram, central venous pressure, respiratory rate, tidal volume, expiratory carbon dioxide concentration, body temperature, brain activity, and degree of neuromuscular relaxation. We developed the technology as a situation awareness tool, analogous to synthetic vision technology in aviation, according to the principles of user-centered design and principles of logic [4,14]. Synthetic vision technology generates a virtual image of the environment from the data measured by the aircraft (eg, airspeed) and Global Positioning System geolocation data and data stored on onboard computers (eg, georeferenced terrain elevation data). For the pilot, the generated virtual image looks like the view from the cockpit in perfect weather conditions. This similarity between the virtual image and reality makes the image intuitively understandable and enables a quick and uncomplicated perception of the flight situation. Visual Patient uses the same logic by creating a virtual image of the patient from vital data. It presents the data in a way that corresponds to the real phenomena that the data would cause in the patient. For example, the pulse rate corresponds to the pulsation of the avatar’s body to represent the pulse wave passing through the body with every heartbeat. High brain activity is represented by open eyes because that is what the care providers expect from a patient with high brain activity according to their mental models. Figure 1 shows a monitoring scenario in routine conventional form and as a Visual Patient representation.

Figure 1. Patient monitoring scenario used in this study: (A) presented in routine conventional number and waveform-based format; (B) presented in avatar-based format (Visual Patient). SpO_2: Peripheral oxygen saturation; Art BP: arterial blood pressure; $EtCO_2$: end-tidal carbon dioxide concentration; Resp rate: respiratory rate; ST V: electrocardiogram ST segment of lead number 5; CVP: central venous pressure; Temp: body temperature; BIS: bispectral index system (brain activity); TOF: train of four (neuromuscular relaxation).



This direct presentation of information eliminates the need to calculate the relevant information (eg, “What is the current anesthesia depth?”) from lower-level data (eg, bispectral index=85) [15]. In addition to this direct presentation of information, the other two main features of the avatar technology are the preprocessing of data for each vital sign into categories (“no data measured,” “too low,” “normal,” or “too high”) and the presentation of vital parameter information in multiple visualizations simultaneously. For example, the caregiver can evaluate the respiratory rate based on the respiratory rate of the avatar lung and the formation rate of the carbon dioxide cloud exhaled by the avatar.

These combined functions translate a large number of numerical values into an animated model of the patient situation, which the caregiver can evaluate and memorize at a glance. The vital parameters are translated into the avatar model in real time from the monitoring data. If no data are measured for a particular vital sign, the corresponding visualization in the Visual Patient remains gray and framed with dashes. We have described the validation and evaluation process of the technology in detail in previous studies [8,16].

Study Participants

The participants in this study were attending and resident physician anesthesiologists and specialist anesthesia nurses from the anesthesia departments of the University Hospital Zurich and the Cantonal Hospital of Winterthur. The University Hospital Zurich is one of the largest university hospitals in Switzerland, where more than 30,000 operations are performed per year; the Cantonal Hospital Winterthur is a large regional teaching hospital where approximately 10,000 operations are performed per year.

Participation in this study was voluntary, and there was no monetary compensation for the participants. We recruited colleagues who responded to an institutional invitation and recruited additional colleagues according to availability. At both centers, we included equal numbers of male and female participants and participants from three professional groups: (1) senior anesthesia physicians, (2) resident physicians, and (3) anesthesia nurses.

Study Setting

Before data collection, the participants received training in avatar-based monitoring through a 6-minute educational video (Multimedia Appendix 1). The participants also familiarized themselves with the layout of the conventional monitoring used in the study: a simulation of a GE Datex Ohmeda Monitors (General Electric Company, Boston, MA, USA) recorded with the SimMon App (Castle 2 Andersen ApS, Hillerød, Denmark), which was equivalent to patient monitoring in routine use in the two centers. There was no additional training in conventional patient monitoring because all participants had at least one full year, some even decades, of anesthesia experience. The eye-tracking data were recorded during the evaluation of patient monitoring scenarios. In random order, we presented participants with 3- and 10-second prerecorded videos of patient monitoring scenarios shown in avatar and state-of-the-art number and waveform format. Multimedia Appendix 2 provides examples of a conventional and an avatar-based scenario. Each participant rated four videos in sequence. These videos consisted of a 3- and a 10-second monitoring scenario, each of which was shown twice, once with either technology. The scenarios came from a pool of four total scenarios, as outlined in Multimedia Appendix 3. The scenarios were designed with unambiguously safe or unsafe vital sign values and contained random vital sign abnormalities to avoid pattern recognition (ie, inferring the status of vital signs based on the status of the other vital signs).

To blind the participants to the fact that they were evaluating the same scenarios twice (once with either technology), we showed the scenarios in alternating order, starting with a random first scenario. Multimedia Appendix 3 shows a flowchart detailing this procedure. The scenario playback was performed on an Aspire V15 Nitro 15-inch laptop computer (ACER, Inc, Taipei, Taiwan) in ultra-high resolution (3840×2160 pixels) at 60 frames per second. The conventional monitoring scenarios included a standard audio signal with frequency and pitch for heart rate and oxygen saturation.

After brief time intervals, the screens darkened, and the participants indicated how they had perceived the 11 vital signs displayed in the scenarios as either “too low,” “too high,” “safe,” or “no recall.” We based this method on the Situation Awareness Global Assessment Tool developed by Endsley [4,17]. After each scenario, for every vital sign, the participants indicated how confident they felt that they had perceived it correctly. Furthermore, they were asked to rate their subjectively perceived workload for each scenario using the NASA (National Aeronautics and Space Administration) Task Load Index [18,19]. Data were collected using an iPad-based (Apple Inc, Cupertino, CA, USA) data collection tool [20].

Recording and Analysis of Eye-Tracking Data

We evaluated the eye-tracking data for this study according to the physiological principles of the human eye and neurophysiological principles of human vision outlined in the Introduction. We used a stationary eye tracker (Gazepoint GP3, Gazept, Vancouver, BC, Canada) to capture visual fixations and saccades of participants observing conventional and avatar-based patient monitoring scenarios. The eye tracker recorded the position of the foveal vision on the screen 60 times per second and with 0.5 to 1 degree of visual angle accuracy.

Outcome Measures

Vital Signs Fixated Per Scenario (Primary Outcome)

We chose to compare visual fixations in this study because we regarded them as a relevant requirement for perception. Based on the anatomic and physiologic principles outlined in the Introduction, we analyzed each visual fixation longer than 50 milliseconds in the eye-tracking recordings of each participant and scenario; for each visual fixation, we counted the vital signs that were within 2 centimeters of the fixation. Using this method, we identified the vital signs that participants could potentially have read during the recording because they were within the potentially readable visual area. A video demonstrating this method is available in Multimedia Appendix 2. Information can only reach the brain for processing after reading, which requires a visual fixation.

Conventional patient monitoring shows the vital signs on the screen in the form of numbers or waveforms. In conventional monitoring, if a participant had a visual fixation within 2 centimeters of the number or waveform representing a certain vital sign (eg, pulse rate), we counted a visual fixation for this vital sign. By comparing the numbers of vital sign visual fixations between the two technologies, we wanted to find out whether the participants could visually fixate on more vital signs with either one of the two technologies.

Visual Fixations Per Vital Sign

We also compared visual fixations for each of the 11 vital signs individually. This allowed us to determine whether vital signs were visually fixated more often with either one of the two monitoring technologies. We expected these findings to provide an explanation for the improved perceptive performance in avatar-based patient monitoring found in previous studies [8,16].

Duration of Visual Fixations Per Vital Sign

Analogous to the visual fixations per vital sign, we also compared the time durations of the visual fixations of each of the 11 vital signs with both monitoring technologies. In doing so, we evaluated whether either one of the two monitoring technologies would cause the vital signs to be visible for a longer time per observation. We analyzed this outcome measure because longer availability of the vital sign information could explain why participants’ perceptual performance was improved with avatar-based patient monitoring in previous studies [8,16].

Correlation of Vital Sign Visual Fixations With Correct Perception

To evaluate the association of visual fixation of a vital sign and its correct perception, we calculated phi correlation coefficients. If visual fixation correlated with correct perception, and the avatar enabled more vital signs to be seen per time interval, these results could validate both the study method and the avatar concept.

Correlation of Vital Sign Visual Fixations With Perceived Confidence

We calculated coefficients between visual fixation and diagnostic confidence to evaluate whether the visual fixation of a vital sign correlated with the subjectively perceived confidence in the correctness of one’s own diagnosis.

Statistical Analysis

Sample Size Calculation

Before starting the study, we conducted a pilot study with five participants. We calculated the sample size using the effect size of 1.23 measured in the pilot study. Assuming a clinically relevant difference of one vital sign and an observed standard deviation of 0.81, the post hoc power analysis for a paired *t* test resulted in a sample size of eight participants, for an alpha error probability of 5% and a power of 0.8. To achieve this sample size in both centers and all four scenarios, we had to include at least 32 participants.

Descriptive Statistics and Normality Tests

Distribution of variables is expressed using medians and interquartile ranges (IQRs) regardless of normality. Normality was assessed with the Shapiro-Wilk test and visual inspection of quantile-quantile plots of dependent variables.

Univariate Statistics

Participants watched and evaluated the same monitoring scenarios with both monitoring technologies; therefore, we were able to perform intraparticipant comparisons. Depending on normality, we used either paired Student *t* tests or Wilcoxon signed rank tests to compare the number of vital sign visual fixations with both monitoring technologies. We calculated the 95% confidence interval (95% CI) of the median of differences using the Hodges-Lehmann estimate. To test the differences in visual fixations and duration of visual fixations per vital sign for statistical significance, we used Mann-Whitney tests. In this study, we performed multiple comparisons; therefore, we considered *P* values between .05 and .01 as trends and *P* values of <.01 as statistically significant.

Multivariable Linear Regression

Multivariable linear regression was performed with number of visually fixated vital signs between the monitoring technologies and its differences as dependent variables. Scenario duration, order of scenarios, center, profession, and gender of the participant served as possible predictors. Clustering of observations within the same participant was addressed using cluster robust standard errors.

Correlation Analyses

To test for associations between visual fixation of a vital sign and its correct perception as well as participants’ subjectively perceived confidence in the correctness of the diagnosis, we calculated chi-square tests for association and Pearson phi coefficients between visual fixation, accurate perception, and diagnostic certainty. The phi coefficient corresponds to a Pearson correlation coefficient estimated for two binary variables. We considered “very unconfident” and “unconfident” as 0 and “confident” and “very confident” as 1. If the frequency of an event was less than five, we used the Fisher exact test to assess statistical significance.

Statistical Software

We used Q*Power 3 (Heinrich-Heine-University, Düsseldorf, Germany) [21], Prism 8.1.1. (GraphPad Software, La Jolla, CA, USA), IBM SPSS Statistics 24 (International Business Machines Corporation, Armonk, NY, USA), and Stata 13.1 (StataCorp, College Station, TX, USA) for statistical analyses.

Results

Study and Participant Characteristics

Table 1 shows the characteristics of the study and the participants in detail. A total of 32 participants participated in the two study centers. Had we been able to record data from all 32 participants and scenarios, a theoretical maximum of 64 direct comparisons between avatar-based and conventional patient monitoring would have been possible. However, we were unable to record the eye-tracking data of two participants. For four other participants, we were only able to record one of two monitoring scenarios they watched due to technical problems. Despite the missing data, we were still able to evaluate 56 within-subject comparisons of eye-tracking data (88% of all 64 theoretically possible comparisons).

Table 1. Study and participant characteristics.

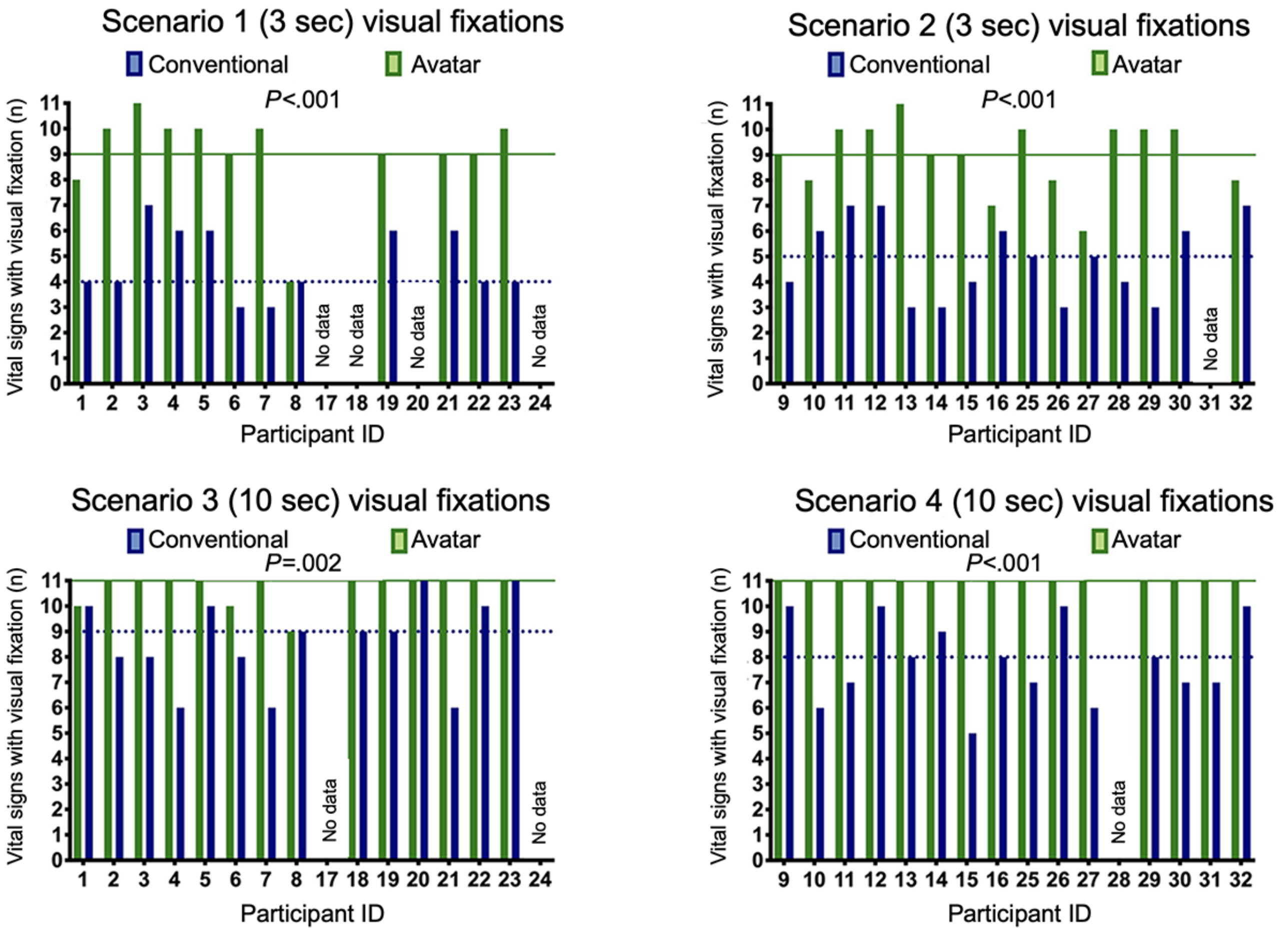
Characteristic	Study center		Total (N=32)
	University Hospital Zurich (n=16)	Cantonal Hospital Winterthur (n=16)	
Participants with successful eye-tracking recording, n	16	14	30
Direct comparisons, n	32	24	56
Participants, n (%)			
Staff members	6 (37)	6 (37)	12 (37)
Residents	4 (25)	4 (25)	8 (25)
Nurse anesthetists	6 (37)	6 (37)	12 (37)
Gender, n (%)			
Female	7 (44)	10 (62)	17 (53)
Male	9 (56)	6 (37)	15 (47)
Age group, n (%)			
25 to 34 years	10 (63)	6 (38)	16 (50)
35 to 44 years	6 (38)	2 (13)	8 (25)
45 to 54 years	0 (0)	6 (38)	6 (19)
55 to 65 years	0 (0)	2 (13)	2 (6)
Anesthesia experience, n (%)			
<1 year	1 (6)	1 (6)	2 (6)
1 to 5 years	5 (31)	4 (25)	9 (28)
5 to 10 years	9 (56)	1 (6)	10 (31)
>10 years	1 (6)	10 (63)	11 (34)
Monitors from different manufacturers previously used, median (IQR)	2 (2-3)	2 (1-4)	2 (1-3)
Duration of data collection session (minutes), median (IQR)	32 (28-35)	35 (32-41)	33 (30-39)
Duration of study (days), n	20	2	22

Outcome Measures

Vital Signs Fixated Per Scenario

With the avatar-based monitoring, all participants in all scenarios were able to visually fixate on more vital signs than with conventional monitoring (Figure 2). In the short 3-second scenarios, the median numbers of vital sign fixations with avatar-based monitoring were approximately twice as high as conventional patient monitoring. In scenario 1, the avatar-based median was 9 (IQR 9-10) versus the conventional median of 4 (IQR 4-6, *P*<.001; median of differences=3, 95% CI 3-4). In scenario 2, the avatar median was 9 (IQR 8-10) versus the conventional median of 5 (IQR 3-6, *P*<.001; median of differences=5, 95% CI 2-6). In scenario 3, the first of the longer 10-second scenarios, the median number of vital sign fixations for avatar-based monitoring was 11 (IQR 11-11) versus the conventional median of 9 (IQR 6-10, *P*=.002; median of differences=2, 95% CI 0-4). In scenario 4, the second 10-second scenario, vital sign visual fixations were a median of 11 (IQR 11-11) for avatar versus the conventional median of 8 (IQR 7-10, *P*<.001; median of differences=3, 95% CI 1-4). Figure 1 shows these results on an individual participant level.

Figure 2. Avatar-based monitoring compared with conventional patient monitoring for vital signs visually fixated on by individual participants. Scenario 1 (3 seconds): n=12; scenario 2 (3 seconds): n=15; scenario 3 (10 seconds): n=14; and scenario 4 (10 seconds): n=15. The dotted lines indicate the medians. Participants 1-8 (University Hospital Zurich) and 17-24 (Cantonal Hospital Winterthur) rated scenarios 1 and 3; participants 9-16 (University Hospital Zurich) and 25-32 (Cantonal Hospital Winterthur) rated scenarios 2 and 4.



In a multivariable linear regression adjusted for scenario duration, order of scenarios, center, profession, and gender of the participant, the technology (conventional versus avatar-based monitoring) had a significant effect on the number of vital sign fixations: difference between technologies=−3.28, 95% CI −3.86 to −2.69, $P<.001$ ($F_{6,30}$=145, Prob>F<.001, R^2=.56, adjusted for clusters within participants). Table 2 shows the results of the multivariable linear regression for the difference of numbers of visually fixated vital signs between the technologies. In this analysis, only scenario duration affected the difference in vital sign fixations between technologies. The difference was less prominent in the 10-second scenarios (difference between scenario durations=−1.46, 95%CI −2.84 to −0.07, P=.04). Study center, profession, gender, and scenario order did not influence the differences between conventional and avatar-based monitoring.

Table 2. Multivariable linear regression for the difference in numbers of visually fixated vital signs between conventional and avatar-based monitoring.[a]

Variable	Difference (95% CI) (multivariable linear regression)	Standard error	Difference / standard error (multivariable linear regression)	*P* value
Scenario duration (3 versus 10 seconds)	−1.46 (−2.84 to −0.07)	0.68	−2.15	.04
Profession	0.33 (−0.87 to 1.52)	0.59	0.57	.58
Study center	−0.41 (−1.95 to 1.14)	0.76	−0.54	.60
Gender	0.01 (−1.20 to 1.22)	0.59	0.02	.99
Order of scenarios	0.10 (−1.51 to 1.70)	0.78	0.12	.90
Technology (conventional versus avatar [intercept])	−3.28 (−3.86 to −2.69)	0.29	−11.47	<.001

[a] Clustering of observations within the same participant was addressed using cluster robust standard errors.

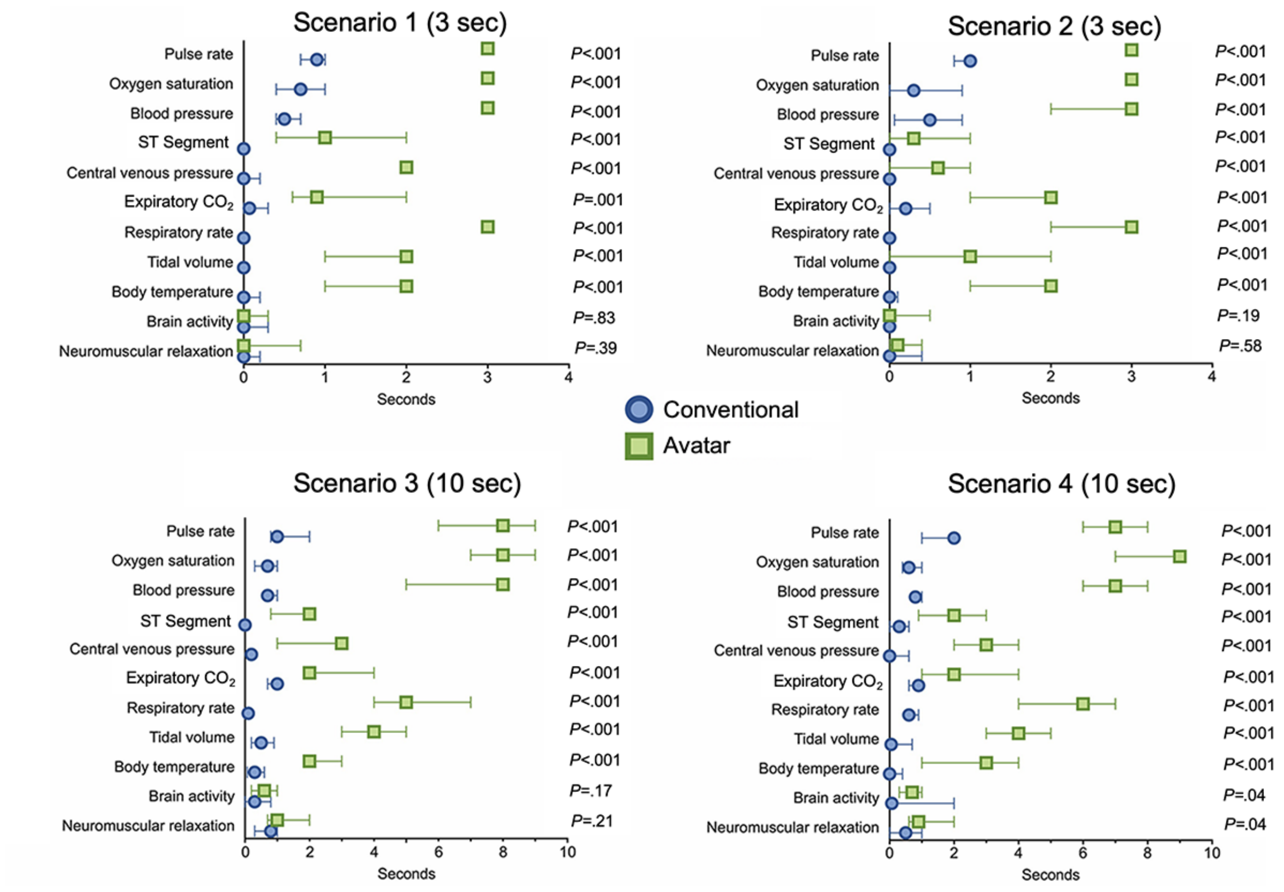
Visual Fixations Per Vital Sign

The analyses for each vital sign individually (Multimedia Appendix 3) showed that with avatar-based monitoring, 9 of 11 vital signs were fixated statistically significantly more often per scenario than with conventional patient monitoring. The vital signs for pulse rate, oxygen saturation, and blood pressure were visible in almost every fixation of participants in all four scenarios with avatar-based monitoring. In comparison, with conventional monitoring, each vital sign was readable only during a small number of visual fixations per observation.

Duration of Visual Fixations Per Vital Sign

Similar to the number of fixations per vital sign, in all four scenarios, 9 of 11 vital signs were visually fixated significantly longer with the avatar than with conventional patient monitoring (Figure 3). With avatar-based monitoring, four critical vital signs (pulse rate, blood pressure, oxygen saturation, and respiratory rate) were visible to users during almost the entire time of the scenarios. This was because, in the avatar's design, this information is displayed in the form of large anatomical objects, which extend across large parts of the screen. For example, the body of the avatar and the exhaled CO_2 cloud (Figure 1).

Figure 3. Avatar-based monitoring compared with conventional patient monitoring: median (with interquartile range) durations of visual fixations for each vital sign, scenario, and technology. Scenario 1 (3 seconds): n=12; scenario 2 (3 seconds): n=15; scenario 3 (10 seconds): n=14; and scenario 4 (10 seconds): n=15.

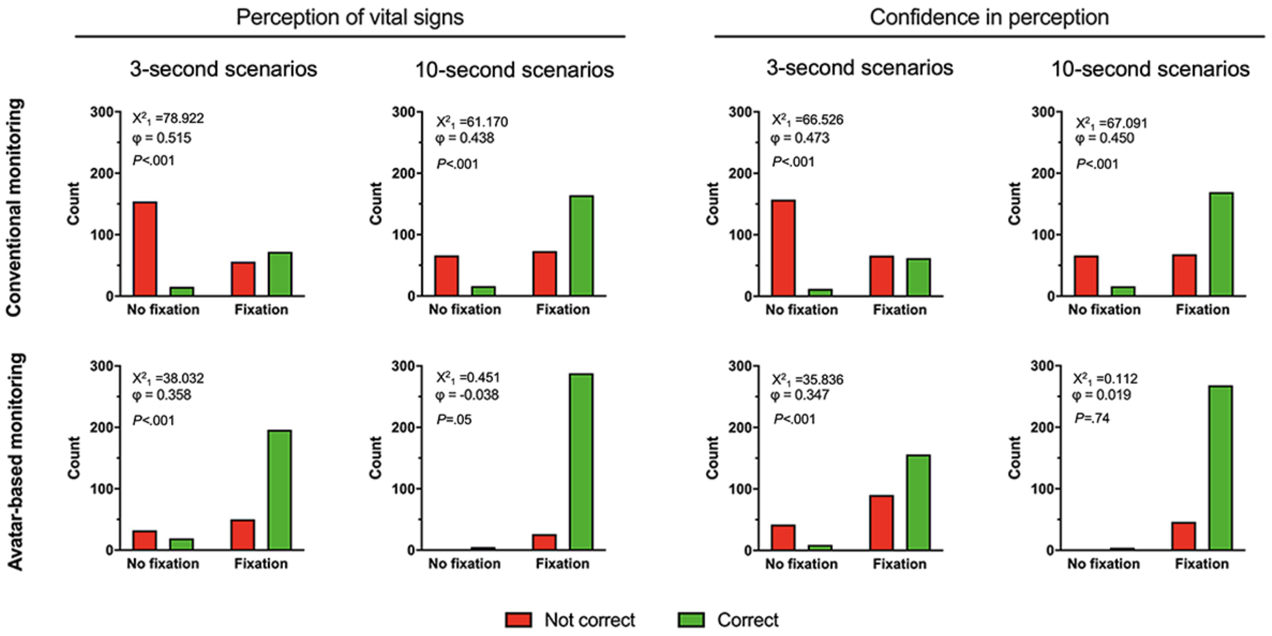


Correlation of Vital Sign Visual Fixations With Correct Perception and Perceived Confidence

A chi-square test for association was conducted between visual fixation of a vital sign and the correct perception of this vital sign. All expected cell frequencies were greater than five, except in the 10-second scenario with avatar-based monitoring. In this scenario, most participants were able to fixate on every vital sign and perceive it correctly. Accordingly, there was a statistically significant association between visual fixation of a vital sign and the correct perception of said vital sign in the 3- and 10-second scenarios with conventional monitoring (3-second scenario: χ^2_1=78.9; 10-second scenario χ^2_1=61.1, both $P<.001$) and in the 3-second scenario with the avatar-based monitoring (χ^2_1=38, $P<.001$). When significant, the association was moderately strong (Figure 4).

Similar results were achieved by a chi-square test for association between visual fixation of a vital sign and the participants' confidence in having perceived it correctly (Figure 4).

Figure 4. Cross tabulation bar graphs with chi-square tests for associations between visual fixation of a vital sign and the correct perception of the vital sign and the confidence in the correct perception.



Discussion

Overview

Patient monitoring is a central part of modern surgery, anesthesia, and intensive care [1,22]. Currently available monitors enhance perioperative safety [23,24]; however, they mainly show vital sign information as numbers and waveforms, which is not an ideal format for quick and easy interpretation [4,5]. An alternative monitoring technique, using an avatar-based representation of vital signs, has been found to enable anesthesiologists to grasp more vital sign information in a shorter time, resulting in improved diagnostic confidence and diminished perceived workload [8,16].

Principal Findings

In this study, we evaluated eye-tracking data collected in two groups of anesthesiologists at two study centers. We recorded these data while the anesthesiologists were given the task to perceive vital sign information from patient monitoring scenarios presented in the two technologies (ie, conventional and avatar-based). Specifically, we evaluated how many vital signs and for how long these vital signs could have potentially been read by the participants according to the paths of their foveal or sharp vision across the screen. We found that participants were able to visually fixate more vital signs during the same time with avatar-based monitoring than with conventional patient monitoring. Nine of 11 vital signs were fixated more frequently per observation with avatar-based monitoring. Moreover, with avatar-based monitoring, participants fixated the vital signs for longer time intervals per recording, which might give them more time to process the information. More time to perceive the information may have been responsible for the reductions in perceived workload. In short, with the avatar, users see more information for a longer time. These findings were a consequence of the design of the avatar, with many of the vital signs spread out over a large part of the screen, and some visualized multiple times. For example, the vital sign “respiratory rate” can be interpreted by looking at the expiratory carbon dioxide “cloud” of the avatar and in the excursions of its lungs. The number of correctly perceived vital signs without a visual fixation accounted for less than 10% of the correctly perceived vital signs in all scenarios and with both technologies. This may have been influenced by the audio signal played in the conventional monitoring scenarios, which contained information on pulse rate and oxygen saturation. There may also have been some correct guesses without actual perception. In the avatar scenarios, some of the vital signs may have been perceived through peripheral vision, which we found to be an additional advantage of avatar-based monitoring [16].

The cases of visually fixated vital signs that were not correctly detected accounted for between 0% and 20% of all vital signs, depending on the scenario and technology, which might be explained by losses during processing after visual fixation, such as when a vital sign is forgotten or confused for another vital sign before being recalled. Numbers are glyphs that cannot be attributed solely to one vital sign; that is, it might be possible that a participant, although remembering the value of a number correctly, may misattribute the number to another vital sign with a similar range. Indeed, our data showed that when participants had to remember more than just a few vital signs in the more extended 10-second scenarios, the number of vital signs with a visual fixation that participants could not recall was more than twice as high in conventional monitoring than in avatar-based monitoring. These results correlate with research on the holding capacity of our working memory, which has shown that people can only store seven plus or minus two digits in short-term memory [25].

With avatar-based monitoring, nearly all participants were able to visually fixate and correctly perceive almost all the vital sign information in the longer (10-second) scenarios. This study shows the limitations of the single-sensor, single-indicator design of conventional patient monitoring, in which a single sensor on the patient feeds a single indicator on the patient monitor. The numbers must be individually read one after the other and then interpreted before meaning and a global mental picture of the situation can be derived [5,15]. With avatar-based monitoring, we found that the four critically important vital signs (pulse rate, blood pressure, oxygen saturation, and respiratory rate) remained perceptible for almost the entire duration of the monitoring scenarios. To perceive this same information, conventional monitoring requires four visual fixations, the eye movements in between them, and brainwork to interpret the meaning of the values. Avatar-based monitoring facilitates the interpretation work by the use of vital sign visualizations that have a logical commonality with the real phenomena they mirror and therefore do not require further mental translation by the user to be understood. The principle that a good model reflects the reality it represents is found both in principles of logic [14] and in situation awareness design principles, in which it is known as “presenting information directly” [4]. Anesthesiologists have mentioned information overload as a common problem in connection with patient monitoring [26]. In the future, more and more inexperienced users will likely monitor patients; therefore, ease of information transfer will be of paramount importance [27]. The ultimate benefit of Visual Patient should be an increase in patient safety. Although at this stage of its development we are not yet able to evaluate patient outcomes, the results of this and our previous studies fit into the context of situation awareness, decision making, and performance. Care providers must perceive and understand the available information before they can confidently make the correct decision and take the right measure [3,28]. Situation awareness failures have been identified as root causes of critical anesthesia events [29,30].

Limitations

This study has some important limitations. For one, self-enrollment, based on interest in the technology, could have led to a selection bias. Less technology-savvy care providers may have achieved different results. Secondly, we recorded the source data in a simulated environment. The operating room and intensive care unit environment are very complex in real life; therefore, it is impossible to predict precisely how substantial the effects of an avatar-based monitor would be in these settings. However, it is plausible that effects would persist when used as a real patient monitor because the general physiological specifications of information intake do not change. A study with a high-end patient simulator in a realistic environment with the technology must be carried out as the next step of scientific evaluation on the way to producing a commercial product. Another potential limitation is the versatility of the eye-tracking method. Although we were able to validate the method through the positive correlation between visual fixation and correct perception, there are influences in perception that are not fully detectable by eye tracking, such as the influences of the audio signal, peripheral vision, and working memory. Particular strengths of this study include its multicenter design and the balanced enrollment of the different occupational groups and genders—a multivariate regression analysis rendered significant local effects, gender, profession, and scenario ordering effects unlikely. The within-subject study design minimizes the impact of interviewer variability of the eye-tracking method and other interparticipant variabilities [31,32]. The sample size adequately powered the analyses, and the significant magnitude and consistency of the differences observed between the two monitoring technologies increase the internal validity of the study.

Limitations of Visual Patient Technology

Visual Patient has some inherent limitations. The technology simplifies vital signs into categories (ie, “too low,” “normal,” or “too high”). This preprocessing leads to improved intelligibility and diagnostic certainty but also reduces data accuracy (three discrete categories versus 300 different numbers in the case of pulse rate). Another limitation of the Visual Patient version used in this study is that it cannot yet display trends. This aspect is important because trend displays of conventional patient monitors can help care providers detect changes over time. In this context, it is important to note that we are developing avatar-based monitoring to improve information transfer, but not as a replacement of the conventional monitoring streams. Successful integration of the two technologies will be key for the success of Visual Patient, as with synthetic vision technology and numerical flight data.

Conclusions

This study analyzed eye-tracking data to explain the improved information transfer with avatar-based patient monitoring. The avatar’s design, in which the vital sign information is presented as large-scale, integrated, colorful, and direct visualizations, allows users to see information about more vital signs with every glance and also see the vital sign information for a longer time with every glance at the monitor. In short, the way the avatar presents the information enables parallel perception of multiple vital signs at the same time, thereby increasing the number of visually fixated vital signs and the time available to view each vital sign. This study provides important groundwork for the future clinical validation of the concept. Future studies should examine the technology’s performance in simulator-based and then real-life studies.

Acknowledgments

The authors are thankful to the study participants for their time and effort. The Institute of Anesthesiology and University Hospital Zurich provided institutional funding; the University of Zurich provided proof-of-concept funding (UZ16/288PoC); and the University of Zurich provided author DWT with a filling-the-gap career development grant.

Authors' Contributions

DWT, LH, BS, DRS, and CBN helped to design the study; DWT, LH, and CBN helped to collect the data; DWT, JR, LH, BS, DRS, and CBN helped to analyze the data; and DWT, JR, LH, BS, DRS, and CBN helped to write the manuscript and approved the final version.

Conflicts of Interest

The University of Zurich, Zurich, Switzerland, and Koninklijke Philips NV, Amsterdam, Netherlands, entered a joint development and licensing agreement to develop avatar-based monitoring software based on technology owned by the university and described in this paper. As part of their contract with the University, as designated inventors, the authors DWT and CBN could receive royalties. All other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Visual Patient educational video shown to the participants in this study.
[MOV File , 52399 KB - jmir_v22i3e15070_app1.mov]

Multimedia Appendix 2
Video demonstrating method used in study. Visual fixations longer than 50 milliseconds were analyzed and the vital signs within 2 centimeters of the fixation were counted. Therefore, the vital signs that participants could potentially have read during the recording (located within the potentially readable visual area) were identified.
[MP4 File (MP4 Video), 20569 KB - jmir_v22i3e15070_app2.mp4]

Multimedia Appendix 3
Supplementary Figures 1 and 2, and Supplementary Table 1.
[DOCX File , 355 KB - jmir_v22i3e15070_app3.docx]

References

1. WHO Guidelines for Safe Surgery. Geneva, Switzerland: World Health Organization; 2009.
2. Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, Uribe-Leitz T, et al. Size and distribution of the global volume of surgery in 2012. Bull World Health Organ 2016 Mar 01;94(3):201-209F [FREE Full text] [doi: 10.2471/BLT.15.159293] [Medline: 26966331]
3. Schulz CM, Endsley MR, Kochs EF, Gelb AW, Wagner KJ. Situation awareness in anesthesia: concept and research. Anesthesiology 2013 Mar;118(3):729-742. [doi: 10.1097/ALN.0b013e318280a40f] [Medline: 23291626]
4. Endsley M, Jones D. Designing For Situation Awareness: An Approach To User-Centered Design. Second Edition. Boca Raton, FL: CRC Press; 2011.
5. Drews FA, Westenskow DR. The right picture is worth a thousand numbers: data displays in anesthesia. Hum Factors 2006;48(1):59-71. [doi: 10.1518/001872006776412270] [Medline: 16696257]
6. Ford S, Birmingham E, King A, Lim J, Ansermino JM. At-a-glance monitoring: covert observations of anesthesiologists in the operating room. Anesth Analg 2010 Sep;111(3):653-658. [doi: 10.1213/ANE.0b013e3181e627d4] [Medline: 20581165]
7. Loeb RG. Monitor surveillance and vigilance of anesthesia residents. Anesthesiology 1994 Mar;80(3):527-533. [Medline: 8141449]
8. Tscholl DW, Handschin L, Neubauer P, Weiss M, Seifert B, Spahn DR, et al. Using an animated patient avatar to improve perception of vital sign information by anaesthesia professionals. Br J Anaesth 2018 Sep;121(3):662-671. [doi: 10.1016/j.bja.2018.04.024] [Medline: 30115265]
9. Tscholl DW, Weiss M, Handschin L, Spahn DR, Nöthiger CB. User perceptions of avatar-based patient monitoring: a mixed qualitative and quantitative study. BMC Anesthesiol 2018 Dec 11;18(1):188 [FREE Full text] [doi: 10.1186/s12871-018-0650-1] [Medline: 30537934]
10. Traquair H. An Introduction to Clinical Perimetry. London: Kimpton; 1927.
11. Holmqvist K. Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford, UK: Oxford University Press; 2011.
12. Neilson J, Pernice K. Eyetracking Web Usability. Berkeley, CA: New Riders; 2009.
13. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull 1998 Nov;124(3):372-422. [Medline: 9849112]
14. Wittgenstein L. Tractatus Logico-Philosophicus. London, UK: Routledge & Kegan Paul; 1922.
15. Degani AJ, Iverson D, Shafto M, Olson L. On Organization of Information: Approach and Early Work. Moffett Field, CA: NASA; 2009.

16. Pfarr J, Ganter MT, Spahn DR, Noethiger CB, Tscholl DW. Avatar-based patient monitoring with peripheral vision: a multicenter comparative eye-tracking study. J Med Internet Res 2019 Jul 17;21(7):e13041 [FREE Full text] [doi: 10.2196/13041] [Medline: 31317870]
17. Endsley M, Garland D. Situation Awareness Analysis and Measurement. Boca Raton, FL: CRC Press; 2000.
18. Hart S, Stavenland L. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Advances in Psychology: Human Mental Workload. Amsterdam: Elsevier; 1988:139-183.
19. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 years later. Proc Hum Fact Soc An 2016 Nov 05;50(9):904-908. [doi: 10.1177/154193120605000909]
20. Tscholl DW, Weiss M, Spahn DR, Noethiger CB. How to conduct multimethod field studies in the operating room: the iPad combined with a survey app as a valid and reliable data collection tool. JMIR Res Protoc 2016 Jan 05;5(1):e4 [FREE Full text] [doi: 10.2196/resprot.4713] [Medline: 26732090]
21. 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. [Medline: 17695343]
22. Gelb AW, Morriss WW, Johnson W, Merry AF, Abayadeera A, Belîi N, International Standards for a Safe Practice of Anesthesia Workgroup. World Health Organization-World Federation of Societies of Anaesthesiologists (WHO-WFSA) International Standards for a Safe Practice of Anesthesia. Anesth Analg 2018 Jun;126(6):2047-2055. [doi: 10.1213/ANE.0000000000002927] [Medline: 29734240]
23. Moller JT, Johannessen NW, Espersen K, Ravlo O, Pedersen BD, Jensen PF, et al. Randomized evaluation of pulse oximetry in 20,802 patients: II. Perioperative events and postoperative complications. Anesthesiology 1993 Mar;78(3):445-453. [doi: 10.1097/00000542-199303000-00007] [Medline: 8457045]
24. Maheshwari K, Khanna S, Bajracharya GR, Makarova N, Riter Q, Raza S, et al. A randomized trial of continuous noninvasive blood pressure monitoring during noncardiac surgery. Anesth Analg 2018 Aug;127(2):424-431 [FREE Full text] [doi: 10.1213/ANE.0000000000003482] [Medline: 29916861]
25. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. 1956. Psychol Rev 1994 Apr;101(2):343-352. [Medline: 8022966]
26. Tscholl DW, Handschin L, Rössler J, Weiss M, Spahn DR, Nöthiger CB. It's not you, it's the design-common problems with patient monitoring reported by anesthesiologists: a mixed qualitative and quantitative study. BMC Anesthesiol 2019 May 28;19(1):87 [FREE Full text] [doi: 10.1186/s12871-019-0757-z] [Medline: 31138143]
27. Michard F, Sessler DI. Ward monitoring 3.0. Br J Anaesth 2018 Nov;121(5):999-1001 [FREE Full text] [doi: 10.1016/j.bja.2018.07.032] [Medline: 30336872]
28. Endsley M, Jones WM. Situation Awareness Information Dominance & Information. Wright-Patterson Air Force Base, OH: United States Air Force Armstrong Laboratory; 1997.
29. Schulz CM, Burden A, Posner KL, Mincer SL, Steadman R, Wagner KJ, et al. Frequency and type of situational awareness errors contributing to death and brain damage: a closed claims analysis. Anesthesiology 2017 Dec;127(2):326-337 [FREE Full text] [doi: 10.1097/ALN.0000000000001661] [Medline: 28459735]
30. Schulz CM, Krautheim V, Hackemann A, Kreuzer M, Kochs EF, Wagner KJ. Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. BMC Anesthesiol 2016 Jan 16;16:4 [FREE Full text] [doi: 10.1186/s12871-016-0172-7] [Medline: 26772179]
31. Engelken EJ, Stevens KW, McQueen WJ, Enderle JD. Eye tracking performance variability in a homogeneous population. Biomed Sci Instrum 1995;31:229-233. [Medline: 7654968]
32. Coey C, Wallot S, Richardson M, Orden G. On the structure of measurement noise in eye-tracking. J Eye Movement Res 2012;5(4):e1 [FREE Full text] [doi: 10.16910/jemr.5.4.5]

Abbreviations

IQR: interquartile range

Edited by N Zary; submitted 17.06.19; peer-reviewed by YR Park, R Haase, JJ Mira, I Kedan, C Fernández; comments to author 20.08.19; revised version received 14.11.19; accepted 16.12.19; published 16.03.20.

Please cite as:
Tscholl DW, Rössler J, Handschin L, Seifert B, Spahn DR, Nöthiger CB
The Mechanisms Responsible for Improved Information Transfer in Avatar-Based Patient Monitoring: Multicenter Comparative Eye-Tracking Study
J Med Internet Res 2020;22(3):e15070
URL: http://www.jmir.org/2020/3/e15070/
doi:10.2196/15070
PMID:32175913

Original Paper

Support System for Early Diagnosis of Chronic Obstructive Pulmonary Disease Based on the Service-Oriented Architecture Paradigm and Business Process Management Strategy: Development and Usability Survey Among Patients and Health Care Providers

Alberto De Ramón Fernández[1*], PhD; Daniel Ruiz Fernández[1*], Prof Dr; Diego Marcos-Jorquera[1*], Prof Dr; Virgilio Gilart Iglesias[1*], Prof Dr

Department of Computer Technology, University of Alicante, San Vicente del Raspeig, Alicante, Spain

[*]all authors contributed equally

Corresponding Author:
Daniel Ruiz Fernández, Prof Dr
Department of Computer Technology
University of Alicante
Carretera San Vicente s/n
San Vicente del Raspeig, Alicante, 03690
Spain
Phone: 34 965903400 ext 3331
Email: druiz@dtic.ua.es

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is a chronic respiratory disease with a high global prevalence. The main scientific societies dedicated to the management of this disease have published clinical practice guidelines for quality practice. However, at present, there are important weaknesses in COPD diagnosis criteria that often lead to underdiagnosis or misdiagnosis.

Objective: We sought to develop a new support system for COPD diagnosis. The system was designed to overcome the weaknesses detected in current guidelines with the goals of enabling early diagnosis, and improving the diagnostic accuracy and quality of care provided.

Methods: We first analyzed the main clinical guidelines for COPD to detect weaknesses that exist in the current diagnostic process, and then proposed a redesign based on a business process management (BPM) strategy for its optimization. The BPM system acts as a backbone throughout the process of COPD diagnosis in this proposed approach. The newly developed support system was integrated into a health information system for validation of its use in a hospital environment. The system was qualitatively evaluated by experts (n=12) and patients (n=36).

Results: Among the 12 experts, 10 (83%) positively evaluated our system with respect to increasing the speed for making the diagnosis, helping in interpreting results, and encouraging opportunistic diagnosis. With an overall rating of 4.29 on a 5-point scale, 27/36 (75%) of patients considered that the system was very useful in providing a warning about possible cases of COPD. The overall assessment of the system was 4.53 on a 5-point Likert scale with agreement to extend its use to all primary care centers.

Conclusions: The proposed system provides a functional method to overcome the weaknesses detected in the current diagnostic process for COPD, which can help foster early diagnosis, while improving the diagnostic accuracy and quality of care provided.

(J Med Internet Res 2020;22(3):e17161) doi:10.2196/17161

KEYWORDS

COPD; misdiagnosis; business process management strategy, support system; service-oriented architecture; hospital information system; early diagnosis

Introduction

Chronic Obstructive Pulmonary Disease

Chronic diseases represent the leading cause of mortality worldwide, and are currently responsible for almost 60% of all deaths [1]. These diseases are of long duration and usually of slow progression, significantly affecting quality of life. In particular, cancer, diabetes, and heart and respiratory diseases are chronic diseases with relatively higher prevalence, and chronic obstructive pulmonary disease (COPD) is categorized in the latter group.

The Global Initiative for Chronic Obstructive Lung Disease describes COPD as “a common, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases” [2]. According to the latest Global Health Observatory data, COPD ranked as the third leading cause of death worldwide, responsible for approximately 5% of all deaths globally in 2015 (3.17 million deaths) [3]. The prevalence and burden of COPD are projected to increase over the coming decades due to the gradual aging of the population, cumulative smoking exposure, and an increase in underdiagnosis [4,5].

The main national and international scientific societies dedicated to the management of COPD developed consensus regulations and clinical practice guidelines, including recommendations for quality clinical practice [4,6-10]. However, the processes described in the current guidelines suffer from important weaknesses for each of the different subprocesses, especially in the diagnostic subprocess. Numerous studies have demonstrated that COPD is often misdiagnosed, leading to inappropriate treatments [11,12]. Accordingly, there has been substantial research effort dedicated to improving diagnosis based on the use of clinical decision support systems, mainly involving machine-learning techniques to make an accurate diagnosis [13-15]. However, since none of these initiatives is based on a standardized process, they have only offered partial solutions to resolving this problem.

Since COPD is a chronic disease, constant supervision is required throughout the patient’s life to obtain greater control and avoid possible exacerbations; thus, management of the disease imposes high costs to the health care system and population [7,16-19]. In this context, early diagnosis of COPD is essential to prevent its progression, improve the quality of life of patients, and reduce the economic impact borne by public health systems.

Business Process Management

Fields such as manufacturing and business have long been developing strategies and paradigms related to process optimization for the continuous improvement of processes aimed at customer satisfaction and execution of these processes, resulting in models with demonstrated success. Business process management (BPM) is one of the most recent process management strategies with the greatest impact, which is focused on the continuous improvement of business processes using information technology as one of its fundamental principles for process execution [20]. BPM consists of a set of methodologies and technologies for the identification, modeling, analysis, execution, control, and improvement of business processes. The BPM strategy seeks to achieve flexibility and agility in the evolution and dynamism of business processes and their associated computer systems. Although the BPM strategy was initially applied in the health field for improvement of administrative tasks, in the last decade, such strategies and associated BPM systems (BPMSs) have begun to be used for the partial management of clinical processes [18,21-24] with successful outcomes. A BPMS includes tools for process modeling such as BPM notation (BPMN), which is a standard notation based on flowcharts that facilitate process modeling. This notation has been specifically designed to coordinate the sequence of processes and messages that flow between the participants for different activities [25]. BPMN provides a common language so that all parties involved can communicate the processes clearly, completely, and efficiently. For any BPMS, the BPMN can be directly translated into a format that best supports process execution.

In this study, we developed a support system for the early diagnosis of COPD based on the BPM approach, which was designed to foster early diagnosis, and improve diagnostic accuracy and the quality of care provided. These objectives will also help to indirectly achieve other secondary goals such as shortening diagnostic times, avoiding unnecessary visits to health care centers, creating awareness about the disease, and helping to reduce associated costs.

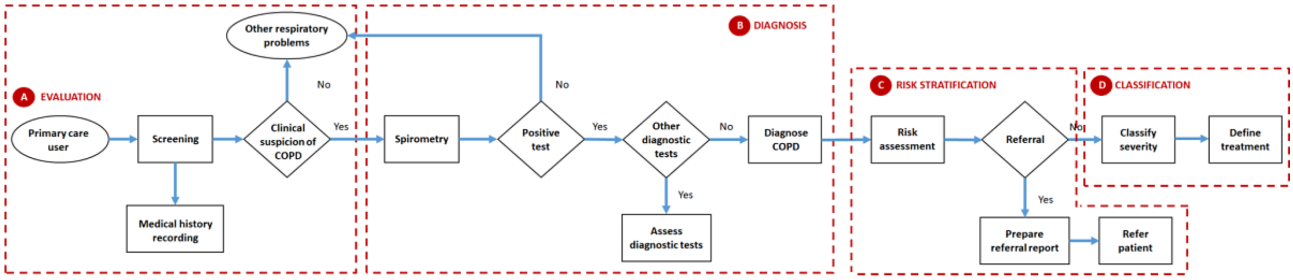
We first analyzed the current diagnostic process to identify its main weaknesses, and developed a redesign for its optimization. We then conducted a questionnaire-based survey in a hospital setting with experts and patients for a qualitative assessment of the architecture of the proposed support system and the developed prototype.

Methods

Current Chronic Obstructive Pulmonary Disease Diagnostic Process Analysis

The consensus regulations and clinical practice guidelines of the main national and international scientific societies include recommendations for quality clinical practice in the management of COPD [4,6-10]. In general, the diagnostic process for COPD is based on 4 steps (Figure 1): evaluation (Figure 1A), diagnosis (Figure 1B), risk stratification (Figure 1C), and classification (Figure 1D).

Figure 1. COPD diagnostic process in a primary care center.



The current process begins in primary care centers when the patient is referred because of respiratory symptoms. The doctor is responsible for carrying out an initial screening to detect key indicators for considering a diagnosis of COPD. Clinical suspicion is recommended for any patient older than 35-40 years who has symptoms such as dyspnea, chronic cough, chronic sputum production, recurrent lower respiratory tract infections, exposure to risk factors, or a family history of COPD.

The doctor can perform a physical examination to detect, for example, edema or anomalous values of blood pressure, or respiratory and cardiac frequency. However, this test will rarely have diagnostic value, since the numerous physical signs of COPD typically do not appear until there is significant lung function impairment. In the case of clinical suspicion, the doctor must prepare a detailed medical history including information such as the patient's exposure to risk factors, medical history of COPD or other chronic respiratory diseases, history of exacerbations, or presence of comorbidities.

To confirm the diagnosis, a postbronchodilator forced spirometry test must be performed. Airflow obstruction is considered to exist if the quotient between the forced expiratory volume in the first second (FEV1) and the forced vital capacity (FVC) is lower than 0.7. In case of doubt, and mainly to rule out alternative diagnoses or establish the presence of comorbidities, the doctor may request other complementary diagnostic tests, which generally include an X-ray, chest tomography, blood analysis, and pulse oximetry. With all this information, the doctor must be able to diagnose the patient a priori. Otherwise, the patient must be referred to hospital care to be further evaluated by a specialist [4].

The next step is to assess the severity of the disease. Since COPD is a heterogeneous condition, no single measure can adequately assess disease severity in an individual. In general, the degree of affectation is estimated by taking into account the airflow limitation (measured through FEV1), number of exacerbations, and degree of dyspnea, which allows for determination of severity among four stages or levels: mild, moderate, severe, and very severe. If the patient is determined to be at an advanced stage of the disease, they must be referred to hospital care for treatment by a pulmonologist. Once diagnosed, the doctor will define a treatment that allows for reducing the symptomatology to improve the patient's quality of life, reduce the frequency of exacerbations, and control disease progression.

Identification of Main Weaknesses in the Diagnostic Process

The diagnosis of COPD in primary care centers has inefficiencies and weaknesses that directly affect the patients' quality of life and raise the economic cost borne by public health systems. In general, the diagnostic process starts when the patient suffering from respiratory problems arrives for a consultation. However, an "opportunistic" search is more profitable in a primary care setting [4,8]; that is, a physician takes the opportunity to assess whether the patient is among the at-risk population when they arrive for a consultation for other reasons. This is partly due to the fact that a large number of patients with COPD can remain asymptomatic until reaching advanced stages of the disease, which increases the ratio of undiagnosed individuals [5].

Standardization of respiratory function tests is necessary to achieve early detection and secondary prevention of the disease, including identification of all affected individuals and assessment of the severity of each patient. However, there are currently numerous nonunified clinical guidelines with differences both in the initial screening and in the assessment of the severity of the disease, resulting in the use of different diagnostic criteria in different countries. Some of these guidelines suggest that any adult with a respiratory symptom is considered to be at risk for COPD [10,26,27], whereas others discard individuals under 35 [7,8] or 40 years [4,6]. Other guidelines subordinate the clinical suspicion to the condition of smoker or ex-smoker [8,9], or to the score obtained in a screening test [28].

Some clinical guidelines establish the stage of the disease based solely on the FEV1, which measures airway obstruction [9,10]. Other guidelines apply multidimensional indices [6,8] that have demonstrated proven risk predictive capacity [29,30] taking into account FEV1 in addition to factors such as dyspnea, number of exacerbations, body mass index, or even the cardiopulmonary response to exercise. However, some authors absolutely reject the use of such factors, arguing that they do not provide better prediction outcomes than FEV1 alone, and require more time to classify and several resources that are not available in all primary care centers [7].

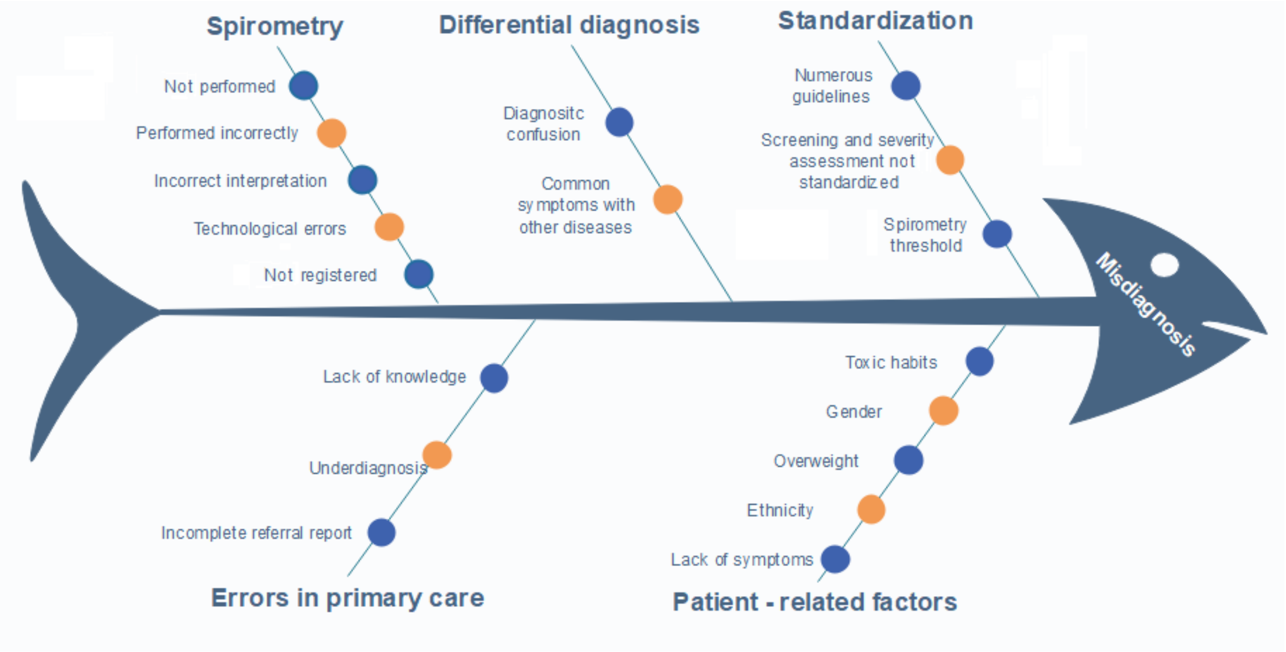
Spirometry is considered to be the most reproducible, standardized, and objective way of measuring airflow limitation; however, this can also be an important weak point in the diagnostic process. In many cases, this test is not performed or the quality of the test is very low (it is performed inefficiently) for different reasons, including lack of an available spirometer, the technique is not known or there is no experienced staff to perform the test, the results are not interpreted correctly, technological errors (calibration errors), errors made by the patient, or the results are not recorded in the medical history. Moreover, the fixed value threshold of 0.7 suggested by the main clinical guidelines can lead to overdiagnosis, and therefore a variable threshold based on the age and sex of the patient has been proposed [31,32].

Misdiagnosis of COPD also occurs due to errors made in primary care. Some studies have shown that professionals in primary care centers make more diagnostic errors than specialists, mainly due to lack of awareness of the disease, which often leads to the criteria for hospital referral to be ignored or the associated reports not including all of the recommended information [33].

A further challenge in COPD diagnosis is the similar symptomatology to other diseases such as asthma, lung cancer, acute coronary syndrome, and ischemic heart disease [34-38]. Finally, misdiagnosis can occur owing to factors related to the characteristics of the patient. For example, women in general and asymptomatic patients are at higher risk of underdiagnosis, as well as overweight or obese patients. In addition, different ethnic groups have different lung capacities that can also lead to misdiagnosis [33].

These inefficiencies result in substantial levels of misdiagnosis, underdiagnosis, and poor diagnostic accuracy. Some studies have shown that primary care physicians have problems in diagnosing up to 19.8% of patients with chronic respiratory symptoms and that the underdiagnosis rate exceeds 70% [12,39]. Figure 2 shows a cause-effect diagram based on Ishikawa et al [40] with a summary of the main weaknesses detected. All of these inefficiencies consequently result in increased costs associated with COPD diagnosis, mainly because the patients must return for several consultations until an accurate diagnosis is obtained, or due to overuse of hospital resources (diagnostic tests). Therefore, an early diagnosis would help to significantly reduce the costs associated with the treatment of a patient with COPD.

Figure 2. Ishikawa diagram of COPD misdiagnosis.

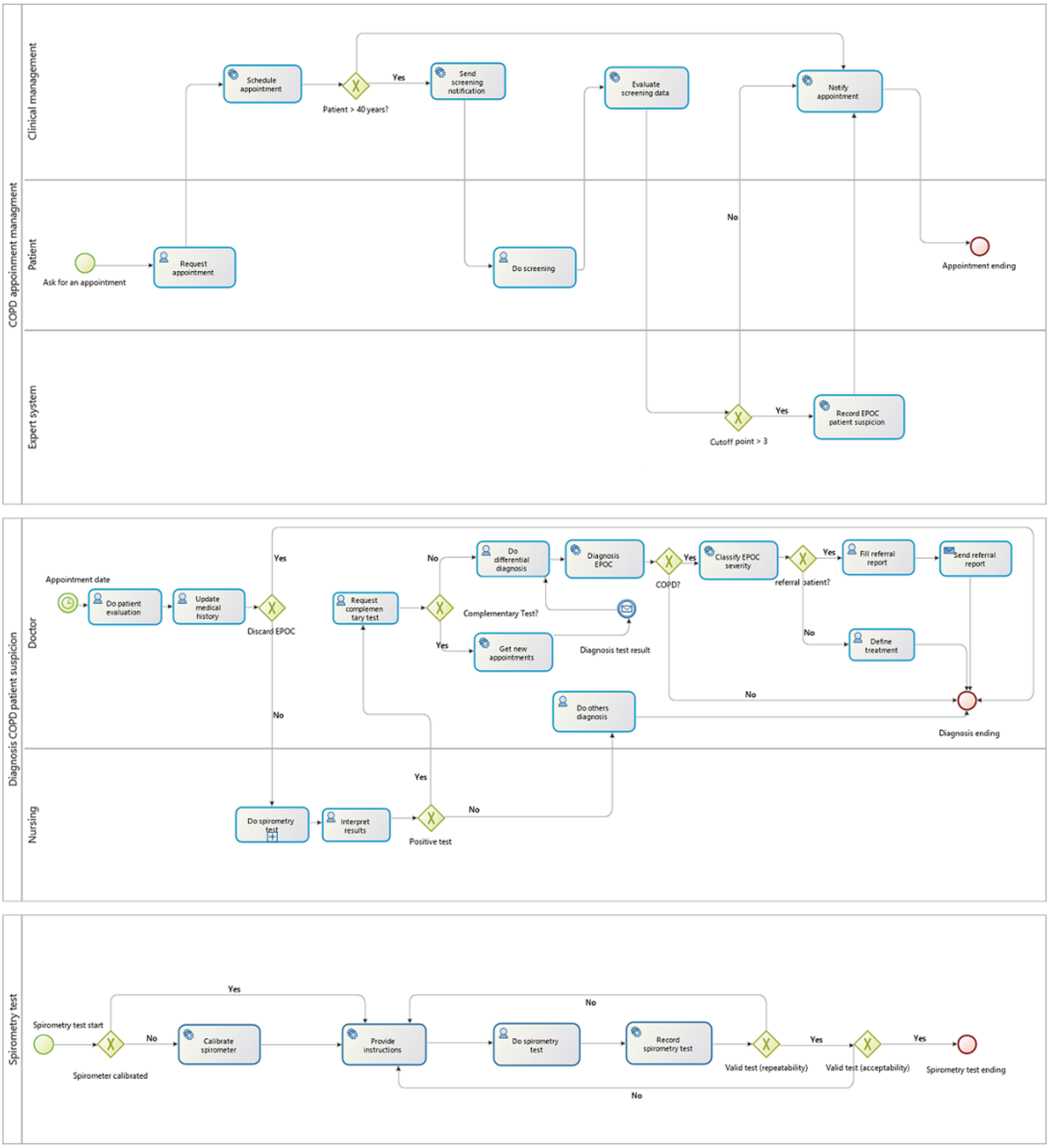


Business Process Management Redesign Process

Based on the detected weaknesses described above, we here propose a redesign of the COPD diagnostic process to facilitate its comprehensive management and optimization. This redesign is based on a BPMS that serves as a backbone throughout the process and allows achievement of the main objectives: foster early diagnosis, improve diagnostic accuracy, and improve the quality of care provided. These objectives will be achieved from process standardization, the traceability of the tasks, notification to the patients, verification of the correct performance and correct interpretation of the spirometry test, diagnosis suggestion, and provision of all of the necessary information to assist the doctor during the diagnosis. The system functionalities are discussed below.

BPMS allows for modeling the process graphically using the BPMN that serves as a reference for standardization of the diagnosis, and to clearly define the tasks to be performed and the responsibility of the actors involved (doctor, nurse, patient, or the health information system [HIS]) in performing them (Figure 3). The modeling of the process allows the BPMS to ensure the traceability of all tasks and appropriate communication in real time to the professionals involved. The BPMS is responsible for controlling the process, guiding the professional, and ensuring that all of the tasks assigned are performed. Another important factor of a BPMS is the recording of the realization time of each task. This aspect is fundamental within a continuous improvement approach since it can help to detect which tasks are slowing/dilating the process or consuming more time, and the current process can be modified without the need to reprogram the entire system.

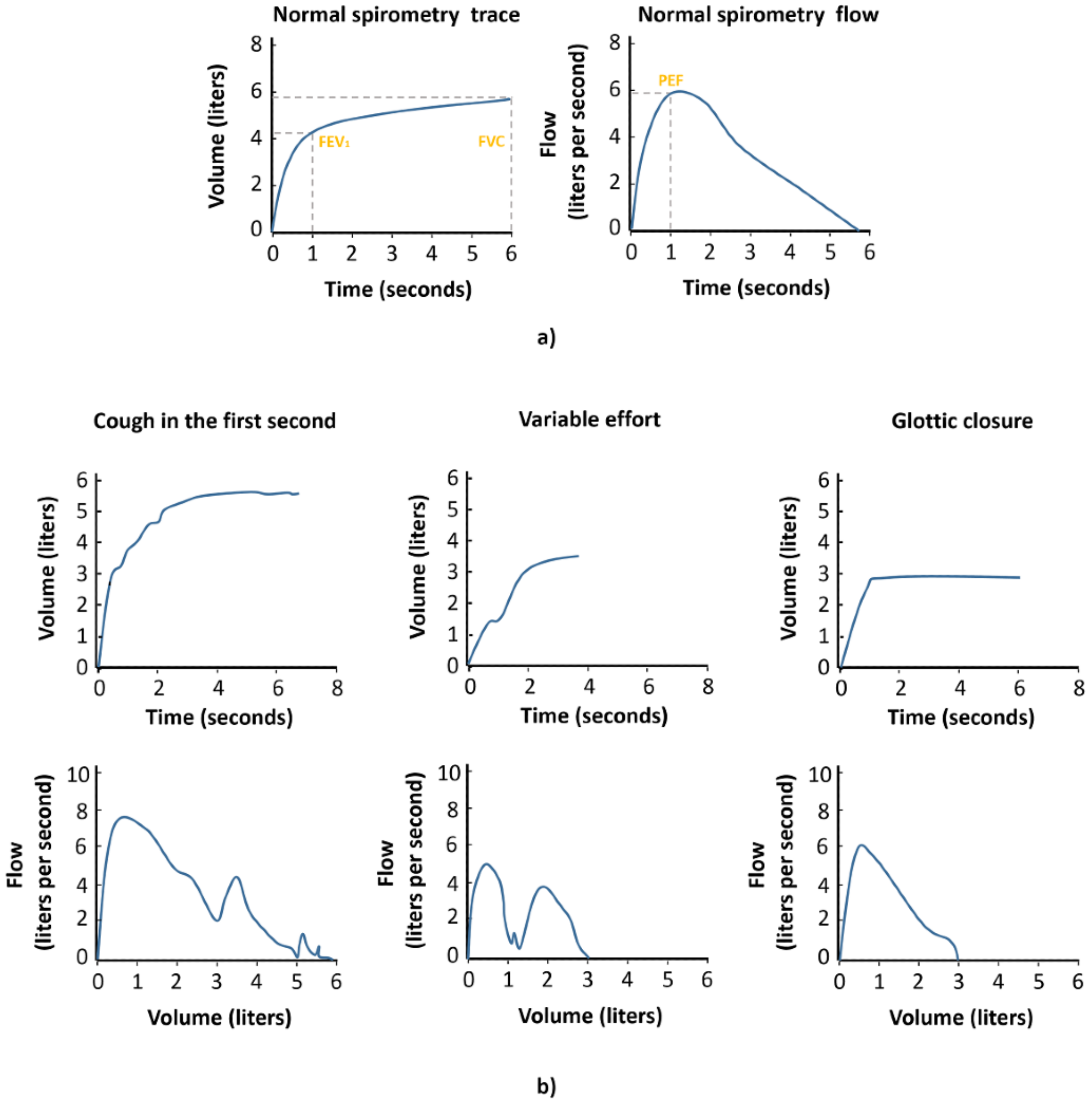
Figure 3. Chronic obstructive pulmonary disease diagnosis clinical process redesign using business process management notation. EPOC: excess postexercise oxygen consumption.



The process in the diagnosis of COPD begins at a preconsultation stage when a patient requests a medical appointment through an HIS. This appointment does not necessarily have to be motivated by a problem relating to COPD. The BPMS can interact with legacy systems, in this case, a hospital management system. If the patient requesting the appointment meets certain indications such as being over 40 years of age, the BPMS captures this information, classifies the patient within the population at risk, and sends a message to the patient so that they perform an initial screening (see Figure 3). This notification will only be sent to patients who have not already carried out this evaluation in the last 2 years. The screening allows for opportunistically detecting patients who are among the population at risk of COPD and, in many cases, in the early stages of the disease. The BPMS is responsible for collecting and evaluating the results of the screening. If the result is positive, both the patient and the doctor are notified of a probable case of COPD and the clinical suspicion is recorded in the system. In this way, the doctor knows the patient's risk situation before the medical appointment. To prevent the patient from forgetting the appointment and improve the absenteeism rates of patients in primary care consultations, which have been estimated between 10% and 15% [38], the system sends a notification with a reminder 24 hours in advance of the appointment.

The process continues in the consultation on the day of the appointment with a more detailed evaluation of the patient by the doctor (symptomatology, risk factors, previous exacerbations, presence of comorbidities, family history, and impact on quality of life), along with an update of the patient's medical history. If a case of COPD is suspected, the spirometry test is performed by the nurse. First, the BPMS is responsible for verifying if the spirometer is calibrated (usually by checking the last calibration date), and otherwise provides information and assistance to the professional for performing the calibration. Before performing the test, the system reproduces an audio file with instructions to the patient for the correct performance of the maneuver. The results are either collected automatically by the system or imputed manually by the professional, depending on the level of automation and integration of the spirometer. The test is supervised at all times by the nurse verifying its acceptability. To do this, the nurse must correctly interpret the spirometric diagrams shown by the spirometer. To assist the professional in this task, a template is shown (see Figure 4) with a normal spirometric volume flow curve and others incorrectly performed for different reasons (eg, slow start, early termination, glottis closure, or variable effort).

Figure 4. (a) Diagrams for normal spirometry; (b) Examples of incorrect spirometries. FEV1: forced expiratory volume (first second); FVC: forced vital capacity.



Next, the system checks whether the test meets the reproducibility criteria. For this step, the difference between the two best FVC and FEV1 values of the three attempts made must be less than or equal to 0.15 L. Otherwise, it is reported that the test is not valid and must be repeated. If both criteria are met, spirometry is validated and recorded, and the test result is analyzed based on the normal lower limit calculated for each patient according to their age, sex, and race. In this work, we have chosen to use this metric since some studies have shown that this approach has better diagnostic accuracy than the use of the fixed limit of 0.7 proposed by the main clinical guidelines [41,42]. If the spirometry result is negative, the patient is ruled out as having COPD and other diagnoses are assessed. In the case of a positive result, the doctor may require additional tests (eg, chest X-ray, blood count, oximetry) to confirm the diagnosis. With the results of the initial screening, medical history, spirometry results, and complementary tests, the doctor must make a differential diagnosis to rule out other respiratory diseases. This is a key step, since COPD may have common symptomatology to other diseases that induce diagnostic errors. At this point, the system provides key information on pathologies with a similar clinical picture to help the doctor make the differential diagnosis of COPD (see Table 1). Once other possible diagnoses have been ruled out, the doctor makes a diagnosis of COPD and proceeds to assess the severity of the disease.

Table 1. Differential diagnoses for respiratory diseases.

Diagnosis	Suggested Features
COPD[a]	Onset in midlife Symptoms slowly progressive History of tobacco smoking or exposure to other types of smoke
Asthma	Onset early in life (often in childhood) Symptoms vary widely from day to day Symptoms worse at night/early morning Allergy, rhinitis, and/or eczema also present Family history of asthma Obesity coexistence
Congestive heart failure	Chest X-ray shows dilated heart, pulmonary edema Pulmonary function tests indicate volume restriction, not airflow limitation
Bronchiectasis	Large volumes of purulent sputum Commonly associated with bacterial infection Chest radiograph/CT[b] scans show bronchial dilatation, bronchial wall thickening
Tuberculosis	Onset at all ages Chest X-ray shows lung infiltrate Microbiological confirmation High local prevalence of tuberculosis
Obliterative bronchiolitis	Onset at younger age, nonsmokers May have history of rheumatoid arthritis or acute fume exposure Seen after lung or bone marrow transplantation CT on expiration shows hypodense areas
Diffuse panbrochiolitis	Predominantly seen in patients of Asian descent Most patients are male and nonsmokers Almost all cases involve chronic sinusitis Chest X-ray and HRCT[c] show diffuse small centrilobular nodular opacities and hyperinflation

[a] COPD: chronic obstructive pulmonary disease.

[b] CT: computed tomography.

[c] HRCT: high-resolution computed tomography.

Classification of the severity of the disease is based on the automatic calculation of the exacerbation of the body mass index, airflow obstruction, dyspnea, and exercise (BODEx) [43] and/or (BODE) [44] indices together with evaluation of the impact of the disease on the patient's quality of life. To do this, together with the FEV1 value obtained in the spirometry, the body mass index, degree of dyspnea, number of exacerbations, and scoring of the 6-minute walk test (if applicable) and the COPD assessment test [39] are recorded.

The system classifies the patient according to four levels of severity (mild, moderate, severe, and very severe) and four levels of impact on their quality of life (low, moderate, high, and very high), and refers the case to a specialist if the degree of affectation is severe or very severe, or if frequent exacerbations are noted. In this case, the doctor must complete the referral report proposed by the system. Otherwise, the doctor defines the most appropriate treatment and ends the process.

Evaluation

The experimentation phase of the proposal was approached from two different but complementary perspectives corresponding to the two types of users involved in the system: the health professionals who are the users and managers of the system, and the patients who are the direct beneficiaries of the execution of the proposed system. It is important to emphasize that both types of users are active users; that is, they are both involved and affected by the execution of the system. Therefore, we consider that the opinion of both types of users is the best result of experimentation that can be obtained, since validation with clinical experience would correspond to a project focused on clinical validation and not on the tool itself as in the present case.

To validate the achievement of the objectives proposed in the design of the system, a Likert-type survey [45] was conducted with 16 questions focused on the objectives associated with the project for the health care professional, and another survey of the same type with 6 targeted questions on the perception of quality of care and empowerment aimed at patients. Each question is associated with a numerical value from 1 to 5, with 1 being “completely disagree” and 5 being “totally agree.” In addition, another open question was incorporated so that the respondents can assess the system qualitatively. It was emphasized to the respondents to focus on evaluating the system comprehensively and not on its separate parts. The questions were designed so as to cover the objectives proposed in this work to provide a tool that can offer an idea of the degree of achievement of the objectives. The issues raised in the questionnaire were related to the main objectives of promoting early diagnosis, improving diagnostic accuracy, and improving quality of care.

Results

System Architecture and Prototype

To validate the proposed redesign for COPD diagnosis, a technical architecture was developed that allows for the design of realistic systems. A prototype of the architecture was developed to demonstrate its suitability. Specific technological tools were selected for the prototype, although these could be replaced by similar tools provided they are in line with the architecture.

Since this is a distributed architecture, in which different and geographically distributed users (patients and professionals), apps, and software modules are identified, we adopted a service-oriented architecture (SOA) that allows for proper integration of these aspects. Specifically, we selected RESTFul type services, which provide characteristics that are fundamental to the proposal such as reusability, scalability, low coupling, interoperability, and security [46].

The designed architecture (Figure 5) incorporates the following components as key points:

- An information system (1 in Figure 5), which includes all of the necessary information for the proposed system. The information system is composed of two elements, a database server and an application server, that were run on a single device in the prototype but could also be implemented on different computers.
- A database server (2 in Figure 5), in which all data necessary for the system will be saved. In the prototype, MariaDB is used as the database manager, which is a fork opensource of the MySQL relational database [47].
- An application server that implements a Representational State Transfer (REST) application programming interface (API) [48] (A in Figure 5) for access to the information system (3 in Figure 5). In the prototype, this server was developed using NodeJS as a platform together with the Express module for the development of HTTP services, and the MySQL module was used for access to the database.
- A BPMS to execute the processes defined in the proposed redesign (4 in Figure 5). In the prototype, Bonita Software Community Edition (version 7.9.4; Bonitasoft, San Francisco, CA, USA) was used that allows access through a REST API (D in Figure 5). Toward this end, several conductors were developed for the different tasks identified in the process (Figure 3) that allow access to other elements of the architecture.
- An electronic health records system (EHRS), in which the developed system is integrated. This is a key aspect of the proposal since it allows validating its use in a real environment (5 in Figure 5). For this purpose, we selected OpenEMR [49], which is a widely used open-source EHRS. This system uses a REST API (B in Figure 5) for access to its information (eg, patients, appointments, treatments). This API does not contain any notification mechanism; thus, the BPMS is not able to know when an appointment is requested by the patient. To solve this problem, at the end of each day, all of the new appointments made are consulted, those meeting the established criteria are analyzed, and, if applicable, a notification is sent to the patient to perform the screening. The following services are used for this process: auth (for authorization), appointment (for access to appointments), and patient (for access to patient data).
- A spirometer (6 in Figure 5) with which the doctor can perform spirometry to the patient in consultation. This may be connected to the system or can operate independently, in which case the doctor must enter the data into the system manually (G in Figure 5). An AirNext model spirometer (NuvoAir AB, Stockholm, Sweden) was used in the prototype.
- A mobile app for the patient to perform the screening on COPD (7 in Figure 5 and Figure 6d). This app is operated by an app server that collects the test results using the defined REST API (C in Figure 5).
- The app that the patient uses to request the initial appointment (8 in Figure 5) that is the trigger of the COPD diagnosis process. In the prototype, the OpenEMR - Patient Portal app was used for this purpose (Figure 6a).
- An app to assist the doctor in the diagnostic process (9 in Figure 5). For the prototype, an app was designed using Bonita Software forms (Figures 6b and 6c) that use the Bonita API (D in Figure 5).
- The app with which the doctor manages the clinical process (10 in Figure 5). The Web app provided by OpenEMR was used in the prototype.

After development of the prototype, its functional validation was conducted, including complete cycles of COPD diagnoses from the request for an appointment by a patient to the final diagnosis by a professional.

Figure 5. Proposed architecture and prototype. COPD: chronic obstructive pulmonary disorder; BPM: business process management; PMS: process management strategy.

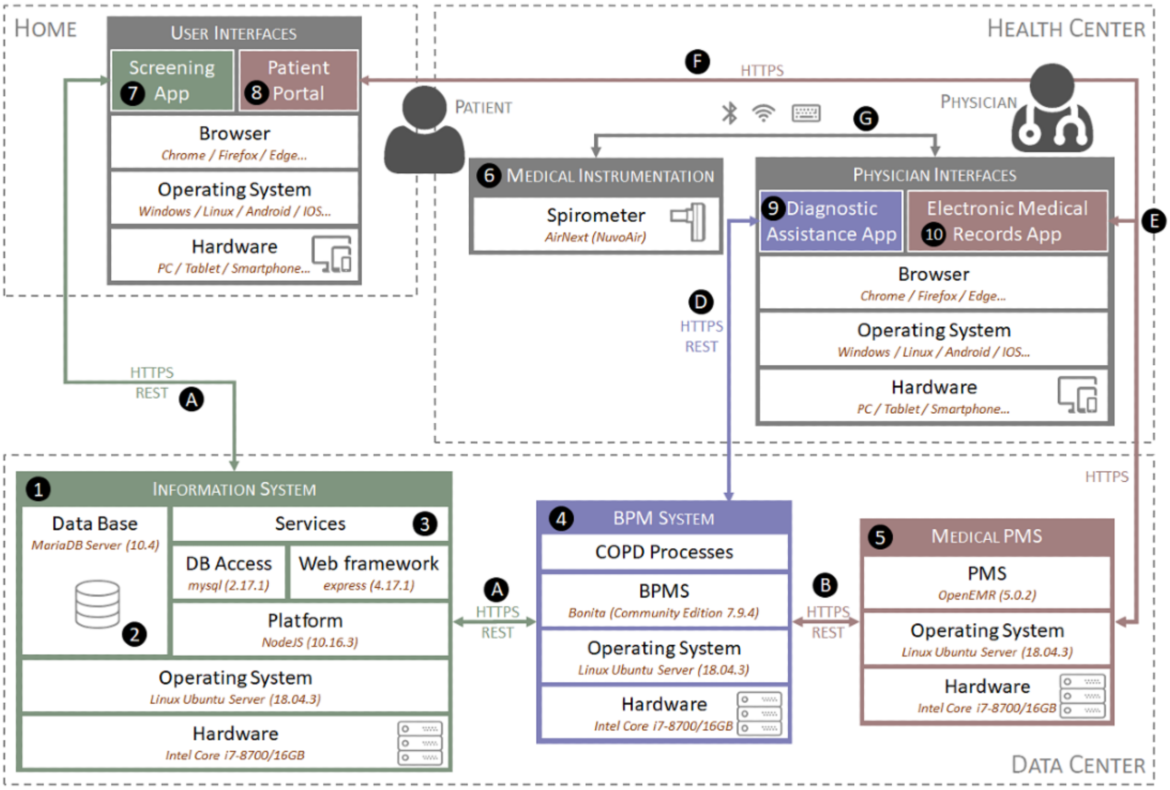
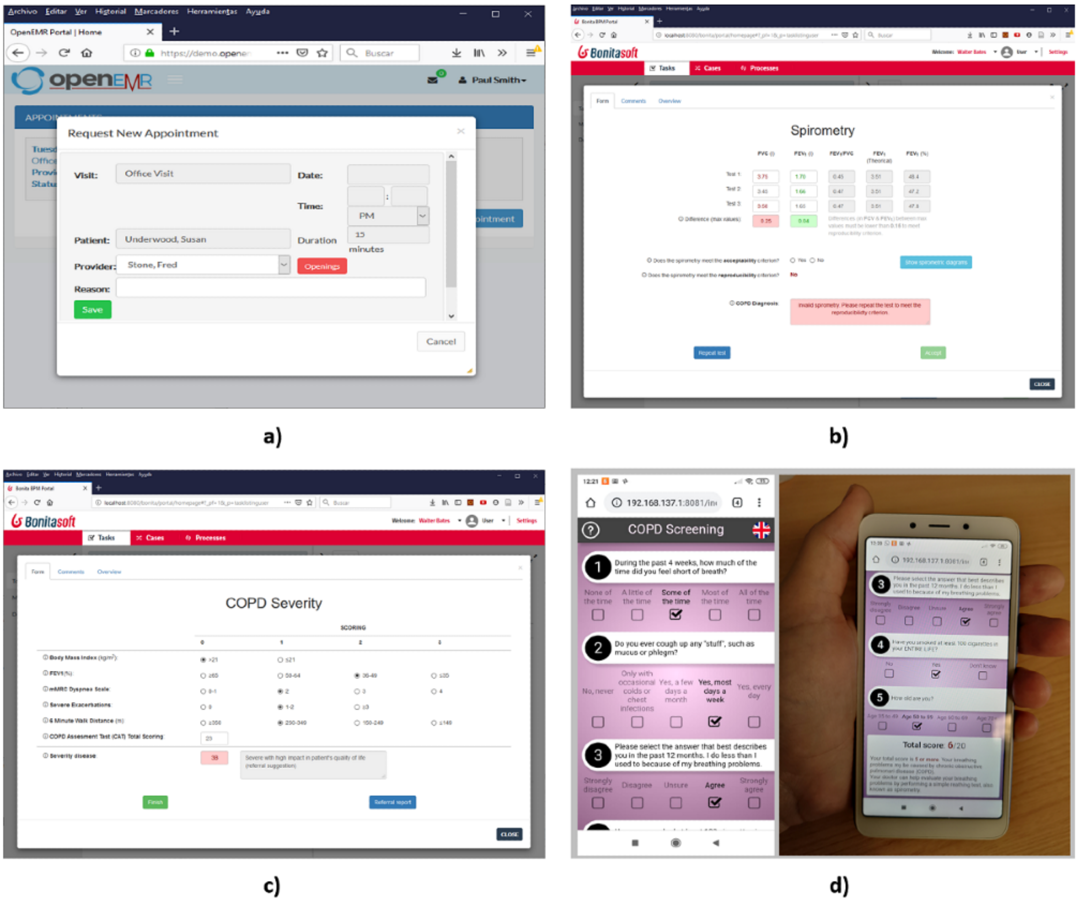


Figure 6. Prototype screenshots. (a) Appointment management from the OpenEMR Patient Portal. (b) Test spirometry interface. (c) Severity assessment interface. (d) Screening app.



Request New Appointment

Visit: Office Visit

Date:

Time: : PM

Patient: Underwood, Susan

Duration 15 minutes

Provider: Stone, Fred

Reason:

a)

Spirometry

b)

COPD Severity

Severity disease: 38

Severe with high impact in patient's quality of life (referral suggestion)

c)

1 During the past 4 weeks, how much of the time did you feel short of breath?

None of the time ☐ A little of the time ☐ Some of the time ☑ Most of the time ☐ All of the time ☐

2 Do you ever cough up any "stuff", such as mucus or phlegm?

No, never ☐ Only with occasional colds or chest infections ☐ Yes, a few days a month ☐ Yes, most days a week ☑ Yes, every day ☐

3 Please select the answer that best describes you in the past 12 months. I do less than I used to because of my breathing problems.

Strongly disagree ☐ Disagree ☐ Unsure ☐ Agree ☑ Strongly agree ☐

Total score: 6/20

d)

Evaluation Outcomes

Table 2 shows a list of the questions asked to the clinical staff, as well as the mean and median obtained for each question from the 5-point Likert scale.

Table 2. Questionnaire and result from the clinical perspective (N=12).

Questions	Mean (SD)	Median
Q1. The system will facilitate increasing the detection of COPD[a] cases by opportunistic diagnosis	4.56 (0.51)	5
Q2. The system will allow improvements in COPD diagnosis	4.01 (0.79)	4
Q3. Thanks to the system, the patient will be aware of their pathology	3.24 (0.77)	4
Q4. The system will help avoid diagnostic confusion	3.09 (0.71)	3
Q5. The system will correctly assist health personnel in the diagnosis of COPD	4.31 (0.49)	4
Q6. The use of the system will help to avoid mistakes made in primary care centers due to lack of specialized personnel	4.47 (0.52)	5
Q7. The system will avoid unnecessary visits by patients with COPD	4.64 (0.49)	5
Q8. The system will help to standardize a protocol for the diagnosis of the disease	2.29 (1.00)	3
Q9. Thanks to the system, the interpretation of the spirometric test results will be improved	4.26 (0.77)	5
Q10. The use of the system will improve the coordination, monitoring, and traceability of the diagnostic process	2.45 (0.98)	3
Q11. The system will help speed up the diagnostic process	4.64 (0.49)	5
Q12. The system will improve the information provided by primary care centers to hospital centers	3.30 (0.49)	3
Q13. The system will help assess alternatives in the diagnosis	3.36 (1.00)	4
Q14. The use of the system will reduce the number of visits due to issues associated with COPD	4.31 (0.49)	4
Q15. The system will help improve spirometry performance	4.64 (0.49)	5
Q16. Overall assessment of the system	4.29 (0.65)	4

[a] COPD: chronic obstructive pulmonary disease.

The questionnaire was completed by 12 medical professionals, including family doctors (n=6) who monitor patients with COPD, emergency doctors (n=2) who care for patients when they have exacerbations, and nursing staff (n=4) who perform diagnostic tests. As shown in Table 2, the overall assessment of the system was generally good, highlighting the issues from Q7, Q11, and Q15 with mean scores above 4.5, indicating that health personnel strongly agree that the system can avoid unnecessary visits by patients with COPD, expedite the diagnostic process, and help improve the performance of spirometry. By contrast, the health personnel did not generally agree that the system would help to standardize a protocol for the diagnosis of the disease or that it would contribute to improving patient adherence to treatments. In this sense, it is important to highlight that these objectives would not be among the main associates of the system. In qualitative assessment, all respondents agreed on the need to use information and communications technologies to improve the quality of care, and this system is an example of this. Overall, 10/12 (83%) of the experts positively evaluated our tool to help speed up the diagnosis, help interpret the results, and also encourage opportunistic diagnosis, recommending its use in health centers, with an overall assessment of the system of 4.29 out of 5.

Table 3 shows the questions provided in the survey to the patients as well as the mean and median of their answers. The survey was completed by 36 patients with suspected COPD between 40 and 65 years of age, including 20 (55%) men and 16 (45%) women. Although this is a small sample, it serves as a proof of concept to present the overall perspective of patients regarding the benefits of the system. In general, the patients had a positive opinion about the use of the system; 27/36 (75%) of the patients considered that the system was very useful to warn about possible cases of COPD (Q1), highlighting their interest in the system being extended to different health centers (Q5). By contrast, patients did not view the system as a key factor in improving their awareness of the disease (Q4), matching with the opinion of the clinical staff on this aspect. The patients provided a global rating of the system of 4.53 out of 5.

Table 3. Questionnaire and results from the patient perspective (N=36).

Questions	Mean (SD)	Median
Q1. The system can help detect the disease	4.42 (0.50)	4
Q2. You feel more confident in the diagnosis thanks to the system	4.28 (0.67)	4
Q3. The use of the system helps improve your quality of life as a patient	4.04 (0.74)	4
Q4. The system helps you to be more aware of your disease	3.16 (0.63)	3
Q5. The use of the system should be extended to all health centers	4.76 (0.42)	5
Q6. Overall assessment of the system	4.53 (0.50)	5

Discussion

Principal Findings

COPD is a chronic respiratory disease that is associated with high morbidity worldwide. The main scientific societies dedicated to the care of patients with COPD have proposed different clinical guidelines over the years to help with its diagnosis and subsequent treatment. However, the COPD diagnostic process continues to present important weaknesses that cause late diagnosis or misdiagnosis. This has a direct impact on patient quality of life and the cost borne by health systems.

In this work, a support system for COPD diagnosis based on the BPM paradigm was developed in order to foster early diagnosis, and to improve diagnostic accuracy and the quality of care provided. The BPM strategy pursues, among other objectives, the optimization and standardization of processes, and allows the integration of human resources and information technology solutions through a BPMS that acts as a backbone throughout the diagnostic process.

To instantiate the redesign model raised for the diagnosis optimization, an architecture based on the SOA paradigm was designed that allows integration with characteristics of low coupling, reusability, scalability, interoperability, and security, typical of this type of architecture. The proposed architecture was designed to assist health care professionals during the diagnosis of COPD through the acquisition of patient information, storage and processing of data, and provision of the necessary clinical information for correct interpretation of the results.

From the definition of the architecture, a prototype was designed for a functional validation of the system. It is important to note that the system developed within an HIS was integrated into the prototype, which validates that the proposal can be integrated into a real hospital environment.

The system was also qualitatively validated by both clinical experts and patients. Overall, 83% of the experts surveyed positively evaluated our tool to help speed up the diagnosis, help interpret the results, and also encourage opportunistic diagnosis, recommending its use in health centers, with an overall assessment of the system of 4.29 on a 5-point scale. In addition, 75% of patients considered that the system was very useful to warn about possible cases of COPD, especially those who presented symptomatology compatible with the disease, and they agreed to extend its use to all primary care centers, with an overall assessment of the system of 4.53 on a 5-point scale.

Limitations

This study has some limitations that must be addressed in the next steps of development. First, the implementation of external systems in a real hospital environment is quite restricted, which requires bureaucratic procedures and authorizations that affect different aspects of implementation. In addition, each hospital has its own hospital management system, and therefore it would be necessary to adapt our system to meet the integration requirements in each case. Second, and as a consequence of the first limitation, although the developed system overcomes the current weaknesses of guidelines for COPD diagnosis, it has not been possible to quantitatively validate how our system will improve the current diagnosis rates.

Conclusions

This study highlights the difficulties that currently exist in the diagnosis of COPD in primary care centers. From a functional point of view, the proposed system can help to overcome the weaknesses detected in the current diagnostic process through integration of a mobile app so that patients can refer their symptoms, a spirometer to measure the patient’s lung capacity, and a Web app for physicians that allows them to consult all of the information provided by both the patients and the system. Implementation of this system is expected to help foster the early diagnosis, and improve the diagnostic accuracy and quality of care provided for COPD.

Acknowledgments

We thank Dr Belén Sánchez Ruiz from Ciudad Jardín Health Center in Alicante (Spain) and Dr Ana Fernández Frías from Hospital of Torrevieja (Spain) for their support in this research. AR received financial support from the Ministry of Economy and Competitiveness (BES-2015-073611).

Conflicts of Interest

None declared.

References

1. World Health Organization. Chronic diseases and health promotion URL: https://www.who.int/chp/about/integrated_cd/en/ [accessed 2019-10-10]
2. Fabbri LM, Hurd SS. Global Strategy for the Diagnosis, Management and Prevention of COPD: 2003 update. Eur Respir J 2003 Jul;22(1):1-2 [FREE Full text] [doi: 10.1183/09031936.03.00063703] [Medline: 12882441]
3. World Health Organization. Chronic obstructive pulmonary disease (COPD) URL: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) [accessed 2020-02-18]
4. Parra Hidalgo P, de la Torre Álvaro JL, Sanz Mateo GR. Recommendations for care of patients with chronic obstructive pulmonary disease. Atencion Primaria 2015 Oct;28(7):491-500.
5. Miravitlles M, Soriano JB, García-Río F, Muñoz L, Duran-Tauleria E, Sanchez G, et al. Prevalence of COPD in Spain: impact of undiagnosed COPD on quality of life and daily life activities. Thorax 2009 Oct 23;64(10):863-868. [doi: 10.1136/thx.2009.115725] [Medline: 19553233]
6. Pauwels R, Buist A, Calverley P, Jenkins C, Hurd S, GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. Am J Respir Crit Care Med 2001 Apr;163(5):1256-1276. [doi: 10.1164/ajrccm.163.5.2101039] [Medline: 11316667]
7. NICE Guideline Updates Team. Chronic obstructive pulmonary disease in over 16s: diagnosis and management. In: National Institute for Health and Care Excellence: Clinical Guidelines. UK: National Institutes for Health and Care Excellence; Dec 5, 2018:1-65.
8. Ministerio de Sanidad, Servicios Sociales e Igualdad. 2014. Actualización de la Estrategia en EPOC del Sistema Nacional de Salud Informe URL: http://www.msssi.gob.es/organizacion/sns/planCalidadSNS/pdf/excelencia/Resumen_Evaluacion.pdf [accessed 2020-02-18]
9. Peces-Barba G, Albert Barberà J, Agustí A, Casanova C, Casas A, Luis Izquierdo J, et al. Guía clínica SEPAR-ALAT de diagnóstico y tratamiento de la EPOC. Arch Bronconeumol 2008 May;44(5):271-281. [doi: 10.1157/13119943]
10. Celli BR, MacNee W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J 2004 Jun;23(6):932-946 [FREE Full text] [doi: 10.1183/09031936.04.00014304] [Medline: 15219010]
11. Llauger Roselló MA, Pou MA, Domínguez L, Freixas M, Valverde P, Valero C, Grup Emergent de Recerca en Malalties Respiratòries (Unitat de Suport a la Recerca de Barcelona-IDIAP Jordi Gol). Treating COPD in chronic patients in a primary-care setting. Arch Bronconeumol 2011 Nov;47(11):561-570 [FREE Full text] [doi: 10.1016/j.arbres.2011.10.001] [Medline: 22036593]
12. Miravitlles M, Andreu I, Romero Y, Sitjar S, Altés A, Anton E. Difficulties in differential diagnosis of COPD and asthma in primary care. Br J Gen Pract 2012 Feb 01;62(595):e68-e75. [doi: 10.3399/bjgp12x625111]
13. Anakal S, Sandhya P. Clinical decision support system for chronic obstructive pulmonary disease using machine learning techniques. 2017 Dec 15 Presented at: International Conference on Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT); December 15-16, 2017; Mysuru, India.
14. Velickovski F, Ceccaroni L, Roca J, Burgos F, Galdiz JB, Marina N, et al. Clinical Decision Support Systems (CDSS) for preventive management of COPD patients. J Transl Med 2014;12(Suppl 2):S9. [doi: 10.1186/1479-5876-12-s2-s9]
15. Mudura V, Frosini F, Iadanza E. Clinical decision support systems for COPD: A general overview. In: Eskola H, Väisänen O, Viik J, Hyttinen J, editors. EMEC 2017, NBC 2017, IFMBE Proceedings. Singapore: Springer; Jun 13, 2017:936-939.
16. Hilleman DE, Dewan N, Malesker M, Friedman M. Pharmacoeconomic evaluation of COPD. Chest 2000 Nov;118(5):1278-1285. [doi: 10.1378/chest.118.5.1278] [Medline: 11083675]
17. Guarascio AJ, Ray SM, Finch CK, Self TH. The clinical and economic burden of chronic obstructive pulmonary disease in the USA. Clinicoecon Outcomes Res 2013;5:235-245 [FREE Full text] [doi: 10.2147/CEOR.S34321] [Medline: 23818799]
18. Stevovic J, Li J, Nezhad HRM, Casati F, Armellin G. Business process management enabled compliance-aware medical record sharing. IJBPIM 2013;6(3):201-223. [doi: 10.1504/ijbpim.2013.056961]
19. Andersson F, Borg S, Jansson SA, Jonsson AC, Ericsson A, Prütz C, et al. The costs of exacerbations in chronic obstructive pulmonary disease (COPD). Respir Med 2002 Sep;96(9):700-708 [FREE Full text] [doi: 10.1053/rmed.2002.1334] [Medline: 12243316]
20. Smith H, Fingar P. Business Process Management: The Third Wave. Tampa, FL: Meghan-Kiffer Press; 2003.
21. Barbagallo S, Corradi L, de Ville de Goyet J, Iannucci M, Porro I, Rosso N, et al. Optimization and planning of operating theatre activities: an original definition of pathways and process modeling. BMC Med Inform Decis Mak 2015 May 17;15:38 [FREE Full text] [doi: 10.1186/s12911-015-0161-7] [Medline: 25982033]

22. Ajmi I, Zgaya H, Gammoudi L, Hammadi S, Martinot A, Beuscart R, et al. Mapping patient path in the Pediatric Emergency Department: A workflow model driven approach. J Biomed Inform 2015 Apr;54:315-328 [FREE Full text] [doi: 10.1016/j.jbi.2014.12.004] [Medline: 25554685]
23. De Ramón Fernández A, Marcos Jorquera D, Soriano Payá A, Gilart Iglesias V, Ruiz Fernández D, Ramirez Navarro J. Business process management for the Crohn's disease clinical process. In: Lecture Notes in Computer Science, vol 10069. 2016 Nov 2 Presented at: Ubiquitous Computing and Ambient Intelligence. UCAmI 2016; November 2, 2016; Gran Canaria, Spain p. 74-79. [doi: 10.1007/978-3-319-48746-5_8]
24. Moon J, Kim D. Context-Aware Business Process Management for Personalized Healthcare Services. 2013 Oct 31 Presented at: 2013 IEEE International Conference on Services Computing; 2013; Santa Clara, USA p. 757-758. [doi: 10.1109/SCC.2013.88]
25. Shapiro R. Bpmn 2.0 Handbook Second Edition. Florida, USA: Future Strategies; 2011.
26. Qaseem A, Wilt TJ, Weinberger SE, Hanania NA, Criner G, van der Molen T, American College of Physicians, American College of Chest Physicians, American Thoracic Society, European Respiratory Society. Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. Ann Intern Med 2011 Aug 02;155(3):179-191. [doi: 10.7326/0003-4819-155-3-201108020-00008] [Medline: 21810710]
27. McKenzie DK, Frith PA. The COPDX Plan: Australian and New Zealand Guidelines for the management of Chronic Obstructive Pulmonary Disease 2003. Med J Australia 2003 Mar 17;178(S6):S1-S39. [doi: 10.5694/j.1326-5377.2003.tb05213.x]
28. Martinez FJ, Raczek AE, Seifer FD, Conoscenti CS, Curtice TG, D'Eletto T, COPD-PS Clinician Working Group. Development and initial validation of a self-scored COPD Population Screener Questionnaire (COPD-PS). COPD 2008 Apr 02;5(2):85-95 [FREE Full text] [doi: 10.1080/15412550801940721] [Medline: 18415807]
29. Ong K, Earnest A, Lu S. A multidimensional grading system (BODE index) as predictor of hospitalization for COPD. Chest 2005 Dec;128(6):3810-3816. [doi: 10.1378/chest.128.6.3810] [Medline: 16354849]
30. Celli BR, Cote CG, Marin JM, Casanova C, Montes de Oca M, Mendez RA, et al. The Body-Mass Index, Airflow Obstruction, Dyspnea, and Exercise Capacity Index in Chronic Obstructive Pulmonary Disease. N Engl J Med 2004 Mar 04;350(10):1005-1012. [doi: 10.1056/nejmoa021322]
31. Hnizdo E, Glindmeyer HW, Petsonk EL, Enright P, Buist AS. Case definitions for chronic obstructive pulmonary disease. COPD 2006 Jun 02;3(2):95-100. [doi: 10.1080/15412550600651552] [Medline: 17175672]
32. Aggarwal AN, Gupta D, Agarwal R, Jindal SK. Comparison of the Lower Confidence Limit to the Fixed-Percentage Method for Assessing Airway Obstruction in Routine Clinical Practice. Resp Care 2011 Nov 01;56(11):1778-1784. [doi: 10.4187/respcare.01160]
33. Hangaard S, Helle T, Nielsen C, Hejlesen O. Causes of misdiagnosis of chronic obstructive pulmonary disease: A systematic scoping review. Respir Med 2017 Aug;129:63-84 [FREE Full text] [doi: 10.1016/j.rmed.2017.05.015] [Medline: 28732838]
34. Tinkelman D, Price D, Nordyke R, Halbert R. Misdiagnosis of COPD and asthma in primary care patients 40 years of age and over. J Asthma 2006;43(1):75-80. [doi: 10.1080/02770900500448738] [Medline: 16448970]
35. Minasian AG, van den Elshout FJ, Dekhuijzen PR, Vos PJ, Willems FF, van den Bergh PJ, et al. COPD in chronic heart failure: less common than previously thought? Heart Lung 2013 Sep;42(5):365-371 [FREE Full text] [doi: 10.1016/j.hrtlng.2013.07.002] [Medline: 23998385]
36. Jagana R, Bartter T, Joshi M. Delay in diagnosis of chronic obstructive pulmonary disease: reasons and solutions. Curr Opin Pulm Med 2015 Mar;21(2):121-126. [doi: 10.1097/MCP.0000000000000133] [Medline: 25575363]
37. Hawkins NM, Petrie MC, Jhund PS, Chalmers GW, Dunn FG, McMurray JJ. Heart failure and chronic obstructive pulmonary disease: diagnostic pitfalls and epidemiology. Eur J Heart Fail 2009 Feb 28;11(2):130-139. [doi: 10.1093/eurjhf/hfn013] [Medline: 19168510]
38. Fromer L. Diagnosing and treating COPD: understanding the challenges and finding solutions. Int J Gen Med 2011;4:729-739. [doi: 10.2147/IJGM.S21387] [Medline: 22114517]
39. Llauger Roselló MA, Pou MA, Domínguez L, Freixas M, Valverde P, Valero C, Grup Emergent de Recerca en Malalties Respiratòries (Unitat de Suport a la Recerca de Barcelona-IDIAP Jordi Gol). Treating COPD in chronic patients in a primary-care setting. Arch Bronconeumol 2011 Nov;47(11):561-570 [FREE Full text] [doi: 10.1016/j.arbres.2011.10.001] [Medline: 22036593]
40. Voelkel JG. Guide to Quality Control. Technometrics 1989 May;31(2):260-260. [doi: 10.1080/00401706.1989.10488523]
41. Wang Y, Xiao W, Ma DD, Jiang YY. Predicted lower limit of normal reduces misclassification risk of airflow limitation in asymptomatic elderly never-smokers. Chin Med J (Engl) 2013;126(18):3486-3492. [Medline: 24034095]
42. Hwang YI, Kim CH, Kang HR, Shin T, Park SM, Jang SH, et al. Comparison of the prevalence of chronic obstructive pulmonary disease diagnosed by lower limit of normal and fixed ratio criteria. J Korean Med Sci 2009 Aug;24(4):621-626. [doi: 10.3346/jkms.2009.24.4.621] [Medline: 19654942]
43. Soler-Cataluña JJ, Martínez-García MA, Sánchez LS, Tordera MP, Sánchez PR. Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients. Respir Med 2009 May;103(5):692-699 [FREE Full text] [doi: 10.1016/j.rmed.2008.12.005] [Medline: 19131231]

44. Celli BR, Cote CG, Marin JM, Casanova C, Montes DOM, Mendez RA, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med 2004 Mar 04;350(10):1005-1012. [doi: 10.1056/NEJMoa021322] [Medline: 14999112]
45. Beglar D, Nemoto T. Developing Likert-scale questionnaires. In: JALT2013 Conference Proceedings. 2013 Presented at: JALT2013; 2013; Tokyo, Japan p. 1-8.
46. Erl T, Carlyle B, Cesare P, Balasubramanian R. SOA with REST: principles, patterns & constraints for building enterprise solutions with REST. In: ACM SIGSOFT Software Engineering Notes. New York, USA: Association for Computing Machinery; 2013:32-33.
47. MariaDB. URL: https://mariadb.org/ [accessed 2020-02-18]
48. Masse M. REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces. Sebastopol, CA: O'Reilly Media; 2011.
49. OpenEMR. URL: https://www.open-emr.org/ [accessed 2020-02-18]

Abbreviations

API: application programming interface
BODE: body mass index, obstruction, dyspnea, and exercise index
BODEx: exacerbation of body mass index, obstruction, dyspnea, and exercise index
BPM: business process management
BPMN: business process management notation
BPMS: business process management system
COPD: chronic obstructive pulmonary disease
EHRS: electronic health records system
FEV1: forced expiratory volume (first second)
FVC: forced vital capacity
HIS: health information system
REST: Representational State Transfer
SOA: service-oriented architecture

Edited by G Eysenbach; submitted 25.11.19; peer-reviewed by P Sánchez González, M Maximiano; comments to author 23.12.19; revised version received 06.01.20; accepted 26.01.20; published 17.03.20.

Please cite as:
De Ramón Fernández A, Ruiz Fernández D, Marcos-Jorquera D, Gilart Iglesias V
Support System for Early Diagnosis of Chronic Obstructive Pulmonary Disease Based on the Service-Oriented Architecture Paradigm and Business Process Management Strategy: Development and Usability Survey Among Patients and Health Care Providers
J Med Internet Res 2020;22(3):e17161
URL: https://www.jmir.org/2020/3/e17161
doi:10.2196/17161
PMID:32181744

Original Paper

Postoperative Remote Automated Monitoring and Virtual Hospital-to-Home Care System Following Cardiac and Major Vascular Surgery: User Testing Study

Michael McGillion[1,2], RN, PhD; Carley Ouellette[3], RN, BSc, MSc; Amber Good[2], MSc; Marissa Bird[1], RN, BA, BSN; Shaunattonie Henry[1], MN, MA Ed; Wendy Clyne[4], PhD; Andrew Turner[5], PhD; Paul Ritvo[6], PhD, CPsych; Sarah Ritvo[6], MA; Nazari Dvirnik[2,3], BHSc, MD, MS; Andre Lamy[2,7], MD, MSHc, FRCSC; Richard Whitlock[2,7], MD, PhD, FRCSC; Christopher Lawton[8], BMET; Jake Walsh[8], MCSE; Ken Paterson[8], BMET; Janine Duquette[9], BScN, MN, ACNP, DNP; Karla Sanchez Medeiros[2], BSc; Fadi Elias[2,3], MD, MS; Ted Scott[8], MAppSC, PhD; Joseph Mills[10], BChir, MD, MA, MB, FRCP; Deborah Harrington[10], FRCS; Mark Field[10], FRCS; Prathiba Harsha[3], MBBS, MSc; Stephen Yang[11], MD, CM, FRCPC; Elizabeth Peter[12], RN, PhD; Sanjeev Bhavnani[13], MD; PJ Devereaux[2,14], MD, PhD, FRCPC

[1]School of Nursing, McMaster University, Hamilton, ON, Canada
[2]Population Health Research Institute, Hamilton, ON, Canada
[3]McMaster University, Hamilton, ON, Canada
[4]Hope for the Community CIC, Coventry, United Kingdom
[5]Coventry University, Coventry, United Kingdom
[6]York University, Toronto, ON, Canada
[7]Department of Surgery, McMaster University, Hamilton, ON, Canada
[8]Hamilton Health Sciences, Hamilton, ON, Canada
[9]Cardiac and Vascular Program, Hamilton Health Sciences, Hamilton, ON, Canada
[10]Liverpool Heart and Chest Hospital, Liverpool, United Kingdom
[11]McGill University, Montreal, ON, Canada
[12]Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada
[13]Scripps Clinic & Research Foundation, San Diego, CA, United States
[14]Departments of Health Research Methods, Evidence, and Impact (HEI) and Medicine, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Michael McGillion, RN, PhD
School of Nursing
McMaster University
1280 Main Street West, HSC 2J40E
Hamilton, ON, L8S 4K1
Canada
Phone: 1 9055259140 ext 20275
Email: mmcgill@mcmaster.ca

Abstract

Background: Cardiac and major vascular surgeries are common surgical procedures associated with high rates of postsurgical complications and related hospital readmission. In-hospital remote automated monitoring (RAM) and virtual hospital-to-home patient care systems have major potential to improve patient outcomes following cardiac and major vascular surgery. However, the science of deploying and evaluating these systems is complex and subject to risk of implementation failure.

Objective: As a precursor to a randomized controlled trial (RCT), this user testing study aimed to examine user performance and acceptance of a RAM and virtual hospital-to-home care intervention, using Philip's Guardian and Electronic Transition to Ambulatory Care (eTrAC) technologies, respectively.

Methods: Nurses and patients participated in systems training and individual case-based user testing at two participating sites in Canada and the United Kingdom. Participants were video recorded and asked to think aloud while completing required user tasks and while being rated on user performance. Feedback was also solicited about the user experience, including user satisfaction and acceptance, through use of the Net Promoter Scale (NPS) survey and debrief interviews.

Results: A total of 37 participants (26 nurses and 11 patients) completed user testing. The majority of nurse and patient participants were able to complete most required tasks independently, demonstrating comprehension and retention of required Guardian and eTrAC system workflows. Tasks which required additional prompting by the facilitator, for some, were related to the use of system features that enable continuous transmission of patient vital signs (eg, pairing wireless sensors to the patient) and assigning remote patient monitoring protocols. NPS scores by user group (nurses using Guardian: mean 8.8, SD 0.89; nurses using eTrAC: mean 7.7, SD 1.4; patients using eTrAC: mean 9.2, SD 0.75), overall NPS scores, and participant debrief interviews indicated nurse and patient satisfaction and acceptance of the Guardian and eTrAC systems. Both user groups stressed the need for additional opportunities to practice in order to become comfortable and proficient in the use of these systems.

Conclusions: User testing indicated a high degree of user acceptance of Philips’ Guardian and eTrAC systems among nurses and patients. Key insights were provided that informed refinement of clinical workflow training and systems implementation. These results were used to optimize workflows before the launch of an international RCT of in-hospital RAM and virtual hospital-to-home care for patients undergoing cardiac and major vascular surgery.

(J Med Internet Res 2020;22(3):e15548) doi:10.2196/15548

KEYWORDS

monitoring, physiologic; postoperative care; user testing

Introduction

Background

Cardiac and major vascular surgeries are common surgical procedures associated with high rates of postsurgical complications and related hospital readmission [1,2]. A North American prospective cohort study (involving 5158 patients) by the National Institutes of Health and Canadian Institutes of Health Research Cardiothoracic Surgical Trials Network found that 18.7% of cardiac surgery patients were readmitted within 60 days [3]. The most common drivers of first readmission included infection, arrhythmia, and fluid volume overload [3]. Data from the US 2014 registry (N=11,246) reported comparable rates of unplanned 30-day readmission among major vascular surgery patients, for example, infrainguinal bypass: 15.7% [4]. Recent retrospective data from Boston Medical Centre, reporting on patients (N=649) having a range of major vascular surgeries (eg, endovascular lower extremity procedures and carotid or cerebrovascular procedures), demonstrated that 21% of patients were readmitted within 30 days. Postsurgical complications accounted for 35.5% of these readmissions, driven most commonly by surgical site infections, graft failures, and bleeding [5].

A factor contributing to high postsurgical complications and readmission rates following cardiac and major vascular surgeries is inadequacy of current systems for patient monitoring in hospital and at home [6-10]. Routine nursing surveillance of patients on hospital surgical wards includes manual vital signs assessments every 4 to 12 hours [6,10]. On the basis of such infrequent vital signs measurements, extrapolations are made about the stability of patients’ physiologic status for extended periods of time [11]. As a result of these practices, the incidences of patient hemodynamic compromise and instability are often missed, as are opportunities to facilitate timely clinician response and early intervention [8-12]. In a study examining the incidence of postoperative hypotension, Turan et al [13] found that 18% of patients on surgical wards had an episode of mean arterial pressure <65 mm Hg for a minimum of 15 min. When taking routine, manual vital signs observations every 4 hours, nurses missed approximately half of all these episodes.

The problem is further compounded once patients are discharged home without surveillance or health professional support—a significant number of patients sustain complications that their surgical teams are unaware of. A prospective study (N=328) in the United Kingdom found that 28% of cardiac surgery patients required urgent physician or district nurse intervention within the first 6 weeks of recovery at home [6]. Of these patients, 21% required hospital readmission because of major complications including cardiac arrhythmia, pneumonia, renal failure, or sternal wound infections. Patients’ respective surgical care teams were unaware of any such complications requiring treatment [6].

Increasing efforts are being made to implement postoperative remote automated monitoring (RAM) and surveillance systems to improve patient outcomes through facilitation of continuous patient monitoring, early detection of deterioration, and remote patient support [9,11]. Sophisticated RAM and virtual care deployments include wireless sensors worn by the patient and supported by network infrastructure to acquire, transmit, and integrate continuous physiologic data [9,11]. Synthesis of this information is typically driven by the hospital early warning systems that direct action of frontline nursing staff, including escalation of care to the most responsible physician or rapid response team [9].

RAM surveillance systems have major potential, but the science of implementing and evaluating these systems is complex and still at an early stage [9]. Although a number of recent studies report that nurses and physicians support the need for RAM technologies on surgical wards and into the home setting, there are conflicting views which convey both excitement and apprehension about the consequences of such systems on clinical workflows and outcomes, as well as the experience of patient care [14-17]. In a study to solicit clinical staff perspectives of the introduction of RAM on general and surgical wards, Prgomet et al [14] found that while RAM technologies were viewed by nurses and physicians as potentially advantageous to the identification of early patient deterioration, a number of concerns were raised about possible drawbacks. Clinicians expressed worry that RAM technologies would decrease meaningful patient contact, reduce flexibility about the use of personal clinical judgement, and result in unwanted patient anxiety and discomfort [14]. Other studies have supported similar results, stressing the need for targeted training, educational opportunities, and pilot user testing studies to allow clinicians and patients to get familiar and comfortable with RAM systems and provide feedback [15,17].

Objectives

Our team is conducting a randomized controlled trial (RCT) of a RAM and virtual hospital-to-home intervention entitled, TecHnology-Enabled remote monitoring and Self-MAnagemenT—VIsion for patient EmpoWerment following Cardiac and major VasculaR surgery (THE SMArTVIEW, CoVeRed) [9,18,19]. The SMArTVIEW intervention combines RAM and virtual hospital-to-home support using Philips monitoring technologies.

The current RCT (N=800) [1,9] will examine the impact of SMArTVIEW on patients aged 65 years and older at two hospital sites (Canada and the United Kingdom) on an array of clinical and feasibility outcomes, including a composite of 45-day hospital readmission and emergency department and urgent care center visits; postoperative complications; patient-reported outcomes; and intervention adherence [9,18,19].

Recognizing that implementation of RAM represents a change to typical postoperative care [9,14-17], the purpose of this study was to examine user performance and acceptance of RAM and virtual care technologies planned for use in the SMArTVIEW trial [9,18,19].

Methods

The TecHnology-Enabled Remote Monitoring and Self-MAnagemenT—VIsion for Patient EmpoWerment Remote Automated Monitoring and Virtual Care Intervention

This user testing study (and the subsequent SMArTVIEW trial) was in response to a call for applications to the Canadian Institutes of Health Research eHealth Innovation Partnerships program. This funding opportunity was designed to facilitate experimental, *real-world*, large-scale implementations, focusing on the integration of existing innovations that are beyond the prototyping stage and ready for deployment in the real-world conditions. To this end, we extended an invitation to vendors, through various hardware and software consortiums in North America, to showcase their market-ready solutions that could support our RAM and virtual care needs. Vendor choice was based on ability to contribute equipment and personnel time for training study staff, as well as the availability of products that were at a minimum *technology readiness level* 7 to 9 (0=early prototyping, 9=ready for use under operational conditions), according to Innovative Solutions Canada.

Our aim was to implement an end-to-end solution that incorporates both RAM in hospital and virtual hospital-to-home recovery support for the first 30 days at home, following cardiac and major vascular surgery. Although multiple vendors came forward with various solutions, Philips was in a position to provide market-ready, configurable technology solutions for both hospital RAM and hospital-to-home virtual care that could be packaged together in an *end-to-end* solution for immediate deployment.

On the surgical ward, RAM is supported by the Philips Guardian solution as illustrated in Figure 1 [1,9,18]. Guardian includes a central trending monitor at the nursing station; a bedside, portable spot-check vital signs monitor (MP5); and three wireless wearable patient sensors, which communicate with the MP5 bedside monitor and the Guardian central monitor via short-range radio and Wi-Fi, respectively. A wireless sensor applied to the index finger and wrist monitors continuous blood oxygen saturation (SpO_2) and pulse rate, an inflatable cuff module measures noninvasive blood pressure, and a small pod applied to the left costal arch measures respiration rate and patient position as shown in Figure 1. The Philips Guardian solution is programmable according to hospital early warning score parameters—vital signs data are integrated automatically to calculate the patient's early warning score [1,9,18]. Hospital warning scores are used to identify patients with early signs of clinical deterioration to facilitate prompt intervention and prevent a major adverse event [9]. In the event that a patient's early warning score triggers the need for prespecified clinical action, a notification is sent to the ward nurse via a handheld device (eg, Android phone), calling for early attention to care [1,9,18]. To safeguard against notification fatigue, the system features built-in trend analyses and reassurance measurements in order to verify signals that are indicative of early patient deterioration. For more information on hospital early warning systems, the readers are referred to McGillion et al [9].

Figure 1. The Philips Guardian Solution. (A) MP5 spot-check monitor, (B) wireless blood pressure monitor, (C) wireless continuous pulse oximetry monitor, and (D) wireless respiratory sensor. Reproduced with permission from Philips Canada (Markham, ON) (reprinted with copyright permission from the publisher).

Hospital-to-home remote monitoring and virtual care is supported by the Philips Electronic Transition to Ambulatory Care (eTrAC) Program as shown in Figure 2, a tablet-based solution that combines clinical software for remote patient management with Bluetooth-enabled, vital signs monitoring equipment to measure SpO_2, blood pressure, weight, core temperature, and blood glucose [1,9,19]. The nurse interface for eTrAC is eCare Coordinator (eCC). Through eCC secure video visits, nurses review patients' vital signs and weight and conduct remote postoperative assessments daily. These standardized, daily assessments are designed to detect early signs of postoperative complications that may require medical intervention and to address patient concerns during recovery (eg, unrelieved pain in the moderate-to-severe range). Customizable daily patient symptom surveys (ie, general health, wound care, nutrition, medication reconciliation, sleep, functional status, and depression) are also collected via eTrAC [1,9,19]. Together, patient vital signs and survey responses factor into a weighted algorithm, which generates a daily triage score to prioritize nursing assessment and facilitate timely escalation of care to the most responsible physician member of the surgical team [1,9,19].

Figure 2. The Philips electronic Transition to Ambulatory Care system, featuring tablet interface and Bluetooth-enabled vital signs monitors. Reproduced with permission from Philips (reprinted with copyright permission from the publisher).

Design

Our approach to Guardian and eTrAC system usability testing was guided by Wiklund et al’s *Usability Testing of Medical Devices* [20] and conducted with an *out-of-the-box* orientation. Out-of-the-box usability testing involves observing participants interacting with ready-for-use medical devices to perform required tasks, according to laid out instructions, in simulated real-world use case scenarios. As such, our intent was to conduct formative user tests, that is, tests focused on refining our approach to nurse and patient training, as well as refining system workflows with participant input [20].

Setting and Recruitment

Participants included surgical ward nurses and patients recovering from cardiac or major vascular surgery. This study was conducted at two hospital sites, one in Ontario, Canada, and one in Liverpool, the United Kingdom. Nurse participants were recruited through brief presentations at staff meetings and nursing rounds, as well as through emails sent by ward managers. Nurses were invited to participate in either Guardian or eTrAC user testing, but not both, to avoid possible confounding influences of cross-system testing. Patients were invited to participate in eTrAC user testing only, given that Philips’ Guardian system does not involve active workflows for patients wearing the wireless sensors. Included patients were ambulatory, recovering from a cardiac or major vascular surgery, and were able to read, speak, and understand English. Patients who exhibited signs of postoperative delirium (via confusion assessment method) were excluded. The research personnel identified and approached eligible nurses and patients to participate in the study, obtained informed consent, collected baseline demographic information, and scheduled user testing sessions.

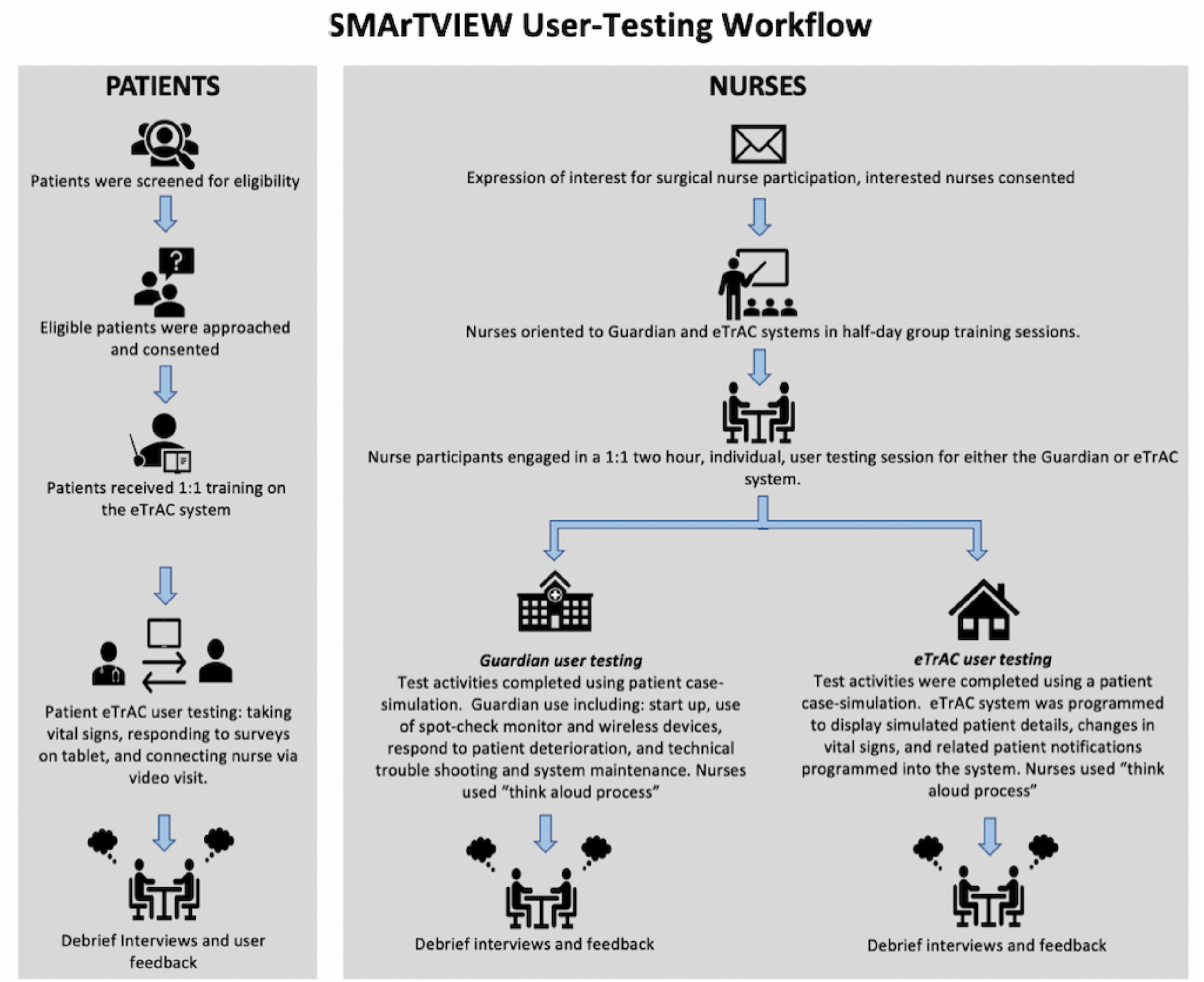
To account for site differences, we aimed to recruit a minimum of 6 participants per site for each system-related user test, for a total of 12 nurses for Guardian testing, 12 nurses for eTrAC testing, and 12 patients for eTrAC testing. Our total sample size of 36 participants was informed by Wiklund et al’s guidance [20] that five to six user test sessions per device are typically sufficient to identify usability issues in the context of a simulated out-of-the-box user testing.

The Hamilton Integrated Research Ethics Board approved the study (project reference number 2332). For the UK arm of the study, Coventry University Ethics Committee granted the ethical approval (project reference ID P50671). Research Governance approval was also granted by Dr Jay Wright, Liverpool Heart and Chest Hospital National Health Service Foundation Trust Research Committee as a Chairman’s action.

Study Procedures

Usability testing of Guardian and eTrAC systems was conducted in two stages: hands-on systems training (Stage 1), followed by individual user testing (Stage 2; Figure 3). Individual user testing occurred as soon as possible after hands-on system training within 24 hours.

Figure 3. Nurse and patient user testing workflow diagram.



Hands-On Systems Training

Nurses

Nurse participants were oriented to Guardian and eTrAC systems during half-day group training sessions. These sessions included a combination of didactic and hands-on learning activities to contextualize and apply content. Topics included nurse and patient onboarding, system navigation, device management, monitoring functions, and technical troubleshooting and related communication. At the end of each session, nurse participants received system user guides for their review and ongoing reference. Nurse participants were asked to review these guides in preparation for their individual usability testing sessions.

Patients

Participating patients received individualized, 1-hour training sessions on the eTrAC system, also featuring didactic and hands-on learning. These sessions focused on daily use of the tablet and vital signs equipment and communicating with a nurse remotely via a secure video link. The training took place at a time convenient to the patient that did not interfere with their routine care. An eTrAC user guide was also provided.

Individual User Testing

Nurses

Following the systems training sessions, nurse participants took part in a 2-hour, individual, usability-testing session for either the Guardian or eTrAC system. The test activities were completed within the context of a simulation featuring one of two patient-cases, depending on nursing area of specialty—cardiac surgery (coronary artery bypass graft) or major vascular surgery (femoropopliteal bypass). Patient cases spanned postoperative days 1 through 4 and featured vital signs–related deteriorations that were programmed into the system remotely; the actors' true vital signs were concealed. These cases were created based on real patient data by staff surgeons, residents, and nurses with minimum 5 years' experience in managing patient deterioration at our study sites. The case scenarios were written in collaboration with the McMaster Standardized Patient Program, Centre for Simulation-Based Learning. A standard template was used to clarify required nurse end-user tasks, the setup of the simulation, and details of the patient cases including surgical details, past medical history, social and family history, and patient-related thoughts, feelings, and concerns that are common in the immediate postcardiac and postvascular surgery context. To provide a high degree of realism, 2 standardized patient actors were trained to portray each case; these actors were placed in an extra bed on the surgical unit at each site, in a separate area designated for the user testing [20].

Guardian User Testing

Following an introduction and review of the system, nurse participants completed the test activities, which represent core nursing competencies for Guardian use including start up, use of the MP5 spot-check monitor and wireless devices, ongoing monitoring and responding to patient deterioration, and technical trouble shooting and system maintenance.

The research assistant moderated the simulation by introducing the background information on the patient and providing instructions and cues to indicate advances in case timelines (eg, “It is now postop day 3 and you return to the unit and reassess your patient.”). The actors were trained to interact with the nurse participants to add realistic distractions and stressors similar to real-world working conditions on the surgical unit.

eTrAC User Testing

The eTrAC user testing followed procedures similar to the Guardian user testing with hospital-to-home patient cases. The eTrAC system was programmed to display simulated patient details, changes in vital signs, and related patient notifications programmed into the system. Patient cases (cardiac and vascular surgery) spanned postdischarge days 1 through 5 and featured vital signs–related deteriorations (eg, high temperature), indicative of common postoperative adverse events (eg, infection) while recovering at home. The research assistant prompted participants with details of these cases as they unfolded to indicate the passage of time and recovery-related circumstances for the patient at home.

Patients

The patient eTrAC user testing focused on completing system tasks required for using the eTrAC system at home, after hospital discharge. The research assistant asked participants to imagine they were now home and undertake the required daily patient-user test activities including using the vital signs equipment, responding to surveys on the tablet, and connecting with a nurse through a secure video visit. The eTrAC system was programmed to display simulated vital signs in demonstration mode. The portable vital signs monitors used by the participants did not display their actual vital signs.

Think Aloud Process

Across all user testing sessions, the research assistant asked participants to *think aloud* as they worked through required tasks [20]. Probing questions were asked to solicit participant reaction to each system and scenario, depending on participants’ apparent level of ease or difficulty as they worked through required tasks (eg, “How are you feeling at this point?”). While thinking aloud, nurse participants were asked to share any critical thinking they were engaged in as they responded to simulated patient deterioration, including problem solving and making decisions about escalating patient care to a physician team member as needed. Patient participants were asked to offer their reflections on their user experience. The research assistant did not respond to participant comments during the *think aloud* process [20].

Outcome Data Collection

User Performance

The research assistant used an observation rubric to evaluate participants’ performance of test activities, according to the following designations: *completed independently*, *completed with difficulty or need for additional prompting*, or *not completed*. Additional observations recorded while participants engaged in the *think aloud* process included task completion time, distraction points, system navigation problems, and any areas of frustration or confusion about workflow when using the systems [20]. A second, silent observer recorded observational field notes to corroborate the research assistant’s observations. A mounted video camera was also used to record the simulations for verifying user performance and any discrepancies noted between participants’ subjective remarks and the documented observations [20].

User Acceptance

Perceived User Satisfaction

Participants were asked to rate their perceived satisfaction with the Guardian or eTrAC system user experience in the context of the workflow training we provided. To solicit this rating, we used the 11-point, single-item Net Promoter Scale (NPS) [21], which asks, “How likely is it that you would recommend the system to a friend or colleague”? We asked participants to respond to this question, while reflecting on the ease of system use during user testing. The NPS includes a score from 0 (*not at all likely*) to 10 (*very likely*). On the basis of response, individuals are classified as potential system *promoters* (response score: 9 or 10), *passives* (response score: 7 or 8), or *detractors* (response score: 0 to 6). A total NPS score, ranging from −100 to +100, can be calculated for each system tested by subtracting the percentage of those classified as *detractors* from the percentage of those classified *promoters*; scores above 0 indicate overall system-related satisfaction [21]. The NPS is widely used for system and health services tests [21-24], with moderate to strong correlation with other measures of patient and user satisfaction [25].

Debrief Interviews

Immediately following each user testing session, the research assistant conducted a 60-min semistructured debrief interview with participants to further identify root causes of any observed difficulties during the simulation, discuss required tasks that may have been missed, and solicit the participants’ overall impressions of their experience using the systems [20].

Data Management and Analyses

Descriptive statistics were used to summarize participants’ demographic characteristics and user performance and acceptance of RAM systems and workflows. All qualitative data, including documented observations and audio recordings of the debrief interviews, were transcribed verbatim. Qualitative analysis was conducted employing inductive, thematic content analysis methods [26], with NVivo 10.0 software (QRS International). Data were coded based on the frequency, extensiveness, and specificity of participants' comments as they related to the usability of Guardian and eTrAC systems [26]. These codes were altered and refined through a recursive process from the data to analyst-generated categorical and conceptual definitions. Revisions to the codebook reflected emerging themes. Constant comparative methods were used to examine individual participants' responses in relation to responses from their respective user groups. New codes occurred less frequently as more transcripts were analyzed, and thematic saturation occurred for each user group when no new codes were generated. Rigor was maintained by completing a reflexive journal and audit trail [26].

Results

Demographics

Nurses

A total of 26 nurses from Canada (CAN group, n=15) and the United Kingdom (UK group, n=11) participated in the Guardian or eTrAC user testing. The majority of nurses were white females, possessing a bachelor's degree in nursing, and employed full time. On average, these nurses had been practicing nursing for over 16 years, in cardiac and vascular surgery and other acute care settings as shown in Table 1.

Table 1. Nurse participant characteristics (n=26).

Nurse characteristics	Values
Sex, n (%)	
Male	4 (15)
Female	22 (85)
Ethnicity, n (%)	
White	21 (81)
African decent	4 (15)
Asian	1 (4)
Education, n (%)	
Professional degree	6 (23)
Bachelor's degree	11 (42)
Masters' degree	7 (27)
Diploma	2 (8)
Employment status, n (%)	
Full-time	19 (73)
Part-time	7 (28)
Number of practicing years, mean (SD)	16.5 (12)

Patients

A total of 11 patients (CAN: n=6; UK: n=5), participated in the eTrAC user testing across both study sites. The majority of patients were male, either married or widowed, and retired. All patients were over 65 years of age; the majority had undergone coronary artery bypass graft or valve replacement surgery as shown in Table 2).

Table 2. Electronic Transition to Ambulatory Care patient participant characteristics (N=11).

Characteristics	Values, n (%)
Sex	
Male	7 (64)
Female	4 (36)
Ethnicity	
White	11 (100)
Marital status	
Married	7 (64)
Widowed	3 (27)
Divorced or separated	1 (9)
Education	
Some high school—no diploma	5 (46)
High school diploma	3 (27)
Trade, technical, vocational training	1 (9)
Professional degree	2 (18)
Employment status	
Full-time	2 (18)
Part-time	3 (27)
Retired	6 (55)
Procedure	
Coronary artery bypass graft	6 (55)
Abdominal aortic aneurysm repair	1 (9)
Heart valve replacement	4 (36)

User Performance

User performance, expressed as percentage of participants observed by task category, that is, *completed*, *completed with difficulty or additional prompting*, and *not completed*, and median task completion times are presented in Table 3.

Table 3. User performance.

User (N) and task	Completed, n (%)	Completed with difficulty or additional prompting, n (%)	Not completed, n (%)	Task completion time (mm:ss), median (IQR)
Guardian nurse: 8 CAN[a]; 6 UK[b] (N=14)				
Nurse pairing the patient to the monitor	11 (79)	3 (21)	0 (0)	00:10 (00:05-00:11)
Assign wireless devices to the patient	9 (64)	5 (36)	0 (0)	04:30 (03:10-06:16)
Complete full set of vital signs	14 (100)	0 (0)	0 (0)	01:04 (00:40-01:47)
Validate EWS[c]	13 (93)	0 (0)	1 (7)	00:03 (00:02-00:05)
Review and manage the patient's vital sign trends	8 (58)	6 (42)	0 (0)	00:26 (00:17-00:47)
Wireless device management	12 (87)	2 (13)	0 (0)	02:08 (01:10-03:17)
Infection control procedures	14 (100)	0 (0)	0 (0)	01:03 (00:30-01:50)
eTrAC[d] patient: 6 CAN; 5 UK (N=11)				
Turn on device	9 (82)	1 (9)	1 (9)	00:08 (00:05-00:15)
Take vital signs (BP[e], SpO_2[f], weight, HR[g], temperature)	8 (73)	3 (27)	0 (0)	01:01 (00:45-01:14)
View scheduled appointment in the calendar	11 (100)	0 (0)	0 (0)	00:22 (00:18-00:43)
Engage in follow-up surveys	9 (82)	2 (18)	0 (0)	00:16(00:12-00:20)
Interface with nurse	10 (91)	1 (9)	0 (0)	00:16 (00:06-00:39)
eTrAC nurse: 7 CAN; 5 UK (N=12)				
Log in and enroll new patient	11 (92)	1 (8)	0 (0)	00:41 (00:31-01:18)
Assign H2H[h] protocol	6 (50)	5 (42)	1 (8)	01:55 (01:14-02:21)
Assign BTE[i] devices	0 (0)	11 (92)	1 (8)	03:35 (03:10-04:47)
Review score and triage the patient	5 (42)	6 (50)	1 (8)	00:47 (00:28-01:07)
Video call and patient wound photo	12 (100)	0 (0)	0 (0)	02:35 (01:57-03:49)
Appropriate escalation of care	12 (100)	0 (0)	0 (0)	Verbal response (not timed)
Add clinical notes	12 (100)	0 (0)	0 (0)	01:12 (00:46-01:34)

[a] CAN: Canada.

[b] UK: United Kingdom.

[c] EWS: early warning score.

[d] eTrAC: electronic transition to ambulatory care.

[e] BP: blood pressure.

[f] SpO_2: blood oxygen saturation.

[g] HR: heart rate.

[h] H2H: hospital-to-home.

[i] BTE: Bluetooth-enabled.

Guardian System

The majority of nurse participants were able to complete most required tasks independently, demonstrating comprehension and retention of required Guardian system workflows, for example, prompting the system to complete a set of *on demand* vital signs. Tasks, which required additional prompting by the facilitator, for some, were related to the use of system features that enable continuous patient biometric data transmission [9,18] and active clinical management of this information (eg, assigning wireless devices to the patient and displaying continuous patient vital signs trends on the central monitor).

eTrAC System

Nurse users demonstrated confidence with eTrAC workflows related to onboarding patients onto the system, as well as direct patient interaction and remote patient management, including assessment, documentation of independent nursing actions, and escalation of care to the most responsible physician in the patient case scenario. These users were less confident in working with system protocols for remote wireless patient vital signs transmission, such as assigning the Bluetooth-enabled vital signs devices to the patient for home use and assigning the appropriate hospital-to-home monitoring regimen based on surgical procedure.

The majority of patient users demonstrated ease and independence with all required eTrAC tasks. Similar to nurses, some required additional prompting to work with the Bluetooth devices to take their vital signs and navigate aspects of the system interface related to remote self-monitoring (eg, responding to symptom survey).

User Acceptance

User Satisfaction—Net Promoter Scale score

Individual and mean NPS ratings, by user group, are presented in Table 4. Mean scores indicate a high degree of likelihood that each user group would recommend the Guardian and eTrAC systems to others, based on their training and user test experience. Overall NPS scores, by user group (nurses and patients), also indicate overall user satisfaction with each system.

Table 4. User satisfaction—Net Promoter Scale score.

User (N) and group	Raw scores (range 0-10)	Value, mean (SD)	NPS[a] score (% of promoter−% of detractors)
Guardian nurses: 8 CAN[b]; 6 UK[c] (N=14)		8.8 (0.89)[d]	64
Nurse 1	8		
Nurse 2	8		
Nurse 3	8		
Nurse 4	8		
Nurse 5	7		
Nurse 6	9		
Nurse 7	10		
Nurse 8	10		
Nurse 9	10		
Nurse 10	9		
Nurse 11	9		
Nurse 12	9		
Nurse 13	9		
Nurse 14	9		
eTrAC[e] nurses: 7 CAN; 5 UK (N=12)		7.7 (1.4)[d]	25
Nurse 1	4		
Nurse 2	7		
Nurse 3	8		
Nurse 4	7		
Nurse 5	8		
Nurse 6	8		
Nurse 7	8		
Nurse 8	8		
Nurse 9	9		
Nurse 10	9		
Nurse 11	9		
Nurse 12	9		
eTrAC patients: 6 CAN; 5 UK (N=11)		9.2 (0.75)	82
Patient 1	8		
Patient 2	8		
Patient 3	9		
Patient 4	10		
Patient 5	10		
Patient 6	10		
Patient 7	9		
Patient 8	9		
Patient 9	9		
Patient 10	9		
Patient 11	10		

[a] NPS: Net Promoter Scale.

[b] CAN: Canada.

[c] UK: United Kingdom.

[d] This is the average score.

[e] eTrAC: electronic transition to ambulatory care.

Debrief Interviews

Posttest debrief interviews provided opportunities for users to reflect on their own performance and how they felt during the test simulations, any areas of difficulty that they had, and what (if any) improvements to system workflow training could be made. The participant's overall accounts of their user experience—positive or negative—were also solicited.

Nurses

Key themes that emerged from the interviews of nurses engaged in the Guardian user testing related to system *ease of use*, vital signs *trend monitoring*, and *wireless device management*.

In terms of *ease of use,* most users spoke of the simplicity of the Guardian interface and feeling confident about navigating the system after a short while. Participants also commented on their perception of Guardian as an aide to day-to-day nursing work on busy surgical wards:

> *It does the job...more frequently than a nurse can. We are only one nurse taking care of 4 to 5 patients, so if our monitor can do continuous monitoring and alert us when our patient is [deteriorating], that's unbelievable.* [UK Nurse, participant 002-009]

Some nurses cited unfamiliarity with more advanced aspects of the system as a barrier to engaging in vital signs *trend monitoring* during the simulations. However, no participants expressed that changes would be needed to the system or related workflow training. Rather, they emphasized the importance of more opportunities to practice and get comfortable with software features that enable management and visualization of continuous patient vital signs data. Others commented that having the ability to examine vital signs trends remotely would be invaluable to patient management:

> *I thought that was the coolest thing. You can monitor your patient from [your handled device], the desktop central station, or from the monitor, at the bedside. So you don't have to be at the bedside all the time to know [patient status].* [CAN Nurse, participant 001-005]

Reflections on *wireless device management*, were also indicative of excitement about the potential for remote monitoring to improve patient safety and create efficiency in nursing work through automation of time-consuming processes, such as manual data entry:

> *I think it's going to be really safe. If I can constantly know that they're [patients] going to be okay if I leave them and do other things and it's just very quick and easy.* [UK Nurse, participant 002-009]
>
> *I think it's awesome that it is barcodes and scanning and no data entry. I think half our shifts are wasted with data entry—we don't need any more of that!* [CAN Nurse, participant 001-003]

Nurse participants who were debriefed following the eTrAC user testing reflected on this system as an enabler of their *clinical nursing skills*, as well as an opportunity and need to learn new *technical skills*. Nurses found that the hospital-to-home simulation allowed them to use their critical thinking skills by assessing the patient directly through the eTrAC video feature, and by reviewing vital signs and patient survey results before they conducted their assessments. Many felt that the clinical interface was well designed and simple to navigate, allowing for a complete picture of the patient and timely decisions about clinical action:

> *It's quite cool to see all your patients and the alerts. Because, you know, you could have twenty patients on there, and they're all fine, making a good recovery. I like the way the alerts and the scores are visible and you can act on it straight away.* [UK Nurse, participant 002-014]

Most participants commented that they needed additional prompting to assign hospital-to-home protocols and to assign the Bluetooth-enabled vital signs devices to the patient and tablet. Lack of familiarity and confidence with the technical aspects of the eTrAC interface were discussed as key challenges to completing these tasks. When asked, all participants said they would need and would welcome the opportunity to develop these technical skills further, and that with additional support, they could see themselves becoming proficient in these aspects of eTrAC use.

Patients

Patient debrief interviews revealed an overall positive experience with the eTrAC user testing, with *ease of system use* and *recovery progress* as key themes that emerged. When asked for feedback on navigating the patient tablet, participants commented that it was both pleasing aesthetically and uncomplicated operationally. Specifically, patients indicated that the interface was clear and easy to use, and that the font was legible. Despite half of these patients noting that they had initial anxiety about using the technology, most remarked feeling comfortable once they began user testing. As one participant commented:

> *At first I thought I couldn't do it but again, like I said, it is very straightforward. Very easy, you start at the top and just finish the temperature or the height, or your weight, and you just follow the tablet.* [CAN patient, participant 001-009]

All participants remarked that they felt the system would be invaluable for helping them through recovery. There was a high degree of enthusiasm about connecting with a nurse daily and being monitored, as it would offer a sense of security after hospital discharge:

> *I think that's the best thing of all... there's somebody at the end of that—just like telephone line, there's*

> *somebody at the end of that you can talk to and you can see them... that's really good.* [CAN patient, participant 001-017]

Participants also commented on the value of engaging in eTrAC patient surveys to give a monitoring nurse more information and ensure that they are on track with recovery:

> *I would give it [survey feature] a 10 plus, plus. I think it is just comforting to know that you are on the right track and can tell the nurse what's happening.* [CAN patient, participant 001-021]

In summary, participant debrief interviews indicated a high degree of acceptance among users. Nurses expressed the importance and potential of remote monitoring and virtual hospital-to-home care as means to improve efficiency of clinical workflows, enhance patient safety, and facilitate timely clinical action. Patient users spoke to the security that these systems can offer through daily connection with a nurse while recovering from surgery. Both user groups stressed the need for additional opportunities to practice in order to become comfortable and proficient in the use of the Guardian and eTrAC systems, with respect, in particular, to mastering more technical aspects related to enabling remote connectivity and assigning and engaging in monitoring protocols.

Discussion

Principal Findings

This study addressed user performance and acceptance of Philips' Guardian and eTrAC systems designed to support in-hospital RAM and virtual care from hospital-to-home, respectively [9,18,19]. The planned use of these systems in combination succession, within a trial intervention, is a unique approach to studying the effects of enhanced patient surveillance and remote recovery support following cardiac and major vascular surgery [1,9]. *Out-of-the-box* user testing [20] uncovered overall strong user performance by both nurses and patients for the majority of required user tasks. Testing also uncovered important areas where our approach to systems training and implementation needed to be strengthened to better support users. With respect to Guardian, a number of participants required additional support from the facilitator to complete more advanced system-related tasks that would enable acquisition and transmission of continuous patient vital signs data monitoring, such as pairing of the wireless sensors to the MP5 spot-check monitor. Results of eTrAC testing were similar, suggesting that nurse users were less confident after training when it came to pairing Bluetooth-enabled monitoring devices to the patient tablet, as well as assigning specific remote monitoring protocols (related to surgical type).

These user performance results were used to enhance our approach to systems training during SMArTVIEW trial start up at participating hospital sites. Initial systems in-services for nursing staff were followed by individual facilitated practice sessions both in classroom settings and on the surgical wards. These applied learning opportunities allowed for the development of ward nurses' required technical skills to become proficient in the use of Guardian, while they transitioned from case-based learning to live systems use. Some nurse participants also became designated Guardian *champions* at SMArTVIEW trial sites, acting as resources for other ward nurses to support ongoing learning and technical trouble shooting.

Given that the use of the Philips eTrAC hospital-to-home system [19] represents a skill set unique from standard ward nursing, it was decided that deployment of the system, at each study site, would be preserved for a designated subteam of nurses (seconded from ward duties), who focus on hospital-to-home virtual care. These specialized nurses, referred to as SMArTVIEW nurses, take responsibility for onboarding study patients to the eTrAC system and monitoring them for the first 30 days at home following hospital discharge [9]. These nurses provide daytime hospital-to-home service for patients allocated to the intervention arm of the SMArTVIEW trial, 7 days per week, from a designated space on the surgical ward. In this role, they also assist ward nurses with Guardian implementation at the beginning of each shift to reinforce systems training and optimize adoption of RAM workflows.

Comparison With Prior Work

This study demonstrated a high degree of acceptance in terms of user satisfaction and communication of a positive learning experience, highlighting the value of providing risk-free opportunities to learn RAM and virtual care technologies, before implementation, to ease end-user apprehensions and achieve buy-in. A few other studies have examined specifically user experience in the context of RAM and virtual care technology planning or pilot testing in surgical settings. In their recent multimethod study to examine nurse and physician perceptions of a planned introduction of continuous RAM on general hospital wards, Prgomet et al [14] found that hospital staff first expressed apprehension and beliefs about RAM technologies that would likely counter successful implementation. Nurses were concerned about the potential for staff overreliance on RAM technology and, hence, reduced bedside patient interaction and examination. Physicians were concerned about inappropriate care escalations based on false-positive RAM notifications and subsequent desensitization to alerts. Both groups expressed similar concerns over the potential for alarm fatigue [14].

As was the case in our study, however, the opportunity to trial the monitoring devices and engage in dialogue about their impact on clinical workflows and patient care gave rise to perceptions that focused training, featuring educational opportunities to address pre-existing attitudes and beliefs about the incorporation of RAM technologies into clinical practice, would be an important prerequisite to successful implementation [14]. Despite initial concerns, there was also acknowledgment by nurses of the potential for RAM devices to enhance early detection of patient deterioration and provide supporting evidence when communicating concerns about patient status to physician colleagues [14].

In a prospective study (N=443), McElroy et al [27] incorporated a digital health kit—featuring a patient tablet and Bluetooth-enabled vital signs monitors (with software enabling abnormal vital signs to trigger automated alerts to clinicians)—into a 30-day readmission reduction program following cardiac surgery. Posttest user satisfaction survey results were positive among patient and nurse respondents. Similar to this study, the high degree of user satisfaction observed was attributed to the simplicity of the digital kit system, including easy-to-use tablet software and system configuration that allowed for easy patient remote connection to a nurse by video [27].

The VItal siGns monitoring with continuous puLse oximetry And wireless cliNiCal notification aftEr surgery study investigators [28] conducted an evaluation report to uncover nurse and patient user perspectives on RAM implementation challenges during a recent RCT (N=2049) of continuous vital signs monitoring with alerts to nursing staff on the incidence of respiratory resuscitations, code blues, and intensive care unit transfers in patients undergoing noncardiac surgery. A key challenge to implementation was nurse adherence to required clinical workflow changes to accommodate RAM. Nearly 23% of records used to track nursing compliance with the RAM intervention were missing, indicating compliance issues despite ongoing ward training [28]. Restriction to ambulation imposed by the continuous pulse oximetry cable and sensor-related discomfort were noted as common reasons for patient withdrawal (10.68%) from the RAM intervention. A key recommendation of the evaluation report was that pre-emptive user testing would be important for future RAM studies and clinical applications to achieve stakeholder buy-in and co-design and refinement of clinical workflows, as well as to establish RAM champions during implementation [28]. Other studies soliciting nurse perspectives following RAM pilots report similar recommendations [15,29] and the need for constant attention to change management.

Limitations

Potential limitations of this study include our homogenous participant sample, as well as our approach to measuring task completion times and level of interaction with participants during the *think aloud* process [20]. Our user testing procedures relied on a convenience sample of nurses and patients who were in-hospital and agreed to participate at the time of testing. Those enrolled were white individuals who spoke English. Although user testing studies are not designed for generalization to broad populations per se, our results speak to user acceptance and performance in a limited subpopulation of end users. It should also be noted that of the 11 patients enrolled, just 1 eligible patient had undergone vascular surgery at the time of study enrollment—the remainder of the patient sample was cardiac surgery patients. Nonetheless, the equipment and workflows tested were designed to be identical for both patient groups.

As is common during formative types of user tests that are intended to reveal shortcomings in systems training or workflows [20], some of our recorded individual task completion times may have been distorted. Task time distortions occur when test participants pause to reflect on completion of their user tasks while thinking aloud [20]. An approach to remedying task time distortion is to require participants to remain silent during user testing. However, this approach is more amendable to summative user tests [20], which place greater importance on measurement precision over participant narrative about their user experience. In this study, we favored concurrent participant observation and moderation over achieving recorded task time precision; this approach allowed us to identify clear gaps in systems training and aspects of workflow implementation that needed refining before launch of the SMArTVIEW trial.

We also undertook a flexible approach to participant moderation during the *think aloud* process. During test simulations, the facilitator prompted participants with key words or instructions if they were clearly struggling with certain required tasks. This strategy allowed us to include a *completed with difficulty or need for additional prompting* category in our user performance rubric. Although less conventional than traditional formative user testing [20], we again placed high value on the solicitation of user narrative. As with task completion times, we were more concerned with participant involvement in as many aspects of Guardian and eTrAC workflows as possible (and related reflection) than with the generation of conventional user performance metrics.

Conclusions

The inadequacy of current systems for postsurgical patient monitoring in hospital and at home is a major factor contributing to postoperative complications, death, and unplanned hospital readmissions [6-10]. Although RAM and virtual care technologies have high potential for improving postoperative patient outcomes, the science of implementing and evaluating these technologies is complex and still at an early stage. This formative *out-of-the-box* user testing study indicated a high degree of user acceptance of Philips' Guardian and eTrAC systems among nurses and patients. Key insights were also provided that informed refinement of clinical workflow training and systems implementation, including clear division of responsibilities, before launch of the international SMArTVIEW trial. This trial, underway, is designed to examine the effectiveness of postoperative RAM and virtual hospital-to-home recovery support on health system and patient-related outcomes following cardiac and major vascular surgery. Practical implementation issues will also be explored, including the need for specialized training of subteams of nurses to deploy RAM and virtual hospital-to-home recovery support at trial sites.

Acknowledgments

This study was funded by the Canadian Institutes of Health Research (grant no. 348440), the Ministry of Health, and the Hamilton Health Sciences Strategic Research Initiative. In-kind industry support was provided by Philips. MM holds the Heart and Stroke Foundation/Michael G DeGroote Endowed Chair in Cardiovascular Nursing.

Authors' Contributions

MM is the principal investigator of this study and has contributed toward conceptualization, study design, protocol writing, user testing, data analysis, and manuscript writing. CO has contributed toward user testing, data collection, data analysis, and manuscript writing. AG has contributed toward study coordination, user testing, and data management. MB has contributed toward data analysis and manuscript writing. SH has contributed toward user testing and data collection. WC has contributed toward user testing, data collection, and manuscript review. AT has contributed toward user testing, data collection, and manuscript review. PR has contributed toward study design and protocol writing. SR has contributed toward study design and protocol writing. ND has contributed toward protocol writing and user test cardiac patient case writing. AL is the Canadian site principal investigator and has contributed toward user test case review and approval. RW has contributed to development of participant recruitment process. CL has contributed toward installation and management of the RAM system and technical study setup. JW has contributed toward installation and management of the RAM system and technical study setup. KP has contributed to installation and management of the RAM system and technical study setup. JD has contributed to the development of the posttest interview guide. KS has contributed to data analysis and manuscript writing. FE has contributed to protocol writing and user test vascular patient case writing. TS has contributed to study conceptualization and hospital technical services procurement. JM is the UK site principal investigator and has contributed toward user test case review and approval. DH has contributed to user test cardiac patient case writing in the United Kingdom. MF has contributed toward user test vascular patient case writing in the United Kingdom. PH has contributed toward data analysis and manuscript writing. SY has contributed toward data analysis and manuscript writing. EP has contributed to development of qualitative analysis methods within the protocol. SB has contributed toward study conceptualization and principal investigator and team training in out-of-the-box usability testing methods. PD is the coprincipal investigator of this study and has contributed toward conceptualization, study design, protocol writing, user testing, data analysis, and manuscript writing.

Conflicts of Interest

MM and PD are members of a research group that does not accept honorariums or other payments from industry for personal financial gain. They do accept honorariums or payments from industry to support research endeavors and costs to participate in meetings. On the basis of study questions PD has originated and grants he has written, he has received grants from Abbott Diagnostics, Boehringer Ingelheim, Covidien, Octapharma, Philips Healthcare, Roche Diagnostics, and Stryker. PD has participated in a consultancy advisory board meeting for Boehringer Ingelheim.

References

1. McGillion M, Yost J, Turner A, Bender D, Scott T, Carroll S, et al. Technology-enabled remote monitoring and self-management - vision for patient empowerment following cardiac and vascular surgery: user testing and randomized controlled trial protocol. JMIR Res Protoc 2016 Aug 1;5(3):e149 [FREE Full text] [doi: 10.2196/resprot.5763] [Medline: 27480247]
2. McGillion MH, Turner A, Carroll SL, Furze G, Busse JW, Lamy A. Optimising self-management to reduce chronic pain and disability after cardiac surgery. Br J Cardiol 2015;22(1):38. [doi: 10.5837/bjc.2015.011]
3. Iribarne A, Chang H, Alexander JH, Gillinov AM, Moquete E, Puskas JD, et al. Readmissions after cardiac surgery: experience of the National Institutes of Health/Canadian Institutes of Health research cardiothoracic surgical trials network. Ann Thorac Surg 2014 Oct;98(4):1274-1280 [FREE Full text] [doi: 10.1016/j.athoracsur.2014.06.059] [Medline: 25173721]
4. Gupta PK, Fernandes-Taylor S, Ramanan B, Engelbert TL, Kent KC. Unplanned readmissions after vascular surgery. J Vasc Surg 2014 Feb;59(2):473-482 [FREE Full text] [doi: 10.1016/j.jvs.2013.09.002] [Medline: 24239519]
5. Tahhan G, Farber A, Shah NK, Krafcik BM, Sachs TE, Kalish JA, et al. Characterization of planned and unplanned 30-day readmissions following vascular surgical procedures. Vasc Endovascular Surg 2017 Jan;51(1):17-22. [doi: 10.1177/1538574416682176] [Medline: 28100157]
6. Efthymiou CA, O'Regan DJ. Postdischarge complications: what exactly happens when the patient goes home? Interact Cardiovasc Thorac Surg 2011 Feb;12(2):130-134. [doi: 10.1510/icvts.2010.249474] [Medline: 21123196]
7. Leuvan CH, Mitchell I. Missed opportunities? An observational study of vital sign measurements. Crit Care Resusc 2008 Jun;10(2):111-115. [Medline: 18522524]
8. McGain F, Cretikos MA, Jones D, van Dyk S, Buist MD, Opdam H, et al. Documentation of clinical review and vital signs after major surgery. Med J Aust 2008 Oct 6;189(7):380-383. [Medline: 18837681]
9. McGillion MH, Duceppe E, Allan K, Marcucci M, Yang S, Johnson AP, PROTECT Network Investigators. Postoperative remote automated monitoring: need for and state of the science. Can J Cardiol 2018 Jul;34(7):850-862 [FREE Full text] [doi: 10.1016/j.cjca.2018.04.021] [Medline: 29960614]
10. Sun Z, Sessler DI, Dalton JE, Devereaux PJ, Shahinyan A, Naylor AJ, et al. Postoperative hypoxemia is common and persistent: a prospective blinded observational study. Anesth Analg 2015 Sep;121(3):709-715 [FREE Full text] [doi: 10.1213/ANE.0000000000000836] [Medline: 26287299]
11. Sessler DI, Saugel B. Beyond 'failure to rescue': the time has come for continuous ward monitoring. Br J Anaesth 2019 Mar;122(3):304-306. [doi: 10.1016/j.bja.2018.12.003] [Medline: 30770047]

12. Izmailova ES, Wagner JA, Perakslis ED. Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 2018 Jul;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]
13. Turan A, Chang C, Cohen B, Saasouh W, Essber H, Yang D, et al. Incidence, severity, and detection of blood pressure perturbations after abdominal surgery: a prospective blinded observational study. Anesthesiology 2019 Apr;130(4):550-559. [doi: 10.1097/ALN.0000000000002626] [Medline: 30875354]
14. Prgomet M, Cardona-Morrell M, Nicholson M, Lake R, Long J, Westbrook J, et al. Vital signs monitoring on general wards: clinical staff perceptions of current practices and the planned introduction of continuous monitoring technology. Int J Qual Health Care 2016 Sep;28(4):515-521. [doi: 10.1093/intqhc/mzw062] [Medline: 27317251]
15. Jeskey M, Card E, Nelson D, Mercaldo ND, Sanders N, Higgins MS, et al. Nurse adoption of continuous patient monitoring on acute post-surgical units: managing technology implementation. J Nurs Manag 2011 Oct;19(7):863-875. [doi: 10.1111/j.1365-2834.2011.01295.x] [Medline: 21988434]
16. Downey CL, Croft J, Buckley H, Randell R, Brown JM, Jayne DG. Trial of Remote Continuous versus Intermittent NEWS monitoring after major surgery (TRaCINg): protocol for a feasibility randomised controlled trial. Pilot Feasibility Stud 2018;4:112 [FREE Full text] [doi: 10.1186/s40814-018-0299-z] [Medline: 29992041]
17. Downey C, Randell R, Brown J, Jayne DG. Continuous versus intermittent vital signs monitoring using a wearable, wireless patch in patients admitted to surgical wards: pilot cluster randomized controlled trial. J Med Internet Res 2018 Dec 11;20(12):e10802 [FREE Full text] [doi: 10.2196/10802] [Medline: 30538086]
18. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
19. Philips Enterprise Telehealth. Remote Patient Monitoring URL: https://www.usa.philips.com/healthcare/services/population-health-management/patient-engagement/remote-patient-monitoring [accessed 2020-01-22]
20. Wiklund PE, Kendler J, Strochlic A. Usability Testing of Medical Devices. Boca Raton: CRC Press; 2016.
21. Sheard L, O'Hara J, Armitage G, Wright J, Cocks K, McEachan R, Yorkshire Quality & Safety Research Group. Evaluating the PRASE patient safety intervention - a multi-centre, cluster trial with a qualitative process evaluation: study protocol for a randomised controlled trial. Trials 2014 Oct 29;15:420 [FREE Full text] [doi: 10.1186/1745-6215-15-420] [Medline: 25354689]
22. Lawton R, O'Hara JK, Sheard L, Armitage G, Cocks K, Buckley H, et al. Can patient involvement improve patient safety? A cluster randomised control trial of the Patient Reporting and Action for a Safe Environment (PRASE) intervention. BMJ Qual Saf 2017 Aug;26(8):622-631 [FREE Full text] [doi: 10.1136/bmjqs-2016-005570] [Medline: 28159854]
23. NHS England. 2015. The Friends and Family Test URL: https://www.england.nhs.uk/wp-content/uploads/2015/07/fft-guidance-160615.pdf [accessed 2020-01-22]
24. Day FC, Pourhomayoun M, Keeves D, Lees AF, Sarrafzadeh M, Bell D, et al. Feasibility study of an EHR-integrated mobile shared decision making application. Int J Med Inform 2019 Dec;124:24-30. [doi: 10.1016/j.ijmedinf.2019.01.008] [Medline: 30784423]
25. Krol MW, de Boer D, Delnoij DM, Rademakers JJ. The Net Promoter Score--an asset to patient experience surveys? Health Expect 2015 Dec;18(6):3099-3109 [FREE Full text] [doi: 10.1111/hex.12297] [Medline: 25345554]
26. Creswell JW. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks: SAGE Publications; 2014.
27. McElroy I, Sareh S, Zhu A, Miranda G, Wu H, Nguyen M, et al. Use of digital health kits to reduce readmission after cardiac surgery. J Surg Res 2016 Jul;204(1):1-7. [doi: 10.1016/j.jss.2016.04.028] [Medline: 27451860]
28. Harsha P, Paul JE, Chong MA, Buckley N, Tidy A, Clarke A, et al. Challenges with continuous pulse oximetry monitoring and wireless clinician notification systems after surgery: reactive analysis of a randomized controlled trial. JMIR Med Inform 2019 Oct 28;7(4):e14603 [FREE Full text] [doi: 10.2196/14603] [Medline: 31661079]
29. Gazarian PK. Nurses' response to frequency and types of electrocardiography alarms in a non-critical care setting: a descriptive study. Int J Nurs Stud 2014 Feb;51(2):190-197. [doi: 10.1016/j.ijnurstu.2013.05.014] [Medline: 23810495]

Abbreviations

eCC: eCare Coordinator
eTrAC: electronic transition to ambulatory care
NPS: Net Promoter Scale
RAM: remote automated monitoring
RCT: randomized controlled trial
SpO_2: blood oxygen saturation
THE SMArTVIEW, CoVeRed: TecHnology-Enabled remote monitoring and Self-MAnagemenT—VIsion for patient EmpoWerment following Cardiac and major VasculaR surgery

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by T Jamieson, H Mufti, T Hale; comments to author 16.08.19; revised version received 31.10.19; accepted 16.12.19; published 18.03.20.

Please cite as:
McGillion M, Ouellette C, Good A, Bird M, Henry S, Clyne W, Turner A, Ritvo P, Ritvo S, Dvirnik N, Lamy A, Whitlock R, Lawton C, Walsh J, Paterson K, Duquette J, Sanchez Medeiros K, Elias F, Scott T, Mills J, Harrington D, Field M, Harsha P, Yang S, Peter E, Bhavnani S, Devereaux PJ
Postoperative Remote Automated Monitoring and Virtual Hospital-to-Home Care System Following Cardiac and Major Vascular Surgery: User Testing Study
J Med Internet Res 2020;22(3):e15548
URL: https://www.jmir.org/2020/3/e15548
doi:10.2196/15548
PMID:32186521

Original Paper

Use of Notification and Communication Technology (Call Light Systems) in Nursing Homes: Observational Study

Haneen Ali[1], PhD; Huiyang Li[2], PhD

[1]Health Services Administration Program & Department of Industrial and Systems Engineering, Auburn University, Montgomery, AL, United States
[2]Binghamton University, Binghamton, NY, United States

Corresponding Author:
Huiyang Li, PhD
Binghamton University
4400 Vestal Pkwy E
Binghamton, NY, 13902
United States
Phone: 1 734 546 1910
Email: hli@binghamton.edu

Abstract

Background: The call light system is one of the major communication technologies that link nursing home staff to the needs of residents. By providing residents the ability to request assistance, the system becomes an indispensable resource for patient-focused health care. However, little is known about how call light systems are being used in nursing homes and how the system contributes to safety and quality of care for seniors.

Objective: This study aimed to understand the experiences of nursing home staff who use call light systems and to uncover usability issues and challenges associated with the implemented systems.

Methods: A mix of 150 hours of hypothetico-deductive (unstructured) task analysis and 90 hours of standard procedure (structured) task analysis was conducted in 4 different nursing homes. The data collected included insights into the nursing home's work system and the process of locating and responding to call lights.

Results: The data showed that the highest alarm rate is before and after mealtimes. The staff exceeded the administration's expectations of time to respond 50% of the time. In addition, the staff canceled 10.0% (20/201) of call lights and did not immediately assist residents because of high workload. Furthermore, the staff forgot to come back to assist residents over 3% of the time. Usability issues such as broken parts, lack of feedback, lack of prioritization, and low or no discriminability also contributed to the long response time. More than 8% of the time, residents notified the staff about call lights after they waited for a long time, and eventually, these residents were left unattended.

Conclusions: Nursing homes that are still using old call light systems risk the continuation of usability issues that can affect the performance of the staff and contribute to declining staff and resident outcomes. By incorporating feedback from nurses, nursing home management will better understand the influence that the perceptions and usability of technology have on the quality of health care for their residents. In this study, it has been observed that the call light system is perceived to be an important factor affecting the outcomes of the care process and satisfaction of both residents and staff as well as the staff's performance. It is important to recognize that communication and notification technology contributes to the challenges the staff faced during their work, making their working conditions more difficult and challenging.

(J Med Internet Res 2020;22(3):e16252) doi:10.2196/16252

KEYWORDS

communication systems; nursing home; response time; safety; quality of health care; observational study

Introduction

Background

Long-term care in nursing homes provides a wide range of services for older people, especially for those who have either cognitive or physical problems, or both. The number of residents at such facilities is expected to increase from 15 million in 2000 to 27 million in 2050 in the United States [1]. As a result, nursing homes will serve an even more essential function in providing long-term geriatric services [2]. However, a recent study [3] suggests a growing concern regarding the overall safety and quality of care in nursing homes [4]. Nursing home residents usually have cognitive deficits and complex health conditions, and they often take more than one medication, which can increase the risk of medication errors [5]. According to the Department of Health and Human Services, 1 in 3 residents are harmed due to a medication error, an infection, or miscellaneous circumstances related to their treatment. These issues are widespread in nursing homes because of the lack of appropriate technologies [3,6]. Furthermore, most of the attempts to improve health care in nursing homes are solely focused on aspects such as improving medical, functional, physical, cognitive, and physiological care, in addition to providing more training for the staff [7], but not on the technological aspects.

Among the technologies used in nursing homes is the call light system, which plays a key role in communication. The call light system is critical for interactions between the nursing home staff and their residents. Research conducted in other health care settings has reported that nurse call systems significantly influence overall resident satisfaction in the delivery of their health care by creating a communication link between residents and nursing home staff [8,9]. Nurse call light systems also help to ensure the safety of patients [10]. Meade et al [10] also described it as a *lifeline* for patients because it is linked with patients' needs and alerts the staff to the situations in which patients may ask for help. Nurse call light systems in nursing homes are associated with many ergonomic concerns and usability issues [3,6,11]. Malfunctions of the nurse call light system can cause negative medical outcomes, as the literature identifies other health care settings whereby a relationship exists between the time it takes to respond to a call light system and adverse events such as falls [12,13]. Certified nurse assistants (CNAs) contribute to more than 80% of direct care to residents. However, they must also respond to different types of alarms such as bed exit, chair exit, and clip alarms [14], as they are ultimately responsible in cases of adverse events [6]. However, their perceptions are rarely considered by the administration in the development and selection of new technology [15,16].

This study not only acknowledges the role of communication technologies as a potential solution for the challenges in nursing homes but also recognizes the importance of considering its influence on nursing practices [17,18]. This paper, therefore, provides a methodological approach to study usability challenges in the call light technologies in nursing homes. The study focuses on usability issues and challenges the staff members face while interacting and using the system to understand the barriers to productive, efficient, and safe use of these systems. The insights of the study sought to inform nursing home management on ways to improve the safety and quality of care for nursing home residents. The theoretical and methodological framework for usability evaluation is detailed in the Methods section.

Nursing Home Challenges

The literature attributes the poor quality of care and patient safety in nursing homes to many reasons and challenges, such as lack of staff knowledge, high turnover rates, understaffing, poor management, and lack of attention to technology [19-21]. Many approaches were mentioned in the literature to overcome these problems, such as improving the quality and safety of care of nursing home residents by increasing the staffing ratio [14,20,22-24], staff training [25,26], and using technological tools such as electronic health record (EHR) and electronic medication administration [27-31]. However, no study to date has focused on the overall impact of technology on nursing homes' work systems, and no studies have analyzed the nursing homes' work system and how the work system elements interact with each other. All the challenges mentioned above, in addition to the lack of technology, were identified as barriers to improvement in nursing home practice [19-21]. Nursing homes need to translate the research on best practices into day-to-day practice to improve the quality of care and patient safety [19]. This can be done by analyzing the process of care and understanding the barriers to changes and how these changes will affect the outcomes of nursing homes [19].

Human factors design is increasingly recognized as an important discipline in improving health care quality and patient safety [32]. More light is being shed these days on the importance of human factors in many fields of health care, such as the usability and design characteristics of information technology systems [11,33], identifying hazards in health care [34], performance process and obstacles [35,36], and system resilience [37]. Human factors engineering approaches have also been used to analyze and describe the work system in health care settings (H Ali and A Ahmed, unpublished data, 2020), [35,38]. To overcome the challenges faced by nursing homes that impede their functionality, it is important to understand the potential benefits of technology and pay attention to how the functionality of technological innovations is improving the work system. Very few studies have examined and investigated the process of care in nursing homes or proposed approaches and practices to overcome the challenges. In addition, there is a lack of empirical studies examining technologies such as call light systems in nursing homes and their interaction with other components in the system. A thoughtful, evidence-based, and large-scale improvement that involves a new intervention is what nursing homes need to improve processes and outcomes [19].

Methods

Research Design

The research presented in this paper employed a cognitive engineering approach to study the interaction surrounding resident-initiated call light processes. Cognitive engineering is concerned with the analysis, design, and evaluation of complex systems of people and technology [39]. The paper presented iterative observations of continuous interaction with call light technology—beginning with the staff goal (responding to call lights in <5 min), leading to an action (being notified, locating, and addressing notifications), and resulting in a change of the system (canceling the alarm after assisting the resident). The research was conducted based on a two-pronged methodological approach to the study of human-computer interaction, qualitative research (specifically ethnography), participant observation, and interaction analysis. In stage 1, a hypothetico-deductive approach was used with unstructured observations to identify and assess the process of using the call light system. These observations took place in 4 different nursing homes to identify the challenges and usability issues associated with the different call light systems in use and to understand the barriers to productive use.

In stage 2, a detailed task analysis was used to understand and learn about the process, structure, flow, and attributes of tasks and was conducted in one of the nursing homes. The goals of this second stage were to understand the frequency, sequence, and complexity of call lights and to generate detailed and precise information on the performance of the staff and the process of being notified and responding to call lights. This stage was conducted by defining specific tasks, subtasks, and actions to be taken and by using the structured, systematic observation approach to collect the sample data.

Procedure

Approximately 25 hours of semistructured interviews and more than 150 hours of unstructured observations were conducted by both researchers in 4 different nursing homes located in upstate New York. The goal of this stage was to understand and evaluate the different call light systems being used in nursing homes. Furthermore, the research design seemed to determine how these systems affect other elements in the nursing homework systems (person, tasks, tools and technology, organization, and physical environment), processes, and outcomes [40]. The semistructured interviews were conducted with 3 nursing home administrators, 2 unit managers, 2 registered nurses (RNs), 6 CNAs, 4 licensed practical nurses (LPNs), a program coordinator for elder service, and 1 nursing faculty with extensive experience of working in and with nursing homes. Interview topics included the following: different roles of the staff, relationships between the staff, teamwork, shifts, the technology used in nursing homes and their effectiveness, limitations, and challenges the staff experienced when using them. The study expects that how the staff is notified about call lights will affect not only their overall response time but also resident satisfaction. Furthermore, the study will identify the average time taken to respond to a call light, the effect of other alarms in the facility, and how the staff interacts and responds to the alarm. As nursing staff often considers call light notifications as interruptions, adverse events in the facility, how the staff is being notified about adverse events, delays in responding, the reasons for the delay, rounds, and how often the staff does rounds were observed to analyze how nursing home staff interacts with the call light technology (H Ali, A Ahmed, unpublished data, 2020), [9,41].

A hypothetico-deductive approach was used with the unstructured observation stage. A nonparticipant technique was employed where the observer watched from a distance for this stage of the study [42,43]. Observations were conducted in all units of the 4 nursing homes. The observation aimed to collect contextual information about the daily routines of nursing staff, the overall nursing home environment, and the tasks the staff has to do. Furthermore, the study seeks to gather information about the call light technology and other technologies that are implemented in nursing home settings and to conduct an analysis of ergonomic challenges these implementations faced (see Table 1).

Table 1. Methods.

Method	Aims	Setting/participants	Procedure
Interviews	To understand the routine tasks, workflow, units, incidents (adverse events), teamwork and communication strategies, technologies, strategies implemented to providing care, limitations, and challenges	Professionals from 4 different nursing homes: administrators, unit managers, program coordinators, RNs[a], CNAs[b], LPNs[c], and nursing faculty	Researchers met with nursing home professionals and took notes
Stage 1 observation	To collect information about the nursing homes' *work system* process and outcomes	Four nursing homes' staff (RNs, CNAs, and LPNs)	The hypothetico-deductive approach was used with a nonparticipant technique. Systems Engineering Initiative for Patient Safety was used to organize the observations
Stage 2 observation (task analysis)	To observe the process of locating and responding to call lights (interaction with the technology)	One nursing homes' staff (CNAs and LPNs)	A systematic approach observation was used, and a standard procedure of 4 steps was developed (see the Methods section)

[a] RN: registered nurse.

[b] CNA: certified nurse assistant.

[c] LPN: licensed practical nurse.

The second stage was conducted in one of the nursing homes for the task analysis. Approximately 90 hours of a structured, systematic observation approach was employed to collect data related to the process of using the call light system and to record and observe the process of locating and responding to call lights [42,43].

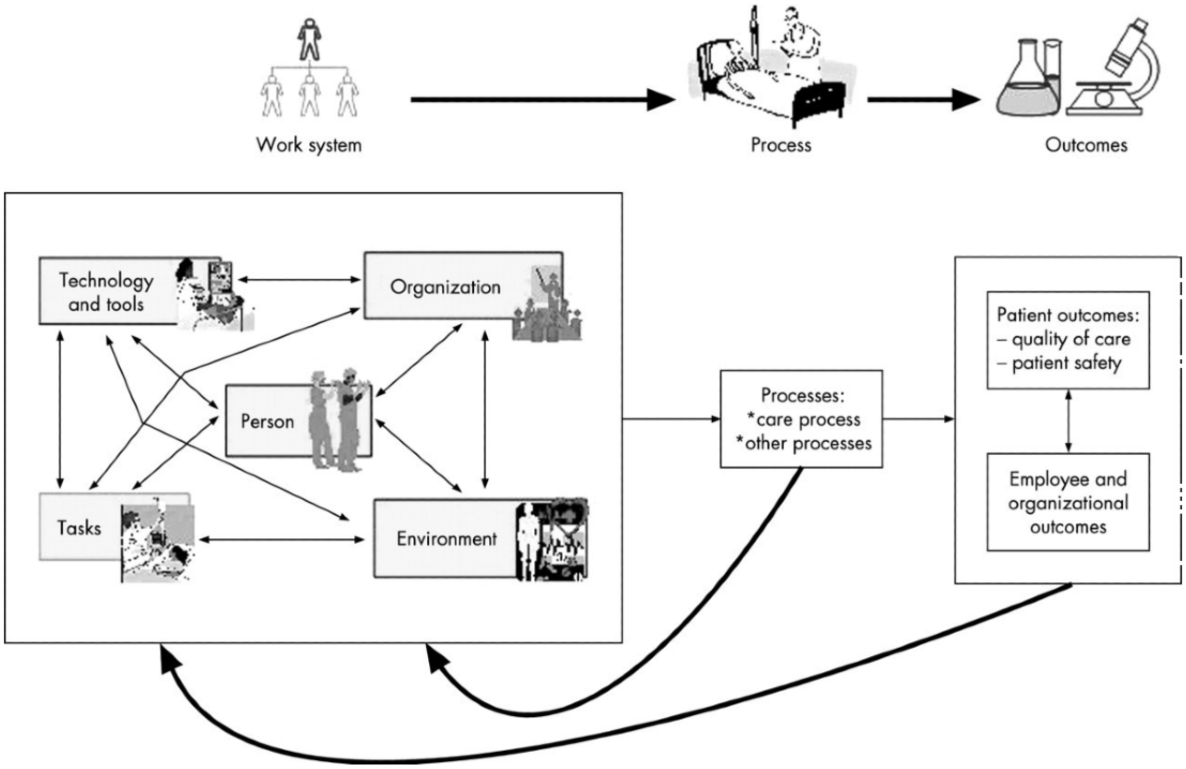
A standard procedure was used to observe how staff responded to call lights where the observers would stand near the nurses' station from where they could see the system displays. The observers would take notes and use a stopwatch to record the time for each event and observe any action the staff took after a call light alarm was triggered. The study observed (1) the actual nurse response time for each call bell (the time from the initiation of the call to the time of entering the room); (2) how the nurse located the resident—using the nurses' station display, using the light above the resident's door, or by the resident's verbalizations; (3) whether the CNA assisted the resident within the allotted response time (how long the resident had to wait); and (4) whether the CNA canceled the nurse call light at the time of response or forgot to cancel it— causing another CNA (or the same CNA) to respond to the same alarm. A total of 201 call light interactions were observed in approximately 3 months. The data were collected in both units: short-term care unit (south unit) and long-term care unit (north unit). The data were also collected during different time intervals, including before and after breakfast, before and after lunch, and before dinner. Of the 201 call light interactions, 97 were collected in the south unit, and 104 were collected in the north unit. Next, analysis of variance (ANOVA) was conducted to test whether the time of collecting the data (eg, before/after breakfast and before/after lunch) and unit type have a significant effect on the call light response time. Both stages were conducted after receiving the oral consent of the staff working at the time of data collection. The study was approved by the nursing homes and the university institutional review boards.

Settings and Participants

The observations took place in 4 skilled nursing homes located in the southern tier of upstate New York. All the nursing homes included regular long-term care units and memory units (serving residents with Alzheimer's disease and dementia or other patients deemed in need of the special care provided in such units). The number of beds in the facilities ranged from 150 to 381. Moreover, 2 of the facilities were private care units, 1 was public (by county), and a nonprofit organization ran the fourth facility. In total, 150 hours of observations were conducted during the 3 different shifts (morning: typically, 7 AM-3 PM; afternoon: 3 PM-11 PM; and night: 11 PM-7 AM). The participants in this study were nursing home staff and residents. CNAs and LPNs working in the different units of the nursing homes were all participants. Residents of nursing homes were also observed to study how they influence interactions with different technologies. The task analysis stage was conducted in 1 public 300-bed nursing home. This nursing home was chosen for this stage because there were many complaints from residents and residents' families about long response times and because of accessibility to the research team.

All relevant aspects of observations were categorized into groups and subgroups using the Systems Engineering Initiative for Patient Safety (SEIPS) model of work system and patient safety as a guide [40] (Figure 1). Observations about behavior, linguistic aspects (eg, speech), extralinguistic aspects (eg, sounds/loudness), and staff's and residents' comments were all coded and categorized using the SEIPS model. Moreover, 2 of the researchers categorized the interviews and recorded observations individually. Finally, multiple focus sessions were conducted to discuss any disagreements on the placement of notes into the main categories.

Figure 1. Systems Engineering Initiative for Patient Safety model of work system and patient safety.



Results

Work System

Staff

Nursing home staff, the primary drivers of care and daily treatments, are essential people at a nursing care facility. RNs are usually the unit managers or supervisors. They graduate from a state-accredited nursing program and have been licensed by the State Board of Nursing. They are the ones responsible for residents' care plans and the ones who assess residents in case of a fall or any other type of injury. They are also the ones to start and manage intravenously and administer any medication by vein. LPNs were originally educated and trained to work as bedside nurses in hospitals and are now working in nursing homes, rehabilitation centers, physicians' offices, schools, and clinics. Becoming a licensed practical or vocational nurse requires the completion of a formal training program plus supervised clinical instruction. They all have high school diplomas or their equivalent and have graduated from a program with a license granted by the State Board of Nursing. Their main task is to administer medication and take vital signs, but they are also required to respond to the various alarms in the unit. LPNs work under the close supervision of RNs. CNAs completed a training program at a state-approved facility. They provide most of the direct care and assist residents with their daily activities such as toileting, shower, changing clothes, transportation, and mealtime.

Organizational Characteristics

Organizational policies and standards are also important to keep residents safe. According to nursing home administrators, all staff working in the unit are required to respond to all bed and chair exit alarms and must respond to the call lights within 5 min. It was observed that LPNs often ignore call light notifications, in many cases, walking past residents in need. It was also observed that it takes more than 5 min to respond to a call light; in some cases, these responses can take up to 20 min.

In many cases, the staff enters the resident's room to ask about their needs and asks them to wait and forgets to come back, in which case, the resident had to use the call light again (see the Processes section). Many complaints were observed in one of the nursing homes regarding the long response time. Nursing home administration was considering the option of upgrading their current call light system to a system capable of monitoring response times of the staff. This information would be used to reduce the number of complaints received by reporting response times to the residents' families to assure families that the response time is within an acceptable range.

Communication across the staff working in the same unit is important for the safety of residents. However, there is no system for the staff to communicate in the same unit or between the staff in different units. If 1 CNA was busy with a resident and another resident needed help, the CNA could not leave the resident to assist another resident. In addition, CNAs, in some cases, cannot provide care to residents without having 2 people involved (eg, using a lifting machine). With the lack of communication systems, it is often difficult for the CNAs to find help; as a result, many CNAs resort to running across the unit to find an available CNA or LPN. This can cause a delay in assisting residents and might lead to an adverse event. Vertical communication within a working system is also important. CNAs have reported limited communication with the RNs on the unit because of their high workload.

Between 2 shifts, there is a 30-min overlap. The second shift, for example, starts at 2:45 PM, whereas the first shift ends at 3:15 PM. During this overlap, staff members are supposed to hand off reports (CNAs to CNAs and LPNs to LPNs) and to make their rounds. Rounding is a means of transferring information to the staff coming on shift and giving them updates regarding the residents' current health care plan. Rounding usually takes place during the overlap between shifts. Staff members are supposed to check safety mechanisms during rounding and report any changes in a resident's condition, such as changes that might require immediate attention and responses to call lights from these residents. However, the checklist is not always followed. Normally, only 2 to 3 people perform their rounds because the others are busy with tasks such as charting. CNAs, who are typically the busiest, tend to underestimate the importance of rounding and do not do it. A lack of direct communication between the 2 shifts was observed, which might result in the lack of feedback regarding the residents' conditions.

Tasks

Residents' rooms are divided into sections, and certain rooms are assigned to teams. This assignment remains unchanged for the whole month to ensure the consistency of service provided to residents. Each CNA oversees 8 to 13 residents, and each LPN oversees 9 to 14 residents. There are 3 shifts per day with 8 hours each. Day shifts have a higher level of staffing than evening and night shifts.

As a policy, when an alarm goes off, whether it is a call light, chair exit alarm, bed exit alarm, or a clip alarm, everyone is expected to respond, including housekeepers. There might be multiple staff members responding to an alarm until the assigned team is assisting the resident. Staff members communicate primarily in person. For example, 2 staff members may go to the same room to address an alarm where they may also discuss another ongoing alarm and notify each other about their plans to deal with it.

Even if the staff is notified on time, they may not be able to address the resident's needs on time. For example, if they were occupied with someone else, they cannot leave that resident until all scheduled tasks related to the resident's treatment and safety have been completed. By the time a staff member reacts to the alarm, the patient may have already fallen [5,17]. This issue is more significant in the morning. Most residents wake up between 6:45 AM and 7:45 AM. Numerous alarms can go off simultaneously when they all try to get out of bed or begin moving at the same time. The staff members must race from one room to another; they rely on their experience and try to start with residents who cannot walk or stand by themselves. However, there are typically more residents needing assistance than available CNAs during these times. At one instance, CNAs asked the observers to watch residents several times that day because of their need to stay with their current resident and the lack of available staff members to help.

The noise from the many alarms can also confuse the staff, as the noise produced from the auditory systems is almost constant. This causes discomfort for both residents and staff members. Alarm fatigue can also develop over time because of the nursing staff receiving too many nonurgent call lights [42]. Throughout their careers, nurses learn how to prioritize tasks because of the need and the time allotted in their shift. As nurses are unable to distinguish urgent alarms from the call light from the nonurgent, calls associated with the call light tend to come from nonurgent needs. This leads the nursing staff to assume that residents do not need urgent help. As nurses begin to feel more confident in how they assess patients and understand their acuity and treatment, they also learn how to work without relying on the call light. This is the beginning of what develops into alarm fatigue, as learning how to work without relying on the system can cause nurses to ignore the alarms. Although using a pager can reduce the noise and display the room number, because of the fast-pace routine performed by nursing home staff members, most are too busy providing round the clock service for residents to grab the pager and read the room number.

Nursing homes are often understaffed, and nurses' stations may not always be occupied. As nursing staff is always on the move to care for residents, this makes using a centralized monitoring system difficult. Owing to the tight schedule and understaffing, although CNAs are working with residents, they may not be able to provide time for rounding as planned. As a result, residents with risks may not be checked as often as prescribed. CNAs must either stay late after their shift is over to enter updates to the EHR system or use their breaks and become exhausted because of working up to 12 hours without a pause.

In addition, a single staff member may be assigned to multiple residents at the same time, and this can be particularly stressful when more than one resident needs attention, such as instances whereby more than one resident must be transported. Although working on their own, these logistical hurdles can create chaotic moments for both the nursing home staff and their residents. Further details are included in the summary of observations in Multimedia Appendix 1.

Tools and Technologies

A total of 3 different call light systems were observed. Included in each system were nurse station consoles, which triggered an alarm to indicate a call light. All systems have lights placed above the residents' doors, and the lights come in 2 colors (white and either red or orange). These lights help in locating the senior using the call light. The white light indicates that the resident is asking for assistance from the bedroom, whereas the red or orange light indicates that the resident is asking for assistance from the bathroom.

The first call light system did not display the room number; instead, it only triggered an auditory alarm that is also broadcasted over speakers throughout the units. The auditory alarm uses a series of *beeping* at 2 levels of speed: slow beeping indicates that the call is coming from the bedroom, and fast beeping indicates that the call is coming from the bathroom. At the ceiling above the nurse station, there is a group of 4 lights, which indicates 4 different areas in the unit. These lights turn on in response to a call light. The staff must locate the area first and then locate the resident by identifying the lights above the doors. If the lights for multiple areas are on, then the staff must check the lights above the doors in those areas.

The second call light system observed was also used at the nurses' station console. The console could display the room number to help the CNA locate the room easily; however, it could only display 1 room number at a time. If a new alarm were triggered, it would override the previous room number. In this system, the only auditory warning triggered is at the nurses' stations. If the staff was working down the hallway or inside the residents' rooms, they could not hear the alarm. Furthermore, there is no auditory distinction between a call bell in the bedroom or the bathroom, with the latter often associated with more urgent needs. It was observed that the console was muted in 1 unit and was covered with a file in another unit in the same nursing home.

The third call light observed was a pager-based system that displays the room numbers and emits an auditory *beeping* sound. In this system, the nurse station console can also display the room number and emit an auditory alarm. The system, however, can only display 1 room number at a time. In cases of more than 1 alarm, the console switches between the room numbers, displaying each room number for approximately 5 seconds before switching to display the next room number. To clear the override, the staff must cancel the alarm from inside the resident's room and from the pager (Multimedia Appendix 2).

Call light systems include visual and auditory alarms. The visual alarms may indicate a location, whereas the auditory alarms are not directional, as the sounds are emitted through loudspeakers located along the hallway walls or broadcasted directly from the nurse station.

Usability Issues of Call Light Systems

There are many usability issues in the current call light system. First, the staff is often unable to find a break in their tasks to contribute to monitoring the central display of the call light. Some systems could only handle 1 alarm at a time. If 2 or more residents triggered an alarm, room numbers would not display until the first one was resolved or the system would display the newest alarm and cover the previous one. In both cases, the nursing home staff have no access to feedback information about the number of alarms in the unit, if the alarms were resolved, and whether the residents were properly assisted. Not distinguishing between the alarms from bedrooms and the alarms from bathrooms (which often reflects more urgent needs) is another usability issue. In addition to reporting to the nurse station to receive the notification, the staff must also look down the hall to distinguish the alarms by using the lights above the residents' room doors.

Furthermore, the nursing home staff are working inside the resident's room most of the time. As a result, they lack access to hear the alarm when the system alerts them from the nurse station. In addition, systems that broadcast alarms using loudspeakers and other alarms in the unit constantly adjust the noise level in the nursing home to a high pitch. Although this is not desirable for the nursing home staff, it is even worse for the residents; the response time to call lights is long because the nurse station is not staffed most of the time. Using a pager can reduce the noise and display the room number, but the nursing home staff is always busy providing service for residents, and they cannot grab the pager to see the room number. Furthermore, the auditory sound is not directional and, thus, is unhelpful for locating the resident's alarm. As a result, many staff members mute their pagers or leave them at the nurse station. CNAs tend to locate the room by looking at lights on the top of the doors, which can also be challenging at times because of the layout. The timing of the alarms also causes problems. Usually, when the alarm goes off, the patient may have already fallen. For example, a resident who tries to stand up but is unable to support himself or herself may lean forward and fall.

Broken parts were also the main usability issue impeding the systems. In some units, there were many broken and nonworking lights. When a resident pushed the call light button, the auditory sound would be triggered, but the light might not have worked. In some cases, because the systems were very old, the lights would turn on, but no auditory alarm would be heard (see the summary of usability issues in Multimedia Appendix 2).

Physical Environment

The unit layouts for the facilities observed in this study were L- or T-shaped with 1 nurse station. Residents who have high risks or more critical conditions stay in the rooms closest to the nurse station.

The auditory alarms are active for most of the day. This can be obnoxious to live-in residents, as it often disrupts their rest and distracts them from their activities. The study finds that this ultimately works against the nursing home's goal of creating a comfortable environment.

The floor layout and other aspects of the physical environment are poorly designed, making it difficult for the staff to know where an alarm (eg, a call bell) is signaling danger. The auditory alarms were broadcasted without any directional information. Beams and doorframes blocked certain alarms from being seen. Furthermore, door lights did not always work, which required staff to walk to the middle of the hallway. More than 90% of times, CNAs tended to locate the room by looking at lights on the top of the door; they have to walk to the middle of the hallway to see the light, which contributes to additional workload caused by usability issues.

Having many auditory notifications in nursing homes, such as bed mats, chair mats, call light systems, and the Wander Gard system, makes nursing home environments noisy and uncomfortable for many residents to live in. During our observations, many residents complained about constantly hearing the notifications and alarms throughout the day. Consequently, the staff members sometimes had to mute the call light system in some units, which posed challenges for notifying and responding to a call light. The impact of this noisy environment is not confined to residents but also reaches and affects the staff's performance (see the summary of the observation in Multimedia Appendix 1).

Processes

The task analysis methodology was used to analyze the task of being notified and responding to the call light system and the actions the staff took to finish the task.

The results showed that the highest response time was before dinner, and this was true in both units. ANOVA results showed that the unit type had no significant effect on overall response time (*P*=.85). However, timing has a significant effect on the overall response time (*P*=.01), whereas the interaction between the unit type and the time interval had no significant effect on the response time (*P*=.90; ANOVA table provided in Multimedia Appendix 3). In the later analysis, the response time data in both units were combined because of there being no significant difference between them.

According to the collected data, the average response time to a call light in both units was 9 min, and these responses were the longest before dinner time. Responses to call lights in the morning were also long; around 50% of the time, CNAs responded after 5 min, which exceeded the administration's expectations or standards.

The staff did not use all the features in the system because of usability issues, as the console was muted in the south unit. CNAs had to track the lights above the rooms' doors all the time. More than 16% of the time, CNAs forgot to cancel the alarm after they responded in this unit because there was no auditory alarm. This could cause redundancy of work if another nurse noticed the light and responded to the same resident. More than 80% of the time, CNAs were notified by the light. All the cases when CNAs were notified by the display or console took place when they were near the nurse station where they could hear and see the display. In more than 7.5% (15/201) of the cases, because of broken parts in the system, CNAs were notified by the residents. After pushing the button for assistance, if no one responded to them, residents tended to stand and attempt to help themselves or to go the room door to look out and shout angrily for a nurse's attention. In more than 10.0% (20/201) of call lights, CNAs responded and asked the resident to wait until they were done with someone else. Around 3% of the time, they forgot to come back to assist the first resident.

Discussion

Communication in Nursing Homes

Approximately 34,000 fatal, life-threatening, or serious adverse events per year occur in nursing home settings, and most of these events are considered preventable [44,45]. A lack of effective communication policies at all levels of nursing homes was found to have contributed to these events [44,46,47]. Research in other health care settings suggests that the call light system is the link between nurses and patients during hospital stays [13]. It is also recognized as a crucial piece of technology for patient-centered care models [48]. It was also found to highly influence the satisfaction levels of nursing home staff and patients [8,9] and highly affect the safety of patients [10,49].

In nursing homes, the call light system is the means of the initial communication between staff and residents. We endeavored to understand how nursing home staff use the system by observing their interactions and identifying the usability issues associated with the call light systems that prevent the staff from providing quality health care.

Principal Findings

In this study, it was observed that nursing home staff that interact with the call light systems are often CNAs. However, according to nursing home administrations, all staff are responsible and supposed to respond to all types of alarms in the unit, such as bed or chair exit, clip alarm, weight sensors, and call light. Device alarms are intended to alert and inform nursing home staff of any changes in the residents' conditions. However, the staff is subjected to too many alarms that disturb their workflow and might lead to many errors, which can contribute to mistrust and long response times [42,50]. CNAs provide most of the direct care to residents and respond to most of the alarms in the unit. However, around 50% of the time, they responded after 5 min, which exceeded the administration's expectations and standards. A high alarm rate limits CNAs' ability to manage the call system properly, resulting in negative perceptions and alarm fatigue by nursing home staff. Having a high number of alarms can cause the staff to ignore the call bell occasionally. The staff was notified about call lights by the residents themselves. Around 7.5% (15/201) of these instances involved the residents getting out of their bed or chair and going to the room door to ask for help. The study finds that this is because the CNAs hold negative perceptions of residents' use of the call light because of frequent, nonurgent use. This ultimately causes misconceptions about the purpose of call lights because of their frequent use [51]. It was also observed that staff members often muted the call light system in some units, which posed a challenge for proper notification and increased the response time to a call light. Many broken parts were observed in 3 facilities, and this may have caused malfunctions in the call light system, resulting in adverse events and contributing to longer overall response time. These delays in responding ultimately increased the risk of residents harming themselves because of an unanswered alarm.

The physical environment of nursing homes was also found to be an issue. Nursing homes attempt to create a home-like environment; however, they are designed in hospital layouts, with L- or U-shaped units and with multiple bedrooms located on long double-loaded hallways. The call light system has one console placed at the nurse station in the units that are unattended most of the time and lacks any directional information. In more than 70% (86/201) of the observed interactions, the staff located the alarms using the lights above the doors because the consoles were not visible or accessible to them. This required that they had walked out of a resident's room or away from the nurse station. In addition, CNAs are spending most of their time in residents' rooms providing care, and as a result, they are oblivious to the alerts of the call light, console, and auditory alarm.

Understaffing is one of the main problems in nursing homes. Thus, the design of technology should try to overcome this problem by providing solutions that are less staff dependent or require less effort from the staff. Due to understaffing, protocols such as rounding may not be followed strictly. Thus, the system should provide important information redundantly to ensure it is received by staff.

The staff did not use all features of the call light systems because of usability issues. CNAs had to look and keep track of the lights all the time. More than 16% of the time, CNAs forgot to cancel the alarm after they responded because there was no auditory alarm because of broken parts or because of continued auditory alarms in the unit. This could cause redundancy of work if another nurse noticed the light and responded to the same resident. More than 80% of the time, CNAs were notified by the light. All the cases when CNAs were notified by the display or console took place when they were near the nurse station area where they could hear and see the display. In many cases, because of broken parts in the system, CNAs were notified by the residents. After pushing the button for assistance, if no one responds to them, they tend to stand to help themselves or go the room door to look and shout angrily asking for help.

In many cases, CNAs had to prioritize when more than one call light was triggered simultaneously or when the interval between call lights was only slightly different. Sometimes, they responded by asking the residents to wait until they were done with someone else. Around 3% of the time, they forgot to come back to assist the first resident. This is because of the staff's high workload and busy schedule [14,52], which might also contribute to the long response time and not following protocols such as rounding and muting the system.

Our analysis of the call lights in nursing homes showed that most call lights were used before and after mealtimes. The average response time to a call light was 9 min, and the wait was the longest before dinnertime. This is because residents were tired at the end of the day and were ready to sleep. Response time was also long in the morning because residents tend to walk up at the same time, and everyone needs to get to the bathroom, get dressed, and have breakfast.

Conclusions

In conclusion, the call light system is critical for interactions between the nursing home staff and residents. Research conducted in other health care settings has demonstrated that the call light system not only significantly improves the communication between staff and patients together but also helps ensure the safety of patients. Nursing homes still using old call light systems risk the continuation of usability issues that can affect the performance of the staff and contribute to a decline in staff and resident outcomes. Although the health care industry has been at the forefront of technological advancements and implementation, it is important to recognize the importance of considering the perceptions of end users in the development of new medical technology and the overall quality of health care service delivery. By incorporating feedback from nurses, nursing home management will better understand the influence that perceptions and usability of technology have on the quality of health care for their residents. In this study, it has been observed that the call light system is perceived to be an important factor affecting the outcomes of the care process and the satisfaction of both residents and staff as well in addition to the staff's performance. Staff shortage, mental and memory conditions of residents, usability issues associated with the call light system, and nursing home layout can also contribute to long response time, which might affect the safety of the residents of nursing homes.

The insights from this work were used to design a smart notification and communication system for nursing homes [3]. The smart system was designed to overcome the challenges the staff members face and aimed to improve their performance. In addition, the effects of a call light system on nursing homework system, the staff's perception about the call light system, and the process of responding to a call light will be thoroughly investigated in a future publication for more understanding and evaluation of the contribution of a call light system to safety and quality of care in nursing homes.

Acknowledgments

This study received funding from the College of Liberal Arts at the Auburn University and the Binghamton University.

Authors' Contributions

HA and HL contributed to the design of the research methodology and conducted the first stage (the observations stage). HA collected the data for the second stage, analyzed all data, wrote the manuscript, submitted the manuscript, and made the critical revisions in the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of observations.
[PDF File (Adobe PDF File), 130 KB - jmir_v22i3e16252_app1.pdf]

Multimedia Appendix 2
Description of the call light systems observed and the usability challenges associated with each system.
[PDF File (Adobe PDF File), 113 KB - jmir_v22i3e16252_app2.pdf]

Multimedia Appendix 3
Analysis of variance.
[PDF File (Adobe PDF File), 144 KB - jmir_v22i3e16252_app3.pdf]

References

1. Centers for Disease Control and Prevention (CDC). United States; 2007. National Center for Health Statistics. Health, United States - 2006 With Chartbook on Trends in the Health of Americans With Special Feature on Pain URL: https://www.cdc.gov/nchs/hus/index.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fnchs%2Fhus.htm [accessed 2019-08-09]
2. Castle NG, Ferguson JC. What is nursing home quality and how is it measured? Gerontologist 2010 Aug;50(4):426-442 [FREE Full text] [doi: 10.1093/geront/gnq052] [Medline: 20631035]
3. Ali H, Li H. Evaluating a smartwatch notification system in a simulated nursing home. Int J Older People Nurs 2019 Sep;14(3):e12241 [FREE Full text] [doi: 10.1111/opn.12241] [Medline: 31099184]
4. Lee HY, Blegen MA, Harrington C. The effects of RN staffing hours on nursing home quality: a two-stage model. Int J Nurs Stud 2014 Mar;51(3):409-417. [doi: 10.1016/j.ijnurstu.2013.10.007] [Medline: 24182619]
5. Krauss MJ, Evanoff B, Hitcho E, Ngugi KE, Dunagan WC, Fischer I, et al. A case-control study of patient, medication, and care-related risk factors for inpatient falls. J Gen Intern Med 2005 Feb;20(2):116-122 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40171.x] [Medline: 15836543]
6. Li H, Ali H. Human factors considerations in the design of falls prevention technologies for nursing homes: a case study. Proc Int Symp Hum Factors Ergon Healthc 2015;4(1):97-102. [doi: 10.1177/2327857915041020]
7. Galinato J, Montie M, Patak L, Titler M. Perspectives of nurses and patients on call light technology. Comput Inform Nurs 2015 Aug;33(8):359-367 [FREE Full text] [doi: 10.1097/CIN.0000000000000177] [Medline: 26176639]
8. Cardoso G, Martin N. The patient connection: improving call bell response at Cambridge Memorial Hospital. Healthc Manage Forum 2003;16(3):47-50. [doi: 10.1016/s0840-4704(10)60235-9]
9. Deitrick L, Bokovoy J, Stern G, Panik A. Dance of the call bells: using ethnography to evaluate patient satisfaction with quality of care. J Nurs Care Qual 2006;21(4):316-324. [doi: 10.1097/00001786-200610000-00008] [Medline: 16985401]

10. Meade CM, Bursell AL, Ketelsen L. Effects of nursing rounds: on patients' call light use, satisfaction, and safety. Am J Nurs 2006 Sep;106(9):58-70; quiz 70. [doi: 10.1097/00000446-200609000-00029] [Medline: 16954767]
11. Ali H, Li H, Wong J. Evaluating a smartwatch-based notification system in a simulated nursing home: the method. Proc Int Symp Hum Factors Ergon Healthc 2017 May 15;6(1):132-138. [doi: 10.1177/2327857917061029]
12. Tzeng HM, Yin CY. The extrinsic risk factors for inpatient falls in hospital patient rooms. J Nurs Care Qual 2008;23(3):233-241. [doi: 10.1097/01.NCQ.0000324588.91837.84] [Medline: 18562866]
13. Tzeng HM, Yin CY. Predicting patient satisfaction with nurses' call light responsiveness in 4 US hospitals. J Nurs Adm 2010 Oct;40(10):440-447. [doi: 10.1097/NNA.0b013e3181f2eb29] [Medline: 20859095]
14. Chung G. Understanding nursing home worker conceptualizations about good care. Gerontologist 2013 Apr;53(2):246-254. [doi: 10.1093/geront/gns117] [Medline: 22936538]
15. Schoville RR, Titler MG. Guiding healthcare technology implementation: a new integrated technology implementation model. Comput Inform Nurs 2015 Mar;33(3):99-107; quiz E1. [doi: 10.1097/CIN.0000000000000130] [Medline: 25799235]
16. Lehoux P, Daudelin G, Williams-Jones B, Denis J, Longo C. How do business model and health technology design influence each other? Insights from a longitudinal case study of three academic spin-offs. Res Policy 2014;43(6):1025-1038. [doi: 10.1016/j.respol.2014.02.001]
17. Ali H, Li H. Designing a Smart Watch Interface for a Notification and Communication System for Nursing Homes. In: International Conference on Human Aspects of IT for the Aged Population. 2016 Presented at: ITAP'16; July 17-22, 2016; Toronto, Canada p. 401-411 URL: https://link.springer.com/chapter/10.1007/978-3-319-39943-0_39#citeas [doi: 10.1007/978-3-319-39943-0_39]
18. While A, Dewsbury G. Nursing and information and communication technology (ICT): a discussion of trends and future directions. Int J Nurs Stud 2011 Oct;48(10):1302-1310. [doi: 10.1016/j.ijnurstu.2011.02.020] [Medline: 21474135]
19. Rahman AN, Applebaum RA, Schnelle JF, Simmons SF. Translating research into practice in nursing homes: can we close the gap? Gerontologist 2012 Oct;52(5):597-606 [FREE Full text] [doi: 10.1093/geront/gnr157] [Medline: 22394494]
20. Wiener JM. An assessment of strategies for improving quality of care in nursing homes. Gerontologist 2003 Apr;43(Spec No 2):19-27. [doi: 10.1093/geront/43.suppl_2.19] [Medline: 12711721]
21. Freedman VA, Calkins M, Derosiers R, van Haitsma K. Office of The Assistant Secretary for Planning and Evaluation. 2005 Dec 10. Barriers to Implementing Technology in Residential Long-term Care Settings URL: https://aspe.hhs.gov/basic-report/barriers-implementing-technology-residential-long-term-care-settings [accessed 2016-05-16]
22. Anderson RA, Issel LM, McDaniel RR. Nursing homes as complex adaptive systems: relationship between management practice and resident outcomes. Nurs Res 2003;52(1):12-21 [FREE Full text] [doi: 10.1097/00006199-200301000-00003] [Medline: 12552171]
23. Castle NG, Wagner LM, Ferguson JC, Handler SM. Nursing home deficiency citations for safety. J Aging Soc Policy 2011 Jan;23(1):34-57 [FREE Full text] [doi: 10.1080/08959420.2011.532011] [Medline: 21207305]
24. Bostick JE, Rantz MJ, Flesner MK, Riggs CJ. Systematic review of studies of staffing and quality in nursing homes. J Am Med Dir Assoc 2006 Jul;7(6):366-376. [doi: 10.1016/j.jamda.2006.01.024] [Medline: 16843237]
25. Bonner AF, Castle NG, Men A, Handler SM. Certified nursing assistants' perceptions of nursing home patient safety culture: is there a relationship to clinical outcomes? J Am Med Dir Assoc 2009 Jan;10(1):11-20. [doi: 10.1016/j.jamda.2008.06.004] [Medline: 19111848]
26. Phillips VL, Roberts DY, Hunsaker AE. Certified nursing aides' and care assistants' views on falls: insight for creation and implementation of fall prevention programs. J Am Med Dir Assoc 2008 Mar;9(3):168-172. [doi: 10.1016/j.jamda.2007.11.012] [Medline: 18294599]
27. Abramson EL, McGinnis S, Moore J, Kaushal R, HITEC investigators. A statewide assessment of electronic health record adoption and health information exchange among nursing homes. Health Serv Res 2014 Feb;49(1 Pt 2):361-372 [FREE Full text] [doi: 10.1111/1475-6773.12137] [Medline: 24359612]
28. Handler S, Sharkey SS, Hudak S, Ouslander JG. Incorporating INTERACT II clinical decision support tools into nursing home health information technology. Ann Longterm Care 2011;19(11):23-26 [FREE Full text] [Medline: 22267955]
29. Munyisia EN, Yu P, Hailey D. Does the introduction of an electronic nursing documentation system in a nursing home reduce time on documentation for the nursing staff? Int J Med Inform 2011 Nov;80(11):782-792. [doi: 10.1016/j.ijmedinf.2011.08.009] [Medline: 21956002]
30. Pillemer K, Meador RH, Teresi JA, Chen EK, Henderson CR, Lachs MS, et al. Effects of electronic health information technology implementation on nursing home resident outcomes. J Aging Health 2012 Feb;24(1):92-112 [FREE Full text] [doi: 10.1177/0898264311408899] [Medline: 21646551]
31. Qian S, Yu P, Hailey DM. The impact of electronic medication administration records in a residential aged care home. Int J Med Inform 2015 Nov;84(11):966-973. [doi: 10.1016/j.ijmedinf.2015.08.002] [Medline: 26358850]
32. Carayon P, Xie A, Kianfar S. Making Health Care Safer II: An Updated Critical Analysis of the Evidence for Patient Safety Practices. 2013. Chapter 31. Human Factors and Ergonomics URL: https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/evidence-based-reports/services/quality/patientsftyupdate/ptsafetyII-full.pdf [accessed 2020-01-27]
33. Institute OM. Health IT and Patient Safety: Building Safer Systems for Better Care. Washington, DC: National Academies Press; 2012.

34. Carayon P. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. Second Edition. Boca Raton, FL: CRC Press; 2012.
35. Carayon P, Hundt AS, Alvarado CJ, Springman S, Borgsdorf A, Jenkins L. Implementing a systems engineering intervention for improving safety in outpatient surgeries. In: AHRQ Henriksen K, AHRQ Battles JB, Marks ES, AHRQ Lewis DI, editors. Advances in Patient Safety: From Research to Implementation (Volume 3: Implementation Issues). Rockville, MD: Agency for Healthcare Research and Quality; 2005:301-321.
36. Karsh B, Holden RJ, Alper SJ, Or CK. A human factors engineering paradigm for patient safety: designing to support the performance of the healthcare professional. Qual Saf Health Care 2006 Dec;15 Suppl 1:i59-i65 [FREE Full text] [doi: 10.1136/qshc.2005.015974] [Medline: 17142611]
37. Jeffcott SA, Ibrahim JE, Cameron PA. Resilience in healthcare and clinical handover. Qual Saf Health Care 2009 Aug;18(4):256-260. [doi: 10.1136/qshc.2008.030163] [Medline: 19651927]
38. Gurses AP, Winters B, Pennathur PR, Carayon P, Pronovost PJ. Human factors and ergonomics in intensive care units. In: Carayon P, editor. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. Second Edition. Boca Raton: CRC Press; 2012.
39. Norman D, Draper SW. User Centered System Design: New Perspectives on Human-Computer Interaction. Boca Raton, Florida: CRC Press; 1986.
40. Carayon P, Hundt AS, Karsh B, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15(Suppl 1):i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
41. Edwards C. Using interdisciplinary shared governance and patient rounds to increase patient safety. Medsurg Nurs 2008 Aug;17(4):255-257. [Medline: 18807861]
42. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010 Jan;19(1):28-34; quiz 35. [doi: 10.4037/ajcc2010651] [Medline: 20045845]
43. Siebig S, Kuhls S, Imhoff M, Gather U, Schölmerich J, Wrede CE. Intensive care unit alarms--how many do we need? Crit Care Med 2010 Feb;38(2):451-456. [doi: 10.1097/CCM.0b013e3181cb0888] [Medline: 20016379]
44. Field TS, Tjia J, Mazor KM, Donovan JL, Kanaan AO, Harrold LR, et al. Randomized trial of a warfarin communication protocol for nursing homes: an SBAR-based approach. Am J Med 2011 Feb;124(2):179.e1-179.e7. [doi: 10.1016/j.amjmed.2010.09.017] [Medline: 21295198]
45. Gurwitz JH, Field TS, Judge J, Rochon P, Harrold LR, Cadoret C, et al. The incidence of adverse drug events in two large academic long-term care facilities. Am J Med 2005 Mar;118(3):251-258. [doi: 10.1016/j.amjmed.2004.09.018] [Medline: 15745723]
46. Husebo BS, Flo E, Aarsland D, Selbaek G, Testad I, Gulla C, et al. COSMOS--improving the quality of life in nursing home patients: protocol for an effectiveness-implementation cluster randomized clinical hybrid trial. Implement Sci 2015 Sep 15;10:131 [FREE Full text] [doi: 10.1186/s13012-015-0310-5] [Medline: 26374231]
47. Kane RA. Definition, measurement, and correlates of quality of life in nursing homes: toward a reasonable practice, research, and policy agenda. Gerontologist 2003 Apr;43(Spec No 2):28-36. [doi: 10.1093/geront/43.suppl_2.28] [Medline: 12711722]
48. Duffy JR, Kooken WC, Wolverton CL, Weaver MT. Evaluating patient-centered care: feasibility of electronic data collection in hospitalized older adults. J Nurs Care Qual 2012;27(4):307-315 [FREE Full text] [doi: 10.1097/NCQ.0b013e31825ba9d4] [Medline: 22677911]
49. Tzeng HM. Using multiple data sources to answer patient safety-related research questions in hospital inpatient settings: a discursive paper using inpatient falls as an example. J Clin Nurs 2011 Dec;20(23-24):3276-3284. [doi: 10.1111/j.1365-2702.2010.03681.x] [Medline: 21585575]
50. Schmid F, Goepfert MS, Kuhnt D, Eichhorn V, Diedrichs S, Reichenspurner H, et al. The wolf is crying in the operating room: patient monitor and anesthesia workstation alarming patterns during cardiac surgery. Anesth Analg 2011 Jan;112(1):78-83. [doi: 10.1213/ANE.0b013e3181fcc504] [Medline: 20966440]
51. Tzeng HM. Perspectives of staff nurses of the reasons for and the nature of patient-initiated call lights: an exploratory survey study in four USA hospitals. BMC Health Serv Res 2010 Feb 26;10:52 [FREE Full text] [doi: 10.1186/1472-6963-10-52] [Medline: 20184775]
52. Chui MA, Mott DA, Maxwell L. A qualitative assessment of a community pharmacy cognitive pharmaceutical services program, using a work system approach. Res Social Adm Pharm 2012;8(3):206-216 [FREE Full text] [doi: 10.1016/j.sapharm.2011.06.001] [Medline: 21824822]

Abbreviations

ANOVA: analysis of variance
CNA: certified nurse assistant
EHR: electronic health record
LPN: licensed practical nurse
RN: registered nurse

SEIPS: Systems Engineering Initiative for Patient Safety

Edited by G Eysenbach; submitted 13.09.19; peer-reviewed by A Ahmed, F Irshaidat, R Cruz Martínez; comments to author 05.10.19; revised version received 16.10.19; accepted 27.12.19; published 27.03.20.

Please cite as:
Ali H, Li H
Use of Notification and Communication Technology (Call Light Systems) in Nursing Homes: Observational Study
J Med Internet Res 2020;22(3):e16252
URL: http://www.jmir.org/2020/3/e16252/
doi:10.2196/16252
PMID:32217497

Original Paper

Effects of Positive Language and Profession on Trustworthiness and Credibility in Online Health Advice: Experimental Study

Lars König[1], Dr rer nat; Regina Jucks[1], Dr phil

[1]Department of Psychology, University of Münster, Münster, Germany

Corresponding Author:
Lars König, Dr rer nat
Department of Psychology
University of Münster
Fliednerstr 21
Münster, 48149
Germany
Phone: 49 2518339459
Email: lars.koenig@uni-muenster.de

Abstract

Background: When searching for health information, many people use the internet as their first source of information. In online health forums, for example, users post their questions and exchange health advice. In recent years, information givers from various professions have begun to use positive language (indicated by the frequent use of positively valenced adjectives) to communicate their information and persuade their audiences.

Objective: The goal of the current study was to answer the following research questions: (1) How does positive language, in comparison to neutral language, influence the trustworthiness of a person arguing in an online health forum and the credibility of their health claims; (2) How does working for a university, compared to working for a lobbying organization, influence the trustworthiness of a person arguing in an online health forum and the credibility of their health claims; and (3) Do the two factors of language style and professional affiliation interact with each other to influence trustworthiness and credibility judgments?

Methods: In a 2 × 2 between-subject experiment, 242 participants read a post from an online health forum and subsequently rated the trustworthiness of the forum post author and the credibility of their information. Within the post, the professional affiliation (scientist vs lobbyist) and language style (neutral vs positive) of the forum post author was varied.

Results: When the forum post author used a positive language style, they were perceived as less trustworthy (high Machiavellianism [P<.001; η^2_p=.076], low Integrity [P=.001; η^2_p=.045], and low Benevolence [P=.02; η^2_p=.025]) and their information was perceived as less credible (low Message Credibility [P=.001; η^2_p=.045]). The professional affiliation of the forum post author did not influence their trustworthiness or the credibility of their information.

Conclusions: When searching for health information, information seekers evaluate the language style of forum posts to decide whether forum post authors are trustworthy and their information is credible.

(J Med Internet Res 2020;22(3):e16685) doi:10.2196/16685

KEYWORDS

health communication; information-seeking behavior; language; occupations; trust

Introduction

Evaluating the Validity of Online Health Information

When searching for health information, many people use the internet as their first source of information [1,2]. In online health forums, for example, users write about their symptoms and get treatment advice from other users [3]. Furthermore, online health forums can provide patients with emotional as well as informational support [4]. However, online health information is not always accurate and often contains misinformation [5-8]. Thus, health information seekers must decide whether they should rely on health claims they encounter online. Based on the reasoning of the content-source integration model [9,10], one strategy to make such decisions is to evaluate the credibility of the provided information (eg, “Is the health claim logical, coherent and compatible with my prior knowledge?”) and the trustworthiness of the information source (eg, “Are there any reasons why the information source might lie?"). Since health information claims can be highly complex, and most people have just a limited understanding of science [11,12], it is often difficult for information seekers to evaluate health claims accurately. In such situations, information seekers base their credibility and trustworthiness judgments on factors that surround health claims. Two factors that seem especially important are the professional affiliation of an information source and the information source's language style [13-23].

For example, one study showed that unfamiliar health information is rated as being more credible if the information source has a professional affiliation indicating they are a medical expert (eg, Dr. William Blake, HIV specialist) rather than a nonexpert (eg, Tim Alster, a high school freshman) [22]. Another study found that health experts are perceived as less trustworthy and their information as less credible if their professional affiliation suggests a potential conflict of interest (eg, experts working for organic food lobbying organizations who argue that organic food is superior to conventional food and base their argumentation on studies that they have conducted themselves) [21]. In the context of language styles, it has been shown that information sources who aggressively communicate health information are judged to be less trustworthy, and their information is deemed less credible [19]. Aggressive language, however, is not the only language style that influences trustworthiness and credibility ratings: enthusiastic [20], technical [17], tentative [14], and conversational [23] language styles also influence the credibility of online health information and the trustworthiness of information sources. Furthermore, it has been shown that language and word choices do not just influence credibility and trustworthiness judgments, but other outcomes as well [24-26]. For example, medical students answer emotional patient queries more emotionally [24], and the use of narratives can influence risk perceptions [25].

How Does Positive Language Influence the Credibility of Health Claims and the Trustworthiness of Health Communicators?

Using positive language, indicated by the implementation of positively valenced adjectives (eg, great, amazing, outstanding), is another language style that seems to be on the rise when it comes to communicating health and science information. For example, in an article for the National Academy of Sciences, the authors did not just neutrally write about an antioxidant. Instead, they chose to write about an "outstanding antioxidant" [27]. Furthermore, in more recent scientific articles, interested readers can learn about "fascinating fasciclins" [28] and "the amazing world of bacteriophage diversity" [29]. These are just three of many examples from scientists who use positive language to communicate their findings. Interestingly, when turning to the popular press, authors seem to become even more eager to use positive language to communicate their advice. Here, health information seekers can read about "amazing diet recipes for weight loss" [30], "genius health tips from around the world" [31], and "an incredible therapy for modern day conditions" [32].

Authors who use such positive language might want to stress the importance of their advice or the quality of their research. However, is this a reasonable strategy in the context of health communication? Moreover, how does positive language influence the trustworthiness of health communicators and the credibility of their health claims? To our knowledge, no research has addressed these questions so far. However, it is known from previous research that businesses often use positive language to generate a favorable view of their prospects and performances [33-35], and this technique seems to be effective in some circumstances. In one study, for example, participants saw hard-to-understand financial disclosure statements that were written in either a positive or a neutral language style [34]. If the participants had little financial knowledge and saw the positively written statement, they indeed thought that the company in question would have higher earnings in the future. Other research that explored the effects of enthusiastic language has shown that expressing too much enthusiasm about a topic (eg, "And what I can tell you at the beginning: I think the topic is fascinating!") might backfire and decrease credibility and trustworthiness [20]. However, another study found that listening to an enthusiastic version of a podcast, in comparison to a neutral version of the same podcast, resulted in more positive instructional quality ratings: Participants who listened to the enthusiastic version rated it as more interesting, and they perceived the podcast host as more trustworthy [36]. Furthermore, in line with language expectancy theory [37], the credibility of an information source might moderate whether using a positive language style is appropriate or not: High-credibility sources like scientists might have the freedom to choose between different language styles without putting their trustworthiness at risk. Low-credibility sources like lobbyists, on the other hand, might not have the same freedom of choice when it comes to language styles.

Hence, diverse effects of positive language seem to be possible. If an information source uses positive language when writing about the effectiveness of a specific drug, information seekers can interpret this language choice in different ways. On the one hand, information seekers might conclude that the positive language style shows that the information source is highly convinced of the effectiveness of the drug and wants to express their excitement. In this case, the positive language style might function as a quality cue that increases credibility and trustworthiness. On the other hand, the positive language style might remind information seekers of commercials that are designed to persuade their audiences and increase sales. In this case, the positive language style might function as a negative cue that decreases credibility and trustworthiness. Furthermore, one might argue that positive language is more likely to be perceived as a quality cue if a scientist uses it because scientists are typically not interested in increasing the sales of a specific drug. However, one could also argue that people expect a scientist to be neutral and objective, and therefore the use of positive language will be perceived as a negative cue.

This example and the previously discussed literature illustrate two things: (1) many authors with different professional affiliations use positive language, indicated by the implementation of positively valenced adjectives, to communicate health information; and (2) little is known about how such positive language influences the trustworthiness of health communicators and the credibility of their health claims. Therefore, we designed an experiment to investigate how different language styles and different professional affiliations influence the trustworthiness of health communicators and the credibility of their health claims. During the experiment, participants read a post from an online health forum and subsequently rated the trustworthiness of the forum post author and the credibility of their information. Within the forum post, we varied the professional affiliation of the forum post author (whether the person was a scientist or a lobbyist) and their language style (whether they used neutral or positive language). Because the previously discussed research has shown that positive language can have either positive or negative effects, no directional hypotheses were stated. Instead, the goal was to answer the following nondirectional research questions: (1) How does positive language, in comparison to neutral language, influence the trustworthiness of a person arguing in an online health forum and the credibility of their health claims; (2) How does working for a university, in comparison to working for a lobbying organization, influence the trustworthiness of a person arguing in an online health forum and the credibility of their health claims; and (3) Do the two factors of language style and professional affiliation interact with each other to influence trustworthiness and credibility judgments?

Methods

Design and Material

A 2 (language style: neutral language vs positive language) × 2 (professional affiliation: scientist vs lobbyist) between-subject experimental design was used, resulting in four experimental conditions. In each experimental condition, participants saw two online forum posts: a question post and an answer post. In the question post, a woman asked whether Batradicum was an effective drug for the treatment of attention deficit hyperactivity disorder (ADHD). The question post was written in a neutral language style and was the same in all four experimental conditions. In the answer post, a man introduced himself and argued that Batradicum is an effective drug for the treatment of ADHD. The experimental manipulations were realized in the answer post. Depending on the experimental condition, the answer post was written either in a neutral language style or a positive language style. Furthermore, the author of the answer post introduced himself either as a scientist who worked for a pharmacological institute at a university, or as a lobbyist who worked at a pharmacological lobbying organization. Textbox 1 shows the question post and the answer post with the experimental manipulation.

As a note, the positive language style version of the answer post contained the words and phrases printed in italics, and the neutral language style version did not contain these words and phrases. For reasons of ecological validity, the name of an existing ADHD medication was used in the original study material. To avoid the impression that the authors endorse or criticize the drug, the fictitious drug name Batradicum is used here. Also, as the table shows an English translation of the German posts, the translated version may not appear as authentic to native English speakers as the original version appears to native German speakers. The original German version of the posts can be obtained from the authors upon request.

Textbox 1. Text of the question and answer post.

Question post:

Dear forum community,

My son is, according to his teachers, hyperactive and cannot concentrate in class. Moreover, his grades are suffering and he might not be allowed to move up to the next grade. The doctor gave him an ADHD diagnosis and offered that he could prescribe him the drug Batradicum. Unfortunately, I am not an expert in this field and I have heard of different studies that either argue for or against the effectiveness of Batradicum.

Does anybody know more about this topic?

Thanks in advance!

Sabine Schneider

Answer post:

Hello Ms. Schneider,

My name is Johannes Becker and since I have been working for the [Scientist Manipulation: *Pharmacological Institute at the University of Bochum*; Lobbyist Manipulation: *Association of Pharmacological Industries in Bochum*] for many years, I have been dealing with the subject of Batradicum for quite some time.

You are right, in the past, many studies on Batradicum have contradicted each other, often because of methodological mistakes.

Recently, however, a *magnificent* study by Mr. Weber has been published, which speaks for the effectiveness of Batradicum. Mr. Weber has *brilliantly* compared different age groups, which many previous studies have not done. In addition, in his *unique* study, he has studied not only the physiological but also the psychological effects of Batradicum, which is rare *and especially praiseworthy*. His *exemplary* methodological approach and *first-class* statistical data analysis speak for the quality of the study. Due to the *really outstanding* study of Mr. Weber and its *excellent* execution, I am convinced that Batradicum is effective.

Yours sincerely,

Johannes Becker

Sample and Procedure

German university students were contacted via email and social network sites and received €5 (US $5.65) for participating in the experiment. Overall, 251 participants completed the study without interruption, but 7 participants were excluded from data analysis because they indicated that they encountered technical problems during the study. Furthermore, 2 participants were excluded from data analysis because they indicated that they did not answer the questions honestly. The final sample contained 242 (165 females, 77 males) students (175 undergraduate students, 67 graduate students) at an average age of 23 years old (mean 22.57; SD 3.05). The average participant had been enrolled in their study program for five semesters (mean 4.81; SD 3.38). The experiment was conducted online using the EFS Survey platform (Questback GmbH, Cologne, Germany). First, participants gave informed consent, provided demographic information, and answered the control measures. They were then randomly assigned to one of the four experimental conditions. After reading the forum posts, participants answered the dependent measures and the manipulation check question. At the end of the experiment, the participants were debriefed.

Control Measures and Manipulation Check

A total of four control measures were included to assess whether the experimental groups differed in characteristics that could bias the study results. Participants answered three questions [General Use: "How often do you visit Internet forums?"; Educational Use: "How often do you visit Internet forums to learn something new or acquire new skills?"; Prior Knowledge: "How much do you know about Batradicum?") and indicated their agreement with one statement (Prior Attitude: "Batradicum is an effective drug for the treatment of ADHD.") on seven-point scales. To check whether the language style manipulation was successful, participants answered the question, "How would you describe Johannes Becker's choice of words?" on a scale ranging from 1 (neutral) to 7 (extremely enthusiastic).

Dependent Measures

To assess the credibility of the provided information, two measures were used, one of which was a general credibility measure, the Message Credibility Scale [38]. This measure was translated and adapted. As a more specific credibility measure, participants were also asked how much they agreed with the main conclusion of the forum post author (Attitude). To assess the trustworthiness of the information source, five measures were used: the German version of the Machiavellianism Subscale from the Dirty Dozen Scale [39,40] was used to assess how manipulative the forum post author was perceived to be (Machiavellianism), the Muenster Epistemic Trustworthiness Inventory [41] was used to assess how knowledgeable (Expertise), sincere (Integrity), and benevolent (Benevolence) the forum post author was perceived to be, and as a more general trustworthiness measure, the Reysen Likability Scale [42] was used to assess how likable the forum post author was perceived to be (Likability). Participants gave their answers on 7-point scales. The original dataset contains further variables that have not been described because they exceed the scope of the present article.

Results

Control Measures and Manipulation Check

For all analyses, the statistical software SPSS Statistics Version 26 (IBM Corp, Armonk, New York, United States) was used. Four one-way between-subject analyses of variance were conducted with experimental condition as the independent variable and the control measures as dependent variables. The results showed that the participants in the four experimental groups did not significantly differ in regard to the control measures of General Use ($F_{3,238}$=1.339; P=.262), Educational Use ($F_{3,238}$=0.784; P=.504), Prior Knowledge ($F_{3,238}$=0.505; P=.679), and Prior Attitude ($F_{3,238}$=1.365; P=.254). Furthermore, participants in the positive language style condition (mean 5.97; SD 1.28) perceived the choice of words as more enthusiastic than participants in the neutral language style condition (mean 3.63; SD 1.67; $tt_{227.769}$=–12.231; P<.001). Thus, the language style manipulation worked as expected.

Dependent Measures

For the analyses of the dependent measures, two-way between-subject analyses of variance were conducted with language style (neutral language vs positive language) and professional affiliation (scientist vs lobbyist) as independent variables.

Main Effects of Language Style

There were significant main effects of language style on Message Credibility ($F_{1,238}$=11.274; P=.001; η^2_p=.045), Machiavellianism ($F_{1,238}$=19.621; P<.001; η^2_p=.076), Integrity ($F_{1,238}$=11.328; P=.001; η^2_p=.045), and Benevolence ($F_{1,238}$=6.036; P=.02; η^2_p=.025). However, there were no main effects of language style on Attitude ($F_{1,238}$=0.785; P=.38; η^2_p=.003], Expertise ($F_{1,238}$=2.045; P=.15; η^2_p=.009], and Likability ($F_{1,238}$=1.721; P=.19; η^2_p=.007).

Main Effects of Professional Affiliation

There were no main effects of professional affiliation on the dependent measures of Message Credibility ($F_{1,238}$=2.852; P=.09; η^2_p=.012), Attitude ($F_{1,238}$=1.122; P=.29; η^2_p=.005), Machiavellianism ($F_{1,238}$=1.144; P=.29; η^2_p=.005), Expertise ($F_{1,238}$=0.400; P=.53; η^2_p=.002), Integrity ($F_{1,238}$=1.401; P=.24; η^2_p=.006), Benevolence ($F_{1,238}$=0.429; P=.51; η^2_p=.002), and Likability ($F_{1,238}$=2.594; P=.11; η^2_p=.011).

Interaction Effects

The two factors of language style and professional affiliation did not interact with each other to influence trustworthiness and credibility judgements for Message Credibility ($F_{1,238}$= 0.091; P=.76; η^2_p<.001), Attitude ($F_{1,238}$=0.112; P=.74; η^2_p<.001), Machiavellianism ($F_{1,238}$=0.397; P=.53; η^2_p=.002), Expertise ($F_{1,238}$=0.559; P=.46; η^2_p=.002), Integrity ($F_{1,238}$=0.725; P=.40; η^2_p=.003), Benevolence ($F_{1,238}$=0.007; P=.93; η^2_p<.001), and

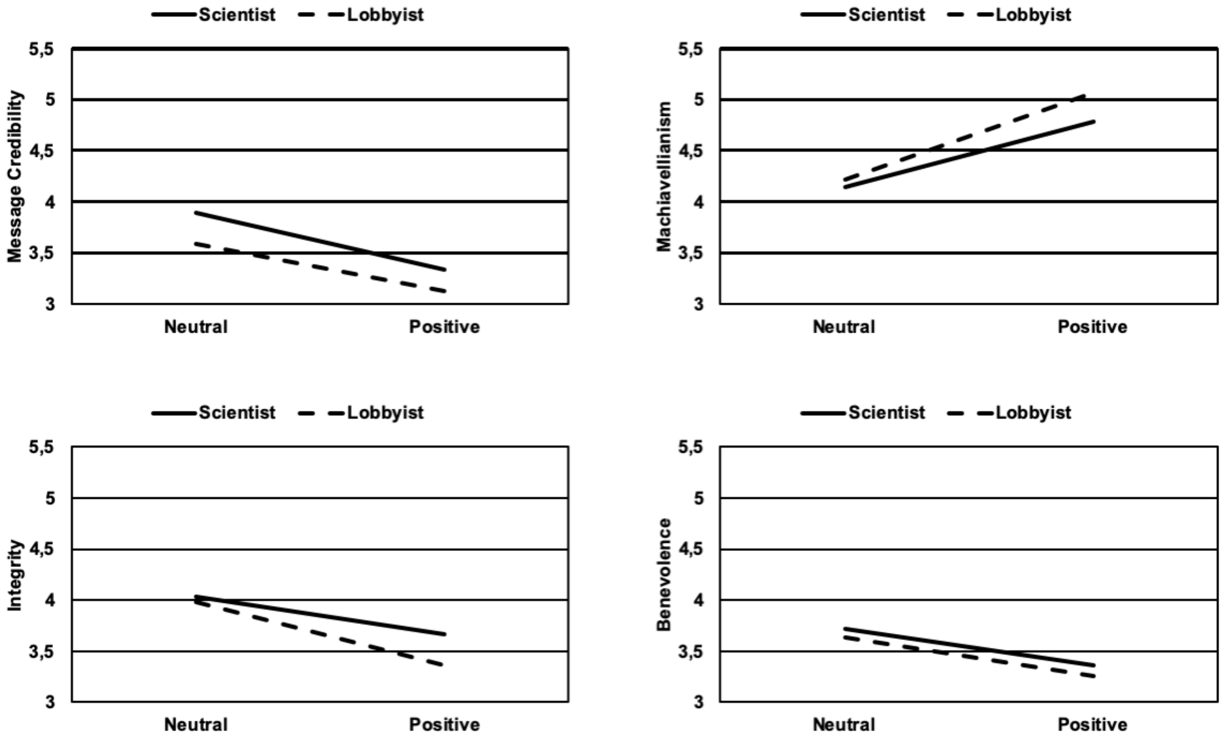
Likability ($F_{1,238}$=3.708; P=.06; η^2_p=.015). Table 1 shows the means and standard deviations of the dependent measures by language style and professional affiliation. Figure 1 shows the dependent measures that were significantly influenced by the language style manipulation.

As a note for Table 1, for the Machiavellianism scale, a low score indicates high trustworthiness and a high score indicates low trustworthiness. For all other scales, a low score indicates low trustworthiness/credibility and a high score indicates high trustworthiness/credibility. All scales ranged from 1-7.

Table 1. Main effects of the dependent measures by language style and professional affiliation.

Dependent measures	Language style			Professional affiliation		
	Neutral (n=123), mean (SD)	Positive (n=119), mean (SD)	*P* value	Scientist (n=124), mean (SD)	Lobbyist (n=118), mean (SD)	*P* value
Credibility						
Message Credibility	3.75 (1.19)	3.24 (1.20)	.001	3.62 (1.25)	3.36 (1.17)	.09
Attitude	3.80 (1.42)	3.95 (1.22)	.38	3.96 (1.29)	3.78 (1.36)	.29
Trustworthiness						
Machiavellianism	4.18 (1.32)	4.93 (1.29)	<.001	4.46 (1.46)	4.64 (1.24)	.29
Expertise	4.45 (1.20)	4.23 (1.14)	.15	4.39 (1.23)	4.29 (1.12)	.53
Integrity	4.00 (1.17)	3.51 (1.10)	.001	3.85 (1.14)	3.68 (1.19)	.24
Benevolence	3.68 (1.20)	3.31 (1.11)	.02	3.54 (1.20)	3.45 (1.13)	.51
Likability	3.37 (0.97)	3.22 (0.97)	.19	3.39 (0.96)	3.19 (0.97)	.11

Figure 1. Dependent measures that were significantly influenced by language style.



Discussion

Discussion of the Research Questions

The goal of the present study was to explore whether the language style (positive vs neutral) and professional affiliation (scientist vs lobbyist) of a person communicating health information in an online forum would influence their trustworthiness and the credibility of their health claims. The results show that if the forum post author used positive language in comparison to neutral language, they were perceived as less trustworthy, and their health claims were deemed less credible. More specifically, if the forum post author used positive language, they were perceived as more manipulative (Machiavellianism), less sincere (Integrity), less benevolent (Benevolence), and their health claims were perceived as less credible (Message Credibility).

In contrast, the professional affiliation of the forum post author (whether they were a scientist or a lobbyist), affected neither their trustworthiness nor the credibility of their information. This result is surprising because it contradicts previous findings [19]. On closer examination, however, the descriptive statistics show that the lobbyist was rated as less trustworthy on every measure, even though these differences did not reach significance. Therefore, one might speculate that no effect of professional affiliation was found because the experimental manipulation was too weak (just one sentence at the beginning of the forum post), or the participants did not identify the Association of Pharmacological Industries in Bochum as a lobbying organization. Lastly, the language style and professional affiliation manipulations did not interact with each other.

Limitations and Future Research Directions

Even though the results of the current study increase the understanding of how information seekers assess the accuracy of health information in online forums, there are limitations to the generalizability of the results. Four limitations seem especially important. First, the study participants were relatively young due to their status as students. Since previous research has found that age might influence source monitoring, suggestibility to misinformation [43], and credibility judgments in online contexts [44], future research should replicate the current study with different age groups. Second, the language style manipulation was realized by incorporating multiple positively valenced adjectives in the forum post. Since the number of adjectives has presumably influenced the language style effect, future research should replicate the current study with varying amounts of adjectives to explore the boundary conditions of the found effect. This suggestion seems especially important because different amounts of adjectives might alter the direction of the language style effect. For example, if an information source uses numerous positive adjectives in a health forum, information seekers might perceive this language use as inappropriate because it reminds them of advertising campaigns that often use extremely positive language. However, if an information source uses just a few positive adjectives, information seekers might perceive this language use as typical for a person who sincerely believes in their position. Third, the manipulation check showed that participants in the positive language style condition perceived the choice of words as more enthusiastic than participants in the neutral language style condition. However, the neutral language style was not perceived as entirely neutral. Instead, it was perceived as moderately enthusiastic. One might argue that the current results, therefore, represent differences between moderately enthusiastic and extremely enthusiastic language styles. Hence, future research should replicate the current study with new experimental material that is perceived as more neutral. Fourth, the current study employed a highly specific health-seeking context: A female information seeker asked for health advice regarding a specific drug and got an answer from a male information source. This specific context might have influenced the results. Previous research has shown that many members of the public have negative attitudes towards the use of drugs to treat ADHD [45]. Therefore, arguing in favor of using drugs to treat ADHD might have intensified the negative effects of the positive language style manipulation.

Consequently, future research should explore the positive language style effect within the context of less controversial topics. Furthermore, previous research suggests that the impact of expert testimony is influenced by the congruency between the gender of the expert and the topic at hand [46]. Thus, it would be interesting to manipulate the gender of the information source and investigate whether this alters the effect of the positive language style manipulation. Furthermore, it would be interesting to investigate whether the gender of the information source and the gender of the information seeker interact with each other when influencing credibility and trustworthiness judgments.

Conclusion

When searching for health information, many people use the internet as their first source of information. When they are confronted with positive language in online health forums, indicated by the frequent use of positively valenced adjectives, they may judge the health communicator as less trustworthy and deem the communicated health claims as less credible. These findings illustrate that health information seekers do not just react to the factual part of health information. Instead, they use the language style that surrounds health claims to evaluate the credibility of the provided information and the trustworthiness of the information source.

Acknowledgments

The authors would like to thank Dr Celeste Brennecka for language editing and their research assistant Hannah Meinert for her support. This work was supported by the German Research Foundation within the framework of the Research Training Group GRK 1712 Trust and Communication in a Digitized World. The funding was granted to the second author. The German Research Foundation had no involvement in study design, data collection, analysis and interpretation, and the decision to submit the study for publication.

Authors' Contributions

LK and RJ contributed to the conceptualization of the study. LK contributed to the study design. LK was responsible for data acquisition. LK and RJ were responsible for analysis and interpretation of data. LK was responsible for drafting the manuscript. LK and RJ were responsible for revising the manuscript critically for important intellectual content. LK and RJ approved the revision of the manuscript to be published.

Conflicts of Interest

None declared.

References

1. Prestin A, Vieux SN, Chou WS. Is Online Health Activity Alive and Well or Flatlining? Findings From 10 Years of the Health Information National Trends Survey. J Health Commun 2015;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
2. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2018-01-16] [WebCite Cache ID 75vv5Hdn7]
3. Sinha A, Porter T, Wilson A. The Use of Online Health Forums by Patients With Chronic Cough: Qualitative Study. J Med Internet Res 2018 Jan 24;20(1):e19 [FREE Full text] [doi: 10.2196/jmir.7975] [Medline: 29367181]
4. Biyani P, Caragea C, Mitra P, Yen J. Identifying Emotional and Informational Support in Online Health Communities. 2014 Presented at: COLING 2014, the 25th International Conference on Computational Linguistics; August; Dublin, Ireland p. 827-836 URL: https://www.aclweb.org/anthology/C14-1079/
5. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. Vaccine 2010 Feb 17;28(7):1709-1716. [doi: 10.1016/j.vaccine.2009.12.022] [Medline: 20045099]
6. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. JAMA 2007 Dec 5;298(21):2482-2484. [doi: 10.1001/jama.298.21.2482] [Medline: 18056901]
7. Miles J, Petrie C, Steel M. Slimming on the Internet. J R Soc Med 2000 May;93(5):254-257 [FREE Full text] [doi: 10.1177/014107680009300510] [Medline: 10884770]
8. Pandey A, Patni N, Singh M, Sood A, Singh G. YouTube as a source of information on the H1N1 influenza pandemic. Am J Prev Med 2010 Mar;38(3):e1-e3. [doi: 10.1016/j.amepre.2009.11.007] [Medline: 20171526]
9. Stadtler M, Bromme R. The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In: Rapp DN, Braasch JLG, editors. Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences. Cambridge, Massachusetts, United States: MIT Press; 2014:379-402.
10. Stadtler M, Winter S, Scharrer L, Thomm E, Krämer N, Bromme R. Selektion, Integration und Evaluation. Psychologische Rundschau 2017 Jul;68(3):177-181. [doi: 10.1026/0033-3042/a000361]
11. 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]
12. Bromme R, Thomm E. Knowing Who Knows: Laypersons' Capabilities to Judge Experts' Pertinence for Science Topics. Cogn Sci 2016 Jan;40(1):241-252 [FREE Full text] [doi: 10.1111/cogs.12252] [Medline: 25974834]
13. Thiebach M, Mayweg-Paus E, Jucks R. “Probably true” says the expert: how two types of lexical hedges influence students’ evaluation of scientificness. Eur J Psychol Educ 2015 Jan 16;30(3):369-384. [doi: 10.1007/s10212-014-0243-4]
14. Mayweg-Paus E, Jucks R. Evident or doubtful? How lexical hints in written information influence laypersons' understanding of influenza. Psychol Health Med 2015;20(8):989-996. [doi: 10.1080/13548506.2014.986139] [Medline: 25482286]
15. Zimmermann M, Jucks R. Investigating the Role of Communication for Information Seekers' Trust-Related Evaluations of Health Videos on the Web: Content Analysis, Survey Data, and Experiment. Interact J Med Res 2018 Dec 21;7(2):e10282 [FREE Full text] [doi: 10.2196/10282] [Medline: 30578181]
16. Jucks R, Linnemann G, Thon F, Zimmermann M. Trust the Words: Insights into the Role of Language in Trust Building in a Digitalized World. In: Blöbaum B, editor. Trust and Communication in a Digitized World. Heidelberg, Germany: Springer International Publishing; 2016:225-237.
17. Thon FM, Jucks R. Believing in Expertise: How Authors' Credentials and Language Use Influence the Credibility of Online Health Information. Health Commun 2017 Jul;32(7):828-836. [doi: 10.1080/10410236.2016.1172296] [Medline: 27466693]
18. Zimmermann M, Jucks R. How Experts' Use of Medical Technical Jargon in Different Types of Online Health Forums Affects Perceived Information Credibility: Randomized Experiment With Laypersons. J Med Internet Res 2018 Jan 23;20(1):e30 [FREE Full text] [doi: 10.2196/jmir.8346] [Medline: 29362212]
19. König L, Jucks R. Hot topics in science communication: Aggressive language decreases trustworthiness and credibility in scientific debates. Public Underst Sci 2019 May;28(4):401-416. [doi: 10.1177/0963662519833903] [Medline: 30843467]
20. König L, Jucks R. Influence of Enthusiastic Language on the Credibility of Health Information and the Trustworthiness of Science Communicators: Insights From a Between-Subject Web-Based Experiment. Interact J Med Res 2019 Aug 12;8(3):e13619 [FREE Full text] [doi: 10.2196/13619] [Medline: 31411138]
21. König L, Jucks R. When do information seekers trust scientific information? Insights from recipients’ evaluations of online video lectures. Int J Educ Technol High Educ 2019 Feb 13;16(1):1-21. [doi: 10.1186/s41239-019-0132-7]
22. Eastin M. Credibility Assessments of Online Health Information: The Effects of Source Expertise and Knowledge of Content. J Comput Commun 2006;6(4):1. [doi: 10.1111/j.1083-6101.2001.tb00126.x]

23. Zimmermann M, Jucks R. With a View to the Side: YouTube’s Sidebar and YouTuber’s Linguistic Style as Hints for Trust-Related Evaluations. International Journal of Human–Computer Interaction 2018 Sep 14;35(14):1279-1291. [doi: 10.1080/10447318.2018.1519165]
24. Bientzle M, Griewatz J, Kimmerle J, Küppers J, Cress U, Lammerding-Koeppel M. Impact of Scientific Versus Emotional Wording of Patient Questions on Doctor-Patient Communication in an Internet Forum: A Randomized Controlled Experiment with Medical Students. J Med Internet Res 2015 Nov 25;17(11):e268 [FREE Full text] [doi: 10.2196/jmir.4597] [Medline: 26607233]
25. Flemming D, Cress U, Kimmig S, Brandt M, Kimmerle J. Emotionalization in Science Communication: The Impact of Narratives and Visual Representations on Knowledge Gain and Risk Perception. Front. Commun 2018 Jan 26;3:1-9. [doi: 10.3389/fcomm.2018.00003]
26. Nau C, Stewart CO. Effects of Verbal Aggression and Party Identification Bias on Perceptions of Political Speakers. Journal of Language and Social Psychology 2013 Dec 04;33(5):526-536. [doi: 10.1177/0261927x13512486]
27. Frei B, England L, Ames BN. Ascorbate is an outstanding antioxidant in human blood plasma. Proc Natl Acad Sci U S A 1989 Aug 15;86(16):6377-6381 [FREE Full text] [doi: 10.1073/pnas.86.16.6377] [Medline: 2762330]
28. Seifert G. Fascinating Fasciclins: A Surprisingly Widespread Family of Proteins that Mediate Interactions between the Cell Exterior and the Cell Surface. Int J Mol Sci 2018 May 31;19(6):1628 [FREE Full text] [doi: 10.3390/ijms19061628] [Medline: 29857505]
29. Hatfull GF. Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity. J Virol 2015 Aug;89(16):8107-8110 [FREE Full text] [doi: 10.1128/JVI.01340-15] [Medline: 26018169]
30. Charushila Biswas. Stylecraze. 2018. 30 Amazing Diet Recipes For Weight Loss URL: https://www.stylecraze.com/articles/amazing-diet-recipes-for-weight-loss/#gref [accessed 2019-02-14]
31. Isadora Baum. Bustle. 2017. 11 Surprising But Genius Health Tips From Around The World URL: https://www.bustle.com/p/11-surprising-but-genius-health-tips-from-around-the-world-52943 [accessed 2019-02-14]
32. Christine Surrago. Houseofcitrine. 2018. Neural Therapy - An Incredible Therapy for Modern Day Conditions URL: https://houseofcitrine.com/journal/neural-therapy-an-incredible-therapy-for-modern-day-conditions [accessed 2019-02-14]
33. Henry E. Are Investors Influenced By How Earnings Press Releases Are Written? Journal of Business Communication 2008 Oct 01;45(4):363-407. [doi: 10.1177/0021943608319388]
34. Tan H, Wang EY, Zhou B. When the Use of Positive Language Backfires: The Joint Effect of Tone, Readability, and Investor Sophistication on Earnings Judgments. Journal of Accounting Research 2014 Jan 21;52(1):273-302. [doi: 10.1111/1475-679x.12039]
35. Larcker D, Tayan B. Stanford Closer Look Series. 2010. Financial Manipulation: Words Don't Lie URL: https://www.gsb.stanford.edu/sites/gsb/files/publication-pdf/cgri-closer-look-07-financial-mainpulation.pdf [accessed 2020-01-27]
36. König L. Podcasts in higher education: teacher enthusiasm increases students’ excitement, interest, enjoyment, and learning motivation. Educational Studies 2020 Jan 07:1-4. [doi: 10.1080/03055698.2019.1706040]
37. Burgoon M, Pauls V, Roberts D. Language Expectancy Theory. In: Dillard JP, Pfau M, editors. The Persuasion Handbook: Developments in Theory and Practice. Thousand Oaks, California, United States: SAGE Publications; 2002:117.
38. Appelman A, Sundar SS. Measuring Message Credibility. Journalism & Mass Communication Quarterly 2015 Oct 05;93(1):59-79. [doi: 10.1177/1077699015606057]
39. Küfner ACP, Dufner M, Back MD. Das Dreckige Dutzend und die Niederträchtigen Neun. Diagnostica 2015 Jan;61(2):76-91. [doi: 10.1026/0012-1924/a000124]
40. Jonason PK, Webster GD. The dirty dozen: a concise measure of the dark triad. Psychol Assess 2010 Jun;22(2):420-432. [doi: 10.1037/a0019265] [Medline: 20528068]
41. Hendriks F, Kienhues D, Bromme R. Measuring Laypeople's Trust in Experts in a Digital Age: The Muenster Epistemic Trustworthiness Inventory (METI). PLoS One 2015;10(10):e0139309 [FREE Full text] [doi: 10.1371/journal.pone.0139309] [Medline: 26474078]
42. Reysen S. Construction of a new scale: The Reysen likability scale. Soc Behav Pers 2005 Jan 01;33(2):201-208. [doi: 10.2224/sbp.2005.33.2.201]
43. Mitchell KJ, Johnson MK, Mather M. Source monitoring and suggestibility to misinformation: adult age-related differences. Appl. Cognit. Psychol 2002 Jan;17(1):107-119. [doi: 10.1002/acp.857]
44. Choi W, Stvilia B. Web credibility assessment: Conceptualization, operationalization, variability, and models. J Assn Inf Sci Tec 2015 May 13;66(12):2399-2414. [doi: 10.1002/asi.23543]
45. Partridge B, Lucke J, Hall W. Public attitudes towards the acceptability of using drugs to treat depression and ADHD. Aust N Z J Psychiatry 2012 Oct 11;46(10):958-965. [doi: 10.1177/0004867412450755] [Medline: 22689334]
46. McKimmie BM, Newton CJ, Terry DJ, Schuller RA. Jurors’ Responses to Expert Witness Testimony: The Effects of Gender Stereotypes. Group Processes & Intergroup Relations 2016 Jul 26;7(2):131-143. [doi: 10.1177/1368430204043724]

Abbreviations

ADHD: attention deficit hyperactivity disorder

Edited by E Borycki; submitted 14.10.19; peer-reviewed by M Eggeling, D Chrimes; comments to author 30.11.19; revised version received 08.12.19; accepted 15.12.19; published 10.03.20.

Please cite as:
König L, Jucks R
Effects of Positive Language and Profession on Trustworthiness and Credibility in Online Health Advice: Experimental Study
J Med Internet Res 2020;22(3):e16685
URL: http://www.jmir.org/2020/3/e16685/
doi:10.2196/16685
PMID:32154786

Original Paper

Sources of Health Information and Their Impacts on Medical Knowledge Perception Among the Saudi Arabian Population: Cross-Sectional Study

Shatha A Alduraywish[1], MBBS, PhD; Lamees A Altamimi[2], MBBS; Rawan A Aldhuwayhi[2], MBBS; Lama R AlZamil[2], MBBS; Luluh Y Alzeghayer[2], MBBS; Futoon S Alsaleh[2], MBBS; Fahad M Aldakheel[3], MSc, PhD; Shabana Tharkar[4], MSc

[1]Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia
[2]College of Medicine, King Saud University, Riyadh, Saudi Arabia
[3]Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia
[4]Prince Sattam Chair for Epidemiology and Public Health Research, College of Medicine, King Saud University, Riyadh, Saudi Arabia

Corresponding Author:
Shatha A Alduraywish, MBBS, PhD
Department of Family and Community Medicine
College of Medicine
King Saud University
Alshikh Hasan Bin Abdullah Alshikh
Riyadh, 12373
Saudi Arabia
Phone: 966 1146 79860
Email: s.alduraywish@gmail.com

Abstract

Background: Having a reliable source for health information is vital to build a strong foundation of knowledge, especially with the current revolution of the internet and social media, which raises many concerns regarding harmful effects on the health of the public. However, there are no studies on how the Saudi Arabian population seeks health information. Details about the most used and trusted sources of health information among the public will help health authorities and public awareness accounts on social media to effectively disseminate health information.

Objective: To investigate the types of sources accessed by the Saudi Arabian population while seeking health information, as well as their level of trust in the sources and to assess the impact of these sources on their perception of medical knowledge and health decision-making.

Methods: A cross-sectional study was conducted to meet the objectives. The study population included both men and women who were aged 16 years or more and visited primary care clinics at King Khalid University Hospital. Four hundred and thirteen participants were sampled using the simple random method, and a self-administered questionnaire was used to collect data. The data were analyzed using SPSS software (IBM Corp, Armonk, New York, USA).

Results: A total of 413 participants were included in this study, and of these, 99 (24.0%) were males and 206 (49.9%) had a bachelor’s degree. Doctors were chosen as the first source of information by 87.6% (283/323) of the participants, and they were completely trusted by most of the population (326/411, 79.3%). The second most commonly used source was pharmacists (112/194, 57.7%), and they were partially trusted by 41.4% (159/384) of the participants. Internet searches, social media, and traditional medicine were not prioritized by most of the participants as the first or second source of health information. The majority of the participants did not trust information obtained from social media, and WhatsApp was the most untrusted source. Almost half of the respondents (197/413, 47.7%) acknowledged that various sources of information can often help them understand their health problems. However, the majority disagreed on substituting a doctor’s prescription with information obtained from the internet or a friend or relative.

Conclusions: Although physicians were preferred and highly trusted, internet sources appeared to impact the medical knowledge of the population. The population still preferred to use internet search to obtain health information prior to a doctor’s visit.

(J Med Internet Res 2020;22(3):e14414) doi:10.2196/14414

KEYWORDS

health information sources; health perception; medical information sources; satisfaction; social media; trust

Introduction

The current sources of health information are diverse, and they influence the perception of medical knowledge among the Saudi Arabian population. Having a reliable source of health information is critical for building a strong foundation of knowledge about health among the public, especially with the current revolution of the internet and social media. According to the latest statistics, more than 50% of the global population uses the internet. Moreover, more than 64% of the Saudi Arabian population uses the internet, and this number has been increasing every year [1]. Although social media, websites, and internet search engines are considered as easily accessible sources of medical information, these sources still contain ambiguities. There is increasing evidence that the instant exchange of news from random resources and the lack of verification and determination of the accuracy and credibility of the information being shared by nonprofessionals raise many concerns about the harmful effects on a person's health [2].

Determining the frequency and characteristics of people who are likely to use the internet and social media to obtain their medical knowledge may guide both clinicians and the public. In addition, it is important from a public health perspective to appreciate the probable subsequent health outcomes. Moreover, the health information sought from different sources and trusted by the public could tremendously influence the quality of health care rendered, as it could affect their judgment of a physician's medical opinion and, in turn, affect their health decision-making. Previous studies have found that the commonly used sources of health information are the internet [1,3,4], physicians [5], social media [6], radio and television [7], pharmacists [8], and parents [9]. It has been reported that more than half of the public widely use the internet to seek health information [1,10,11], and the majority are young adults [1]. In a study conducted in Qatar, it was found that about 37.8% of Qataris seek health information from family and friends and 31.2% rely on primary health care centers as a source of health information [1].

In Australia, the use of the internet as a source of health information by the university population (students and staff) is greater than the use by the low or middle socioeconomic population [12]. It has been reported that gender, age, and educational level play significant roles in predicting the source of medical information. However, irrespective of the source of consultation, women are generally more interested in seeking health information than men, and the majority of those who seek medical information from the internet are women [1]. In comparison with the different available sources of health information, health care providers have been reported to be the most trusted source among different adult populations in the United States [11]. The majority of those who trust health care providers as the source of their health information are female, young, and educated [11], and this is irrespective of their socioeconomic status [12].

The roles of family and friends, pharmacists, and local alternative medicine practitioners are rarely discussed. As such, additional research is needed to evaluate the use of the internet and social media as tools for medical decision-making. Familiarity with the most used and trusted sources of health information among the public will assist health authorities and public awareness accounts on the internet and social media in accurately and carefully disseminating health information among the public, who should be educated and advised about the most credible sources of health information.

It is important to mention that social media platforms in Saudi Arabia, specifically WhatsApp, are not only used as communication tools among family and friends. The easy broadcasting characteristics of WhatsApp allow the population to exchange numerous videos, news, and messages. As these broadcasts continue to circulate among the public, the source of the information can no longer be tracked and verified. Health information is one of the most commonly exchanged messages and are most frequently completely false. This is why we express our concerns in this paper, and we investigated the level of trust of information broadcasted through WhatsApp and assessed whether it affects a patient's health-seeking behavior and medical decisions.

To our knowledge, there have been no studies on how the Saudi Arabian population seeks information regarding their health. Therefore, the purposes of this study were to (1) investigate the different sources that the Saudi Arabian population uses and trusts for medical information and (2) assess the impact of these sources on medical knowledge and the patient's health decision-making. We hypothesized that more than half of the Saudi Arabian population uses the internet and social media to obtain medical information.

Methods

Study Setting and Population

We conducted an observational cross-sectional study to address our research objectives. The study was conducted in primary care clinics at King Khalid University Hospital (KKUH), which is a large 1000-bed tertiary-care hospital in the northern part of Riyadh, Saudi Arabia. KKUH provides primary, secondary, and tertiary care services to a large patient-catchment area and government-funded free preventive and curative services. A random selection was performed of all attending Saudi male and female patients aged 16 years or more, which is the cutoff age for patients attending the adult primary care clinics. There were no exclusion criteria other than age and nationality.

The required ethical approval was obtained from the institutional review board of King Saud University. Informed consent was obtained from the participants, and the confidentiality of the information and privacy of the participants were protected throughout the study.

Sample Size Estimation

According to a recent study from the neighboring gulf region, 71% of the Qatari population uses the internet to seek health-related information. With a 95% CI and precision of 5, the sample size in this study was estimated at 386, which was further increased to 425 considering a nonresponse rate of 10%.

Recruitment of Participants

Four hundred and thirteen participants were recruited using simple random sampling, which was conducted through a random number generator software.

Data Collection Tools

The data were collected using a self-administered questionnaire or an interview (in the case of illiterate participants). The questionnaire had three main parts. The first part assessed the demographic data of the participants, such as age, gender, area of residence, educational level, and current occupation. The second part assessed the different used and trusted sources of medical information. It contained two questions. The first question concerned the ranking of the most used source, and the second question concerned the extent of trust in the sources chosen in the first question. The third part of the questionnaire was designed to explore the impacts of those sources on the participant's medical knowledge and their effects on health decision-making.

Tool Validation

The development of the questionnaire survey was based on a literature review, and some of the questions in the second part of the questionnaire were adapted from the Health Information National Trends Survey (HINTS). We used the information sources mentioned in the HINTS [13]. For the third part, we adopted most of the questions from the behavioral involvement subscale of the Assessment of Preferences for Self-Treatment and Information in Health Care survey [14]. A pilot study was conducted to assess the time needed to complete the questionnaire and the understandability of the included questions.

Statistical Analysis

The data were analyzed using IBM SPSS Statistics for Windows version 23.0 (IBM Corp, Armonk, New York, USA). The data are expressed using frequencies and percentages for categorical variables and means and SDs for continuous variables. The chi-square test was used for categorical variables. The association between the sociodemographic characteristics and the sources of health information was examined using logistic regression, and the results are expressed with ORs and 95% CIs. A two-sided P value of <0.05 was considered statistically significant.

Results

Characteristics of the Study Participants

Of 425 distributed questionnaires, 413 completed questionnaires were obtained (response rate of 88%). Twenty participants were illiterate and thus were interviewed. The majority of the respondents were female (314/413, 76.0%). The study sample was well distributed across all age groups, and most of the population was literate. The overall health status of the participants was good. However, 130 (31.6%) had chronic diseases, and 52 (40.0%) had diabetes, 44 (33.8%) had hypertension, and 34 (26.2%) had asthma. The demographic data are summarized in Table 1.

Table 1. Characteristics of the study participants.

Characteristic	Value, n (%)
Gender	
Male	99 (24.0)
Age (years)	
16-25	98 (24.0)
26-35	138 (33.7)
36-45	91 (22.2)
46-55	64 (15.6)
≥56	18 (4.4)
Residency (province)	
Central	362 (87.7)
Northern	27 (6.5)
Southern	9 (2.2)
Eastern	6 (1.5)
Western	9 (2.2)
Educational level	
Elementary school	16 (3.9)
Intermediate school	22 (5.3)
High school	86 (20.8)
Diploma	32 (7.7)
Bachelor studies	206 (49.9)
Postgraduate studies	31 (7.5)
Illiterate	20 (4.8)
Occupation	
Student	64 (5.6)
Government employee	115 (28.0)
Private sector employee	55 (13.4)
Retired	21 (5.1)
No occupation	155 (37.8)
Field of occupation	
Education	125 (45.6)
Medical	29 (10.6)
Military	9 (3.3)
Business	20 (3.7)
Others	91 (33.2)
Self-assessment of health status	
Very weak	2 (0.5)
Weak	18 (4.4)
Good	253 (61.3)
Excellent	140 (33.9)
Reported history of chronic disease	130 (31.6)

Different Sources and Choice of Preference for Health Information

Doctors were the most favored choice for the majority of the study population (283/323, 87.6%); however, a smaller fraction rated them as second (18/323, 5.6%), third (12/323, 3.7%), and fourth preferences (10/323, 3.1%). Pharmacists were rated as the second most favored choice after doctors by a little more than half of the population (112/194, 57.7%). Social media was least preferred as the first choice (2/105, 1.9%). Even doctors who are on social media were less preferred as the first choice (19/157, 12.1%) (Table 2).

Table 2. Different sources used for health information.

Source of information	Respondents, n or n (%)		Ranking, n (%)							
			First		Second		Third		Fourth or more	
	Total	Female	Total	Female	Total	Female	Total	Female	Total	Female
Doctor	323	240 (74.3)	283 (87.6)	211 (87.9)	18 (5.6)	14 (5.8)	12 (3.7)	9 (3.8)	10 (3.1)	6 (2.5)
Pharmacist	194	137 (70.6)	13 (6.7)	12 (8.8)	112 (57.7)	74 (54.0)	31 (16.0)	24 (17.5)	38 (19.6)	27 (19.7)
Traditional medicine practitioner	78	50 (64.1)	5 (6.4)	2 (4.0)	6 (7.7)	5 (10.0)	15 (19.2)	9 (18.0)	52 (66.7)	34 (68.0)
Social media	105	69 (65.7)	2 (1.9)	2 (2.9)	11 (10.5)	7 (10.1)	17 (16.2)	10 (14.5)	75 (71.4)	50 (72.5)
Doctors who are on social media	157	122 (77.7)	19 (12.1)	16 (13.1)	48 (30.6)	38 (31.1)	36 (22.9)	29 (23.8)	54 (34.3)	39 (32.0)
Family and friends	187	133 (71.1)	14 (7.5)	9 (6.8)	26 (13.9)	21 (15.8)	46 (24.6)	34 (25.6)	101 (54.0)	69 (51.9)
Internet search	184	132 (71.7)	28 (15.2)	20 (15.2)	41 (22.3)	31 (23.5)	45 (24.5)	31 (23.5)	70 (38.0)	50 (37.9)
Articles	95	64 (67.3)	2 (2.1)	1 (1.6)	7 (7.4)	4 (6.3)	15 (15.8)	11 (17.2)	71 (74.7)	48 (75.0)
Television and radio	122	81 (66.3)	5 (4.1)	2 (2.5)	14 (11.5)	9 (11.1)	19 (15.6)	12 (14.8)	84 (68.9)	58 (71.6)
Courses and campaigns	93	66 (70.9)	13 (14.0)	11 (16.7)	19 (20.4)	15 (22.7)	61 (65.6)	40 (60.6)	—[a]	—[a]

[a] None of the participants selected this source.

Level of Trust in Each Source of Health Information

The levels of trust in the sources of information are presented in Table 3. Doctors were the most trusted, and there was either complete trust (326/411, 79.3%) or partial trust (85/411, 20.6%), and no participant reported distrusting them. Pharmacists and traditional practitioners were partially trusted, with similar ratings; however, they were also distrusted by some participants. The majority of the participants did not trust the information obtained from social media. Gender differences were observed in the level of trust in social media, wherein more women showed distrust in social media than did men (*P*=.01) (Table 3).

In addition, the population distributions for the first choice of the source of health information and complete trust in the first choice are presented in Table 4.

Table 3. The level of trust in each source of health information.

Source of information	Number of respondents, n or n (%)		Level of trust, n (%)					
			Completely trusted		Partially trusted		Not trusted	
	Total	Female	Total	Female	Total	Female	Total	Female
Doctor	411	312 (75.9)	326 (79.3)	244 (74.8)	85 (20.6)	68 (80.0)	—[a]	—[a]
Pharmacist	384	287 (74.7)	23 (6.0)	1 (4.3)	159 (41.4)	89 (56.0)	202 (52.6)	197 (97.5)
Traditional medicine	271	209 (66.6)	9 (3.3)	5 (55.6)	115 (42.4)	85 (74.0)	147 (54.2)	119 (81.0)
WhatsApp	331	233 (70.4)	22 (6.6)	3 (13.6)	123 (37.1)	69 (56.1)	186 (56.1)	161 (86.6)
Snapchat	318	219 (68.9)	22 (6.9)	6 (27.3)	127 (39.9)	83 (65.4)	169 (53.1)	130 (77.0)
Twitter	273	205 (74.8)	5 (1.8)	3 (60.0)	139 (50.9)	104 (74.8)	130 (47.6)	98 (75.4)
Family and friends	360	271 (75.3)	27 (7.5)	23 (85.2)	265 (73.6)	199 (75.1)	68 (18.9)	49 (72.1)
Internet search	353	266 (75.4)	28 (7.9)	20 (71.4)	249 (70.5)	187 (75.1)	76 (21.5)	59 (77.6)
Television and radio	314	234 (74.5)	33 (10.5)	22 (66.6)	197 (62.7)	147 (74.6)	84 (26.8)	65 (77.4)

[a] None of the participants selected this option.

Table 4. Distributions of the population for the first choice and complete trust in the information.

	First choice, n (%)			Complete trust, n (%)		
	Medical[a]	Media[b]	Others[c]	Medical	Media	Others
Age (years)						
16-25	73 (83.0)	6 (7.0)	8 (9.0)	77 (93.0)	0 (0.0)	5 (6.0)
26-45	151 (75.0)	38 (19.0)	12 (6.0)	179 (97.0)	3 (2.0)	3 (2.0)
≥46	67 (92.0)	3 (4.0)	3 (4.0)	73 (95.0)	1 (1.0)	3 (4.0)
Gender						
Male	72 (78.0)	10 (11.0)	10 (11.0)	82 (99.0)	0 (0.0)	1 (1.0)
Female	221 (82.0)	37 (14.0)	13 (5.0)	250 (95.0)	4 (2.0)	10 (4.0)
Chronic diseases						
Present	96 (83.0)	18 (16.0)	1 (1.0)	110 (94.0)	0 (0.0)	7 (6.0)
Absent	196 (79.0)	29 (12.0)	22 (9.0)	222 (97.0)	4 (2.0)	4 (2.0)
Occupation						
Employed	49 (86.0)	3 (5.0)	5 (9.0)	47 (96.0)	0 (0.0)	2 (4.0)
Retired	115 (77.0)	23 (15.0)	12 (8.0)	136 (96.0)	3 (2.0)	3 (2.0)
Unemployed	17 (89.0)	1 (5.0)	1 (5.0)	19 (96.0)	0 (0.0)	1 (5.0)
Student	111 (82.0)	20 (15.0)	5 (4.0)	128 (96.0)	1 (1.0)	4 (4.0)
Education						
School	92 (84.0)	11 (10.0)	6 (6.0)	101 (97.0)	1 (1.0)	2 (2.0)
Higher education	185 (78.0)	35 (15.0)	16 (7.0)	215 (96.0)	3 (1.0)	7 (3.0)
Illiterate	17 (89.0)	1 (5.0)	1 (5.0)	17 (89.0)	0 (0.0)	2 (11.0)

[a] Medical includes doctors, pharmacists, and traditional medicine practitioners.

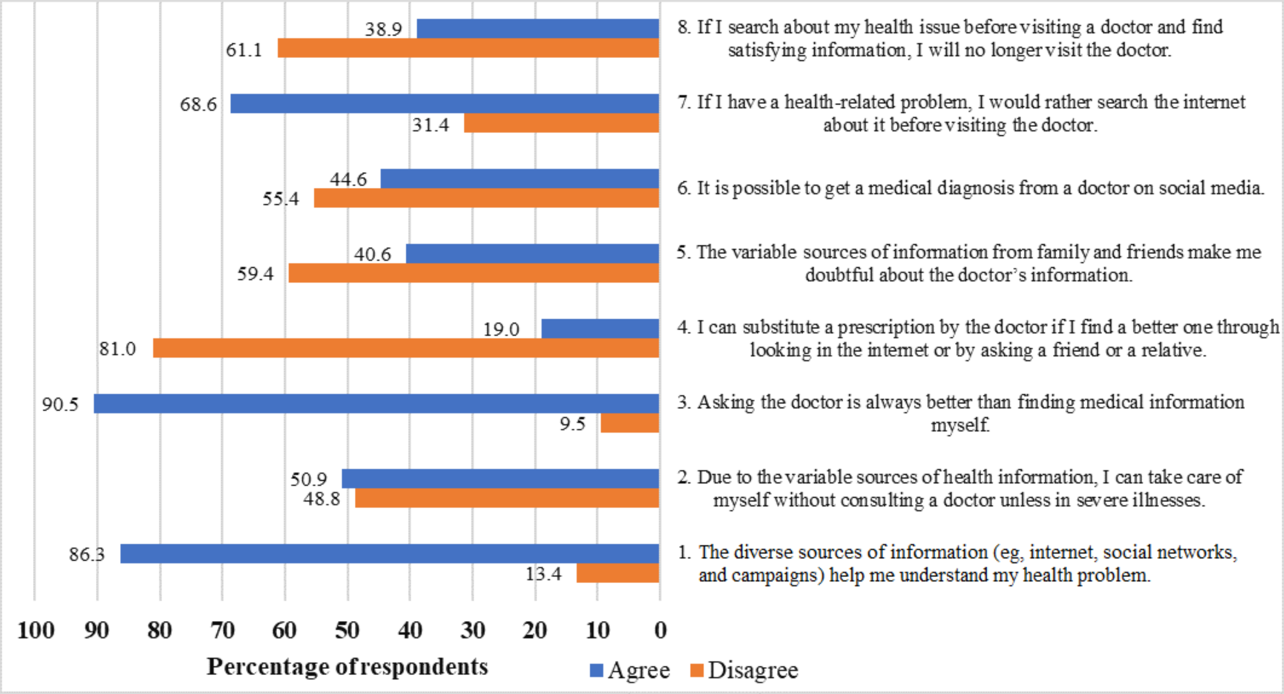
[b] Media includes WhatsApp, Twitter, Snapchat, and the internet.

[c] Others include family, friends, courses, campaigns, television, and radio.

Impacts of the Sources of Health Information on the Individual’s Health Perception and Clinical Decision-Making

Although around 86.0% (355/413) of the population reported using diverse sources for health information, 90.0% (371/413) preferred to seek help from doctors. Additionally, 68.0% (280/413) perceived to seek information from other sources only prior to a doctor’s visit. The majority disagreed on substituting a doctor’s prescription with information obtained from the internet or a friend or relative (Figure 1).

Figure 1. Impacts of sources of health information on an individual’s health perception and clinical decision-making.



Discussion

Principal Findings

This study aimed to investigate the different sources that the Saudi Arabian population uses and trusts for their health information and to assess the impact of these sources on an individual’s medical knowledge and health decision-making. The findings from this study revealed that doctors were the most commonly used source of health information, followed by internet search, whereas courses and campaigns were the least common sources. Doctors were the most trusted source, and unexpectedly, social media was the least trusted source. Although about half of the population admitted that various sources helped them understand their health problems, they prioritized doctor’s advice over other sources.

In this study, among the sources of information, we found that less than half of the Saudi Arabian population used the internet as a source of information. Similarly, more than half of the population did not use social media as one of the principal sources for health information. These findings did not support our study hypothesis. However, we found that more than half of the population (283/323, 87.6%) considered doctors as a primary source of information. Moreover, doctors were completely trusted for medical information by more than half of the population (326/411, 79.3%), which is similar to the findings in the study by Hesse et al, in which health care providers were the most trusted source among different populations [11]. However, 20.6% (85/411) of the population partially trusted their health care providers. This raises many concerns regarding the quality of medical care provided, as there might be an issue with the doctor-patient relationship. Many reasons could contribute to this diminished level of trust. For example, problems in communication between a doctor and patient could result in misunderstanding of the patient’s health issues, which could lead to patient dissatisfaction with the consultation. As a result, complete information will not be disclosed if the patient does not trust the doctor [15]. This affects the efficiency in patient management and consequently the trust in the doctor’s medical judgment.

Furthermore, more than half of the population stated that they will not substitute a doctor’s prescription with the information obtained from other sources. This finding is similar to that in the study by Diaz et al [10], confirming that other sources of information do not affect the certainty of the doctor’s information. In addition, it supports that initially seeking a doctor’s advice is prioritized over searching for other sources of information.

Pharmacists were considered as the second most commonly used source of information (112/194, 57.7%), although they were partially trusted by 41.4% (159/384) of the population. However, only a few respondents considered pharmacists as the first source of information (13/194, 6.7%). The fact that pharmacists follow doctors as a source of medical information is a good indicator of the population’s perception regarding the priorities for obtaining information from health care practitioners.

Despite the common use of traditional medicine by the Saudi Arabian population [16], our data revealed that traditional medicine practitioners were not trusted by the majority of the Saudi Arabian population. As a result, most of the population did not consider traditional medicine practitioners as a preferred source of information. These findings indicate that although individuals believe in traditional medicine, they do not obtain this kind of treatment from its practitioners. The reason for this is that a lot of traditional medicine practitioners in Saudi Arabia are not certified. Instead, most of them are owners of small shops of herbal remedies and practice this field of medicine as a hobby, and they lack a qualification and scientific background in this field.

Social media was not the most used or trusted source of health information by the majority of the participants. However, in the United States, more than half of the population uses social media to obtain health information [17]. Furthermore, we found that among the different social media applications, WhatsApp was the most untrusted source. These findings seem to be a good indicator of the population's awareness that unreliable sources should not be used to obtain health information. Similarly, a large percentage of the population did not rely on doctors who are on social media to obtain diagnoses of their conditions, and instead, they used this source to acquire general health information. This finding reflects the population's awareness of the dangers of a social media diagnosis without a medical consultation, as social media diagnosis can involve many inaccuracies despite the good intentions of the doctors. It emphasizes that doctors who are on social media need to deliver accurate and up-to-date material to avoid potentially harmful effects among their audience.

The majority of the participants did not prioritize internet searching, and more than half partially trusted this source. It is known that internet searches for medical information have well-recognized drawbacks regarding the quality and accuracy of information, as mentioned in the study by Benigeri and Pluye [18]. Surprisingly, courses and campaigns were the least used sources of information, and they were not used at all as the first source of information by the study participants. This reflects a weakness in the role of community awareness of health problems.

Strengths and Limitations

To our knowledge, this is the first study to discuss the use and trust of the Saudi Arabian population regarding different sources of medical information. The strength of this study is that the sample size was relatively large. However, a few limitations exist. The study population included many female individuals (314/413, 76%) and individuals living in the central province (362/413, 87.7%), which may not be a good reflection of the whole Saudi Arabian population. More diverse patient groups with a larger sample size may be needed to generalize the results. In addition, the data were self-reported, which might involve recall bias. Lastly, we did not investigate the types of information searched for in every source, such as sensitive topics, serious conditions, and educational information, which may change the information source preference.

Conclusion and Recommendations

This study is the first in Saudi Arabia to investigate the different sources of medical information that are used and trusted by the Saudi Arabian population. We found that doctors were the most used and trusted source, courses and campaigns were the least used sources, and social media, specifically WhatsApp, was the least trusted source. This draws attention to the need to develop well-structured courses and campaigns that meet the needs of the population in an easily understandable way. It also sheds light on the requirement to increase the quality of information provided in the nonmedical field. In addition, further research is needed to understand why a large number of participants only partially trusted their doctors.

Acknowledgments

The authors are grateful to the Deanship of Scientific Research, King Saud University, Riyadh, Kingdom of Saudi Arabia for funding through the Vice Deanship of Scientific Research Chairs. The authors would like to thank Enago (www.enago.com) for English language review.

Conflicts of Interest

None declared.

References

1. Choudhury S, Arora T, Alebbi S, Ahmed L, Aden A, Omar O, et al. How Do Qataris Source Health Information? PLoS One 2016;11(11):e0166250 [FREE Full text] [doi: 10.1371/journal.pone.0166250] [Medline: 27832112]
2. Rice R. 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]
3. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot IR, 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 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
4. Simou E. Health information sources: trust and satisfaction. IJH 2015 Oct 25;2(1):38. [doi: 10.5430/ijh.v2n1p38]
5. Clarke M, Moore J, Steege L, Koopman R, Belden J, Canfield S, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: A literature review. Health Informatics J 2016 Dec;22(4):992-1016. [doi: 10.1177/1460458215602939] [Medline: 26377952]

6. Laranjo L, Arguel A, Neves A, Gallagher A, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
7. Alishahi-Tabriz A, Sohrabi M, Kiapour N, Faramarzi N. Addressing the changing sources of health information in Iran. Int J Prev Med 2013 Jan;4(1):33-41 [FREE Full text] [Medline: 23412519]
8. Tsai M, Chou Y, Lin S, Lin S. Factors associated with adolescents' perspectives on health needs and preference for health information sources in Taiwan. Arch Dis Child 2013 Jan;98(1):9-15. [doi: 10.1136/archdischild-2012-301629] [Medline: 22820106]
9. Baheiraei A, Khoori E, Foroushani A, Ahmadi F, Ybarra M. What sources do adolescents turn to for information about their health concerns? Int J Adolesc Med Health 2014;26(1):61-68. [doi: 10.1515/ijamh-2012-0112] [Medline: 23625280]
10. Diaz J, Griffith R, Ng J, Reinert S, Friedmann P, Moulton A. Patients' use of the Internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]
11. Hesse B, Nelson D, Kreps G, Croyle R, Arora N, Rimer B, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
12. Dart J, Gallois C, Yellowlees P. Community health information sources--a survey in three disparate communities. Aust Health Rev 2008 Feb;32(1):186-196. [doi: 10.1071/ah080186] [Medline: 18241163]
13. Rutten L, Squiers L, Hesse B. Cancer-related information seeking: hints from the 2003 Health Information National Trends Survey (HINTS). J Health Commun 2006;11 Suppl 1:147-156. [doi: 10.1080/10810730600637574] [Medline: 16641080]
14. Krantz D, Baum A, Wideman MV. Assessment of preferences for self-treatment and information in health care. Journal of Personality and Social Psychology 1980;39(5):977-990. [doi: 10.1037/0022-3514.39.5.977]
15. Dorr Goold S, Lipkin M. The doctor-patient relationship: challenges, opportunities, and strategies. J Gen Intern Med 1999 Jan;14 Suppl 1:S26-S33 [FREE Full text] [doi: 10.1046/j.1525-1497.1999.00267.x] [Medline: 9933492]
16. K Suleiman A. Attitudes and Beliefs of Consumers of Herbal Medicines in Riyadh, Saudi Arabia. J Community Med Health Educ 2013;4(2):269. [doi: 10.4172/2161-0711.1000269]
17. Fox S. Pew Research Center. Washington, DC: Pew Internet & American Life Project The Social Life of Health Information, 2011 URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2019-04-16]
18. Benigeri M, Pluye P. Shortcomings of health information on the Internet. Health Promot Int 2003 Dec;18(4):381-386. [doi: 10.1093/heapro/dag409] [Medline: 14695369]

Abbreviations

HINTS: Health Information National Trends Survey
KKUH: King Khalid University Hospital

Edited by G Eysenbach; submitted 07.05.19; peer-reviewed by A Nguyen, T Fazzino, E Da Silva; comments to author 03.10.19; revised version received 26.02.20; accepted 27.02.20; published 19.03.20.

Please cite as:
Alduraywish SA, Altamimi LA, Aldhuwayhi RA, AlZamil LR, Alzeghayer LY, Alsaleh FS, Aldakheel FM, Tharkar S
Sources of Health Information and Their Impacts on Medical Knowledge Perception Among the Saudi Arabian Population: Cross-Sectional Study
J Med Internet Res 2020;22(3):e14414
URL: http://www.jmir.org/2020/3/e14414/
doi:10.2196/14414
PMID:32191208

Original Paper

Comparative Usability Analysis and Parental Preferences of Three Web-Based Knowledge Translation Tools: Multimethod Study

Harrison Anzinger[1], BSc; Sarah Alexandra Elliott[1,2], PhD; Lisa Hartling[1,2], PhD

[1]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada

[2]Cochrane Child Health, Edmonton, AB, Canada

Corresponding Author:
Lisa Hartling, PhD
Department of Pediatrics
University of Alberta
4-472 Edmonton Clinical Health Academy
11405-87 Avenue
Edmonton, AB, T6G 1C9
Canada
Phone: 1 7804926124
Email: hartling@ualberta.ca

Abstract

Background: Connecting parents to research evidence is known to improve health decision making. However, guidance on how to develop effective knowledge translation (KT) tools that synthesize child-health evidence into a form understandable by parents is lacking.

Objective: The aim of this study was to conduct a comparative usability analysis of three Web-based KT tools to identify differences in tool effectiveness, identify which format parents prefer, and better understand what factors affect usability for parents.

Methods: We evaluated a Cochrane plain language summary (PLS), Blogshot, and a Wikipedia page on a specific child-health topic (acute otitis media). A mixed method approach was used involving a knowledge test, written usability questionnaire, and a semistructured interview. Differences in knowledge and usability questionnaire scores for each of the KT tools were analyzed using Kruskal-Wallis tests, considering a critical significance value of *P*=.05. Thematic analysis was used to synthesize and identify common parent preferences among the semistructured interviews. Key elements parents wanted in a KT tool were derived through author consensus using questionnaire data and parent interviews.

Results: In total, 16 parents (9 female) with a mean age of 39.6 (SD 11.9) years completed the study. Parents preferred the Blogshot over the PLS and Wikipedia page (*P*=.002) and found the Blogshot to be the most aesthetic (*P*=.001) and easiest to use (*P*=.001). Knowledge questions and usability survey data also indicated that the Blogshot was the most preferred and effective KT tool at relaying information about the topic. Four key themes were derived from thematic analysis, describing elements parents valued in KT tools. Parents wanted tools that were (1) simple, (2) quick to access and use, and (3) trustworthy, and which (4) informed how to manage the condition. Out of the three KT tools assessed, Blogshots were the most preferred tool by parents and encompassed these four key elements.

Conclusions: It is important that child health evidence be available in formats accessible and understandable by parents to improve decision making, use of health care resources, and health outcomes. Further usability testing of different KT tools should be conducted involving broader populations and other conditions (eg, acute vs chronic) to generate guidelines to improve KT tools for parents.

(J Med Internet Res 2020;22(3):e14562) doi:10.2196/14562

KEYWORDS

child health; knowledge translation; parents; internet; comparative study; health information,consumer

Introduction

Background

The impact of health research is often minimized by ineffective communication and utilization. Knowledge translation (KT) offers a solution to this problem through raising awareness of evidence and facilitating its use. Impactful KT goes beyond dissemination, and involves engagement, participation, and evaluation by knowledge users [1]. Cochrane Child Health seeks to facilitate the uptake of evidence and evidence-informed decisions by key knowledge users (clinicians, patients, parents, and policy makers) and other stakeholders (eg, health system organizations such as Alberta Health Services) through the development and dissemination of a variety of engaging KT tools. Owing to the unique needs of pediatric patients, Cochrane Child Health aims to develop parent- and guardian-directed KT tools to meet their needs. Although we know connecting parents and caregivers to research evidence has the power to improve health decision making and appropriate access of health care services [2], traditional KT tools directed at health care professionals remain too complex for parents to effectively utilize [3]. Although many previous attempts have been made to create more user-friendly KT tools, finding a method of realizing this goal has been a persistent difficulty in the child health field.

Usability is a concept developed in the software and Web design industry that is increasingly being applied to KT tools to systematically address factors limiting a KT tool from meeting the needs of its audience. Usability aims to develop tools that provide relevant information in a satisfying, effective, and efficient way to the target end user [3]. At the center of improving usability is field testing and iterative design, which has been adopted by the KT field to develop more user-centered tools for consumers [4-8]. Through this method, KT tools are developed, evaluated, redesigned, and reevaluated based on feedback from the end user. Comparative usability analysis offers another method of assessing the usability of a product. In comparative usability analysis, several prototypes or competing products are compared with one another to identify strengths and weaknesses between the products. Allowing participants to see multiple designs allows them to provide comparative feedback and identify specific areas they like or dislike about a product. Doing such work early in design provides an opportunity to create an end product using the best features from each tool and to better understand the underlying concepts leading to end users' preferences for a specific product. Although iterative design has been readily adopted into the KT research, comparative usability analysis remains unused in the broader usability field.

Despite usability importance and integration, many KT tools remain complicated and inaccessible to parents and caregivers, creating a drought of understandable knowledge on many topics. As a result, internet KT tools have emerged as a popular source to fill the public demand for understandable, accessible health care knowledge, with 69% of Canadians reporting using the internet for health-related information [9]. Health information on the internet is used in a multitude of ways, including as a second opinion, to determine when to access care, and to inform lifestyle changes [10]. However, with such a wide range of information available over Web, the quality of health information is often mixed and inaccurate [11,12].

In recent years, Cochrane initiatives have aimed to address this need by developing quality KT tools that are available to consumers over Web. Despite Cochrane systematic reviews being regarded as providing the highest quality evidence to make informed choices about health care treatment; they are often inaccessible and impenetrable to parents and familial caregivers. Cochrane has addressed these usability concerns by developing several consumer-orientated KT tools. There are three such tools: the plain language summaries (PLS), Blogshots, and updated Wikipedia pages through the Cochrane-Wikipedia partnership.

Objectives

Although PLS, Blogshots, and related Wikipedia pages on child health–related topics are readily available to the public, little is known about their usability to parents and familial caregivers. Specifically, we wanted to know more about parents' views of these tools to further increase their usability and inform future KT tool creation. Subsequently, we conducted a comparative usability analysis of three Web-based Cochrane KT tools to identify differences in tool effectiveness, identify which format parents prefer, and better understand what factors affect usability for parents.

Methods

Overview

A mixed method study comprising a knowledge test, written questionnaire, and semistructured interview was conducted to assess the usability of a child health–related Cochrane PLS, Blogshot (developed by Cochrane Child Health), and Wikipedia page on the topic *acute otitis media* (AOM). Study participants were randomly assigned one KT tool to evaluate.

Interview design is demonstrated in the participant flow diagram (Figure 1). Briefly, participants were given a situational prompt and instructed to read or skim the KT tool as they would if they suspected their child had an ear infection. Participants had as long as they wanted to read the KT tool, and reading time was recorded without the participant's knowledge as a measure of tool efficiency. As we define usability as a tool that is satisfying, effective, and efficient to the end user, we felt that recording reading time was an important marker of efficiency [3]. The participants were given a structured questionnaire (including a knowledge test) about the KT tool they had just read (Multimedia Appendix 1). Participants then took part in a one-on-one interview. Half way through the interview, participants were shown the two other KT tools for comparison. Participants were then asked to rank the three KT tools in order of aesthetics, ease of use, credibility, and general preference.

Figure 1. Participant flow diagram. Participants (n=16) were randomized to 1 of 3 groups: Cochrane plain language summary (n=5), Cochrane Blogshot (n=5), or Wikipedia Page (n=6). Participants had unlimited time to read the knowledge translation (KT) tool. Participants were then given a knowledge test without being able to refer back to the KT tool. Participants then completed a written usability questionnaire and semistructured interview focusing on the KT tool randomly assigned. Finally, participants were asked to read the other two KT tools, and a second semistructured interview was completed focusing on comparing the three tools and broader participant preferences for KT tools. All participants completed the study. KT: knowledge translation; PLS: plain language summary.

Morville's Honeycomb Model of User Experience Design [13] (Figure 2) was used to design the questionnaire and interview questions to properly assess the usability of the KT tools and identify ways to improve them. The honeycomb model breaks user experience into seven categories (usability, credibility, usefulness, desirability, findability, value, and accessibility) that can be used to categorize the various aspects of a user-friendly system and has been successfully used and validated by several past KT usability studies to design interviews and organize results [7,14,15].

Figure 2. The honeycomb model of user experience.

Sampling and Recruitment

Participants were eligible for enrollment into the study if they currently were a parent, guardian, or grandparent of a child under 18 years old, were 18 years or older themselves, and could read and speak English. The study was advertised via email to our current Pediatric Parent Advisory Group members and members of the Edmonton Early Childhood Coalition, as well as via Facebook, Twitter, and word of mouth throughout our local community. Ethics approval was received from our institutional ethics board, and all participants gave informed consent before any data collection.

Study Components

Knowledge Questions and Structured Questionnaire

The knowledge test included six short answer questions to assess the effectiveness of the KT tool at accurately communicating health evidence (Multimedia Appendix 2). The participants were also asked to indicate the confidence in their responses using a 5-point Likert scale (very unsure, a little unsure, neither sure nor unsure, a little sure, and very sure) [8]. The questionnaire also asked a series of usability-related questions using an 11-point anchored Likert scale to obtain quantitative usability data regarding the individual KT tools.

Semistructured Interview

The third component of the study was a semistructured interview (Multimedia Appendix 1) consisting of two parts. The first part of the interview focused on the usability of the tool they were randomly assigned. Participants were then asked to read the other two tools. The second part of the interview asked the participants to compare the three KT tools and asked more general questions about KT tool preferences (eg, “What are the key parts of an ideal KT tool?” and “How much information is needed about the condition?”). The interview was field tested and adapted over three nonrecorded interviews using in-house parent volunteers.

Child Health Topic

Antibiotic use to treat AOM in children was selected as the health topic for the study because of the availability of a recent Cochrane review [16], the high prevalence and disease burden of AOM [17,18], high use of antibiotics to treat uncomplicated AOM despite best evidence [19,20], parental misinformation for treatment [21,22], and the identification of AOM as a priority for parental education from previous studies [22,23].

Knowledge Translation Tools Evaluated

A Cochrane PLS [16], Blogshot [24], and Systematic EvidEnce Disseminator (SEED)–updated Wikipedia page [25] on the use of antibiotics to treat AOM in children were recently developed or updated (Multimedia Appendix 1). These tools are publicly accessible over Web, the specifics of which are outlined below.

Plain Language Summary

Cochrane PLS were developed as a standalone summary of Cochrane systematic review findings aimed directly at health care consumers and written in plain language [26]. All Cochrane reviews must have a PLS freely available on Cochrane’s website and are required to follow a set of guidelines aimed to make them easy for the public to understand [27].

Blogshot

Cochrane Blogshots were originally developed by Cochrane UK as another method to effectively communicate recent Cochrane systematic reviews to consumers through social media [28]. Blogshots aim to present the key points in the Cochrane review relevant to consumers in a concise picture that can easily be shared on social media. Cochrane Child Health developed a blog, which houses the Blogshots and describes the results of the corresponding Cochrane review in an easy to read and less formal tone [24]. They are also shared on the Cochrane Child Health Facebook and Twitter accounts.

Systematic EvidEnce Disseminator–Updated Wikipedia Page

The Cochrane-Wikipedia partnership was formalized in 2014 with the aim to include relevant evidence and ensure the accuracy of all Wikipedia medical articles. Recently a novel software, SEED, has been developed to automatically generate summary of findings like tables compatible with Wikipedia from Cochrane Review Manager files [29]. These tables aim to present Cochrane review findings with a short description of the medical context in a way that is understandable and relevant to for health care consumers.

Statistical Analysis

Knowledge test scores, KT tool reading times, and written questionnaire responses were treated as continuous data, and analyzed using a Kruskal-Wallis test because of failure to meet the normality and homogeneity of variance assumptions required for a three-way analysis of variance test. Knowledge test *confidence* scores were treated as ordinal data and analyzed via a Kruskal-Wallis test to identify a difference in confidence between groups. All analyses were performed using SPSS (IBM Corp) or Excel (Microsoft), considering a critical significance value *P*=.05.

Rank data from the semistructured interview were analyzed via Friedman test with a post hoc Wilcoxon signed-rank test to determine if any KT tool consistently ranked better in a given category. For post hoc testing among the three groups, a Bonferroni correction was applied resulting in a corrected significance value of *P*=.02.

Thematic Analysis

Thematic analysis was used to synthesize and identify common parent preferences described in the semistructured interviews. Data management and analysis were facilitated using NVivo 12 Software (v.12, 2017 QSR International PTY Ltd.). Our process of thematic analysis followed the method outlined by Braun and Clarke [30]: familiarization with the data, initial coding, searching for themes among the initial codes, reviewing themes that may fit together as subthemes, and, then, defining and naming final major themes that best represented the data.

Data collection and analysis occurred iteratively, allowing for more precise and purposeful data collection. Data collection continued until saturation of major thematic categories was achieved. In our case, this was 16 interviews. Interviews were coded and categorized to facilitate development of themes. An inductive *bottom up* approach was taken, aiming to strongly link the developed codes and themes to the data themselves. The interviewer became immersed in the data through transcription of the recorded interviews. Initial coding for the first seven interviews was conducted using a *line-by-line* approach. Codes stayed close to the data by using participants’ own words as much as possible. Focused coding, grouping similar codes together, was then used to identify patterns in the data. These focused codes developed from the first seven interviews were then used directly for the remaining eight interview transcripts. To reduce interpretive bias, a second reviewer coded and categorized a random sample of 50% (7/15) of the interview transcripts. Any discrepancies between the reviewers were discussed and resolved via consensus. The focused codes were further refined via collaboration between the 2 reviewers into themes and subthemes that identified common factors contributing to parental preference and usability. All codes and transcripts were then re-examined to ensure consistency and accuracy of the interpretation.

Recommendations

Recommendations for developing KT tools are proposed based on consensus of the authors using data gathered from questionnaire and parent interviews. These recommendations considered the core and subthemes of parent preferences identified through thematic analysis, trends in the questionnaire results, and the authors’ interpretation of the raw interview transcripts. These recommendations serve to create actionable items for researchers developing KT tools but represent a subjective interpretation of the data from the authors.

Results

Study Participants

The usability of three Cochrane KT tools on AOM was assessed by 16 parents. Participant demographics are presented in Table 1. Briefly, all participants were parents or grandparents, with 56% (9/15) of respondents being female. All participants had at least a high school diploma, with 88% (14/16) having a postsecondary degree equivalent or higher. In all, 25% (4/16) of participants self-reported that their child has had AOM in the past, and 19% (3/16) worked in the health care field. Finally, 63% (10/16) had heard of Cochrane before participation in the study, but many were unfamiliar with what research activities Cochrane Child Health and the Cochrane organization carry out.

Table 1. Demographic characteristics of participants in the study randomly assigned to each knowledge translation tool.

Characteristic	Knowledge translation tool evaluated			Total (N=16)
	PLS[a] (n=5)	Blog[b] (n=5)	Wiki[c] (n=6)	
Gender, n				
Female	3	3	3	9
Male	2	2	3	7
Parents' age (years), n				
20-30	3	1	0	4
31-40	0	3	2	5
41-50	1	1	3	5
51+	1	0	1	2
Number of children, n				
1	3	2	2	7
2	2	2	2	6
3	0	1	1	2
4	0	0	1	1
Highest level of education, n				
High school diploma	0	0	0	0
Some postsecondary	1	0	1	2
Postsecondary degree	1	2	1	4
Master's	2	2	4	8
PhD	1	1	0	2
Works in health care, n				
Yes	0	1	2	3
Heard of Cochrane reviews, n				
Yes	3	2	5	10
Child has had AOM[d], n				
Yes	3	0	1	4

[a] PLS: plain language summary.

[b] Blog: Blogshot.

[c] Wiki: Wikipedia page.

[d] AOM: acute otitis media.

Effectiveness of Tested Knowledge Translation Tools

We assessed efficiency through timing parents reading the KT tool and effectiveness through a short answer knowledge retention test in addition to parents self-reported confidence in their answers.

Parents on average spent less time (mean, range) reading the PLS (163 seconds, 129-200 seconds) and Blogshot (168 seconds, 97-257 seconds) than the updated Wikipedia page (275 seconds, 91-519 seconds); however, there was no significant difference among the three groups, χ^2_2=3.2, P=.20 (Table 2, Figure 3). Although parents spent the most time reading the Wikipedia article, the Wikipedia page also had the shortest and longest reading times, 92 seconds and 519 seconds, respectively. There was no difference in efficiency among the three KT tools.

Table 2. Summary of quantitative data from knowledge test and usability questionnaire (Q1-Q9 refer to usability questionnaire included in Multimedia Appendix 2. Likert scales have been flipped so a higher number is always better).

Variables assessed	Knowledge translation tool			*P* value (Kruskal-Wallis)
	PLS[a]	Blogshot	Wikipedia	
Sample size (N=16), n	5	5	6	N/A[b]
Time reading tool (seconds), mean (95% CI)	163 (127-200)	168 (91-244)	275 (114-435)	.20
Knowledge test score, mean (95% CI)	4.2 (2.4-6.0)	4.6 (3.2-6.0)	3.7 (3.1-4.2)	.31
Survey responses, mean (95% CI)				
Q1 Key information	6.2 (3.0-9.4)	8.2 (6.2-10.0)	7.8 (6.0-9.6)	.44
Q2 Easy to remember	3.6 (0.4-6.8)	7.2 (4.2-10.0)	3.8 (0.84-6.8)	.08
Q3 Increased knowledge	8.2 (6.2-10.0)	8.6 (6.9-10.0)	6.5 (2.7-10.0)	.54
Q4 Aesthetically pleasing	4.8 (1.7-7.9)	6.8 (3.4-10.0)	2.5 (0.3-4.7)	.07
Q5 Mentally demanding	2.6 (1.2-4.0)	5.4 (2.2-8.6)	3.2 (1.8-4.6)	.11
Q6 Understandable	5.6 (3.2-8.0)	7.6 (5.3-9.9)	7.7 (5.0-10.0)	.16
Q7 Frustrating to read	4.4 (1.2-7.6)	7.8 (5.0-10.0)	4.5 (0.9-8.1)	.11
Q8 Helps with decision making	6.6 (1.7-10.0)	7.6 (4.2-10.0)	5.7 (4.1-7.2)	.32
Q9 Would recommend to others	4.8 (0.38-9.2)	7 (3.7-10.0)	3.7 (0.8-6.5)	.20

[a] PLS: plain language summary.

[b] Not applicable.

Figure 3. Plain language summary, Blogshot, and Wikipedia ranked in order of most (Rank 1) to least (Rank 3) in categories of aesthetics, ease of use, credibility, and general preference. Dot area indicates number of participants placing each tool at a given rank (N=16). Asterisk indicates statistically significant difference in rank order between groups determined via a Friedman test (*P*<.05). PLS: plain language summary.

Although there was no difference among groups in knowledge test scores (χ^2_2=2.4, *P*=.31), parents on average answered the most questions correctly after reading the Blogshot (4.6/6), followed by the PLS (4.2/6) and Wikipedia page (3.7/6). Parents who were assigned the Blogshot reported feeling slightly more confident in their responses; however, this was not significantly different across groups.

Identifying Knowledge Translation Tool Preference

Parents' ranking results of the three tools for aesthetics, ease of use, credibility, and general preference are presented in Figure 3. There was a significant difference in the ranking of aesthetics (χ^2_2=24.1, *P*<.001), ease of use (χ^2_2=21.9, *P*≤.001), and general preference (χ^2_2=11.6, *P*=.003) among the three tools. Post hoc analysis showed that parents preferred the Blogshot over the PLS or Wikipedia page in the categories of aesthetics (*P*<.001 and *P*<.001, respectively), ease of use (*P*<.001 and *P*=.001, respectively), and general preference (*P*=.004 and *P*=.011, respectively). There was no significant difference in the ranking between the PLS and Wikipedia page in any category. There was no statistically significant difference in the ranking of the credibility of the three tools (χ^2_2=1.4, P=.50). Overall, parents found the Cochrane Blogshot to be the most preferred, the most aesthetic, and the easiest tool to use.

Identifying Factors Leading to Parental Preferences

Survey responses did not show any significant difference in parent preferences among the three KT tools (Table 2). However, the Blogshot was ranked highest in all categories except understandability, where participants narrowly preferred the updated Wikipedia page (7.6 vs 7.7, respectively).

However, four themes were identified through thematic analysis relating to parental preferences (Table 3). A detailed explanation on how we reached these themes is included in the Methods section. In brief, parents want a tool that is (1) simple, (2) trustworthy, and (3) quick to access and use, and which (4) informs how to manage the condition. A detailed description of each theme and supportive illustrative quotes are documented in Multimedia Appendix 2.

Table 3. Themes and subthemes of parental knowledge translation tool preferences.

Theme	Subthemes
Simple	• Understandable language • Nonmedical graphics • Simple and familiar aesthetic • Not interested in study characteristics
Trustworthy	• Evidence based • Recognized source • Cites sources • Peer advice
Quick to access and use	• Easy-to-find electronic tools • Efficient organization • Concise • Usable in stressful scenarios
Informs how to manage the condition	• Describes what to expect • Explains when to seek care • Describes how to manage the child at home

Recommendations

These recommendations (Table 4) take into account both the quantitative and qualitative observations from the gathered data and should be considered when developing child health–related KT tools for parents or caregivers. A detailed explanation for how we arrived at these recommendations can be found in the Methods section.

Table 4. Checklist of recommendations for developing child health knowledge translation tools for parents or caregivers.

Elements	Recommendation
Content	• Introduction should contain symptoms, so parents can determine if the article is relevant. • Include information on how to manage child at home, when to seek care, and what to expect. • Only include statistics from the research if they are meaningful to a caregiver not familiar with the condition.
Format	• Use headings and bullet points to offer a familiar and efficient organization. • Replace words with graphics wherever possible. • Be concise, but link to more information. More details confuse parents and reduce readership.
Language	• Use repetition selectively and sparingly. • Avoid abbreviations, jargon, and polysyllable words to make writing easy to understand. • Write as if you are speaking to a stressed parent.
Branding	• Associate with a recognized brand to increase credibility and findability.

Discussion

Principal Findings

We found no significant differences among the PLS, Blogshot, and Wikipedia pages in terms of reading time and knowledge retention. This finding is consistent with a pervasive difficulty in the KT field of identifying quantitative differences in the effectiveness of different KT tools [2]. An identified flaw contributing to this problem is the lack of adequate power in many studies to identify statistically significant differences. Often, as in our case, the desire to include time-intensive interviews and qualitative analysis makes reaching such statistical power unachievable. Future studies interested in the quantitative differences between tools may consider eliminating or limiting qualitative analysis to sufficiently reach those goals.

Using the honeycomb model of user experience to describe usability, we conducted a comparative usability analysis of three Web-based KT tools. We determined that parents prefer Blogshots as a KT tool compared with the PLS and updated Wikipedia pages, and ultimately found Blogshots to be the most usable tool. Usability remains difficult to define universally in the KT field. Although many previous studies have similarly defined usability through a model such as the honeycomb model of user experience [7,14,15], others have defined it to meet specific objectives of their research [5,31]. Without an agreed upon model of usability within the KT field, the definition ultimately depends on how it is measured in each individual study [32]. Unfortunately, measurement techniques vary significantly between studies as well, often leading to significant bias and results of questionable utility to the larger field [2,32]. As KT usability research continues to develop, it will be vital to move toward a standard definition to allow research with broad generalizable implications to be realized.

Through thematic analysis of interview transcripts, we revealed four emergent themes: parents want KT tools that are (1) simple, (2) easy to access and use, and (3) trustworthy, and which (4) inform how to manage the condition. Although these themes seem simple to understand and implement, many are lacking from currently used pediatric KT tools. Previous KT studies using thematic analysis focused on a single tool, using thematic analysis as a way of recognizing usability hurdles specific to that tool [7,8,14,33]. To our knowledge, no previous study has identified themes leading to KT tool preference and usability that are applicable to a wider range of KT tools. We hope that by using a comparative usability methodology, these themes will be more broadly generalizable to the larger pediatric KT field. Furthermore, using all the data collected and interviews conducted, we developed a set of recommendations for developing KT tools directed at parents. These findings were consistent with previously identified weaknesses in the usability of the Cochrane Library as a whole for health professionals [15].

Through a comparative usability analysis, we present a novel method of identifying factors contributing to usability that are more generalizable than previous KT tool–specific studies. Comparative usability analysis is different from traditional comparative studies where participants are assigned to 1 of typically 2 tools and quantitatively compared. In comparative usability analysis, participants are shown all KT tools and asked to identify what factors, shared or unique, about elements they preferred or disliked. By showing participants many ways of presenting the same information, we theorize that it primes participants to think about new ways of presenting the same material, and can lead to more creative and insightful feedback. While achieving useful insights about each tool, this method also contributes to our understanding of the universal themes leading to parental preferences and ultimately determining the degree of usability across all KT tools. We hypothesize that our recommendations are more methodologically rigorous than conducting a similar analysis where participants are shown only a single tool, yet further studies are needed to develop the comparative usability analysis method in the field of KT.

Another interesting result in our study was that the participants ranked the three KT tools as being similarly credible. We theorized that Wikipedia would be ranked lowest in credibility because of our perceived common understanding that it is editable by anyone. Credibility, however, was very difficult for participants to assess and rank among the three tools. Interestingly, it was difficult to assess not because people's perception of the tools' credibility varied significantly, but because most viewed all three tools as being very credible. Participants often cited familiarity with the source, such as Wikipedia, or traits that make it look official, such as citations, complex terms, and familiar sources contributing to their perception of credibility. Similarly, previous research into credibility identified that consumers chiefly make credibility judgments based on factors visually prominent on the page such as aesthetics and brand recognition [34,35]. With the large availability of poor-quality information over Web, this contributes to the potential of parents accessing misinformation unintentionally. It also further emphasizes the importance of disseminating quality KT tools to consumers to compete with well-designed, but factually incorrect, Web-based health information.

Limitations

Despite having good representation across age groups, number of children cared for, and both mothers and fathers, our sample size and selection was the largest limiting factor in our study. Owing to the qualitative components of the study, we were limited to a relatively small sample size (N=16), which restricts the utility of the quantitative results. It is possible that with a larger sample size, we may have identified statistically significant differences among the effectiveness and efficiency of the three KT tools. Future quantitative studies with larger sample sizes are still needed to address the difference of effectiveness among the three tools. However, our sample size allowed for a more robust and bottom-up approach to the qualitative analysis reducing bias in our interpretation. We found that the interviews were very similar between individuals with the four identified themes being touched on in nearly every interview.

Our sample also consisted of mostly highly educated individuals, with 88% (14/16) participants having a postsecondary degree or higher. Although it is possible that parents or guardians with a lower level of education may prefer information presented differently, some of these concerns were reduced by the participants indicating that they desired simplicity and less technical information.

Finally, as the sample consisting of only parents or guardians, the results should be interpreted with caution for developing adult health care KT tools. Although many of the themes and preferences may be similar, the difference between adult and pediatric medicine, as well as the different dynamics between being a caregiver and a patient, may significantly change what an individual would want from a KT tool. Furthermore, we only conducted research on a single acute condition, and the findings may not be applicable to more urgent or chronic pediatric conditions.

Conclusions

The comparison of three Web-based KT tools on the same pediatric health topic allowed us to identify four underlying themes parents want from a KT tool: simple, easy to access and use, trustworthy, and to be able inform how to manage the condition. The development and exploration of subthemes provide meaningful insight into how to achieve these simple but hard to reach objectives when translating health evidence for general consumption. We identified that Blogshots are the preferred (from the three tools tested) KT tool for parents, and using research consensus, these findings have been translated into a checklist to consider when developing a KT tool aimed at parents and familial caregivers. Our research provides meaningful insight for developing and improving future pediatric KT tools. Further usability testing of different KT tools should be conducted involving broader populations, other conditions (eg, acute vs chronic), and varying decision-making needs to generate guidelines to improve KT tools for parents, ultimately improving child health outcomes.

Acknowledgments

The authors would like to extend their sincerest appreciation to the parents and caregivers involved in this study for their valuable contribution. The work was supported in part by a Foundation Grant from the Canadian Institutes of Health Research and Alberta Health Services, Maternal Newborn Child & Youth Strategic Clinical Network Health Outcomes Improvement Fund. HA was supported by an Alberta Innovates Summer Studentship for the duration of this work. LH was supported by a Canada Research Chair (Tier 1) in Knowledge Synthesis and Translation.

The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Authors' Contributions

SAE and LH were responsible for the study concept. HA and SAE designed the study. HA recruited participants and performed data collection and analysis with input from SAE and LH. All authors have contributed to drafting the manuscript and have approved the final version.

Conflicts of Interest

Two authors (SAE and LH) of this paper are members of the Cochrane Collaboration and Cochrane Child Health. Cochrane Child Health produced the Blogshot. Although Wikipedia is an independent organization, there is an agreement between Wikipedia and the Cochrane Collaboration to allow dissemination of Cochrane studies on Wikipedia articles.

Multimedia Appendix 1
Interview package containing participant instructions, knowledge questions, usability questionnaire, and semi-structured interview prompts.
[DOCX File , 55 KB - jmir_v22i3e14562_app1.docx]

Multimedia Appendix 2
Subtheme description and illustrative quotes for parental knowledge translation tool preference.
[DOCX File , 20 KB - jmir_v22i3e14562_app2.docx]

References

1. Kothari A, Wathen CN. Integrated knowledge translation: digging deeper, moving forward. J Epidemiol Community Health 2017 Jun;71(6):619-623. [doi: 10.1136/jech-2016-208490] [Medline: 28298415]
2. Albrecht L, Scott SD, Hartling L. Knowledge translation tools for parents on child health topics: a scoping review. BMC Health Serv Res 2017 Sep 29;17(1):686 [FREE Full text] [doi: 10.1186/s12913-017-2632-2] [Medline: 28962637]
3. International Organization for Standardization. 2018. Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2019-05-01]
4. Buljan I, Malički M, Wager E, Puljak L, Hren D, Kellie F, et al. No difference in knowledge obtained from infographic or plain language summary of a Cochrane systematic review: three randomized controlled trials. J Clin Epidemiol 2018 May;97:86-94 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.12.003] [Medline: 29269021]
5. Kools M, Ruiter RA, van de Wiel MW, Kok G. Testing the usability of access structures in a health education brochure. Br J Health Psychol 2007 Nov;12(Pt 4):525-541. [doi: 10.1348/135910706X132930] [Medline: 17931471]
6. Marquez C, Johnson AM, Jassemi S, Park J, Moore JE, Blaine C, et al. Enhancing the uptake of systematic reviews of effects: what is the best format for health care managers and policy-makers? A mixed-methods study. Implement Sci 2018 Jun 22;13(1):84 [FREE Full text] [doi: 10.1186/s13012-018-0779-9] [Medline: 29929538]

7. Rosenbaum SE, Glenton C, Nylund HK, Oxman AD. User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews. J Clin Epidemiol 2010 Jun;63(6):607-619. [doi: 10.1016/j.jclinepi.2009.12.013] [Medline: 20434023]
8. Taddio A, Shah V, Leung E, Wang J, Parikh C, Smart S, et al. Knowledge translation of the HELPinKIDS clinical practice guideline for managing childhood vaccination pain: usability and knowledge uptake of educational materials directed to new parents. BMC Pediatr 2013 Feb 8;13:23 [FREE Full text] [doi: 10.1186/1471-2431-13-23] [Medline: 23394070]
9. Statistics Canada. 2012. Internet Use by Internet Activity, Age Group, Sex, Level of Education and Household Income URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2710001601&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=3.4&pickMembers%5B2%5D=4.1&pickMembers%5B3%5D=5.1&pickMembers%5B4%5D=6.12012 [accessed 2019-05-01]
10. Fox S, Rainie L, Horrigan J, Lenhart A, Spooner T, Burke M, et al. Office of The Assistant Secretary for Planning and Evaluation. Washington, DC: Pew Research Center; 2001. The Online Health Care Revolution: How the Web Helps Americans Take Better Care of Themselves URL: https://aspe.hhs.gov/pew-internet-american-life-project-online-life-report-january-11-2001-ncvhs-joint-hearings [accessed 2019-05-01] [WebCite Cache ID 783AGmpc7]
11. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. J Am Med Assoc 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
12. Parvizi M, Talai N, Parvizi Z. Quality of healthcare information on the internet: the case of Apicectomies. Oral Surg 2017;10(4):e35-e39. [doi: 10.1111/ors.12263]
13. Morville P. Semantic Studios. 2004 Jun 21. User Experience Design URL: http://semanticstudios.com/user_experience_design/ [accessed 2019-04-10] [WebCite Cache ID 77X8eZHY6]
14. Fearns N, Graham K, Johnston G, Service D. Improving the user experience of patient versions of clinical guidelines: user testing of a Scottish Intercollegiate Guideline Network (SIGN) patient version. BMC Health Serv Res 2016 Feb 2;16:37 [FREE Full text] [doi: 10.1186/s12913-016-1287-8] [Medline: 26837683]
15. Rosenbaum SE, Glenton C, Cracknell J. User experiences of evidence-based online resources for health professionals: user testing of The Cochrane Library. BMC Med Inform Decis Mak 2008 Jul 28;8:34 [FREE Full text] [doi: 10.1186/1472-6947-8-34] [Medline: 18662382]
16. Venekamp RP, Sanders SL, Glasziou PP, Del Mar CB, Rovers MM. Antibiotics for acute otitis media in children. Cochrane Database Syst Rev 2015 Jun 23(6):CD000219. [doi: 10.1002/14651858.CD000219.pub4] [Medline: 26099233]
17. Dubé E, de Wals P, Gilca V, Boulianne N, Ouakki M, Lavoie F, et al. Burden of acute otitis media on Canadian families. Can Fam Physician 2011 Jan;57(1):60-65 [FREE Full text] [Medline: 21252135]
18. Dubé E, de Wals P, Ouakki M. Quality of life of children and their caregivers during an AOM episode: development and use of a telephone questionnaire. Health Qual Life Outcomes 2010 Jul 26;8:75 [FREE Full text] [doi: 10.1186/1477-7525-8-75] [Medline: 20659326]
19. Wang EE, Einarson TR, Kellner JD, Conly JM. Antibiotic prescribing for Canadian preschool children: evidence of overprescribing for viral respiratory infections. Clin Infect Dis 1999 Jul;29(1):155-160. [doi: 10.1086/520145] [Medline: 10433579]
20. McCaig LF, Hughes JM. Trends in antimicrobial drug prescribing among office-based physicians in the United States. J Am Med Assoc 1995 Jan 18;273(3):214-219. [Medline: 7807660]
21. Finkelstein JA, Stille CJ, Rifas-Shiman SL, Goldmann D. Watchful waiting for acute otitis media: are parents and physicians ready? Pediatrics 2005 Jun;115(6):1466-1473. [doi: 10.1542/peds.2004-1473] [Medline: 15930205]
22. Meherali S, Campbell A, Hartling L, Scott S. Understanding parents' experiences and information needs on pediatric acute otitis media: a qualitative study. J Patient Exp 2019 Mar;6(1):53-61 [FREE Full text] [doi: 10.1177/2374373518771362] [Medline: 31236452]
23. Featherstone RM, Leggett C, Knisley L, Jabbour M, Klassen TP, Scott SD, et al. Creation of an integrated knowledge translation process to improve pediatric emergency care in Canada. Health Commun 2018 Aug;33(8):980-987. [doi: 10.1080/10410236.2017.1323538] [Medline: 28537762]
24. Cochrane Child Health. 2015 Apr 27. Evidence-Based Treatments for Acute Otitis Media (Middle Ear Infection) URL: https://cochranechild.wordpress.com/2015/04/27/evidence-based-treatments-for-acute-otitis-media/ [accessed 2019-04-10] [WebCite Cache ID 77X9Awa8r]
25. Wikipedia. 2019. Otitis Media URL: https://en.wikipedia.org/wiki/Otitis_media [accessed 2019-04-12] [WebCite Cache ID 77a9s2L4Y]
26. Schünemann HJ, Oxman AD, Higgins JP, Vist GE, Glasziou P, Guyatt GH. Presenting results and ‘summary of findings’ tables. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, United States: Wiley; 2008:335-357.
27. Cochrane Methods. London: Cochrane Collaboration; 2013. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Standards for the reporting of Plain Language Summaries in new Cochrane Intervention Reviews URL: https://methods.cochrane.org/sites/default/files/public/uploads/PLEACS_0.pdf [accessed 2019-04-01] [WebCite Cache ID 77X8yWc4c]

28. Arienti C, Gimigliano F, Ryan-Vig S, Kiekens C, Negrini S. Cochrane Rehabilitation blogshots: a modern method to spread Cochrane Evidence. Eur J Phys Rehabil Med 2018 Jun;54(3):466-468 [FREE Full text] [doi: 10.23736/S1973-9087.18.05313-3] [Medline: 29901360]
29. Schmidt L, Friedel J, Adams CE. SEED: a tool for disseminating systematic review data into Wikipedia. Syst Rev 2017 Oct 17;6(1):206 [FREE Full text] [doi: 10.1186/s13643-017-0607-3] [Medline: 29041959]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. MacDermid JC, Uddin Z. A knowledge translation perspective on the two quantitative sensory tests and their usability with clinicians. J Nov Physiother 2015;5:257. [doi: 10.4172/2165-7025.1000257]
32. Hornbæk K. Current practice in measuring usability: challenges to usability studies and research. Int J Hum Comput Stud 2006;64(2):79-102. [doi: 10.1016/j.ijhcs.2005.06.002]
33. Reid K, Hartling L, Ali S, Le A, Norris A, Scott SD. Development and usability evaluation of an art and narrative-based knowledge translation tool for parents with a child with pediatric chronic pain: multi-method study. J Med Internet Res 2017 Dec 14;19(12):e412 [FREE Full text] [doi: 10.2196/jmir.8877] [Medline: 29242180]
34. Fogg BJ, Soohoo C, Danielson DR, Marable L, Stanford J, Tauber ER. How Do Users Evaluate the Credibility of Web Sites?: A Study With Over 2,500 Participants. In: Proceedings of the 2003 conference on Designing for user experiences. 2003 Presented at: DUX'03; June 6 - 7, 2003; California, San Francisco p. 1-15. [doi: 10.1145/997078.997097]
35. Robins D, Holmes J, Stansbury M. Consumer health information on the web: the relationship of visual design and perceptions of credibility. J Am Soc Inf Sci Technol 2010;61(1):13-29. [doi: 10.1002/asi.21224]

Abbreviations

AOM: acute otitis media
KT: knowledge translation
PLS: plain language summary
SEED: Systematic EvidEnce Disseminator

Edited by G Eysenbach; submitted 02.05.19; peer-reviewed by M Mahmic Kaknjo, S Lin, D Scherer, P Santana-Mancilla; comments to author 29.09.19; revised version received 10.12.19; accepted 31.12.19; published 13.03.20.

Please cite as:
Anzinger H, Elliott SA, Hartling L
Comparative Usability Analysis and Parental Preferences of Three Web-Based Knowledge Translation Tools: Multimethod Study
J Med Internet Res 2020;22(3):e14562
URL: http://www.jmir.org/2020/3/e14562/
doi:10.2196/14562
PMID:32167478

Original Paper

Utilization of a Voice-Based Virtual Reality Advanced Cardiac Life Support Team Leader Refresher: Prospective Observational Study

Daniel Katz[1*], MD; Ronak Shah[1*], MD; Elizabeth Kim[2*], BA; Chang Park[1*], MD; Anjan Shah[1*], MD; Adam Levine[1*], MD; Garrett Burnett[1*], MD

[1]Department of Anesthesiology, Pain, & Perioperative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States
[2]City University of New York City Medical School, New York, NY, United States
[*]all authors contributed equally

Corresponding Author:
Daniel Katz, MD
Department of Anesthesiology, Pain, & Perioperative Medicine
Icahn School of Medicine at Mount Sinai
New York, NY
United States
Phone: 1 7323226675
Email: dkatz621@gmail.com

Abstract

Background: The incidence of cardiac arrests per year in the United States continues to increase, yet in-hospital cardiac arrest survival rates significantly vary between hospitals. Current methods of training are expensive, time consuming, and difficult to scale, which necessitates improvements in advanced cardiac life support (ACLS) training. Virtual reality (VR) has been proposed as an alternative or adjunct to high-fidelity simulation (HFS) in several environments. No evaluations to date have explored the ability of a VR program to examine both technical and behavioral skills and demonstrate a cost comparison.

Objective: This study aimed to explore the utility of a voice-based VR ACLS team leader refresher as compared with HFS.

Methods: This prospective observational study performed at an academic institution consisted of 25 postgraduate year 2 residents. Participants were randomized to HFS or VR training and then crossed groups after a 2-week washout. Participants were graded on technical and nontechnical skills. Participants also completed self-assessments about the modules. Proctors were assessed for fatigue and task saturation, and cost analysis based on local economic data was performed.

Results: A total of 23 of 25 participants were included in the scoring analysis. Fewer participants were familiar with VR compared with HFS (9/25, 36% vs 25/25, 100%; P<.001). Self-reported satisfaction and utilization scores were similar; however, significantly more participants felt HFS provided better feedback: 99 (IQR 89-100) vs 79 (IQR 71-88); P<.001. Technical scores were higher in the HFS group; however, nontechnical scores for decision making and communication were not significantly different between modalities. VR sessions were 21 (IQR 19-24) min shorter than HFS sessions, the National Aeronautics and Space Administration task load index scores for proctors were lower in each category, and VR sessions were estimated to be US $103.68 less expensive in a single-learner, single-session model.

Conclusions: Utilization of a VR-based team leader refresher for ACLS skills is comparable with HFS in several areas, including learner satisfaction. The VR module was more cost-effective and was easier to proctor; however, HFS was better at delivering feedback to participants. Optimal education strategies likely contain elements of both modalities. Further studies are needed to examine the utility of VR-based environments at scale.

(J Med Internet Res 2020;22(3):e17425) doi:10.2196/17425

KEYWORDS

video game; experimental game; virtual reality; advanced cardiac life support

Introduction

Background

The incidence of cardiac arrests per year in the United States continues to increase, yet in-hospital cardiac arrest survival rates significantly vary between hospitals. Survival rates are reported between 11% to 35%, and patients in hospitals where clinical staff report adequate resuscitation training have greater odds of survival [1]. Health care professionals are often required to have advanced cardiac life support (ACLS) training depending on the institution, but despite training, survival rates are low. Current methods of face-to-face training are expensive, time consuming, and difficult to scale, which necessitates improvements in ACLS training aimed at improving patient survival. The current gold standard for ACLS training involves face-to-face, high-fidelity exercises that allow clinicians to work together to resolve mock resuscitation codes. An instructor observes the group while trainees perform a setlist of tasks for different clinical scenarios. At the end of the session, the team is given feedback on their performance. Although this is the standard for training, there are several limitations to this modality, including the need for lengthy sessions, expensive durable equipment, need for trained personnel, and difficulty with scale.

Virtual reality (VR) has been proposed as an alternative or adjunct to high-fidelity simulation (HFS) in several environments, including engineering [2], sports [3], and aviation [4]. Although it has been studied in the realm of ACLS education, to date, no evaluations have explored the ability of a VR program to examine both technical and behavioral skills and demonstrate a cost comparison [5]. In addition, previous studies involving VR applications in ACLS education relied on additional peripheral devices and did not utilize a fully immersive VR environment [6,7].

Objectives

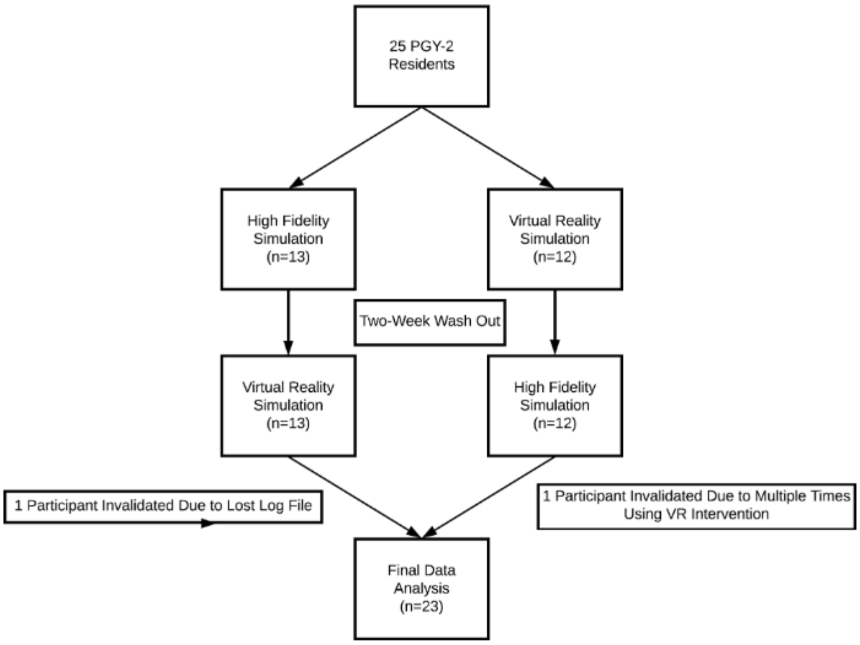
Therefore, we set out to examine the utility of a fully immersive VR-based team leader refresher to enhance ACLS skills and compare it with a traditional HFS team leader refresher.

Methods

Study Design

After obtaining the institutional review board (IRB) approval (IRB 19-02053) and written informed consent, 25 postgraduate year 2 (PGY-2) anesthesiology residents were recruited to participate in our prospective study at the Mount Sinai Human Emulation Education and Evaluation Lab for Patient Safety and Professional Study (HELPS) Center at the Icahn School of Medicine, New York. Each resident was 1 year past their first ACLS certification, had the same clinical rotations the year prior, and had passed the required examinations, and was in good standing with our department. Participants were then randomized into two groups using a balanced random number generator (see Figure 1) based on the first modality utilized. Each group then utilized the alternate modality after a 2-week washout period. A 2-week washout period was chosen to minimize the potential that the residents encountered codes outside of the training environments, yet it still allowed a separation between the two modalities to allow for independent observations of each modality. Study participants and proctors were not told about the purpose of the study. Participants were given no orientation to either modality.

Figure 1. Participant flow diagram.



High-Fidelity Simulation Intervention

HFS sessions were proctored by ACLS-certified instructors who were also board-certified anesthesiologists. All proctors were faculty at the Mount Sinai HELPS Center and had extensive experience with HFS. Instructors were given a rubric of all current American Heart Association (AHA) algorithms to be tested along with a rubric against which to grade in a *Correct, Correct with Assistance, and Incorrect* manner. Algorithms tested included all tachycardia and bradycardia rhythms and algorithms for ventricular tachycardia, ventricular fibrillation, and pulseless electrical activity. Only vocal skills were tested, and only the team leader role was examined. Participants were expected to delegate all manual tasks to other team members, including compressions, airway management, and defibrillation. For example, should the team leader determine compressions were needed, they were expected to tell a team member to begin compressions. After the session concluded, the learner was given a debrief in a structured manner and was given the opportunity to ask any questions. When the debriefing was finished, the session was completed. Simulations were performed on a human patient simulator mannequin (Canadian Aviation Electronics, CAE) using MUSE software (CAE, Montreal, Canada).

Virtual Reality Intervention

The VR intervention was an educational module designed by Health Scholars (Denver, CO) that tested participants on the same AHA algorithms as mentioned above. Sessions were run on Dell and HP laptops and Samsung and HP VR headsets using Windows (Microsoft, Redmond, WA) Mixed Reality Software. The session placed the participant in the role of team leader to take care of a critical patient in a radiology suite. The VR intervention utilized voice controls, with a virtual team to which the participant could delegate tasks (Figure 2). Participants were graded on the same rubric as above, including *Correct, Correct with Assistance, and Incorrect*. All of the same scenarios and algorithms were tested as in the HFS group. When the module concluded and the participant removed the headset, the session was completed. As with the high-fidelity arm, only vocal skills were tested, and only the team leader role was examined. Participants were expected to delegate all manual tasks to other team members, including compressions, airway management, and defibrillation. The participants did not interact with the proctor unless there was an issue with the functionality of the system. Participants were given as much time as needed to complete the module. Independent proctors were given transcripts of the VR sessions in the form of log files to determine if incorrect answers were because of a knowledge deficit or because of the voice recognition system misinterpreting the learner.

Figure 2. Virtual reality participants and refresher course screen shots.

Grading Methodology

Each modality was graded against the same rubric (see Multimedia Appendix 1) using the same criteria. Fifteen seconds were allowed to either make a diagnosis or institute a management plan for each item of the algorithm. The only exception to this was the initiating and resuming of chest compressions, which were required to be within 10 seconds. Incorrect answers or answers given after 15 seconds (after which coaching was provided) were graded as incorrect. The scoring mechanism for the VR arm was programmed into the module and exported as a Microsoft Excel file and analyzed. Voice capture output by the natural language processing (NLP) was manually double-checked for every wrong answer to ensure that an incorrect score was not given because of the failure of the NLP to translate speech. The proctor for the HFS performed the grading for the simulations using the same rubric mentioned above. For nontechnical skills (NTSs), proctors in the simulation were given a behaviorally anchored rating scale (BARS) and a sheet containing examples and expected behaviors for each scoring domain. In the VR intervention, BARS outputs were determined by analyzing the voice outputs and placing them into categories along the scale. For example, the score in communication was determined by the percentage of the time that the participant used team member names when communicating. Scores for each domain were tallied and analyzed. Self-assessment scores and feedback on the comparator arms were obtained immediately after the debrief portion of the exercise through survey administration. Our primary outcome was set to be technical skills scores, as measured by the correct percentage of items within 15 seconds without coaching.

Proctor Assessment

At the end of sessions each day, proctors were asked to fill out a National Aeronautics and Space Administration task load index (NASA-TLX) form indicating their performance and experience throughout the day. The NASA-TLX is a validated instrument for measuring perceived workloads for performing tasks that are graded on a 20-point scale in six domains [8]. Scores were analyzed in each domain. Proctors were allowed to take breaks as needed, including a 45-min lunch break each day. The time required to complete each group was also notated.

Cost Analysis Methodology

Cost data were obtained from purchasing orders for equipment and were based on predicted salaries for personnel involved. Salaries were adjusted based on the minimum certifications and expertise needed to perform the task. For example, even though the ALCS instructor was a board-certified anesthesiologist, the calculated salary line was adjusted to be in line with ACLS instructors in our area (New York Metro Area). Certain items such as insurance and building costs were not included in the assessment.

Statistical Methods

Normally distributed variables are presented as mean (SD) with nonnormally distributed variables reported as median (IQR). Normality testing was performed via Shapiro Wilk testing and visual inspection of histograms. Appropriate statistical tests were performed based on normality and qualifying for assumptions. All tests were performed with SPSS version 24 (IBM).

Results

Participant Results

Of the 25 participants recruited, 23 were included for the final analysis. One participant erroneously went through the VR intervention multiple times in one sitting, and another participant's VR log file was lost during the study (see Figure 1). All study participants were aged between 25 and 35 years, and 68% (17/25) of them were male (see Multimedia Appendix 1). There was no difference in baseline comfort or experience with leading or participating in codes (see Table 1). It should be noted that after 1 year of training, no participant had experience in leading a code. Significantly more participants had prior simulation experience (25/25, 100%) compared with those with prior VR experience (9/25, 36%; *P*<.001, Fischer exact test). There was no difference between groups with regard to reality or enjoyment of the experience. Similarly, there was no difference in preference for using either modality to train, and the willingness to use either modality every 6 months was high: 100% (25/25) HFS vs 96% (24/25) VR; *P*=.72 (Fischer exact test; see Table 2). Significantly more participants rated the HFS debrief as providing better feedback (HFS: median 99.0, IQR 89.0-100.0 vs VR: median 79.0, IQR 71.0-88.0; *P*<.001), and there was a nonsignificant trend toward participants rating HFS as more useful in teaching ACLS skills (HFS: median 90.0, IQR 83.0-99.5 vs VR: median 83.0, IQR 80.0-90.5; *P*=.080). Our primary outcome, as measured by scores in technical domains, as measured in percentage correct without assistance, was significantly lower in the VR group than in the HFS group (HFS: median 72.7, IQR 60.0-78.2 vs VR: median 47.0, IQR 40.0-58.0; *P*<.001; Mann-Whitney U Test). Scores were not dependent on the first modality encountered (VR first: median 40.5, IQR 35.5-42.75 vs HFS first: median 38.0, IQR 32.0-44.5; *P*=.810). In nontechnical domains, scores in decision making and communication were no different; however, situational awareness scores were rated lower in the VR group (see Table 2). The overall accuracy of the voice recognition system was very good, with less than 2% of scoring modifications based on incorrect interpretations.

Table 1. Baseline information of prior experiences.

Variable	HFS[a] first (n=13)	VR[b] first (n=12)	*P* value
I feel comfortable running a code (out of 100), median (IQR)	12 (0-21.5)	18.5 (10.25-26.5)	.27
How many codes have you run, median (IQR)	0 (0-0)	0 (0-0)	.54
How many codes have you participated in, median (IQR)	6 (5-10)	9 (5-15)	.32
Have you participated in HFS? (Yes), n (%)	13 (100)	12 (100)	>.99
Have you used VR before? (Yes), n (%)	4 (31)	5 (42)	.57

[a] HFS: high-fidelity simulation.

[b] VR: virtual reality.

Table 2. Self-reported results and session scores.

Variable	High-fidelity simulation (n=25)	Virtual reality (n=25)	*P* value
How real was the experience, median (IQR)	62.0 (50.5-70.0)	50.0 (44.5-66.0)	.13
How useful was the experience in teaching you how to run a code, median (IQR)	90.0 (83.0-99.5)	83.0 (80.0-90.5)	.08
How useful was the feedback received, median (IQR)	99.0 (89.0-100.0)	79.0 (71.0-88.0)	<.001
I enjoyed the experience (Yes), n (%)	23 (92)	22 (88)	.63
I would like to use this as a way to recertify my ACLS[a] (Yes), n (%)	25 (100)	23 (92)	.14
Was this experience as valuable as your live training for Mega Code the last time you had to recertify? (Yes), n (%)	25 (100)	22 (88)	.25
I would do this once every 6 months to refresh my skills if it was NOT required but I was given time to do so (Yes), n (%)	25 (100)	24 (96)	.72
Scored domains[b]			
Total correct percentage technical domains, median (IQR)	72.7 (60.0-78.2)	47.0 (40.0-58.0)	<.001
Nontechnical domains, median (IQR)			
Situational awareness	6.0 (5.0-7.0)	1.0 (1.0-1.0)	<.001
Decision making	6.0 (4.0-6.0)	6.0 (4.0-6.0)	.52
Communication	5.0 (4.0-6.0)	4.0 (1.0-6.0)	.09

[a] ACLS: advanced cardiac life support.

[b] For scored domains, n=23 each for high-fidelity simulation and virtual reality.

Instructor Results

The instructor task load was significantly higher for the HFS proctors in every domain tested (see Table 3) in the NASA-TLX. On average, VR sessions were 21 (IQR 19-24) min shorter than the HFS sessions and were US $103.68 less expensive. Table 4 demonstrates the estimated cost difference depending on the number of learners and sessions for a variety of theoretical institutions. Including the time taken to logistically organize the participants to come to the simulation laboratory, each VR group required less than 1 day to complete. Each simulation group required 2 or 3 working days to complete the exercise.

Table 3. National Aeronautics and Space Administration task load index data.

Variable	High-fidelity simulation proctors (n=20), median (IQR)	Virtual reality proctors (n=5), median (IQR)	*P* value
Mental demand	12.5 (9.6-15.0)	2 (2-2.5)	<.001
Physical demand	10.0 (5.0-13.7)	3.0 (2.0-3.5)	<.001
Temporal demand	14 (10.2-16.0)	3.0 (2.0-4.0)	<.001
Performance impact	8.0 (6.2-12.0)	2.0 (1.0-3.0)	<.001
Effort	13.0 (12.2-15.0)	2.0 (2.0-3.0)	<.001
Frustration	14.0 (6.7-16.0)	2.0 (2.0-2.5)	<.001

Table 4. Time and cost analysis.

Variable	High-fidelity simulation group	Virtual reality group	Difference[a]	Percentage difference
Time per session (min), median (IQR)	42 (38-44)	20 (18-21)	21 (19-24)	50
Cost for single learner, single session (US $)[b]	193.00	89.32	103.68	54
Cost for 50 learners, single session (US $)[b]	9650.00	4466.15	5183.85	54
Cost of 1000 learners, single session (US $)[b]	193,000.00	89,322.92	103,677.08	54
Cost for single learner, four sessions (US $)[b]	772.00	132.29	639.71	83
Cost for 50 learners, four sessions (US $)[b]	38,600.00	6614.58	31,985.42	83
Cost for 1000 learners, four sessions (US $)[b]	772,000.00	132,291.67	639,708.33	83

[a] Median difference in time calculated via Hodges-Lehman median difference.

[b] Cost estimates are based on purchase orders for equipment and salaries for New York Metro Area and are in US $.

Discussion

Technical and Nontechnical Skills

The AHA scientific statement by Cheng et al [9] demonstrated that the current strategy for teaching and maintaining ACLS skills must change. It is clear that the frequency of refresher training is inadequate to maintain skills and that our current teaching modalities may have negative impacts on survival [9]. Our study supports previous studies demonstrating that learners believe that HFS is a great teaching modality for ACLS skills [10]. Despite this, HFS as a refresher modality is prohibitively expensive, time consuming, and places a large burden on instructors. This is where new technologies such as VR can be implemented. Our study is not the first to explore VR as a means of teaching ACLS skills, but it does provide new insight into the topic.

Studies by Khanal et al [5], Khanal and Kahol [11], Creutzfeldt et al [12,13], and Semeraro et al [6] have demonstrated the effectiveness of virtual environments to train ACLS skills; however, our study differs in the study subjects and methodology. The largest study was performed by Khanal et al [5] and included 148 participants. Their team demonstrated that VR could lead to enhanced performance in simulated scenarios graded by ACLS experts at varying levels of user feedback. The level of experience of the subjects was not disclosed, and the intervention was a team-based VR experience with peripheral equipment such as joysticks. Our study included PGY-2 residents and was a standalone VR experience. It did not require a team to play, nor did it require controls or joysticks that would enhance the utilization and uptake and make scaling easier. The second study by Khanal and Kahol [11] had 11 participants and was a mixed reality environment (VR and a haptic device). Their team demonstrated the effectiveness of the experience; however, its generalizability is limited by the need of a haptic device. The studies by Creutzfeldt et al [12,13] and Semeraro et al [6] again demonstrated proof of concept; however, their VR experiences required the use of peripheral devices or mannequins in conjunction with VR.

Our study is the first to compare a fully immersive, stand-alone, voice-controlled experience to HFS, and it has demonstrated important findings. Our primary outcome, technical scores for the algorithms, were lower in VR than those in HFS. The source of this difference is likely multifactorial. First, this may have been an artifact because of a lack of familiarity with the environment. All of our residents have experience and comfort with HFS; however, the minority had experience with VR. Furthermore, we did not design or mandate an orientation to the VR module. This unfamiliarity may have contributed to the lower scores. In addition, although both systems used the same rubric for scoring, it may be that in the VR, grading was more stringent. A human grader might interpret *almost* correct responses as correct. Further, human graders may be giving subtle feedback to learners within the grading timeframe. This could be demonstrated by body language cues, a change in voice pitch or timing, or through some other mechanism. Moreover, the inability of the VR system to recognize subjects' vocal responses could potentially have limited scoring when compared with HFS; however, on manual recheck, this impact was minimal, and as such, this was not likely contributory. VR also offers a level of granularity of assessment in real time that would be impossible for a human proctor to detect, especially when also driving the experience. This might result in more stringent grading, as mentioned above. For example, the VR system can note the difference between taking 30 vs 31 seconds to respond and grade accordingly, whereas a human proctor could not. We opted not to use a third party to grade the HFS group for this reason, as it would be impractical and not congruent with current practice and would have further inflated the cost of the HFS group. Further works are needed to elucidate the differences between human and computerized grading schema for technical domains.

In terms of baseline comparisons, it was interesting but not surprising that none of the participants had code leader experience after 1 year in practice. Second, although all participants in the study had experience with HFS, only 36% (9/25) of participants had experience with VR. There was no orientation to VR for the study, which may have put the VR experience at a disadvantage and may have partially explained the reason for the decrease in scoring, as discussed above. Satisfaction with both experiences was very high, but there was a clear advantage to HFS in the feedback domain. This is not surprising because the HFS group had a full formal debrief, whereas the VR group received only the feedback on items missed in a binary manner. The less effective debriefing could have an impact on knowledge retention. Our study was not designed to evaluate retention; however, it would be important to examine this in future studies. In this regard, VR lags behind HFS and more work needs to be done to enhance the feedback given to learners. Experienced debriefers can tailor the feedback type and technique to individual learners, whereas the VR experience can only deliver feedback one way in its current form. Although this is a clear limitation of VR in this module, there are ways in which this can be improved in the future with better software. Despite this limitation, participants believed the VR experience was as valuable as their live training for recertification. More importantly, the participants were very willing (24/25, 96%) to use VR as a refresher even if it was not required, despite the aforementioned disadvantages.

There were similarities in NTSs as well, which are an important component of acute care and have been shown to impact outcomes in other areas [14]. Our VR experience was able to grade participants in three nontechnical domains [15]. When compared with HFS scores, both decision making and communication scores were not different. This is one of the first indicators that we can use virtual environments to identify and stratify participants by NTSs. However, not every domain correlated, such as situational awareness. On the basis of our study design, we cannot know which of these assessments (HFS or VR) is most correct for this domain; similarly, we can only hypothesize why the situational awareness scores were much lower in the VR group than in the HFS group. It is our belief that this is most likely because of the mechanism by which the VR experience is graded on this domain and not that a VR-based experience is unable to grade a domain such as situational awareness.

Instructor Fatigue and Cost

Our analysis also demonstrated what is known anecdotally about trying to scale HFS to accommodate the need for frequent refreshers. As demonstrated, proctoring HFS is demanding on proctors, as evidenced by high NASA-TLX scores in the HFS group. It would be difficult to sustain a model of multiple refreshers in a year without causing burnout and fatigue of the staff. Proctors for the VR sessions experienced minimal fatigue. Furthermore, although all proctors were ACLS instructors, proctors for future VR sessions would only need training in using the VR system and not require a mastery of the content. The use of a VR proctor with minimal training would further decrease the cost associated with ACLS refresher courses.

Our time and cost analysis demonstrated that VR sessions can accomplish learning objectives in a shorter time than an HFS session. An advantage like this could be extremely important, as a VR session could be completed during a short coffee break, whereas an HFS session would take twice as long and require practitioners to give up their lunch breaks or come in on off hours to train. Finally, as the frequency of refresher training increases or the number of learners per session increases, the cost savings amplify. One proctor could supervise multiple VR sessions; however, the converse is not true for HFS.

Advantages of Each Modality

On the basis of our analysis, there are apparent advantages of HFS and VR: HFS provides very high-quality education but at high costs and low scale, and VR-based education currently may lag behind HFS but is more cost-effective and more easily scaled. It is the opinion of the authors that the interpretation of our findings is that each modality has its strengths and weakness, and neither of them is a panacea. As it currently stands, we would recommend that practitioners could utilize a VR medium as a means of a refresher or to screen for those who need more in-depth retraining. If during the VR module, it is discovered that the learner needs more in-depth remediation, one could then deploy the more expensive HFS alternative. In this manner, an institution can provide quality education at scale, while allowing for more targeted programs for those that need extra attention. Another option would be to develop some methodology for reviewing how a participant performed utilizing screen sharing technology; however, this would increase the cost, and it may be more effective simply to hold another session. Finally, as technology improves, it may be possible to provide HFS levels of directed feedback to learners and close this gap. Further development and research are needed before a conclusion should be drawn.

Limitations

Our study has several limitations. First, as the number of subjects is relatively low, the findings should be interpreted with caution. As stated, our study is exploratory in nature. Second, we utilized a specific VR experience (ACLS VR) by one company (Health Scholars). Other VR experiences for ACLS exist, but our findings may be specific to this experience, which may limit their generalizability. Our participants were all PGY-2 residents at a major academic center. Although we do not believe that our findings would not translate to other practitioners, we cannot say for certain. Finally, a major limitation of our work is that this VR experience does not include testing for hands-on skills such as the ability to perform chest compressions. Although this was not the goal of this experience, it would not be able to replace current training methodologies unless it was paired with a part-task trainer capable of grading these techniques.

Conclusions

Our study highlights some of the differences and similarities between VR and HFS for team leader refresher training. Scores were lower in the VR module, although the implications of this are unknown. NTSs were similar in some domains but different in others. The VR module was more cost-effective and was easier to proctor; however, HFS was better at delivering feedback to participants. Both modalities demonstrated high levels of satisfaction and a similar willingness for participants to use each modality. Further studies are needed to examine the utility of VR-based environments at scale.

Acknowledgments

The authors would like to acknowledge the department of anesthesiology at the Icahn School of Medicine at Mount Sinai and the Mount Sinai HELPS Center staff for providing proctors for our study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ACLS Scoring Matrix.
[PNG File , 150 KB - jmir_v22i3e17425_app1.png]

References

1. Ramachandran SK, Mhyre J, Kheterpal S, Christensen RE, Tallman K, Morris M, American Heart Association's Get With The Guidelines-Resuscitation Investigators. Predictors of survival from perioperative cardiopulmonary arrests: a retrospective analysis of 2,524 events from the Get With The Guidelines-Resuscitation registry. Anesthesiology 2013 Dec;119(6):1322-1339 [FREE Full text] [doi: 10.1097/ALN.0b013e318289bafe] [Medline: 23838723]
2. Wang P, Wu P, Wang J, Chi H, Wang X. A critical review of the use of virtual reality in construction engineering education and training. Int J Environ Res Public Health 2018 Jun 8;15(6):pii: E1204 [FREE Full text] [doi: 10.3390/ijerph15061204] [Medline: 29890627]
3. Pagé C, Bernier P, Trempe M. Using video simulations and virtual reality to improve decision-making skills in basketball. J Sports Sci 2019 Nov;37(21):2403-2410. [doi: 10.1080/02640414.2019.1638193] [Medline: 31280685]
4. Hawkins D. US Air Force. 2019 Aug 30. Pilot Training Next Graduates 14 in Learning Experiment's Second Iteration URL: https://www.af.mil/News/Article-Display/Article/1948666/pilot-training-next-graduates-14-in-learning-experiments-second-iteration/ [accessed 2019-09-30]
5. Khanal P, Vankipuram A, Ashby A, Vankipuram M, Gupta A, Drumm-Gurnee D, et al. Collaborative virtual reality based advanced cardiac life support training simulator using virtual reality principles. J Biomed Inform 2014 Oct;51:49-59 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.005] [Medline: 24732098]
6. Semeraro F, Frisoli A, Bergamasco M, Cerchiari EL. Virtual reality enhanced mannequin (VREM) that is well received by resuscitation experts. Resuscitation 2009 Apr;80(4):489-492. [doi: 10.1016/j.resuscitation.2008.12.016] [Medline: 19203823]
7. Jiang B, Ju H, Zhao Y, Yao L, Feng Y. Comparison of the efficacy and efficiency of the use of virtual reality simulation with high-fidelity mannequins for simulation-based training of fiberoptic bronchoscope manipulation. Simul Healthc 2018 Apr;13(2):83-87. [doi: 10.1097/SIH.0000000000000299] [Medline: 29621098]
8. Human Systems Integration Division @ NASA Ames. NASA TLX URL: https://humansystems.arc.nasa.gov/groups/TLX/tlxpaperpencil.php [accessed 2019-10-01]
9. 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 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 7;138(6):e82-122. [doi: 10.1161/CIR.0000000000000583] [Medline: 29930020]
10. Au K, Lam D, Garg N, Chau A, Dzwonek A, Walker B, et al. Improving skills retention after advanced structured resuscitation training: a systematic review of randomized controlled trials. Resuscitation 2019 May;138:284-296. [doi: 10.1016/j.resuscitation.2019.03.031] [Medline: 30928503]
11. Khanal P, Kahol K. Interactive haptic virtual collaborative training simulator to retain CPR skills. In: Elias D, Gabrielli S, Kahol K, editors. Ambient Media and Systems. Berlin, Heidelberg: Springer; 2011:70-77.
12. Creutzfeldt J, Hedman L, Medin C, Heinrichs WL, Felländer-Tsai L. Exploring virtual worlds for scenario-based repeated team training of cardiopulmonary resuscitation in medical students. J Med Internet Res 2010 Sep 3;12(3):e38 [FREE Full text] [doi: 10.2196/jmir.1426] [Medline: 20813717]
13. Creutzfeldt J, Hedman L, Medin C, Wallin CJ, Felländer-Tsai L. Effects of repeated CPR training in virtual worlds on medical students' performance. Stud Health Technol Inform 2008;132:89-94. [Medline: 18391263]
14. Pucher PH, Aggarwal R, Batrick N, Jenkins M, Darzi A. Nontechnical skills performance and care processes in the management of the acute trauma patient. Surgery 2014 May;155(5):902-909. [doi: 10.1016/j.surg.2013.12.029] [Medline: 24468038]
15. Watkins SC, Roberts DA, Boulet JR, McEvoy MD, Weinger MB. Evaluation of a simpler tool to assess nontechnical skills during simulated critical events. Simul Healthc 2017 Apr;12(2):69-75. [doi: 10.1097/SIH.0000000000000199] [Medline: 28704284]

Abbreviations

ACLS: advanced cardiac life support
AHA: American Heart Association
BARS: behaviorally anchored rating scale
CAE: Canadian Aviation Electronics
HELPS: Human Emulation Education and Evaluation Lab for Patient Safety and Professional Study
HFS: high-fidelity simulation
IRB: institutional review board
NASA-TLX: National Aeronautics and Space Administration task load index
NLP: natural language processing
NTS: nontechnical skill
PGY: postgraduate year
VR: virtual reality

Edited by G Eysenbach; submitted 11.12.19; peer-reviewed by I DeSouza, S Bond; comments to author 13.01.20; revised version received 18.01.20; accepted 27.01.20; published 12.03.20.

Please cite as:
Katz D, Shah R, Kim E, Park C, Shah A, Levine A, Burnett G
Utilization of a Voice-Based Virtual Reality Advanced Cardiac Life Support Team Leader Refresher: Prospective Observational Study
J Med Internet Res 2020;22(3):e17425
URL: http://www.jmir.org/2020/3/e17425/
doi:10.2196/17425
PMID:32163038

Original Paper

Understanding How Virtual Reality Can Support Mindfulness Practice: Mixed Methods Study

Elizabeth Seabrook[1], PhD; Ryan Kelly[2], PhD; Fiona Foley[1], BSc (Hons); Stephen Theiler[3], PhD; Neil Thomas[1], DClinPsy; Greg Wadley[2], MSc, PhD; Maja Nedeljkovic[1], PhD

[1]Centre for Mental Health, Swinburne University of Technology, Hawthorn, Melbourne, Australia
[2]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia
[3]Department of Psychological Sciences, Swinburne University of Technology, Melbourne, Australia

Corresponding Author:
Elizabeth Seabrook, PhD
Centre for Mental Health
Swinburne University of Technology
1 John Street
Hawthorn, Melbourne, 3122
Australia
Phone: 61 392145266
Email: eseabrook@swin.edu.au

Abstract

Background: Regular mindfulness practice has been demonstrated to be beneficial for mental health, but mindfulness can be challenging to adopt, with environmental and personal distractors often cited as challenges. Virtual reality (VR) may address these challenges by providing an immersive environment for practicing mindfulness and by supporting the user to orient attention to the present moment within a tailored virtual setting. However, there is currently a limited understanding of the ways in which VR can support or hinder mindfulness practice. Such an understanding is required to design effective VR apps while ensuring that VR-supported mindfulness is acceptable to end users.

Objective: This study aimed to explore how VR can support mindfulness practice and to understand user experience issues that may affect the acceptability and efficacy of VR mindfulness for users in the general population.

Methods: A sample of 37 participants from the general population trialed a VR mindfulness app in a controlled laboratory setting. The VR app presented users with an omnidirectional video of a peaceful forest environment with a guided mindfulness voiceover that was delivered by a male narrator. Scores on the State Mindfulness Scale, Simulator Sickness Questionnaire, and single-item measures of positive and negative emotion and arousal were measured pre- and post-VR for all participants. Qualitative feedback was collected through interviews with a subset of 19 participants. The interviews sought to understand the user experience of mindfulness practice in VR.

Results: State mindfulness (P<.001; Cohen d=1.80) and positive affect (P=.006; r=.45) significantly increased after using the VR mindfulness app. No notable changes in negative emotion, subjective arousal, or symptoms of simulator sickness were observed across the sample. Participants described the user experience as relaxing, calming, and peaceful. Participants suggested that the use of VR helped them to focus on the present moment by using visual and auditory elements of VR as attentional anchors. The sense of presence in the virtual environment (VE) was identified by participants as being helpful to practicing mindfulness. Interruptions to presence acted as distractors. Some uncomfortable experiences were discussed, primarily in relation to video fidelity and the weight of the VR headset, although these were infrequent and minor.

Conclusions: This study suggests that an appropriately designed VR app can support mindfulness practice by enhancing state mindfulness and inducing positive affect. VR may help address the challenges of practicing mindfulness by creating a sense of presence in a tailored VE; by allowing users to attend to visual and auditory anchors of their choice; and by reducing the scope of the content in users' mind-wandering. VR has the unique capability to combine guided mindfulness practice with tailored VEs that lend themselves to support individuals to focus attention on the present moment.

(J Med Internet Res 2020;22(3):e16106) doi:10.2196/16106

KEYWORDS
virtual reality; mindfulness; emotion; pilot projects; acceptability; evaluation

Introduction

Background

Originating in Buddhist traditions, mindfulness involves an individual focusing his or her attention on the present moment and approaching experiences with a nonjudgmental, nonreactive, and accepting attitude [1,2]. Mindfulness often involves bringing one's attention to an anchor, most commonly the breath, to facilitate the awareness of a moment-by-moment experience [3].

Mindfulness practice has been demonstrated to be beneficial for mental health by helping people strengthen attention flexibility and adopt an orientation toward experiences that reduces the reliance on automatic thoughts or maladaptive emotion regulation strategies [2,4-6]. However, mindfulness requires conscious effort and can be difficult to maintain, particularly for novice meditators who already expend greater cognitive resources to control their self-regulatory skills [7-9]. Anderson et al [10] recently summarized a range of experiential challenges that can arise during mindfulness practice. These include affective demands such as cognitive effort and frustration, task demands introduced by the physical environment (eg, noisy surroundings and people), and negative emotional or psychological outcomes, such as boredom, upsetting thoughts, and emotions. Adapting mindfulness practice to reduce the likelihood or impact of these challenges may be an important consideration for improving the success of mindfulness-based interventions.

In recent years there has been an interest in using digital technology to support mindfulness practice [11,12]. For example, Web-based interventions have been designed to improve mindfulness skills [13-15], and smartphone apps now deliver guided audio practices that support user engagement with mindfulness through habit formation [11,16,17]. Although these delivery platforms improve the accessibility of mindfulness-based interventions, research has shown that adherence with Web and mobile programs is often low [18], suggesting they may not be engaging over the long term [19]. It is possible that these technologies do not meaningfully assist users with overcoming some of the experiential challenges associated with mindfulness practice, which may contribute to disengagement or even adverse effects [20,21].

Virtual reality (VR) has recently been proposed as a medium to support mindfulness [22,23]. VR technologies may pragmatically address the challenges related to environmental distraction by providing an immersive, engaging, and controlled (ie, predictable) visual and auditory sandbox in which one could rehearse mindfulness skills [12,24,25], shifting attention away from the real-world environment. VR refers to the use of a headset with a display that projects an interactive, audiovisual 360-degree virtual environment (VE) to the user [26-28]. Although VEs can be presented via other mediums (eg, a computer screen), VR systems have greater immersive capacity, which aids in stimulating multiple senses and may create a sense of "presence," ie, a feeling of *being there* in a simulated environment [27-29]. Presence contributes to an "illusion of reality" in which the user behaves as if the environment were real even though it is computer generated [30,31].

A number of studies have evaluated VR-supported mindfulness practice in relation to its impact on mental health or state mindfulness [12,24,32-34]. In an evaluation with expert meditators from a nonclinical population, Navarro-Haro et al [12] demonstrated that a VR-supported mindfulness practice increased state mindfulness from pretest to posttest, reduced feelings of sadness and anxiety, and increased feelings of relaxation [12]. More recently, Chandrasiri et al [24] evaluated an omnidirectional video recording of a beach scene delivered in VR, which was matched with a breath-focused mindfulness practice in a novice, nonclinical adult sample. Consistent with the findings of Navarro-Haro et al [12], Chandrasiri et al [24] found that state mindfulness significantly increased after using VR. However, their work provides limited insight into how and why the features of VR supported mindfulness. It also does not explore the challenges that VR may introduce for individuals practicing mindfulness.

Although VR has the capacity to deliver support for mindfulness practice, it may also introduce its own unique challenges. For example, simulator sickness is a frequently reported negative side effect of VR, which produces symptoms similar to motion sickness [35,36]. Researchers have also speculated that headset discomfort may be a barrier to mindfulness in VR [24], but little is known about how specific elements of headset design (eg, comfort, weight, and degree of immersion) may impact the potential benefits of VR-supported mindfulness.

The existing evidence suggests that VR is a promising tool for supporting mindfulness. VR mindfulness interventions can lead to increased state mindfulness in both new and experienced meditator samples [12,22,24]. However, little is known about the end-user experience of using VR for mindfulness, in particular, user perceptions of *how* VR may aid or detract from a mindfulness practice. This is an important issue to explore when considering the range of potential mechanisms through which mindfulness may be supported in VR. A better understanding of mechanisms may inform design decisions (eg, environment types, the need for guidance, and target cohort) and subsequently impact the efficacy and potential use cases for such systems.

This Study

This study aimed to explore how VR can support mindfulness practice and to understand user experience issues that may affect the acceptability and efficacy of VR-supported mindfulness. To achieve these aims, we conducted a pilot study in which 37 participants from the general population used a VR mindfulness app in a controlled laboratory setting. We sought to:

- Quantitatively assess changes in state mindfulness, emotion, and simulator sickness following the use of the VR app.

- Qualitatively explore the user experience of mindfulness in VR, with a specific investigation of whether particular features are helpful or disruptive to mindful awareness in this context.

Methods

Ethics Approval

The Swinburne University of Technology Human Research Ethics Committee (SHR 2018/256) and the University of Melbourne Human Research Ethics Committee (ID# 1852613.2) approved all the procedures.

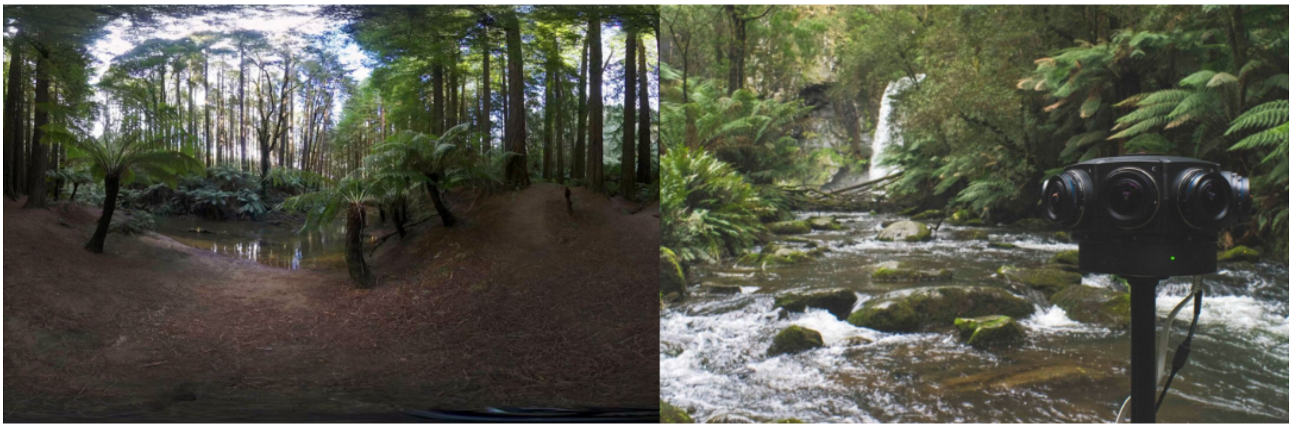
Description of the Virtual Reality Mindfulness App

This study used a VR mindfulness app that was designed for use in self-guided contexts. The app delivered a 15-min program of guided, focused-attention mindfulness within a VE created from an omnidirectional video footage of a forest. The VE included ambient audio (ie, sound originating from the forest) and a guided mindfulness voiceover. The app did not require the use of a hand controller, encouraging intuitive exploration between the user and the environment. The app was designed by the authors in collaboration with a commercial software company and it was developed in the Unity platform for use with the Oculus Go VR headset. During the design process, the app was refined through 3 focus groups (not reported here) in which 9 prospective users trialed the experience and gave feedback that helped address usability and content-related issues.

Omnidirectional Video and Audio

The VE comprised forest scenes captured in the Great Otway National Park in Australia. The footage was recorded at two different sites in the forest, presented in Figure 1. The first site was a clearing near a river, and the second site was at the river's edge. Both sites were filmed in 4K resolution using a Z Cam V1 Pro. The camera height was set at 1.3 meters, giving the impression that the user was experiencing the environment from a seated position. The footage did not include any visible people or animals, giving the impression that the user was alone in the environment. Ambient sounds of the natural environment were captured at the time of video recording, using omnidirectional and stereo microphones (Zoom H6 and Zoom H2n). This sound was overlayed onto the video footage during postprocessing.

Figure 1. Filming locations for the virtual reality mindfulness app. The image on the left shows the environment presented for the first half of the mindfulness practice. The image on the right shows the second environment as well as the camera used to capture the omnidirectional footage.



Guided Mindfulness Voiceover

The VR app incorporated a guided mindfulness voiceover tailored to the VE that was delivered by a male narrator. The narrator is an experienced counseling and clinical psychologist with expertise in mindfulness-based interventions. The voiceover delivered a focused-attention mindfulness practice that used invitational language to guide the user's attention to different parts of the VE (eg, "If you would like to...look at the rocks, you can explore their colour and texture") and different physical sensations from the body (eg, "bring your attention to the breath"). Guidance around thoughts and feelings was also provided (eg, "If you notice the mind has wandered...just gently bring your attention back to the moment"). The guided voiceover also allowed for periods of unguided practice (ie, without the voiceover), lasting up to 70 seconds.

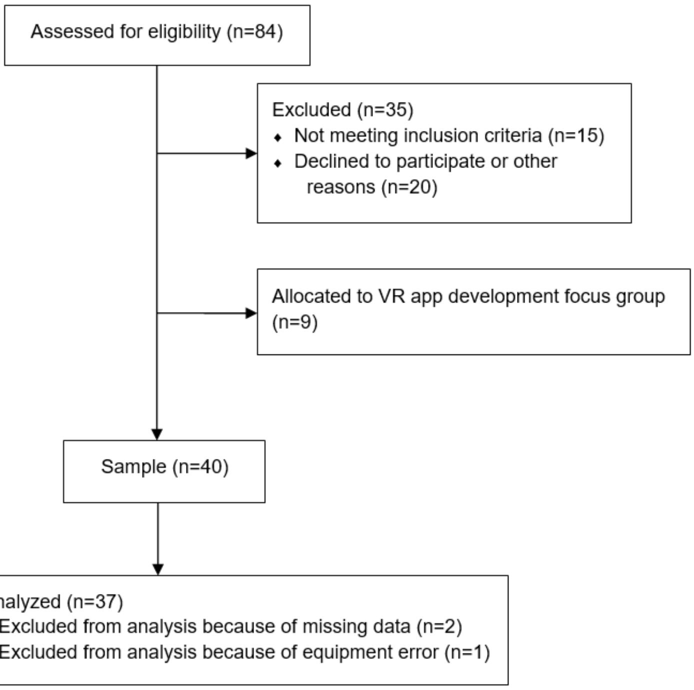
Participants

Participants were recruited from the general population using a Facebook advertisement, a promotion on the website Mental Health Online, and flyers distributed around the Swinburne University campus in Melbourne, Australia. Potential participants completed a Web-based screening survey that assessed their eligibility for inclusion. The inclusion criteria required participants to be over the age of 18 years, have normal or corrected-to-normal hearing and vision, have no history of photosensitive epilepsy or previous experience of severe simulator sickness, and not be currently taking psychotropic medication or experiencing serious mental illness (eg, schizophrenia or psychosis). Figure 2 illustrates the recruitment process.

A sample of 40 participants trialed the VR app. As shown in Figure 2, a total of 3 participants were excluded from quantitative and qualitative analyses as they did not complete the full procedure (1 participant encountered a technical error with the VR app, and 2 participants did not complete posttest measures). Of the remaining 37 participants, 13 were men and 24 were women, with an average age of 37.86 years (SD 14.56; n=1 missing). A total of 4 participants had no previous experience of practicing mindfulness, but they had heard of it before. A total of 33 participants had previous experience of practicing mindfulness. Of these, 17 participants had tried mindfulness one to five times, and 16 participants reported regular mindfulness practice at the following frequencies: monthly (n=7), weekly (n=6), and daily (n=3). A total of 23 participants had never used VR or had only experienced it once, whereas 12 participants had tried VR several times and 2 participants reported using VR regularly, either less than monthly (n=1) or weekly (n=1).

A total of 19 of the 37 participants opted to contribute qualitative feedback at the time of recruitment. No significant differences, between participants who provided qualitative feedback and those who did not, were revealed in age (U=115.5; P=.15) or gender (χ^2_1=.1; P=.82). Owing to violations of chi-square test assumptions (expected cell count <5), further demographic comparisons were not performed. A visual inspection of the level of education, previous experience with mindfulness, or previous experience with VR suggested a similarity between groups, although the interviewees included all participants (n=4) without previous mindfulness experience.

Figure 2. Flowchart illustrating participant recruitment to data analysis. VR: virtual reality.



Materials

Measures

Demographics and Baseline Characteristics

Questionnaires were used to collect each participant's gender, age, highest level of education, and previous experience with both VR and mindfulness—Response options: Not at all, I haven't heard of it; Not at all, but I have heard of it; I have tried it once; I have tried it a few times (eg, two to five times); and I do mindfulness/VR regularly. Those who reported regular use of either VR or mindfulness were prompted to complete an additional item addressing frequency of use or practice—Response options: Less than monthly; Monthly; Weekly; Daily; and More than once a day).

Trait mindfulness was measured using a short-form of the Five Facet Mindfulness Questionnaire (FFMQ-15) [37,38]. The FFMQ-15 comprises 15 items drawn from the 39-item original measure, rated on a 5-point scale from "1=Never or very rarely true" to "5=Very often or always true." A total score reflecting the tendency to be mindful ranges from 15 to 75. As a total scale, the FFMQ-15 demonstrated good reliability in this sample (Cronbach alpha=.83).

Participant baseline general mental health distress was assessed using the short-form Depression Anxiety and Stress Scale (DASS-21) [39]. In this sample, the DASS-21 showed good internal reliability (Cronbach alpha=.92).

Primary Outcomes

Change in state mindfulness was assessed using the 21-item State Mindfulness Scale [40], which was delivered before and after using the VR mindfulness app. Items are rated on a 5-point Likert scale from "0=Not at all" to "4=Very much" and address an individual's level of perceived mindfulness in relation to the individual's attention and meta-cognitive orientation across a recent period. The scale has a total score obtained by summing the item ratings (range 0-84). In this study, the State Mindfulness Scale had good internal reliability both before (Cronbach alpha=.95) and after (Cronbach alpha=.93) the use of the VR mindfulness app.

Emotion was measured dimensionally pre- and postapp use via 3 items—1 (Positive or pleasant), 2 (Negative or unpleasant), and 3 (Active or alert)—drawn from the circumplex model of emotion [41]. Participants were asked to consider how they were feeling “right now” and rated items on a 7-point Likert scale from “0=Not at all” to “6=Extremely.”

The Simulator Sickness Questionnaire (SSQ) [35] was used to detect changes in symptoms of simulator sickness. The SSQ comprises 16 items (eg, Eyestrain, Difficulty Focusing, and Fatigue), which are rated from “0=None” to “3=Severe,” and addresses the domains of nausea, oculomotor symptoms, and disorientation. It is important to note that there is considerable overlap between preexisting physiological and psychological symptoms (eg, anxiety symptoms) with the items on the SSQ [42]. For this reason, the symptoms of simulator sickness were measured pre and post use of the VR mindfulness app to determine a change in the symptoms relevant to the experience. In this sample, the internal reliability of the total score of the SSQ was poor both pre and post use of the VR mindfulness app (Cronbach alpha=.68 and Cronbach alpha=.51, respectively). As such, we considered the items as a checklist of symptoms rather than as a total scale, and we examined these individually as indicators of adverse responses.

Presence and General System Feedback

Presence was measured using a 22-item adapted version of the Presence Questionnaire 3.0 [29,43], which was shortened to reduce participant burden and remove items related to haptic senses and object manipulation, which were not features of the current system. As such, the total scale ranged between 22 and 154, with higher scores indicating a greater sense of presence in the VE. In this sample, the total scale demonstrated good internal reliability (Cronbach alpha=.84).

A total of 7 general system feedback items were also presented, relating to the perceived level of engagement, risk, quality, and future use (eg, “How engaging was the virtual reality mindfulness experience?”). These items were rated on a 5-point Likert scale, ranging from “0=Not at all” to “4=Extremely.”

Procedure

All testing took place in a soundproof room at Swinburne University of Technology. Following informed consent, participants completed baseline questionnaires for approximately 15 min. Participants were then invited to wear an Oculus Go headset, which had the VR app preinstalled. Participants were seated in a swivel chair, giving them the ability to rotate their body to view the omnidirectional forest environment. A verbal check of visual clarity was conducted, and participants were assisted with adjusting the headset to improve clarity if required. The researcher then used the Oculus Go hand controller to start the VR mindfulness app (duration of approximately 15 min) and left the room. At the end of the experience, the participants removed the headset and the researcher reentered the room, at which time participants completed follow-up measures addressing state mindfulness, emotion, symptoms of simulator sickness, presence, and general feedback items (taking approximately 15 min).

For the 19 participants who opted to provide qualitative feedback, a semistructured interview, lasting an average of 26 min (range 16-46 min), was conducted. The interview questions focused on eliciting the participants’ views about the VR experience, positive and negative features of the app, and features of VR and VE, that helped or hindered the mindfulness practice.

All participants were thanked for their time and reimbursed with an Aus $10 (US $6.60) gift voucher (Aus $30 [US $19.79] if they had also provided qualitative feedback).

Data Analysis

All quantitative data were processed and analyzed in IBM SPSS Statistics version 23.0. Descriptive statistics were used to describe the sample and calculate the ratings of general system feedback. Changes in mean scores from pre- to post-VR use on the State Mindfulness Scale [40] and emotion items were assessed using two-tailed repeated-measures *t* tests with an alpha level of .05. A Bonferroni-adjusted alpha level of .003 was applied to the analysis of the simulator sickness items on the SSQ [35]. Normality checks were conducted to confirm the appropriateness of a parametric model. Wilcoxon signed-rank tests were conducted where the assumptions were violated. Effect sizes were interpreted by following the guidelines from Cohen [44].

Qualitative data were collected by ES and 2 trained research assistants. The interviews were coded using a general inductive approach [45]. Both ES and RK conducted independent parallel coding by reading the transcripts separately and applying labels to the text through open coding. These codes were subsequently compared, discussed, and developed into a final set of 9 themes that characterized participants’ experiences [45].

Results

Quantitative Findings

Table 1 provides the mean scores and standard deviations of the primary outcome measures (state mindfulness, emotion, and arousal) both before and after using the VR mindfulness app. Trait mindfulness was normally distributed in the sample, ranging from 37 to 65 (mean 52.70, SD 7.50). On average, participants had low levels of general mental health distress, as measured by the DASS-21 (mean 9.19, SD 6.95; range 1-32).

Table 1. Descriptive statistics of primary outcomes pre- and post use of the virtual reality mindfulness app.

Variable	Pre-VR[a] app use		Post-VR app use	
	Mean (SD)	Range	Mean (SD)	Range
State mindfulness	38.76 (18.04)	11-77	64.32 (13.34)	30-84
Positive emotion	4.32 (0.88)	2-6	4.78 (0.79)	3-6
Negative emotion	0.76 (0.90)	0-3	0.49 (0.61)	0-2
Arousal	4.00 (0.91)	2-5	4.19 (0.97)	2-6
Presence[b]	N/A[c]	N/A	114.83 (14.06)	81-149

[a] VR: virtual reality.

[b] Presence was measured after use of the virtual reality mindfulness app. This variable had n=1 missing observation because of an incomplete response.

[c] N/A: not applicable.

Primary Outcomes

A two-tailed repeated-measures *t* test was used to evaluate the change in state mindfulness following the use of the VR app. There was a statistically significant increase in state mindfulness scores from pre-VR to post-VR use (t_{36}=–10.97; *P*<.001; 95% CI –30.30 to –20.84). The size of this effect was large, Cohen *d*=1.80.

Owing to non-normality, Wilcoxon signed-rank tests were conducted to examine the change in subjective positive and negative emotion as well as levels of arousal. There was a statistically significant increase in the median positive emotion rating from pre-VR (median=4.00; interquartile range [IQR] 4.00-5.00) to post-VR use (median=5.00; IQR 4.00-5.00; *Z*=–2.75; *P*=.006; *r*=.45). No significant changes were detected for negative emotion (pre-VR: median=1.00; IQR 0.00-1.00; post-VR: median=0.00; IQR 0.00-1.00; *Z*=–1.85; *P*=.06) or for arousal (pre-VR: median=4.00; IQR 3.00-5.00; post-VR: median=4.00; IQR 3.00-5.00; *Z*=–0.91; *P*=.36).

Table 1 also shows high average ratings of presence following the use of the VR mindfulness app, indicating that participants experienced a strong sense of *being there* in the forest environment.

Simulator Sickness

Mean scores of simulator sickness symptoms were low across all items, both before and after VR app use. All median scores were at or below the mild symptom awareness threshold (a score of 1 on the SSQ). Bonferroni-adjusted Wilcoxon signed-rank tests did not reveal significant differences in the SSQ items pre- to post-VR app use (all *P*>.003). At an individual level, no participant rated *severe* effects post-VR app use. Participants who reported symptoms in the *moderate* range post-VR app use included stomach awareness (n=2, change from *not at all* preintervention), fatigue (n=2, change from *not at all* and *mild* preintervention), general discomfort (n=1, change from *mild*), and difficulty concentrating (n=1, change from *mild*).

General Feedback Ratings of the Virtual Reality Mindfulness App

Overall, most participants reported that the VR mindfulness app was very engaging (22/37, 60%) or extremely engaging (10/37, 27%), that it had very intuitive interactions (19/37, 51%) or extremely intuitive interactions (13/37, 35%), and that it was very easy to use (16/37, 43%) or extremely easy to use (19/37, 51%). Most participants rated the quality of the visual and auditory experience as moderate (15/37, 41%) or very good (18/37, 49%). Participants also thought that the VR mindfulness app would benefit their mental health at a moderate level (11/37, 30%) to high level (14/37, 38%), with 19% (7/37) indicating that the VR mindfulness app would benefit their mental health extremely. Most participants indicated that they felt using the VR mindfulness app was not at all risky (35/37, 95%).

Qualitative Findings

A total of 9 themes were produced from the participants' qualitative feedback. These encapsulate descriptions of the general experience of mindfulness in VR, the main ways that VR helped or hindered mindfulness practice (presence and interruptions in presence, engagement, providing a scope for mind-wandering, directing and anchoring attention, and personalization), and perceptions of video fidelity and the headset.

Experiencing Mindfulness in Virtual Reality

Participants were positive about the mindfulness app, describing the experience it created as "comforting," "calming," "relaxing," and "peaceful." Expressions of positive emotion were common and often related to the feeling of being situated in a forest environment: "that specific stimulus is really comforting for me" [P20]. *Relaxing* was one of the most common descriptions: "I felt myself get really comfortable and kind of relaxed and engaged with the environment" [P30]. Participants commented on the tranquility of the VE: "I like sitting in the middle of a river or next to a creek bed or in the middle of the forest. It feels safe" [P7].

Another participant explained that a feeling of safety was related to being in an environment that "doesn't change" and that "doesn't feel like there's all these people there" [P6].

Although most participants found the experience straightforward, some described taking time to adjust to VR. For some, more time was needed between putting on the headset and starting the practice:

> *there needed to be just a little more space between that direction and then...getting started because I was*

> *still fiddling around with the mask and trying to feel comfortable.* [P9]

For others, there was some apprehension around what they expected to see and do, although this period of apprehension did not appear to impact their practice:

> *[I]t was just that adjusting, but it wasn't really negative, it was just sort of a little bit of anxiousness. Am I doing it right? It took a few minutes to get used to it, but the scene...drew me in and I just forgot about it.* [P2]

Presence in the Forest Environment

Participants frequently mentioned experiencing a sense of presence in the virtual forest, which facilitated their ability to be aware of the present moment:

> *...to be able to be in a place that helps your mind focus in the present and in the now, something that's pleasant to look at, something that is sensory...it was actually helpful.* [P9]
>
> *it was just purely I was there and that was it.* [P2]

Some participants linked their sense of presence to feeling that their real-world concerns were reduced:

> *I think because you think you're there...it is hard to think about what's going on back in reality.* [P27]

Presence was often discussed in terms of being "transported" to the forest environment, which allowed participants to take a step back from their thoughts and daily life experiences. A participant described how:

> *it was great because it really kind of took me away from my thoughts...and just put me in this different world.* [P30]

Participants also noted that the *virtuality* of the environment helped facilitate this:

> *If I were sitting in that same environment in reality I would be thinking...are there other people there... is the car there. But knowing that this environment was virtual, I was able to simply enjoy it.* [P9]

Interrupted and Variable Presence

Although presence was an important precursor to engaging with mindfulness in VR, participants' sense of presence appeared to vary dynamically while using the VR mindfulness app. A participant explicitly described how changes in their sense of presence acted as a distractor from the practice and sometimes caused brief disengagement from the experience:

> *parts of it felt like it was seamless. And then there's other parts where it felt like I could see that it was a projection. And then I was like oh that's really distracting because I'm looking at that as a projection.* [P6]

A participant described compensating for the loss of presence by actively redirecting attention within the VE, particularly to images that felt "more real" to them:

> *There was one occasion ... [where] there was something about the graininess or something about the image that made me go 'oh yeah I'm not actually at the river'... I just would then focus my attention on something that felt more real...so it wasn't a big deal.* [P42]

The disparity between perceived presence and physical sensation was something participants described becoming aware of across the experience. Although this was not an issue for most participants, 1 individual claimed that this disparity was particularly disengaging:

> *they're not real sensations they're just the video of sensations...when you're in that place and you look down you don't see your body...that's unnatural...and it's a bit shocking and it kind of pulls you out of it a bit.* [P39]

The real-world environment also contributed to interruptions in presence and may have acted as a distractor to the mindfulness practice. This was primarily related to comfort within the physical space:

> *the chair was uncomfortable...sometimes the environment makes me forget about it. But then I was moving or the sound of the chair legs...it was kind of a break of my experience.* [P18]
>
> *if I was in like maybe the comfort of my own home...I'd be able to focus on my breathing.* [P17]

Engagement

Participants described the environment as being "engaging," and this was discussed as a useful feature of the VE for practicing mindfulness:

> *I found the engagement in the images improved my mindfulness because I really did try to be present in the moment...it filled my mind.* [P42]

Presence was a prominent cause of being engaged:

> *because I thought 'oh this is a real place' I kind of was more involved in the little details of this place.* [P30]

A sense of being connected to the virtual experience also emerged from participants descriptions of being engaged:

> *I...felt really connected to it when I was watching really small details. ...the visual was really important for me.* [P41]

The experience of being engaged in the VE did not mean that it was "too grabby or like showbiz" [P24]. Rather, participants described the VE as having "not too much going on so it would distract you but enough to keep it interesting" [P3]. This appeared to support mindfulness practice by providing interesting stimuli in the environment, which helped to maintain attention in the present moment but without becoming boring or distracting:

> *I never felt like...I was getting restless or I needed to look at something else. I was quite happy where I was.* [P34]

Scope for Mind-Wandering

Part of the process when practicing mindfulness is the ability to become aware of mind-wandering and learning to direct attention back to the present moment [2]. Some participants described how their mind wandered in a conventional sense, involving thoughts related to their daily life or forthcoming events:

> *I went through a whole pile of other kind of recent events or things happening in my life.* [P3]

Other participants discussed how VR seemed to affect the content of their mind-wandering, such that this wandering became a response to stimuli in the VE:

> *[My mind] wandered within that world more than it wandered outside of it...because even though [the] mind wanders it still wanders within the parameters of what you're seeing.* [P27]
>
> *It's like wandering but wandering with focus...It was all very much focused but within the realms of the environment.* [P30]

Directing and Anchoring Attention

Similar to the use of an attentional anchor (eg, the breath) to maintain present-moment awareness in conventional mindfulness practices, participants described how they directed their attention to focus on “one thing” by anchoring to the elements of the VE: “Initially I'd have a look around and then I'd just focus on one particular bit I liked” [P42]. Different visual and auditory stimuli were described as helpful for anchoring to the present moment, with some participants describing how they focused solely on visual aspects of the VE to reduce distractions: “when I did find my mind-wandering...the visual side of it allowed me to...rein it in quicker” [P42]. Visual elements were cited as especially useful by participants who had previously struggled with mindfulness because of the difficulty of imagining a visual environment:

> *I've done meditation before and I just zone out to what they are saying...because your mind's working to picture something it then is working to daydream as well...Whereas, when it was just there in front of you, I think that it took a bit of pressure off of thinking, and you could be in the present.* [P24]

Two participants discussed how the omnidirectional video provided a prompt to “recenter” themselves:

> *My mind drifted less because I could just look at something then re-centre myself.* [P42]
>
> *I think being able to watch the water running when I was in that environment and then to be able to look up at the sky - that was really helpful in terms of a grounding experience.* [P7]

Variety in the visual environment was also perceived to be useful as it allowed participants to shift their focus:

> *You just notice different things because you have that space to kind of focus on that, or the ability to kind of go oh I'm bored of that rock, I'll go check out the tree over there.* [P26]

Other participants described how the environment’s sounds were more useful to them. For some, this occurred within a process of becoming aware of the full environment:

> *I love to hear those sounds so sometimes just bringing my attention back to the sounds, which then brought my attention back to the visuals and then I was in that entire environment again.* [P9]

Some even chose to close their eyes to focus on the sounds:

> *I actually shut my eyes a lot...I felt like I'd focused on myself and then focused on the environment and then I like to focus on the sounds because they kept going and that was really good for me.* [P6]

Regarding the guided mindfulness voiceover, 15 participants found this to be a useful feature of the VR app. They described how the narrator’s suggestions “provided ways to sort of dig deeper into the experience” [P27] and how they were able to return to the present moment during mind-wandering so they would not “drift into that space where you're just observing it just in awe” [P2]. However, 5 participants who found the voiceover useful also commented that it became a source of distraction. A total of 2 others described the voiceover as entirely distracting. One issue was that the frequency of suggestions did not always allow for the time to explore and attend to the VE:

> *it'd be like “Okay now look at this, now look at this”...I [had] just started looking at one thing, now I can't not focus on the other thing.* [P3]

Another was that some people did not need the voiceover because they had previous experience with mindfulness: “I do my own silent practice...the audio guide of meditation now I find distracting” [P34].

Interestingly, participants sometimes described becoming aware of what was *not* in the VE (eg, a sense of touch and smell congruent with a forest) and how the absence of information directed their attention during the practice. A participant used audio as a substitute for touch:

> *I wish I could touch the water just to feel if it's cold ... I guess that was the reason that I focused more on the audio.* [P18]

Another participant described missing sensations and how this was interwoven with the process of guiding attention back to the present moment:

> *there were moments where I was able to bring myself back to the actual moment...and I had this sense that I really wanted to...actually feel the breeze. I was missing that sensation...I actually was very aware of the fact that sensation...was missing...* [P9]

Participants also commented on the way in which features of the VR app combined to guide attention back to the present moment. By providing many sources of detail to which to attend, the app’s “multi-modal” nature was seen as a benefit, and participants discussed the complex and evolving integration of multiple sources of stimuli:

> *It brings the ambience of the environment and the mindfulness into your actual body as opposed to just*

> *your mind, so you're focusing not only on the VR experience but also what your body is doing and then how it relates to the VR experience as well, so it's all connected.* [P26]

This process of becoming aware of and incorporating the experience of multiple senses was also described in relation to understanding and shifting between the different sources of information presented by VR and felt by the body:

> *You can feel your feet on the ground, you can feel the back of the chair, you know you're not in this environment. And because you're being present in the moment you've kind of got this contrast of sensations...On the other hand you open your eyes and...you feel part of the environment again...perhaps when you close your eyes...you become more aware of the chair again....* [P20]

Personalization

Participants' responses on the use of attentional anchors highlighted the potential of VR to support users in personalizing their practice. Participants frequently described how they chose to attend to parts of the experience that they enjoyed. In this way, personal preferences for visual and auditory stimuli may have been directing attention and maintaining engagement with mindfulness practice:

> *I could acknowledge and recognise in myself the parts of the environment that I really enjoyed and I sort of stuck to looking at them and looking at all the detail in them.* [P27]

The freedom to focus on particular elements of the VE also appeared to support a sense of agency:

> *...I'd find moments where I was staring up at the sky...counting my breaths...It felt like it was an active choice...engaging with this in your own time and of your own volition.* [P7]

Agency was considered a positive aspect of the VR mindfulness app, and it incorporated discussion of emotion:

> *If there's an area I don't really feel like it's calming me or I don't really feel a connection to...I just like twirl around and move somewhere else. So I just felt like I was able to control my own experience...I could...control what I was feeling.* [P17]

However, participants who reported practicing mindfulness regularly stated that the features of VR sometimes competed for their attention:

> *I found it quite distracting from what I was used to. Because it was an additional sensory distraction rather than trying to find a single point to focus. So it was a little confusing for a lot of it.* [P34]

This preference for a *single point to focus* was reinforced by another participant who valued having a "simple focus of attention...something to crystallise my attention around...[to] forget about the periphery" [P20]. These comments may suggest the need for variable stimulus levels or options within a VE, particularly when considering a user's previous mindfulness experience.

Perceptions of the Video Fidelity

Visual quality was an important feature to participants, which impacted their ability to focus on the mindfulness practice. Although our app used professional-quality video, 2 participants noted discomfort created by bright or blurry images:

> *like when you look into the sun and how it burns your eyes it's kind of like that kind of type of effect.* [P3]
>
> *The one thing that I found that kind of like impeded the experience was that when you try to look closely at something, the resolution of the display doesn't really allow you.* [P40]

A participant described how sections of the environment, which appeared to be of lower quality (eg, as they were further away or blurry), were distracting and how they actively redirected attention to parts of the environment with higher fidelity:

> *I kind of just stopped paying as much attention [to] things which were distracting...I feel like [things that] were kind of higher quality or less far away...wasn't as much of an issue.* [P3]

Perceptions of the Headset

In reflecting on the experience, participants indicated that the distractions introduced by the headset were both infrequent and minor. A majority of participants reported that it was "not really uncomfortable" [P34] or that "You kind of stop paying attention to it" [P3] when engaging with the VE. This may be because our app was delivered using a commodity headset (Oculus Go) that has been designed for user comfort.

The participants who did mention distraction from the headset noted issues of physical discomfort and disruption to immersion. Regarding physical discomfort, 5 participants described the headset as feeling heavy or tight on the head. A total of 2 participants described becoming aware of this sensation during the mindfulness practice. The first participant described how the awareness of the headset introduced a challenge for engaging with the VE:

> *it wasn't a case of trying to pull myself back to being mindful because of my thoughts, it was pulling myself back...into the virtual world away from the thought of oh this thing is heavy on my face.* [P9]

The second participant mentioned how the tension between being in the VE and "attempting to ignore the headset...puts a bit of a load on your mental state" [P40]. Other participants seemed to actively incorporate their awareness of the headset as a focus for their attention: "I was sort of focusing on [the headset] sometimes instead of my breathing..." [P14].

Regarding disruption to immersion, participants commented on the way the headset sat on the head and how this allowed light from the external environment to seep in to the bridge of the nose. This was described as a disengaging experience, making it "really hard...to concentrate" [P18] for one participant. For another, the ability to see his or her (real) legs when looking down was "a bit off-putting...[and] distracting" [P32]. They discussed a set of pragmatic actions to reduce the impact of the headset, which included “ma[king] a conscious effort to look left, look right, look up, and look straight ahead but not really right down” [P32].

Discussion

Principal Findings

This study aimed to explore how VR can support mindfulness practice and to understand user experience issues that may affect the acceptability and efficacy of VR mindfulness in the general population. To achieve these aims, we conducted a pilot study in which 37 participants used a VR app in a controlled laboratory setting.

Our first main finding was a large and statistically significant increase in state mindfulness, indicating that the VR app we created was successful in supporting mindful awareness in our sample. This finding is consistent with previous studies [12,24], and this provides further empirical support for the ability of an appropriately designed VR app to enable mindfulness practice. Qualitative feedback from participants suggested that the features of the VR app, including the VE and guided voiceover, played an active role in bringing their attention to the present moment. Although this points to the utility of VR for supporting mindful awareness, it is important to note that VR provides users the opportunity to engage in two “present moments”—one that is simulated and the other in the real world. Our participants commented that present-moment awareness often shifted between the two worlds, in line with guidance from the mindfulness audio. For some, this was experienced as a distraction, and for others, these dual present-moment realities were actively integrated. This may help to create a mindful state by shifting attention to the present-moment experience both within and outside of the VE, and future research involving comparisons with other groups can help to further unpack the unique role of VR in contributing to changes in state mindfulness [24].

Our second main finding was a statistically significant increase in positive emotion following the use of the VR mindfulness app. No change was indicated for negative emotion, likely because of floor effects. Similarly, there was no change in arousal. However, the participants qualitatively described a sense of “relaxation.” Although mindfulness practice does not seek to engender a positive mood state, the ability of a VR app to facilitate positive emotion may usefully contribute to mindfulness practice by reducing experiential challenges related to negative emotion [46]. This may be an important consideration for novice meditators or for groups of meditators who experience complex or negative emotions [20], and this may further enhance the acceptability of VR for mindfulness practice. However, it is doubtful that positive affect arises simply from the use of VR, and therefore, the content in the VE presented to the user is likely to play an important role. In our VR app, positive affect may have arisen from the simulated forest environment, which participants found to be peaceful and enjoyable. Natural spaces have been shown to provide individuals with well-being benefits [47], and this effect has been observed in the use of VR and other nonimmersive platforms (ie, computer screens) [48,49]. Future work should explore the impact of other kinds of neutral and positive VEs to explore the role of positive affect in supporting mindfulness practice in VR.

Our third main finding was that the VR mindfulness app did not generate any severe adverse experiences. Simulator sickness has been noted as a potential risk factor for VR in general [36], but our quantitative and qualitative data indicated only minor symptom increases for some participants. These increases were related to discomfort, concentration, and fatigue, and these may have been influenced by the image quality (particularly blurriness) or by the intensity of light in some parts of the VE. However, these are modifiable aspects of the experience; thus, these can be alleviated with the careful refinement of a VR app. In addition, we found that only limited discomfort was introduced by the Oculus Go. This indicates that the weight imposed by a commodity VR headset does not act as a substantial barrier to engagement, provided that this weight is minor. These practical insights are important as they emphasize the potential for a well-designed VR app to be a safe, acceptable, and tolerable approach for supporting mindfulness practice; they are also important for highlighting the areas for consideration (ie, image quality and brightness) in risk minimization. It may be especially important to consider these issues when designing VR apps for use with clinical populations.

How Virtual Reality Can Support Mindfulness Practice: Presence, Attention, and Agency

Our qualitative findings provide insight into the mechanisms that may have contributed to an increase in state mindfulness among our sample and into the potential affordances of VR to mindfulness practice. The key features mentioned by participants were the sense of being present in a calm environment that provided a bounded scope for attention, availability of structured guidance along with the freedom to explore, and ability to utilize multiple attentional anchors within the VE and in line with personal preferences.

The sense of presence emerged as a perceived support to mindfulness practice in VR, and its disruption became a temporary distraction. Our participants discussed how being in the VE allowed them to step back from everyday concerns and created a bounded scope around the content of their mind-wandering. Being engaged in the VE was beneficial in that it maintained interest across the practice, and the VE itself was neither overwhelming nor burdensome on concentration for most participants—a challenge often reported with conventional mindfulness practice [10]. Participants described a process of constructing the realness or believability of their experience by attending to parts of the VE that were appealing to them. Even when the sense of presence was interrupted (eg, by real-world intrusions), participants described directing their attention back to the VE and choosing to become involved in the simulation again. It could be that constructing a sense of presence, a process intrinsic to immersive VR, is a useful feature for training present-moment attention by actively involving attention shifting and facilitating engagement in the moment.

The availability of the multiple types of stimuli to act as anchors for present-moment attention was also identified as beneficial by our participants. In addition to the contribution that anchors made to support attention in the present moment, they provided users with an opportunity to personalize their experience by directing attention to their preferred objects of interest, which may have consequently contributed to a feeling of agency. As discussed by Anderson and Farb [3], individuals hold personal preferences for attentional anchors. Catering to preferences or motivations to engage with an attentional anchor may thus improve engagement with and adherence to mindfulness practice [3]. Peters et al [50] argue that designing (digital) interventions with the intention of satisfying basic psychological needs (self-determination: autonomy, competence, and relatedness) [51] at the point of use (ie, the interface) may greatly improve motivation, engagement, and general well-being. Although conventional mindfulness practices may support feelings of agency as an outcome of practice over time and as experience increases, the freedom of movement of attention within VR may uniquely contribute to a sense of agency within a practice (ie, at the point of use). This feature may motivate sustained use [50].

Kaplan’s [52] Attention Restoration Theory (ART) provides a useful framework for understanding how the features of VR, which were discussed by our participants, may be involved in facilitating present-moment awareness and contributing to the positive emotion. ART is a complementary viewpoint to mindfulness, which suggests that directed attention fatigue impacts mental health (eg, stress) and that person-environment interactions contribute to the recovery of attention resources, which may have mental health benefits. Environments that are considered restorative in ART are hypothesized to contain properties that involve (1) “being away” from an individual’s everyday environment (in our data—transportation), (2) containing features that hold attention with little effort via fascination (eg, nature), (3) containing features that maintain engagement and have coherence and scope (in our data—multiple types of stimuli to act as attention anchors), and (4) containing elements that support “what one wants or is inclined to do” (in our data—personalization) [52]. These properties were reflected in participant comments, and taken together, these suggest that the person-VE interaction may support a restorative experience and augment the mindfulness practice.

Considerations for Supporting Mindfulness in Virtual Reality

From the discussion of the mechanisms that supported mindfulness practice in VR, several design considerations emerge, which can inform future VR-supported mindfulness practices. First, an appropriately designed environment that supports the active construction of a sense of presence may be beneficial, both in implicitly guiding attention to the VE and in supporting the properties of a restorative environment. Consistent with ART [52], participants described the environment as having an appropriate degree of engagingness (not too interesting but interesting enough), which aided in their ability to maintain present-moment awareness. Where this consideration may need to be modified, however, is for more experienced meditators, who perceived the VE to be confusing or overstimulating due to the variety of anchors available. Second, the capacity of a VE to support personalization through providing a choice of attentional anchors both supports engagement and mitigates the need for extensive customization of a VR experience beyond an appropriate choice of VE. In the VE, users are likely to gravitate to the elements that best support mindful awareness and their sense of presence.

Design considerations also emerged from our participants’ description of distracting and disengaging experiences. Demands introduced by the physical environment have been discussed in previous research as a prominent challenge in conventional meditation practice [10] (although this should be balanced with an understanding that a part of mindfulness practice is concerned with noticing when the mind is distracted and bringing attention back to the present). In our app, distractions in the physical environment were not necessarily removed by the use of VR, as participants discussed experiences of feeling disengaged when they became aware of real-world stimuli. Despite this, these experiences were not so distracting as to completely disrupt the mindfulness practice, indicating that the system could help reduce potentially disruptive environmental distractors if it were used outside the lab. We also noted that distractions were experienced in the VE (eg, the environment being overwhelming and footage quality requiring greater concentration). This suggests the potential to consider VE design from the perspective of generating environments that are variably distracting in a graded approach for training mindful awareness.

Strengths and Limitations of This Study

A strength of this study was the mixed methods approach to examining the experience of a VR mindfulness app. This approach provided the opportunity to compare the quantitative measures of efficacy with the findings of previous studies utilizing different VR designs [12,24], while also qualitatively exploring the practice of mindfulness in VR and asking, from a user’s perspective, what helps and what hinders.

Our measures of efficacy were limited to state-based change in mindfulness and emotion following a single session of VR use in a controlled lab setting. As such, we cannot generalize the changes in mindfulness and emotion to repeated use contexts (eg, in the real world) or to long-term outcomes relevant to mental health (eg, changes in trait mindfulness and emotion regulation). More research is required to understand the patterns of engagement and mental health outcomes that arise from ongoing use of a VR mindfulness app. Our work provides the grounding for these lines of study.

Furthermore, without a comparison or control group, it is unclear whether the changes we observed in state mindfulness and emotion are primarily driven by contributions from the VR mindfulness app. Present-moment awareness may have increased in response to phenomena outside of the app. Although we argue that VR was integral to shaping participants’ experience, comparing virtual and real-world mindfulness practice may help to further understand the unique contributions a VR system can have for the user. Finally, our capacity to examine a range of moderators (eg, level of previous experience with mindfulness) was limited by sample size. Future work may want to consider how individual characteristics impact the responses to VR mindfulness practice in an effort to identify the most beneficial use cases.

Conclusions

Mindfulness practice can benefit mental health by enabling people to adopt an open, accepting, and nonjudgmental attitude toward present-moment experience. Through the use of a VR app, this study explored how VR can support mindfulness practice, and it investigated user experience issues that may help or hinder the ability to practice mindfulness in VR.

We found that VR can support statistically significant increases in state mindfulness and positive emotion in participants from the general population without fostering negative emotions or heightening arousal. The VR app we studied was well received and did not induce any strong symptoms of simulator sickness. These findings suggest that VR can foster mindfulness in a way that is safe and enjoyable, evidencing the acceptability and efficacy of VR as a platform for mindfulness.

We also found that the sense of presence in the VE and the freedom to explore were key contributors to participants' experience of mindfulness. The VE allowed participants to select anchors of their choice which enabled them to align with or freely deviate from the structured audio guidance provided in the app. The app also played a role in restricting the scope of participants' mind-wandering, enabling them to gently return to the present moment within the virtual forest setting. These features speak to the potential of VR to minimize the potential affective and task demands that are often experienced as challenges to mindfulness, by allowing users to personalize their practice in response to the content of the VE. However, our study revealed that video fidelity, level of guidance, and breaks in presence were sometimes perceived to be disengaging. These issues should be accounted for in future apps that are designed to support VR mindfulness.

Our investigation suggests that VR is able to combine a guided mindfulness practice with exposure to a relaxing, restorative environment. These complementary features may be a unique affordance of a well-designed VR app. Further research is required to compare the utility of this approach with conventional mindfulness practice.

Acknowledgments

The authors wish to acknowledge the contributions of Dr Craig Hassed in the early discussion of applying VR to mindfulness practice. The authors also wish to thank Nastaran Doroud and Lauren Rossi for their contributions to data collection, as well as Frank Vetere and Yaoxi Shi for their contributions to the analysis of the focus group data (not reported here). The authors also acknowledge the contributions of PHORIA in the development of the VR mindfulness app. Finally, the authors acknowledge the constructive feedback of the peer reviewers on this paper and thank reviewer K for providing a useful description of VR.

Authors' Contributions

This work was conceived by ES. GW, ES, FF, and RK led the focus groups and contributed to the iterative design of the VR app. ES and 2 research assistants conducted the second round of user interviews during the evaluation phase. ES and RK conducted the analysis of qualitative and quantitative data. ES and RK drafted the manuscript, and all authors provided input to the final version.

Conflicts of Interest

None declared.

References

1. Bishop SR, Lau M, Shapiro S, Carlson L, Anderson ND, Carmody J, et al. Mindfulness: a proposed operational definition. Clin Psychol 2004 Sep;11(3):230-241. [doi: 10.1093/clipsy/bph077]
2. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York: Delta Trade Paperbacks; 1990.
3. Anderson T, Farb NA. Personalising practice using preferences for meditation anchor modality. Front Psychol 2018;9:2521 [FREE Full text] [doi: 10.3389/fpsyg.2018.02521] [Medline: 30618947]
4. Blanck P, Perleth S, Heidenreich T, Kröger P, Ditzen B, Bents H, et al. Effects of mindfulness exercises as stand-alone intervention on symptoms of anxiety and depression: Systematic review and meta-analysis. Behav Res Ther 2018 Mar;102:25-35. [doi: 10.1016/j.brat.2017.12.002] [Medline: 29291584]
5. Cavanagh K, Strauss C, Forder L, Jones F. Can mindfulness and acceptance be learnt by self-help? A systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. Clin Psychol Rev 2014 Mar;34(2):118-129. [doi: 10.1016/j.cpr.2014.01.001] [Medline: 24487343]
6. Aldao A, Nolen-Hoeksema S, Schweizer S. Emotion-regulation strategies across psychopathology: a meta-analytic review. Clin Psychol Rev 2010 Mar;30(2):217-237. [doi: 10.1016/j.cpr.2009.11.004] [Medline: 20015584]
7. Lutz A, Slagter HA, Dunne JD, Davidson RJ. Attention regulation and monitoring in meditation. Trends Cogn Sci 2008 Apr;12(4):163-169 [FREE Full text] [doi: 10.1016/j.tics.2008.01.005] [Medline: 18329323]
8. Lymeus F, Lundgren T, Hartig T. Attentional effort of beginning mindfulness training is offset with practice directed toward images of natural scenery. Environ Behav 2017;49(5):536-559. [doi: 10.1177/0013916516657390]

9. Lindsay EK, Creswell JD. Mechanisms of mindfulness training: Monitor and Acceptance Theory (MAT). Clin Psychol Rev 2017 Feb;51:48-59 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.011] [Medline: 27835764]
10. Anderson T, Suresh M, Farb NA. Meditation benefits and drawbacks: empirical codebook and implications for teaching. J Cogn Enhanc 2019;3(2):207-220. [doi: 10.1007/s41465-018-00119-y]
11. Little D. UCLIC - UCL Interaction Centre.: University College London; 2016. Supporting or Inhibiting Motivations to Meditate? A Semi-Structured Interview Study of the User Experience of Mobile Mindfulness Apps URL: https://uclic.ucl.ac.uk/content/2-study/4-current-taught-course/1-distinction-projects/2-16/little_david.pdf [accessed 2020-01-14]
12. Navarro-Haro MV, López-Del-Hoyo Y, Campos D, Linehan MM, Hoffman HG, García-Palacios A, et al. Meditation experts try Virtual Reality Mindfulness: a pilot study evaluation of the feasibility and acceptability of Virtual Reality to facilitate mindfulness practice in people attending a Mindfulness conference. PLoS One 2017;12(11):e0187777 [FREE Full text] [doi: 10.1371/journal.pone.0187777] [Medline: 29166665]
13. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. JMIR Ment Health 2018 Sep 25;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]
14. Turner T, Hingle M. Evaluation of a mindfulness-based mobile app aimed at promoting awareness of weight-related behaviors in adolescents: a pilot study. JMIR Res Protoc 2017 Apr 26;6(4):e67 [FREE Full text] [doi: 10.2196/resprot.6695] [Medline: 28446423]
15. Wahbeh H, Oken BS. Internet mindfulness meditation intervention for the general public: pilot randomized controlled trial. JMIR Ment Health 2016 Aug 8;3(3):e37 [FREE Full text] [doi: 10.2196/mental.5900] [Medline: 27502759]
16. Roquet CD, Sas C. Evaluating Mindfulness Meditation Apps. In: Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. New York: ACM; 2018 Apr Presented at: CHI EA'18; April 21 - 26, 2018; Montreal, Canada p. 1-6. [doi: 10.1145/3170427.3188616]
17. Collins E, Cox A, Wilcock C, Sethu-Jones G. Digital games and mindfulness apps: comparison of effects on post work recovery. JMIR Ment Health 2019 Jul 18;6(7):e12853 [FREE Full text] [doi: 10.2196/12853] [Medline: 31322125]
18. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303 [FREE Full text] [doi: 10.2196/jmir.7487] [Medline: 28860106]
19. Kvillemo P, Brandberg Y, Bränström R. Feasibility and outcomes of an internet-based mindfulness training program: a pilot randomized controlled trial. JMIR Ment Health 2016 Jul 22;3(3):e33 [FREE Full text] [doi: 10.2196/mental.5457] [Medline: 27450466]
20. Baer R, Crane C, Miller E, Kuyken W. Doing no harm in mindfulness-based programs: conceptual issues and empirical findings. Clin Psychol Rev 2019 Jul;71:101-114 [FREE Full text] [doi: 10.1016/j.cpr.2019.01.001] [Medline: 30638824]
21. Banerjee M, Cavanagh K, Strauss C. A qualitative study with healthcare staff exploring the facilitators and barriers to engaging in a self-help mindfulness-based intervention. Mindfulness 2017;8(6):1653-1664 [FREE Full text] [doi: 10.1007/s12671-017-0740-z] [Medline: 29201248]
22. Damen KH, van der Spek ED. Virtual Reality as e-Mental Health to Support Starting with Mindfulness-Based Cognitive Therapy. In: Clua E, Roque L, Lugmayr A, Tuomi P, editors. Entertainment Computing – ICEC 2018. Cham: Springer; 2018:241-247.
23. Bruggeman KJ, Wurster SW. The Hiatus System: Virtual Healing Spaces: Low Dose Mindfulness Based Stress Reduction Virtual Reality Application. In: Proceedings of the 2018 Special Interest Group on Computer Graphics and Interactive Techniques Conference. New York, USA: ACM Press; 2018 Presented at: SIGGRAPH'18; August 12 - 16, 2018; British Columbia, Vancouver, Canada p. 1-2. [doi: 10.1145/3213779.3213785]
24. Chandrasiri A, Collett J, Fassbender E, de Foe A. A virtual reality approach to mindfulness skills training. Virtual Real 2019 Feb 18;24:143-149. [doi: 10.1007/s10055-019-00380-2]
25. Cikajlo I, Staba UC, Vrhovac S, Larkin F, Roddy M. A cloud-based virtual reality app for a novel telemindfulness service: rationale, design and feasibility evaluation. JMIR Res Protoc 2017 Jun 5;6(6):e108 [FREE Full text] [doi: 10.2196/resprot.6849] [Medline: 28583904]
26. Steuer J. Defining virtual reality: dimensions determining telepresence. J Commun 1992;42(4):73-93. [doi: 10.1111/j.1460-2466.1992.tb00812.x]
27. Diemer J, Alpers GW, Peperkorn HM, Shiban Y, Mühlberger A. The impact of perception and presence on emotional reactions: a review of research in virtual reality. Front Psychol 2015;6:26 [FREE Full text] [doi: 10.3389/fpsyg.2015.00026] [Medline: 25688218]
28. Slater M. Measuring presence: a response to The Witmer and Singer Presence Questionnaire. Presence 1999;8(5):560-565. [doi: 10.1162/105474699566477]
29. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence 1998;7(3):225-240. [doi: 10.1162/105474698565686]
30. Slater M. University College London. 2003. A note on presence terminology URL: http://s3.amazonaws.com/publicationslist.org/data/melslater/ref-201/a%20note%20on%20presence%20terminology.pdf [accessed 2020-03-04]

31. Slater M. Grand challenges in virtual environments. Front Robot AI 2014;1:1 [FREE Full text] [doi: 10.3389/frobt.2014.00003]
32. Flores A, Linehan MM, Todd SR, Hoffman HG. The use of virtual reality to facilitate mindfulness skills training in dialectical behavioral therapy for spinal cord injury: a case study. Front Psychol 2018;9:531 [FREE Full text] [doi: 10.3389/fpsyg.2018.00531] [Medline: 29740365]
33. Navarro-Haro MV, Modrego-Alarcón M, Hoffman HG, López-Montoyo A, Navarro-Gil M, Montero-Marin J, et al. Evaluation of a mindfulness-based intervention with and without virtual reality dialectical behavior therapy mindfulness skills training for the treatment of generalized anxiety disorder in primary care: a pilot study. Front Psychol 2019;10:55 [FREE Full text] [doi: 10.3389/fpsyg.2019.00055] [Medline: 30745888]
34. Gomez J, Hoffman HG, Bistricky SL, Gonzalez M, Rosenberg L, Sampaio M, et al. The use of virtual reality facilitates dialectical behavior 'therapy observing sounds and visuals' mindfulness skills training exercises for a Latino patient with severe burns: a case study. Front Psychol 2017;8:1611 [FREE Full text] [doi: 10.3389/fpsyg.2017.01611] [Medline: 28993747]
35. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
36. Davis S, Nesbitt K, Nalivaiko E. A Systematic Review of Cybersickness. In: Proceedings of the 2014 Conference on Interactive Entertainment.: ACM Press; 2014 Presented at: IE'14; December 2 - 3, 2014; NSW, Newcastle, Australia p. 1-9. [doi: 10.1145/2677758.2677780]
37. Gu J, Strauss C, Crane C, Barnhofer T, Karl A, Cavanagh K, et al. Examining the factor structure of the 39-item and 15-item versions of the Five Facet Mindfulness Questionnaire before and after mindfulness-based cognitive therapy for people with recurrent depression. Psychol Assess 2016 Jul;28(7):791-802 [FREE Full text] [doi: 10.1037/pas0000263] [Medline: 27078186]
38. Baer RA, Carmody J, Hunsinger M. Weekly change in mindfulness and perceived stress in a mindfulness-based stress reduction program. J Clin Psychol 2012 Jul;68(7):755-765. [doi: 10.1002/jclp.21865] [Medline: 22623334]
39. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
40. Tanay G, Bernstein A. State Mindfulness Scale (SMS): development and initial validation. Psychol Assess 2013 Dec;25(4):1286-1299. [doi: 10.1037/a0034044] [Medline: 24059475]
41. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
42. Pot-Kolder R, Veling W, Counotte J, van der Gaag M. Anxiety partially mediates cybersickness symptoms in immersive virtual reality environments. Cyberpsychol Behav Soc Netw 2018 Mar;21(3):187-193. [doi: 10.1089/cyber.2017.0082] [Medline: 29356575]
43. Witmer BG, Jerome CJ, Singer MJ. The factor structure of the presence questionnaire. Presence 2005;14(3):298-312. [doi: 10.1162/105474605323384654]
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. New York: Lawrence Erlbaum Associates; 1988.
45. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006;27(2):237-246. [doi: 10.1177/1098214005283748]
46. Garland EL, Farb NA, Goldin P, Fredrickson BL. Mindfulness broadens awareness and builds eudaimonic meaning: a process model of mindful positive emotion regulation. Psychol Inq 2015 Oct 1;26(4):293-314 [FREE Full text] [doi: 10.1080/1047840X.2015.1064294] [Medline: 27087765]
47. Bowler DE, Buyung-Ali LM, Knight TM, Pullin AS. A systematic review of evidence for the added benefits to health of exposure to natural environments. BMC Public Health 2010 Aug 4;10:456 [FREE Full text] [doi: 10.1186/1471-2458-10-456] [Medline: 20684754]
48. Ulrich RS, Simons RF, Losito BD, Fiorito E, Miles MA, Zelson M. Stress recovery during exposure to natural and urban environments. J Environ Psychol 1991;11(3):201-230. [doi: 10.1016/s0272-4944(05)80184-7]
49. Valtchanov D, Barton KR, Ellard C. Restorative effects of virtual nature settings. Cyberpsychol Behav Soc Netw 2010 Oct;13(5):503-512. [doi: 10.1089/cyber.2009.0308] [Medline: 20950174]
50. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797 [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
51. 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]
52. Kaplan S. Meditation, restoration, and the management of mental fatigue. Environ Behav 2001;33(4):480-506. [doi: 10.1177/00139160121973106]

Abbreviations

ART: Attention Restoration Theory

FFMQ-15: Five Facet Mindfulness Questionnaire
SSQ: Simulator Sickness Questionnaire
VE: virtual environment
VR: virtual reality

Edited by G Eysenbach; submitted 06.09.19; peer-reviewed by P Navarro, I Cikajlo; comments to author 11.11.19; revised version received 09.12.19; accepted 31.12.19; published 18.03.20.

Please cite as:
Seabrook E, Kelly R, Foley F, Theiler S, Thomas N, Wadley G, Nedeljkovic M
Understanding How Virtual Reality Can Support Mindfulness Practice: Mixed Methods Study
J Med Internet Res 2020;22(3):e16106
URL: http://www.jmir.org/2020/3/e16106/
doi:10.2196/16106
PMID:32186519

Original Paper

Measuring Alliance Toward Embodied Virtual Therapists in the Era of Automated Treatments With the Virtual Therapist Alliance Scale (VTAS): Development and Psychometric Evaluation

Alexander Miloff[1], MA, MPH; Per Carlbring[1], PhD; William Hamilton[2], BSc; Gerhard Andersson[3,4], PhD; Lena Reuterskiöld[4], PhD; Philip Lindner[1,4], PhD

[1]Department of Psychology, Stockholm University, Stockholm, Sweden
[2]Mimerse AB, Stockholm, Sweden
[3]Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden
[4]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

Corresponding Author:
Alexander Miloff, MA, MPH
Department of Psychology
Stockholm University
Frescati Hagväg 8, C402
Stockholm, SE-106 91
Sweden
Phone: 46 08 163680
Email: alexander.miloff@psychology.su.se

Abstract

Background: Automated virtual reality exposure therapies (VRETs) are self-help treatments conducted by oneself and supported by a virtual therapist embodied visually and/or with audio feedback. This simulates many of the nonspecific relational elements and common factors present in face-to-face therapy and may be a means of improving adherence to and efficacy of self-guided treatments. However, little is known about alliance toward the virtual therapist, despite alliance being an important predictor of treatment outcome.

Objective: In this study, we aimed to evaluate the first alliance instrument developed for use with embodied virtual therapists in an automated treatment format—the Virtual Therapist Alliance Scale (VTAS)—by (1) assessing its psychometric properties, (2) verifying the dimensionality of the scale, and (3) determining the predictive ability of the scale with treatment outcome.

Methods: A psychometric evaluation and exploratory factor analysis of the VTAS was conducted using data from two samples of spider-fearful patients treated with VRET and the help of an embodied, voice-based virtual therapist (n=70). Multiple regression models and bivariate correlations were used to assess the VTAS relationship with treatment outcome, according to self-reported fear and convergence with presence and user-friendliness process measures.

Results: The VTAS showed a sound two-factor solution composed of a primary factor covering task, goal, and copresence; adequate internal consistency; and good convergent validity, including moderate correlation (r=.310, P=.01) with outcomes over follow-up.

Conclusions: These preliminary results suggest that alliance toward a virtual therapist is a significant predictor of treatment outcome, favors the importance of a task-goal over bond-factor, and should be explored in studies with larger sample sizes and in additional forms of embodiment.

(J Med Internet Res 2020;22(3):e16660) doi:10.2196/16660

KEYWORDS

alliance; virtual reality; exposure therapy; automated treatment; psychometric; embodiment; virtual therapist; virtual coach; avatar; usability; presence; empathy

Introduction

The relationship between therapist and patient during psychotherapy, referred to as alliance and measured by a number of instruments such as the Working Alliance Inventory (WAI) [1], is considered an important common factor shared by diverse treatments [2]. Alliance is conceptualized as requiring three distinct processes: agreement on therapeutic goals, consensus on tasks that make up therapy, and bond between therapist and patient [3]. Alliance is also understood to occur in stages, first with identifying the therapist as a source of encouragement, warmth, and support and, in later stages, the development of a client's faith and investment in the therapeutic process, a collaborative *working together* [4]. Some alliance measures, such as the Helping Alliance Questionnaire (HAq) [5], are an authoritative operationalization of their authors' key theories; however, the boundaries of the popular WAI were never formalized [2] and researchers continue to refine the measure as they seek to develop a holistic understanding of alliance for their particular treatment.

Virtual reality exposure therapy (VRET) is a technology-based method of delivering exposure-based treatments; it has been tested in clinical trials for some 20 years, with positive results [6], but has only recently become available to consumers [7]. VRET works by using head-mounted displays to present computer-animated graphical stimuli or 360-degree video, interactive to head movements, while occluding the outside world [8]. Few studies, however, have assessed alliance during VRET treatments, perhaps due to an expectation that alliance will be poor in treatments that prevent face-to-face contact with a therapist [9]. Contrary to this expectation, studies of patients treated with VRET and in vivo exposure therapy for social anxiety disorder found no significant difference between interventions as measured by the WAI [10-12]. Using augmented reality exposure therapy (ARET), a related technology [13], the same lack of difference in WAI scores was found following treatment for small-animal phobia [14]. In a systematic review of VRET process and outcome research, only two studies were identified that captured alliance information [15]: a study of fear of flying, but not an acrophobia study, found a significant correlation with outcome.

Automated VRET treatments—replacing in vivo stimuli with virtual stimuli in addition to replacing human therapist with virtual therapist [16]—are an example of autonomous digital treatments, and being self-guided may be among the most scalable of psychological interventions [17,18]. However, previous research suggests that treatments conducted without therapist support suffer from lower adherence and efficacy as compared to treatments that include contact [19,20]. There is also evidence that more contact and better contact (ie, frequency of interactions and persuasive design elements, such as dialogue support) may result in more patients completing treatment [21]. Therefore, simulating key elements of face-to-face therapist interaction with a virtual agent manifesting visually and/or through auditory instruction may improve adherence and bridge the efficacy gap between guided and unguided interventions [22]. This is particularly so if they are successful in replicating common factors in therapy and nonspecific relational elements, such as empathy and therapist attention [23], which some have suggested are better conceived of as an active ingredient and should be intentionally used in treatment to facilitate better outcomes [24]. In low-income regions, they may also offer an opportunity to disseminate evidence-based cognitive behavioral therapy (CBT) where literacy skills are low and even therapist-supported online treatments are not appropriate [25].

To date, three automated VRET treatments have been tested in clinical trials: two for use with acrophobia [26,27] and one for use with spider phobia [28]. A treatment for social anxiety disorder has also been tested as a small pilot study [29]. Freeman et al [27] used what they referred to as a virtual coach, physically embodied in virtual reality (VR) by a trained actor, to assist patients to conduct behavioral experiments, challenge safety behaviors, repeat key learning points, and provide empathic encouragement. In the study by Donker et al [26], patients were given background information about phobias, provided case examples, and given motivation using a 2D, animated, mobile phone app-based avatar called *Tara*. The Hartanto et al [29] system enabled interactive dialogue with an animated virtual health agent via laptop computer to instruct in the use of VRET sessions; provide monitoring data interpretations, based on heart rate, for example; and motivate clients while completing treatments at home. However, the paradigm relied on a personalized treatment plan created by a human therapist versus the three more recent trials, which incorporated game-level-like progression systems similar to an exposure fear hierarchy. We designed our automated treatment for spider phobia [28] with a virtual therapist to deliver guidance and support from within the VRET application using primarily voice-based instructions and, in one version, a graphical representation. Participants were greeted from the beginning of treatment by the virtual therapist, provided information on the purpose of treatment, provided instructions on how to conduct exposure tasks, and were followed throughout treatment with encouragement and positive reinforcement when a task was completed.

To our knowledge, no clinical trial conducted using an automated VRET treatment with a virtual therapist has reported on alliance, despite the therapist-like human qualities their creators designed them with. There are indications, however, that alliance may be possible toward a virtual therapist. One recent trial of an automated online CBT program for insomnia [30] incorporated an audio-based avatar to guide patients through the application and to provide cognitive restructuring for sleep-related concerns. The authors identified comparable or higher goal and task subscale scores using an adapted WAI scale as compared to a therapist-led, Internet-based CBT (ICBT) program for tinnitus and similar task subscale scores as compared to an outpatient treatment group, but lower goal. Affective bond was stable and relatively high over six sessions according to five items from the Bern Post-Session Report. The authors suggested that a questionnaire designed specifically for automated programs and avatars is needed to provide a clearer picture of alliance relationship with symptom change. Correlations with outcome were found but only for the affective bond subscale.

Just one effort has been directed toward developing a WAI scale intended specifically for VRET and ARET treatment: the WAI applied to virtual and augmented reality (WAI-VAR) [31]. The WAI-VAR collected data from across three studies (n=75) for fear of flying, adjustment disorder (VRET), and cockroach phobia (ARET). Clinically significant *recovered* and *improved* participants had large effect size differences with *not changed* patients according to the WAI-VAR (eta-square=0.32, *P*<.001) and large Spearman correlations were noted between alliance and clinically significant change (*r*=.55, *P*<.001). High scores were also found on the non-VR-specific WAI-short form (WAI-S) [32].

The WAI-VAR [31], however, was developed to address comfort and trust with the virtual environment itself, replacing “my therapist” with “virtual environment” for all items of the Spanish version of the WAI-S. Recent research suggests that using conventional measures and simply replacing “therapist” with another term (eg, “app”) may be insufficient for accurately capturing the subtleties of therapeutic alliance [33]. Given the embodied nature of the virtual therapist in this automated treatment, we developed a novel instrument entitled the Virtual Therapist Alliance Scale (VTAS) to gather information on the patient relationship with the virtual therapist themselves. In addition to phrasing WAI items in terms of “the virtual therapist,” modified items from an empathy questionnaire [34] and copresence scale (ie, “the experience of being with another”) [35] were included to capture additional relational characteristics of working with the virtual therapist. A preliminary psychometric evaluation was conducted using data collected from two independent studies (Miloff et al [28] and Lindner et al, under review), convergent validity assessed according to common process measures (ie, presence and user friendliness), and the scale compared to treatment outcome (ie, self-reported fear). Alliance is considered a generic common factor, with meaning and usefulness expected to be preserved across theoretical orientations [2] and evidence suggests that a range of technology-based interventions can facilitate alliance [36]. Therefore, we hypothesized that the therapeutic relationship with a virtual therapist would correlate significantly with treatment outcome and offer insights on what components of the virtual therapist are most important for treatment efficacy.

Methods

Development and Description of the Virtual Therapist Alliance Scale

Inspiration for VTAS items came primarily from the WAI 12-item short report [37] and 32-item long report [1], covering dimensions of bond, task, and goal. Novel items were constructed using pre-existing items from across bond, task, and goal dimensions identified as most relevant to treatments conducted using a virtual therapist. All items were framed in reference to “the virtual therapist.”

Empathy is considered an important component of psychotherapeutic practice and has been shown to strengthen alliance through goals and tasks [38], yet it is not typically associated with artificial human actors. Therefore, categories for measuring empathic behavior from a study entitled *A Virtual Therapist That Responds Empathically to Your Answers* [34] were used to construct novel items.

Presence in virtual reality is the subjective experience of actually being in a virtual environment and that events are really occurring [16]; social presence, or copresence, is a specialized form of presence associated with the experience of *being with another* in a virtual environment [35]. Evidence suggests that presence may be a principle mechanism by which individuals experience emotions in a virtual environment, such as anxiety (eg, during exposure therapy [39]) or comfort and trust (eg, in the presence of a therapist [35]). The Social Presence Survey [40] was used to construct items relevant to work with a virtual therapist.

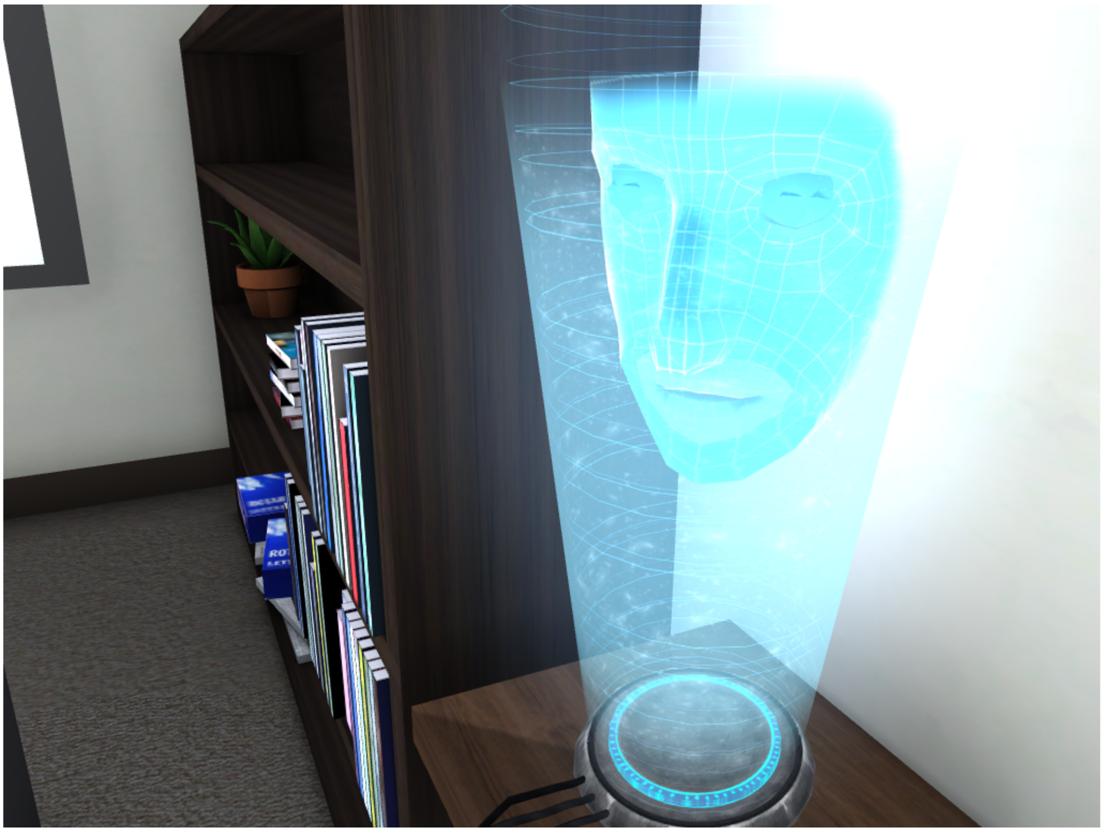
The resulting VTAS has a total of 17 items (see Table 1). Items were originally written in English, before translation into Swedish for data collection. Items were then retranslated into English and then back-translated into Swedish by experienced researchers and a psychotherapist to ensure accuracy. All disagreement was resolved by group consensus among bilingual researchers. All items are visually scored from 0 (*Do not agree at all*) to 4 (*Agree completely*) using the same response format and written anchors. The VTAS total scores were computed by summing all items.

Samples and Procedure

Data for the VTAS psychometric evaluation were collected from two independent samples. The first study consisted of participants meeting criteria for spider phobia and randomized to a gamified VRET treatment, with virtual therapist support provided in Swedish [28]. Participants (n=50 eligible) received the intervention conducted within a single 3-hour period in the presence of a therapist (in case of severe emotional response), who was instructed to act as a computer technician. In the second study (Lindner et al, under review), participants (n=25 eligible) with similar severity of fear of spiders, most of whom also met criteria for spider phobia, were given a nearly identical VRET treatment, except the virtual therapist provided instruction in English. Instead of a technician being present with the patient, the patient received a set of steps to follow in order to complete VRET tasks and a number they could call for assistance.

The virtual therapists in the two studies delivered psychoeducation, treatment instruction, reassurance during the course of a task, and reinforcement when a level was completed, primarily using voiceover instruction. In the second study, the virtual therapist was also given a physical embodiment by means of a holographic image that moved while speaking during the introductory phase of treatment (see Figure 1); instructions were complemented with animated graphics to emphasize psychoeducation lessons. Voiceover commentary was triggered at each stage of treatment in accordance with the initiation and completion of a task, as well as in accordance with participants’ Subjective Unit of Distress Scale (SUDS) ratings: if the SUDS rating by a participant went down over the course of a treatment stage, the patient was provided positive encouragement, and if the SUDS rating went up (ie, the patient became more fearful), the patient was informed that they could slow down and take more time if required. The voice of the virtual therapist was male in both studies and delivered instructions as “your therapist.” Both studies also included instances of an additional individual with a female voice referred to explicitly during treatment as a “spider expert,” who instead of providing therapeutic content, strictly delivered information about the biology and lifecycle of spiders. The language of instruction was Swedish and English in the first and second study, respectively.

Figure 1. Physical embodiment of the primarily voice-based virtual therapist.



All treatments were conducted in a university laboratory environment at Stockholm University. Participants were provided with inexpensive over-ear headphones and virtual reality devices—Gear VR (Samsung) using Galaxy S6 (Samsung) mobile phones—for use during the treatment. In the first study, the VTAS questionnaire was administered in paper format at the postassessment occasion (ie, 1 week following treatment). In the second study, the VTAS was administered along with other questionnaires immediately after treatment via an online survey. Outcome follow-ups were administered at 3 months in the first study and at 6 months in the second study. Outcome data were collected using an online tool [41] for all follow-up in the second study and for follow-up in the first study only if a participant was unavailable to meet in person.

VTAS surveys were not completed by 3 participants in the first study (2 dropouts prior to postassessment and 1 form never completed) and 2 participants in the second (both dropouts prior to treatment), giving a final sample of 70 participants with VTAS data. All items were mandatory, collected in the same order for all participants in both studies, and no individual items were missed. The Stockholm Regional Ethical Review Board approved both study 1 (2015/472-31) and study 2 (2018/1640-32). All participants provided written informed consent.

Regarding demographic characteristics and severity of symptoms, there was no difference in age between individuals in the first (mean 34.1, SD 10.9) and second study (mean 29.7, SD 11.0; t_{73}=1.63, *P*=.11), nor differences in self-reported fear at baseline (t_{73}=-1.32, *P*=.19). Participants were mostly female (42/50, 84%, and 19/25, 76%, respectively), with no significant difference between them (N=75, χ^2_1=0.7, *P*=.40).

Other Measures

Fear of Spiders Questionnaire

The Fear of Spiders Questionnaire (FSQ) [42] is composed of 18 items and evaluates self-reported fear and avoidance of spiders. According to the authors, the FSQ has good internal consistency (Cronbach alpha=.92) and split-half reliability (.89).

System Usability Scale

The System Usability Scale (SUS) is a 10-item scale for assessing user friendliness [43] and appropriateness of a tool for a given context [44]; it is considered technology agnostic, showing utility across a range of interface types, including graphical and speech-based systems. The mean score on the SUS was 82.8 (SD 10.4) (n=70). The SUS was taken as a measure of convergent validity with bond and empathy; the SUS captured characteristics similar to alliance common factors, such as accessibility, empowerment, guidance, and delivery of a sense of security [45].

Igroup Presence Questionnaire

The Igroup Presence Questionnaire (IPQ) [46] is a scale used for measuring the subjective experience of being present in a virtual environment (ie, *presence*). The questionnaire is made up of 14 items, scored from -3 to +3; total scores are a summation of all items, with reverse-coded items compiled such that increased presence was always associated with positive numbers. Item 6 was removed from analysis due to a typo in printing. The mean score on the IPQ was 3.72 (SD 13.2) (n=47) and the Cronbach alpha was .874 (not including item 6). The IPQ was administered in the first study only. Presence according to the IPQ, and the Gatineau Presence Questionnaire (GPQ) below, were taken as a measure of convergent validity with VTAS copresence items.

Gatineau Presence Questionnaire

The GPQ [47] is a scale intended to measure the experience of presence in a virtual environment. In this study, the two positively worded items—"the impression of being there" (item 1) and "appraising the experience as being real" (item 2)—from the 4-item scale were used for analysis. The two negatively worded items (ie, being absent) were not included because they were not expected to correlate with other measures in this study. Items were rated from 0 to 10. The mean score on the first two items of the GPQ was 6.20 (SD 2.31) (n=23). These items had an adequate internal reliability (alpha=.77).

Statistical Analyses

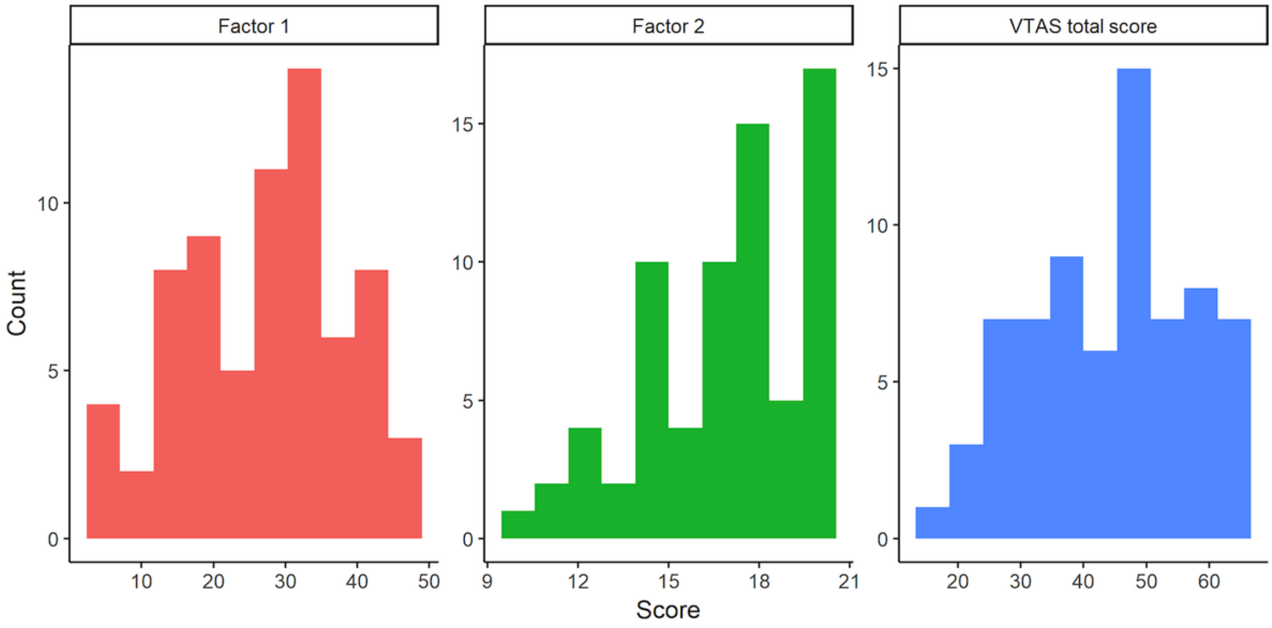
Data cleaning and processing was performed using SPSS for Windows, version 25.0 (IBM Corp), and analyses were conducted using the jamovi, version 1.0.5 (The jamovi project), statistical platform running on an R (The R Foundation) back end. Parallel analysis [48] was used to extract factor loading of items using maximum likelihood explorative factor analysis with oblimin rotation. Factors were identified based on factor loadings; in case of cross-loading (three items), the item was deemed associated with the factor it loaded most heavily on. As recommended by Clark and Watson [49], a sensitivity analysis was carried out that removed similarly cross-loading items, with the threshold set at 0.20 (two items above threshold). VTAS factors used sum scores when conducting correlations with other measures but mean scores when conducting *t* tests. Cronbach alpha and McDonald omega were calculated as measures of internal consistency. Convergent validity was assessed by comparing the VTAS to presence and user-friendliness scores using Pearson correlations. Presence scores were z-transformed since the two studies used different measures. The Steiger test was used to evaluate differences in correlations with other measures between VTAS factors 1 and 2 [50]. Change scores on the symptomatology outcome measure were calculated by subtracting the later time period from the earlier time period (eg, post minus pre). Next, bivariate correlations were conducted between VTAS scores and symptoms change (pre-post and post-follow-up, respectively), presence, and user-friendliness scores. Finally, separate multiple regression models included all these predictors at once, examining the unique associations to VTAS total and subscale scores, while holding the other variables constant. Missing data were handled with case-wise deletion.

Results

Effect of Demographics

There was no effect of sex on VTAS scores ($F_{1,68}$=.269, *P*=.61), nor correlations with age (*r*=-.095, *P*=.43). Mean VTAS scores were 44.0 (SD 12.2) in the first study and 45.0 (SD 13.4) in the second study; VTAS scores followed an approximate normal distribution as evaluated by visual inspection of the VTAS total score histogram (see Figure 2) and computation of the Shapiro-Wilk test (*P*=.11).

Figure 2. Histograms of the Virtual Therapist Alliance Scale (VTAS) total and subscale scores.



Exploratory Factor Analysis

Evaluation of a parallel analysis scree plot for the VTAS indicated a clear, marked reduction of eigenvalues after the first factor, followed by another small reduction after the second factor, remaining consistent thereafter (see Figure 3). This suggests a two-factor solution for the VTAS. Factor 1 consisted of 12 items from task, goal, and copresence categories (except item 3), and factor 2 consisted of five items from bond and empathy (except item 11). Factor 1 explained 38.8% of total variance in the sample and factor 2 explained 14.0% of total variance. All factors had loadings above 0.35—the cutoff recommended by Clark and Watson [49]—and just one item (from factor 2) loaded below 0.40 (item 4, 0.363; see Table 1). The Pearson correlation between factor 1 and factor 2 sum scores was moderate and significant (*r*=.493, *P*<.001), and there was a significant difference between factors 1 and 2 mean scores, with the latter scoring higher (mean -1.13, SE 0.096; t_{69}=-11.8, *P*<.001).

Figure 3. Parallel analysis scree plot of Virtual Therapist Alliance Scale (VTAS) items.

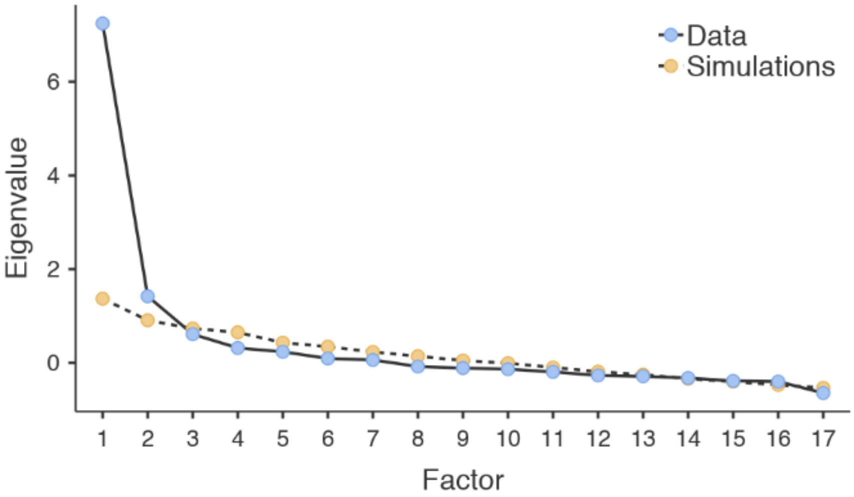


Table 1. Factor loadings for the two factors, including mean item-level scores.

Item	Factor loadings[a]		Mean (SD)
	Factor 1	Factor 2	
1. I experienced the virtual therapist as friendly		0.778	3.81 (0.46)
2. I experienced the virtual therapist as warm		0.643	3.27 (0.90)
3. I felt that the virtual therapist gave clear instructions		0.614	3.63 (0.62)
4. I experienced the virtual therapist as supportive		0.363	3.24 (0.94)
5. The presence of the virtual therapist made the experience more enjoyable	0.492	0.437	3.14 (0.92)
6. It felt like the virtual therapist shared the virtual environment with me	0.580		2.50 (1.28)
7. The virtual therapist appeared alive to me	0.549	0.307	2.19 (1.18)
8. I felt that the virtual therapist and I interacted	0.695		1.41 (1.12)
9. The way that the virtual therapist communicated was captivating	0.659		1.76 (1.21)
10. I felt that the virtual therapist was trustworthy	0.427	0.465	3.04 (0.96)
11. It felt comforting to have the virtual therapist there with me	0.791		2.33 (1.37)
12. The presence of the virtual therapist helped me achieve my goals	0.910		2.30 (1.34)
13. The virtual therapist and I shared common goals	0.648		2.47 (1.21)
14. I felt that the virtual therapist understood my fears	0.710		2.73 (1.17)
15. I felt that the virtual therapist tailored the treatment according to my needs and progress	0.808		1.60 (1.15)
16. The encouragement of the virtual therapist helped me	0.969		2.26 (1.26)
17. The virtual therapist gave me new perspectives on my troubles	0.601		2.61 (1.28)

[a] The *maximum likelihood* extraction method was used in combination with an *oblimin* rotation.

A sensitivity analysis was conducted by removing cross-loaded items. In this case, factors 1 and 2 correlated less strongly (*r*=.320, *P*=.007) and the factor 2 mean score continued to be higher than that of factor 1 (mean -1.29, SE 0.11; t_{69}=-11.7, *P*<.001).

Internal Consistency

The VTAS total Cronbach alpha was .921 (with identical McDonald omega) and was similar across the two subsamples (study 1 alpha=.915, omega=.917; study 2 alpha=.936, omega=.938). Factor 1 Cronbach alpha and McDonald omega was .931 and .932, respectively. This was higher than factor 2, which had Cronbach alpha=.710 and McDonald omega=.759. Factor 2 internal consistency was not improved by dropping the lowest loading item (item 4: alpha=.670, omega=.738). A sensitivity analysis conducted with cross-loaded items removed indicated an internal consistency of alpha=.928 and omega=.930 for factor 1 and alpha=.694 and omega=.751 for factor 2.

Convergent Validity

Presence scores (z-standardized), including data from both studies and assessing convergent validity, correlated strongly with VTAS total scores (*r*>.50; see Cohen 1988 as cited in Hemphill [51]; see Table 2). Factor 1, which included all copresence items, also correlated strongly with presence, while factor 2, which contained no copresence items, had only a weak correlation. A Steiger test indicated a significant difference between the two factors in regard to correlations with presence.

User-friendliness scores, as measured by the SUS, were found to correlate moderately with VTAS total scores. Factor 1 correlated weakly with user friendliness, whereas factor 2 correlated moderately. There was no significant difference, however, between factor correlations with user friendliness according to the Steiger test. A sensitivity analysis conducted with cross-loaded items removed indicated a similar pattern of results with no changes in significance levels.

Table 2. Correlations table between Virtual Therapist Alliance Scale (VTAS) total and factor sum scores as compared to process measures, outcome difference scores, and Steiger test conducted between factor correlations.

Measure	VTAS total		Factor 1 sum		Factor 2 sum		Steiger test	
	r	*P*	*r*	*P*	r	*P*	z	*P*
SUS[a]	.351	.003	.300	.01	.407	<.001	-0.844	.40
Presence *z*-score	.592	<.001	.603	<.001	.298	.01	3.292	<.001
FSQ[b] pre-post	-.213	.08	-.216	.08	-.107	.39	0.720	.47
FSQ post-follow-up	-.310	.01	-.333	.007	-.078	.54	1.824	.07

[a] SUS: System Usability Scale.

[b] FSQ: Fear of Spiders Questionnaire.

Correlations Between Fear of Spiders Questionnaire Change Scores and Virtual Therapist Alliance Scale

Correlations were conducted to evaluate the relationship between VTAS total scores and change in FSQ at postassessment and over follow-up (see Table 2). At postassessment, the VTAS did not correlate significantly with symptom change scores; however, over the follow-up period, the VTAS correlated moderately with change scores.

Neither VTAS factor 1 nor factor 2 sum scores correlated significantly with FSQ change scores pre- to posttreatment. Over follow-up, however, factor 1 moderately correlated with change scores, whereas factor 2 did not and was not significant. Steiger tests did not determine a significant difference of correlations between factor 1 and factor 2 during either change-score time period (*P*>.07 for both). A sensitivity analysis conducted without cross-loaded items identified a similar pattern of results with factor 1 correlating significantly with outcome only over follow-up (*r*=-.338, *P*=.006 versus *r*=-.217, *P*=.08), and factor 2 not correlating with outcome at either time outcome (*P*>.45 for both).

Prediction of Virtual Therapist Alliance Scale Scores

A multiple regression analysis was used that included all predictors of VTAS total scores in a single model (see Table 3). Due to case-wise deletion and missing data across three time periods, a total of 64 out of 70 (91%) participants were included. Of the four covariates, user friendliness, presence, and follow-up FSQ change scores were all significantly associated with the VTAS. Two additional models were included, with each factor as a dependent variable. In the model with factor 1, only presence and follow-up FSQ change scores significantly predicted VTAS factor 1 sum scores; in the model with factor 2, only user friendliness was significant.

Table 3. Multiple linear regression table of Virtual Therapist Alliance Scale (VTAS) total and factor sum values as dependent variables, with included process measure and outcome difference score covariates.

Measure	VTAS total		Factor 1 sum		Factor 2 sum	
	B (SE)	*P*	B (SE)	*P*	B (SE)	*P*
Intercept	14.848 (10.584)	.17	7.248 (9.312)	.44	7.600 (2.630)	.005
SUS[a]	0.311 (0.124)	.02	0.203 (0.109)	.07	0.108 (0.031)	<.001
Presence z-score	5.949 (1.454)	<.001	5.590 (1.280)	<.001	0.359 (0.361)	.32
FSQ[b] pre-post	-0.097 (0.065)	.14	-0.084 (0.057)	.15	-0.014 (0.016)	.41
FSQ post-follow-up	-0.183 (0.069)	.01	-0.177 (0.061)	.005	-0.006 (0.017)	.74

[a] SUS: System Usability Scale.

[b] FSQ: Fear of Spiders Questionnaire.

Discussion

Automated VRET treatments using embodied virtual therapists may be a means of increasing access to exposure therapy by supporting patients to conduct treatments on their own; however, little has been known about patient alliance toward the virtual therapist and its relationship to treatment outcome. This study evaluated a novel scale—the VTAS—using data from two samples of spider-fearful patients treated over a single session.

Psychometric evaluation of the VTAS identified a two-factor solution for the scale, with items based on task, goal, and copresence loading primarily on factor 1, and items based on bond and empathy on factor 2. As summarized in Hatcher and Gillaspy [37], a two-factor solution for the WAI scale is not uncommon. Falkenström, Hatcher, and Holmqvist [52] conducted a confirmatory factor analysis of the scale among three large samples in Sweden and the United States and argued that, given the high intercorrelation between task and goal factors, a two-factor solution was more defensible psychometrically. A recently evaluated WAI scale for guided Internet-based treatments had a similar task-goal and bond two-factor solution [53]. Goal and task factors have been referred to as *agreement-confidence* in the therapist and affective bond as *relationship* [54]. While this terminology may not be appropriate for use with a virtual therapist, if such a therapist can be conceived of in terms of a guide motivating the patient toward behavior change, then a suitable terminology could be task-oriented versus relationship-oriented guidance [55].

It should be noted, however, that not all goal and task items loaded on the first factor and vice versa. For example, item 11—“It felt comforting to have the virtual therapist there with me,” which suggests a bond category—loaded quite highly on factor 1 (0.791), and item 3—“I felt that the virtual therapist gave clear instructions,” suggesting a task category—loaded quite highly on factor 2 (0.614). In the context of using a foreign technical application (ie, VRET), a relationship between clarity of instruction and development of affective bond could be suggested by this outcome (item 3 in factor 2); also, a relationship between receiving comfort and aid and accomplishing the tasks and goals of therapy (item 11 in factor 1). Alternatively, identification of the factor structure could be based on how items were written. Factor 2 items, for the most part, attributed a human quality to the virtual therapist—“I felt/experienced that the virtual therapist was...” *friendly* (item 1), *warm* (item 2), *supportive* (item 4), *trustworthy* (item 10), and *gave clear instructions* (item 3)— whereas factor 1 items were all written as an observation on the nature of the relationship or the benefits accruing from the presence of the virtual therapist. Previous authors of text-based, self-help treatments have argued that alliance is toward the application, as summarized in Heim et al [30]; however, embodied avatars do contain more realistic, human-like qualities and, therefore, the concept of alliance here may be closer to the traditional one. Further research will be needed, such as comparing different forms of embodiment (eg, audio-only instruction versus audio plus physical embodiment), using a rating scale sensitive to this. Preliminary results from this study suggest patients rate factor 2 bond items highly (ie, significantly higher than those of factor 1, $P<.001$); also, factor 2 items explain a sizeable additional total variance in the scale over five items (38.8% vs 14.0%). It is conceivable that bond is strengthened in interventions with an embodied avatar as compared to text-based, self-help applications [35] or, as suggested by Heim et al [30], bond may occur earlier in treatment. In certain treatments, low bond (ie, trust and faith) may be sufficient, whereas in other treatments, such as exposure therapy, high bond may be needed to motivate treatment completion.

Relational bond between patient and mobile self-help apps was explored recently in a mixed qualitative-quantitative study [33]. Qualitative analysis suggested patients do form a personal bond even with their nonavatar, text-based app, referring to it as a “therapist in their pocket” and a friend in the app they could turn to for reassurance and encouragement. Quantitative analysis using a purpose-built questionnaire for measuring alliance in self-guided programs—the mobile Agnew Relationship Measure (mARM)—evaluated one iteration of their scale that simply exchanged “therapist” for “app” on all items of the Agnew Relationship Measure (ARM). Their findings suggest that while certain item terminology could risk inappropriately anthropomorphizing an app—therefore, making it harder to relate to—where human-like qualities are concerned, participants are more likely to endorse a relationship item that includes a qualifying perception (eg, “I feel the app...”) rather than a statement of fact (eg, “The app seems bored or impatient with me”). The former was done for all items attributing human characteristics to the virtual therapist in this study. Internal consistency was lower on factor 2 (Cronbach alpha=.710) and below recent cutoff recommendations by Clark and Watson (alpha=.80) [49]; however, a high coefficient alpha is difficult to obtain in scales with few items, given its relationship to length (ie, 12 versus five items on factor 2) [49]. Factor 1 and overall VTAS consistency was high (Cronbach alpha=.931 and .921, respectively).

Convergent validity findings provide additional support for the independence of factors. The demographic variables age and gender did not have a significant relationship with VTAS scores (*P*>.43 for both). Questionnaires capturing the subjective experience of presence in a virtual environment were found to correlate significantly more with factor 1 (*P*<.001), in which all copresence items loaded. Scores of user friendliness correlated moderately with both factor 1 and factor 2 (*P*<.01 for both). However, of the two measures, only presence significantly predicted factor 1 VTAS results in a multiple linear regression model, other than FSQ follow-up scores, and only user friendliness predicted factor 2 VTAS results. It should be noted that user-friendless scores were between good (71.4) and excellent (85.5), according to generic SUS adjective descriptors (mean 82.8) [56]. This is positive, considering that good system usability may be particularly important in automated treatments that are intended to be self-administered [29].

Evidence for the relationship with outcome indicates that perceived alliance toward a virtual therapist had a small, nonsignificant (*r*=.213) correlation with outcome at posttreatment, but a moderate and significant (*r*=.310) correlation at follow-up. Unlike the study by Heim et al [30], which showed a correlation between bond-factor and outcome and is the only other known study to measure alliance to a therapist avatar in an automated treatment format, only factor 1 (ie, task, goal, and copresence) correlated with outcome in this study. The lack of significant relationship with outcome immediately after treatment is not surprising, given the small-to-moderate sample size and limited power to detect a small association: a post hoc power analysis indicated that 70 participants would have 80% power to detect a correlation of above .33. However, the significant relationship at follow-up may suggest that those who had a good alliance with the virtual therapist better understood and interpreted the treatment instructions correctly or received adequate reinforcement to apply them in their daily lives. Further research is needed here, as well as a better understanding of why particular factors correlate with outcome (eg, as compared to Heim et al [30]). A recent narrative review of therapeutic alliance in Internet interventions found no other study that identified a relationship between the bond-factor and outcome, but many with a meaningful and statistically significant relationship between task-goal and outcome [9]. The author suggests that this may be due to a ceiling effect from high scores on the bond dimension in Internet interventions, similar to this study, or possibly the relative lack of importance of bond in treatments with little-to-no face-to-face interaction. Nevertheless, overall results in this study compare favorably to current meta-analytic best evidence for the relationship between alliance and outcome [57]. Across 39 different alliance measures and 306 studies, the alliance correlation with outcome was .278 (*P*<.0001); over 18 Internet-based studies, the alliance correlation with outcome was .275 (*P*<.0001).

Limitations of this research include the relatively small sample size in the two studies included and small differences between the automated treatments, such as language used and whether the treatments were carried out with a technician present. Together, this meant that measurement invariance had to be assumed but could not be formally tested. The use of a single-session treatment format prevented the exploration of cross-lagged and mediation models to predict symptom reduction following a session [58]. This exploration is important, given the limited capacity of alliance research to show an experimentally causal association with treatment outcome [57]. In the Heim et al [30] study, multiple treatment sessions also provided the authors evidence that a human therapist was increasingly missed over time, although this may be due to the therapeutic techniques used by the virtual avatar in later sessions. Finally, this study explored only client-reported ratings of alliance, versus therapist, observer, and friend or family ratings, whereas other studies, such as a recent blended Internet-based CBT treatment for depression, found only therapist-rated alliance results to be significant with outcome [59].

In the future, research on automated treatments may benefit from evaluating different forms of embodiment for virtual therapists and exploring their effect on alliance, as mentioned earlier, and relationship to treatment outcome and treatment adherence. The virtual therapist in this study primarily used voice-based instruction and was noninteractive; however, it is conceivable that a quality implementation of visual presence, behavioral realism, and interactivity may improve outcomes [35]. Heim et al [30] suggest this may be possible through improved credibility and expectancy, which mediation models indicate function in a bidirectional relationship with alliance to improve outcomes [60]. Previous research on embodied graphically generated agents have included SimCoach [61], which used rich visual representations of humans, gestures, and emotions, as well as natural conversational speech to comfort, guide, and motivate armed service members toward psychological health and to solicit anonymous information. As summarized by Lucas et al [62], rapport may be improved by relying on verbal and nonverbal behavior, such as specific verbal patterns, intonation, and choice of language as well as welcoming gestures, open posture, and attentive eye gaze. Disclosure of personal backstories and intimate personal details by an automated therapist has been shown to increase self-disclosure by highly anxious patients [63]. Relational behaviors, such as calling the patient by name, use of humor when appropriate, including information discussed during past interactions, and engaging in a social chat at the beginning of a new interaction, have all been shown to increase how much a patient feels cared for and how useful the information they received felt to them [25]. Adherence to psychoeducation, operationalized as successful task execution, may be improved by attitude toward a virtual agent, such as their level of trust of the agent, its realism, and its amiableness [64].

In conclusion, this is the first study evaluating an alliance scale for use with virtual therapists in an automated VRET treatment format. Preliminary evidence suggests that this novel VTAS instrument has sound psychometric properties, with one exception being the somewhat low internal consistency of factor 2. Nevertheless, despite this and other limitations, alliance toward a virtual therapist was found to be positively correlated with treatment outcome at follow-up. Further exploration of the predictive abilities of this scale on treatment outcome and adherence is warranted, particularly in conjunction with alternative forms of embodiment, in larger samples of patients, and across different treatment programs and treatment formats.

Conflicts of Interest

WH is the founder and chief technology officer of Mimerse, a company specializing in developing VR interventions for mental health. PL has consulted for Mimerse, but holds no financial stake in the company.

References

1. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Couns Psychol 1989 Apr;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]
2. Horvath AO. Research on the alliance: Knowledge in search of a theory. Psychother Res 2018 Jul;28(4):499-516. [doi: 10.1080/10503307.2017.1373204] [Medline: 28899230]
3. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy 1979;16(3):252-260. [doi: 10.1037/h0085885]
4. Horvath AO, Luborsky L. The role of the therapeutic alliance in psychotherapy. J Consult Clin Psychol 1993;61(4):561-573. [doi: 10.1037/0022-006X.61.4.561]
5. Luborsky L, Barber JP, Siqueland L, Johnson S, Najavits LM, Frank A, et al. The Revised Helping Alliance Questionnaire (HAq-II): Psychometric properties. J Psychother Pract Res 1996;5(3):260-271 [FREE Full text] [Medline: 22700294]
6. 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]
7. Lindner P, Miloff A, Hamilton W, Reuterskiöld L, Andersson G, Powers MB, et al. Creating state of the art, next-generation virtual reality exposure therapies for anxiety disorders using consumer hardware platforms: Design considerations and future directions. Cogn Behav Ther 2017 Sep;46(5):404-420. [doi: 10.1080/16506073.2017.1280843] [Medline: 28270059]
8. Botella C, Baños R, García-Palacios A, Quero S. Virtual reality and other realities. In: Hofmann SG, Asmundson GJG, editors. The Science of Cognitive Behavioral Therapy. San Diego, CA: Academic Press; 2017:551-590.
9. Berger T. The therapeutic alliance in Internet interventions: A narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
10. Anderson PL, Price M, Edwards SM, Obasaju MA, Schmertz SK, Zimand E, et al. Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. J Consult Clin Psychol 2013 Oct;81(5):751-760. [doi: 10.1037/a0033559] [Medline: 23796315]
11. Bouchard S, Dumoulin S, Robillard G, Guitard T, Klinger É, Forget H, et al. Virtual reality compared with exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. Br J Psychiatry 2017 Apr;210(4):276-283. [doi: 10.1192/bjp.bp.116.184234] [Medline: 27979818]
12. Ngai I, Tully EC, Anderson PL. The course of the working alliance during virtual reality and exposure group therapy for social anxiety disorder. Behav Cogn Psychother 2015 Mar;43(2):167-181. [doi: 10.1017/S135246581300088X] [Medline: 24131567]
13. Baus O, Bouchard S. Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: A review. Front Hum Neurosci 2014;8:112 [FREE Full text] [doi: 10.3389/fnhum.2014.00112] [Medline: 24624073]
14. Wrzesien M, Bretón-López J, Botella C, Burkhardt J, Alcañiz M, Pérez-Ara M, et al. How technology influences the therapeutic process: Evaluation of the patient-therapist relationship in augmented reality exposure therapy and in vivo exposure therapy. Behav Cogn Psychother 2013 Jul;41(4):505-509. [doi: 10.1017/S1352465813000088] [Medline: 23410820]
15. Meyerbröker K, Emmelkamp PM. Virtual reality exposure therapy in anxiety disorders: A systematic review of process-and-outcome studies. Depress Anxiety 2010 Oct;27(10):933-944. [doi: 10.1002/da.20734] [Medline: 20734361]
16. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
17. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
18. Lindner P, Miloff A, Hamilton W, Carlbring P. The potential of consumer-targeted virtual reality relaxation applications: Descriptive usage, uptake and application performance statistics for a first-generation application. Front Psychol 2019;10:132 [FREE Full text] [doi: 10.3389/fpsyg.2019.00132] [Medline: 30778311]
19. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: From innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]

20. 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]
21. 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]
22. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: A scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
23. Cavanagh K, Millings A. (Inter)personal computing: The role of the therapeutic relationship in e-mental health. J Contemp Psychother 2013 Jul 17;43(4):197-206. [doi: 10.1007/s10879-013-9242-z]
24. Zilcha-Mano S, Roose SP, Brown PJ, Rutherford BR. Not just nonspecific factors: The roles of alliance and expectancy in treatment, and their neurobiological underpinnings. Front Behav Neurosci 2018;12:293 [FREE Full text] [doi: 10.3389/fnbeh.2018.00293] [Medline: 30760986]
25. Bickmore TW, Pfeifer LM, Jack BW. Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the 27th SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM Press; 2009 Presented at: 27th SIGCHI Conference on Human Factors in Computing Systems; April 4-9, 2009; Boston, MA p. 1265-1274. [doi: 10.1145/1518701.1518891]
26. Donker T, Cornelisz I, van Klaveren C, van Straten A, Carlbring P, Cuijpers P, et al. Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: A randomized clinical trial. JAMA Psychiatry 2019 Jul 01;76(7):682-690. [doi: 10.1001/jamapsychiatry.2019.0219] [Medline: 30892564]
27. Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial. Lancet Psychiatry 2018 Aug;5(8):625-632 [FREE Full text] [doi: 10.1016/S2215-0366(18)30226-8] [Medline: 30007519]
28. Miloff A, Lindner P, Dafgård P, Deak S, Garke M, Hamilton W, et al. Automated virtual reality exposure therapy for spider phobia vs in-vivo one-session treatment: A randomized non-inferiority trial. Behav Res Ther 2019 Jul;118:130-140. [doi: 10.1016/j.brat.2019.04.004] [Medline: 31075675]
29. Hartanto D, Brinkman WP, Kampmann I, Morina N, Emmelkamp PG, Neerincx MA. Home-based virtual reality exposure therapy with virtual health agent support. In: Proceedings of the International Symposium on Pervasive Computing Paradigms for Mental Health, MindCare 2015. Cham, Switzerland: Springer International Publishing; 2016 Presented at: International Symposium on Pervasive Computing Paradigms for Mental Health, MindCare 2015; September 24-25, 2015; Milan, Italy p. 85-98 URL: https://doi.org/10.1007/978-3-319-32270-4_9
30. Heim E, Rötger A, Lorenz N, Maercker A. Working alliance with an avatar: How far can we go with Internet interventions? Internet Interv 2018 Mar;11:41-46 [FREE Full text] [doi: 10.1016/j.invent.2018.01.005] [Medline: 30135758]
31. Miragall M, Baños RM, Cebolla A, Botella C. Working Alliance Inventory applied to virtual and augmented reality (WAI-VAR): Psychometrics and therapeutic outcomes. Front Psychol 2015;6:1531 [FREE Full text] [doi: 10.3389/fpsyg.2015.01531] [Medline: 26500589]
32. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. Psychol Assess 1989 Sep;1(3):207-210. [doi: 10.1037/1040-3590.1.3.207]
33. Berry K, Salter A, Morris R, James S, Bucci S. Assessing therapeutic alliance in the context of mHealth interventions for mental health problems: Development of the Mobile Agnew Relationship Measure (MARM) questionnaire. J Med Internet Res 2018 Apr 19;20(4):e90 [FREE Full text] [doi: 10.2196/jmir.8252] [Medline: 29674307]
34. Pontier M, Siddiqui GF. A virtual therapist that responds empathically to your answers. In: Proceedings of the Intelligent Virtual Agents: 8th International Conference, IVA 2008. Berlin, Germany: Springer; 2008 Presented at: Intelligent Virtual Agents: 8th International Conference, IVA 2008; September 1-3, 2008; Tokyo, Japan p. 417-425. [doi: 10.1007/978-3-540-85483-8_42]
35. Oh CS, Bailenson JN, Welch GF. Frontiers in Robotics and AI. 2018 Oct 15. A systematic review of social presence: Definition, antecedents, and implications URL: https://vhil.stanford.edu/mm/2018/10/oh-frontiers-review-of-social-presence.pdf [accessed 2020-01-23]
36. Lopez A, Schwenk S, Schneck CD, Griffin RJ, Mishkind MC. Technology-based mental health treatment and the impact on the therapeutic alliance. Curr Psychiatry Rep 2019 Jul 08;21(8):76. [doi: 10.1007/s11920-019-1055-7] [Medline: 31286280]
37. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the Working Alliance Inventory. Psychother Res 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
38. McClintock AS, Anderson T, Patterson CL, Wing EH. Early psychotherapeutic empathy, alliance, and client outcome: Preliminary evidence of indirect effects. J Clin Psychol 2018 Jun;74(6):839-848. [doi: 10.1002/jclp.22568] [Medline: 29364509]
39. Ling Y, Nefs HT, Morina N, Heynderickx I, Brinkman WP. A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. PLoS One 2014;9(5):e96144 [FREE Full text] [doi: 10.1371/journal.pone.0096144] [Medline: 24801324]

40. Bailenson JN, Blascovich J, Beall AC, Loomis JM. Interpersonal distance in immersive virtual environments. Pers Soc Psychol Bull 2003 Jul;29(7):819-833. [doi: 10.1177/0146167203029007002] [Medline: 15018671]
41. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for Internet-based psychological treatment. Internet Interv 2016 Nov;6:107-114 [FREE Full text] [doi: 10.1016/j.invent.2016.09.006] [Medline: 30135819]
42. Szymanski J, O'Donohue W. Fear of spiders questionnaire. J Behav Ther Exp Psychiatry 1995 Mar;26(1):31-34. [doi: 10.1016/0005-7916(94)00072-t] [Medline: 7642758]
43. 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]
44. Brooke J. SUS: A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland I, editors. Usability Evaluation In Industry. London, UK: Taylor & Francis; 1996:189-194.
45. Barazzone N, Cavanagh K, Richards DA. Computerized cognitive behavioural therapy and the therapeutic alliance: A qualitative enquiry. Br J Clin Psychol 2012 Nov;51(4):396-417. [doi: 10.1111/j.2044-8260.2012.02035.x] [Medline: 23078210]
46. Schubert T, Friedmann F, Regenbrecht H. The experience of presence: Factor analytic insights. Presence: Teleoperators Virtual Env 2001 Jun;10(3):266-281. [doi: 10.1162/105474601300343603]
47. Laforest M, Bouchard S, Crétu AM, Mesly O. Inducing an anxiety response using a contaminated virtual environment: Validation of a therapeutic tool for obsessive-compulsive disorder. Front ICT 2016 Sep 05;3(18):1-11. [doi: 10.3389/fict.2016.00018]
48. Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organ Res Methods 2016 Jun 29;7(2):191-205. [doi: 10.1177/1094428104263675]
49. Clark LA, Watson D. Constructing validity: New developments in creating objective measuring instruments. Psychol Assess 2019 Dec;31(12):1412-1427. [doi: 10.1037/pas0000626] [Medline: 30896212]
50. Lee IA, Preacher KJ. quantpsy.org. 2013 Sep. Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software] URL: http://www.quantpsy.org/corrtest/corrtest2.htm [accessed 2020-01-23]
51. Hemphill JF. Interpreting the magnitudes of correlation coefficients. Am Psychol 2003 Jan;58(1):78-79. [doi: 10.1037/0003-066x.58.1.78] [Medline: 12674822]
52. Falkenström F, Hatcher RL, Holmqvist R. Confirmatory factor analysis of the patient version of the Working Alliance Inventory-Short Form Revised. Assessment 2015 Oct;22(5):581-593. [doi: 10.1177/1073191114552472] [Medline: 25271007]
53. Gómez Penedo JM, Berger T, Grosse Holtforth M, Krieger T, Schröder J, Hohagen F, et al. The Working Alliance Inventory for guided Internet interventions (WAI-I). J Clin Psychol 2019 Jun 25:1. [doi: 10.1002/jclp.22823] [Medline: 31240727]
54. Andrusyna TP, Tang TZ, DeRubeis RJ, Luborsky L. The factor structure of the Working Alliance Inventory in cognitive-behavioral therapy. J Psychother Pract Res 2001;10(3):173-178 [FREE Full text] [Medline: 11402080]
55. Law M. Examining task and relationship-oriented leadership styles: An in-class application activity. J Acad Bus Ethics 2015 Jan;10(6):1-6 [FREE Full text]
56. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud 2009 May;4(3):114-123 [FREE Full text]
57. 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]
58. Falkenström F, Ekeblad A, Holmqvist R. Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. J Consult Clin Psychol 2016 Aug;84(8):738-751. [doi: 10.1037/ccp0000119] [Medline: 27213493]
59. Vernmark K, Hesser H, Topooco N, Berger T, Riper H, Luuk L, et al. Working alliance as a predictor of change in depression during blended cognitive behaviour therapy. Cogn Behav Ther 2019 Jul;48(4):285-299. [doi: 10.1080/16506073.2018.1533577] [Medline: 30372653]
60. Vîslă A, Constantino MJ, Newkirk K, Ogrodniczuk JS, Söchting I. The relation between outcome expectation, therapeutic alliance, and outcome among depressed patients in group cognitive-behavioral therapy. Psychother Res 2018 May;28(3):446-456. [doi: 10.1080/10503307.2016.1218089] [Medline: 27494662]
61. Rizzo A, Lange B, Buckwalter JG, Forbell E, Kim J, Sagae K, et al. SimCoach: An intelligent virtual human system for providing healthcare information and support. Int J Disabil Hum Dev 2011;10(4):277-281. [doi: 10.1515/ijdhd.2011.046]
62. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Frontiers in Robotics and AI. 2017 Oct 12. Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers URL: https://www.frontiersin.org/articles/10.3389/frobt.2017.00051/pdf [accessed 2020-01-23]
63. Kang S, Gratch J. Exploring users' social responses to computer counseling interviewers' behavior. Comput Human Behav 2014 May;34:120-130. [doi: 10.1016/j.chb.2014.01.006]
64. Tielman ML, Neerincx MA, van Meggelen M, Franken I, Brinkman WP. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. Technol Health Care 2017 Dec 04;25(6):1081-1096 [FREE Full text] [doi: 10.3233/THC-170899] [Medline: 28800346]

Abbreviations

ARET: augmented reality exposure therapy
ARM: Agnew Relationship Measure
CBT: cognitive behavioral therapy
FSQ: Fear of Spiders Questionnaire
GPQ: Gatineau Presence Questionnaire
HAq: Helping Alliance Questionnaire
ICBT: Internet-based cognitive behavioral therapy
IPQ: Igroup Presence Questionnaire
mARM: mobile Agnew Relationship Measure
SUDS: Subjective Units of Distress Scale
SUS: System Usability Scale
VR: virtual reality
VRET: virtual reality exposure therapy
VTAS: Virtual Therapist Alliance Scale
WAI: Working Alliance Inventory
WAI-S: Working Alliance Inventory-short form
WAI-VAR: Working Alliance Inventory applied to virtual and augmented reality

Edited by G Eysenbach; submitted 11.10.19; peer-reviewed by E Carl, M Lippmann; comments to author 25.11.19; revised version received 09.12.19; accepted 15.12.19; published 24.03.20.

Please cite as:
Miloff A, Carlbring P, Hamilton W, Andersson G, Reuterskiöld L, Lindner P
Measuring Alliance Toward Embodied Virtual Therapists in the Era of Automated Treatments With the Virtual Therapist Alliance Scale (VTAS): Development and Psychometric Evaluation
J Med Internet Res 2020;22(3):e16660
URL: https://www.jmir.org/2020/3/e16660
doi:10.2196/16660
PMID:32207690

Original Paper

Attitudes of Nurses Towards Searching Online for Medical Information for Personal Health Needs: Cross-Sectional Questionnaire Study

Avi Zigdon[1], MHA, PhD; Tamar Zigdon[1], BA; Daniel Sender Moran[1], BEd, MSc, PhD

[1] Department of Health Systems Management, School of Health and Medical Sciences, Ariel University, Ariel, Israel

Corresponding Author:
Avi Zigdon, MHA, PhD
Department of Health Systems Management
School of Health and Medical Sciences
Ariel University
Science Park, P O B 3
Ariel, 40700
Israel
Phone: 972 3 907 6571
Email: aviz@ariel.ac.il

Abstract

Background: Use of online clinical health care information has become part of the skill set required by medical teams. Nurses believe that information quality and availability affect nursing care and methods. However, nurses tend not to exploit professional medical databases for evidence-based medical information for their personal needs. This phenomenon has received little research attention.

Objective: This study aimed to address the knowledge gap around nurses' attitudes towards searching online for medical information for their personal needs (ie, for themselves and their families) by (1) evaluating the level of exposure to medical information and the effect on attitudes towards the use of online search options, (2) assessing the effect of the choice of a primary means of searching for medical information on the attitudes towards the use of online search options, and (3) gauging the influence of sociodemographic data and health status on nurses’ attitudes towards searching online for medical information.

Methods: Nurses employed in general departments in a general hospital (34/210, 16.2%), nursing home (42/200, 21.0%), and geriatric medical center (45/180, 25.0%) in Israel were invited to complete the eHealth Impact Questionnaire (alpha=.95). Questionnaires were distributed by nurses in charge of the general hospitalization wards. The data collection period was February to March 2018. The response rate was 40.3% (121/300).

Results: Nurses tended to search for medical information for personal needs on social media (24/121, 19.8%) and TV (eg, health programs, health news; 23/121, 19.0%). Nurses who chose social media as their primary means of receiving general information had a positive attitude about using the online environment as a source for medical information compared to nurses who found information through other means (t_{119}=4.44, P<.001). Nurses exposed to medical information via social media had a positive attitude towards the use of the internet to find medical information compared to nurses who were not exposed to social media (t_{119}=3.04, P=.003). The attitudes of nurses towards the utility of online medical information for personal needs increased with better participant health status ($F_{2,118}$=3.63, P=.03). However, the attitudes of participants with a chronic disease did not differ from those of healthy participants.

Conclusions: Nurses in Israel are less likely to use their professional skills and knowledge to search in professional databases for evidence-based medical information for their personal needs. Instead, they prefer medical information that is easy to access and not evidence-based, such as that on social media and TV. However, these search patterns for personal use may affect their clinical role, impair quality of care, and lead to incorrect medical decisions for their patients in the health care system. Therefore, during nursing education, training for searching skills, retrieval skills, and online search techniques for evidence-based medical information is vital for evidence-based practice.

(J Med Internet Res 2020;22(3):e16133) doi:10.2196/16133

KEYWORDS

information retrieval; social media; evidence-based practice; nursing education; eHealth

Introduction

Background

Searching for information using specialized, online professional databases is a required skill for medical teams in the clinical health care sector [1]. For nurses and physicians, the use of these medical databases is a legitimate part of their clinical function [2], improves patient care [3], and facilitates professional development [4] and decision-making for both patients and health care practitioners [5]. Lack of information retrieval skills and training in online search techniques [6] and the required investment in time [4,7,8] and cost [8] constitute major obstacles to searching for clinical information [7,9]. Nurses are required to possess basic knowledge and multiple skills to perform their clinical role: critical and analytical thinking, searching skills, critical reading skills, and critical evaluation of research [10]. However, nurses prefer to receive clinical information from coworkers, which may not constitute evidence-based medicine [6,11,12], with most accessing medical information in their native language [13]. On one hand, nurses acknowledge that the quality and availability of information affect nursing care and methods [9]. On the other hand, they primarily rely on Google searches [6] and mobile instant messaging applications [14], which provide non-evidence-based medical information. Bibliographic medical databases such as PubMed constitute secondary choices [6]. Nurses use online resources in their daily routines for patient care [15-17], patient training [17-19], medical monitoring [20,21], and patient health tracking [7]. In addition, nurses use virtual communities of health professionals to share professional knowledge [22]. Of course, accessing online medical information is not reserved exclusively for health professionals. Patients are active users of online medical information. With greater online accessibility comes significantly greater consumption of health information by patients [23]. Technology facilitates patient involvement and empowerment in the therapeutic process [24], promotes cooperation between patients and therapists [25], allows the medical team to elicit important feedback on patient opinions and experiences [26], and enables patient management of illness and lifestyle modification during treatment [27].

Attitudes of Nurses Towards Searching Online for Medical Information for Personal Health Needs

Little research has focused on the use of online medical information for the personal needs of nurses, although some research has examined nurses’ use of electronic personal health records (ePHR) as health consumers [28] and social media for health needs [29-31]. In a study on factors related to the use of ePHR by nurses to manage their own health, only a third of 664 registered nurses used ePHR. This research did not find differences in demographic information, career characteristics, or healthcare experience between ePHR users and non-users. Nurses who accessed the internet for general needs used ePHR more, and electronic health (eHealth) literacy was not significantly different between ePHR users and non-users. ePHR non-users were more concerned about their privacy than ePHR users. However, a significant correlation was found between nurses who were ePHR users and nurses who had a chronic illness or underwent a drug therapy regimen [28]. Other studies found that most nurses tend not to search for health information or services when they are sick, with self-treatment very common [32] and professional roles becoming blurred with private life [1-5].

An exploratory study examining technology, internet, and social media use among nurses for personal and professional needs identified significant correlations between the likelihood of nurses recommending searching online for medical information to their patients and family members and age, level of education, and experience. An analysis between age groups found that the older group had a higher probability of recommending internet use to patients and family members. Nurses older than 30 years with formal training were less likely to recommend medical websites as an information source, while those older than 30 years without formal training were more likely to recommend internet use. Nurses with advanced nursing degrees were more likely to suggest using the internet than nurses with a bachelor's degree. Experience also had a role. Nurses with ≥31 years of reported experience had a higher chance of recommending medical websites than nurses with ≤30 years of reported experience. Only 15 nurses reported recommending their patients to “only surf” or use Google (or some other general search engine) to find medical information, while only 4 nurses suggested using .gov, .org, or .edu sources. In fact, social media use may impact the health of both individual nurses and their workplaces. Many nurses use social media for both personal and professional reasons [31], although most nurses tend not to search for health information or services when they themselves are sick, while self-treatment is very common [32].

Nurses often use social media to communicate with peers and track health-related milestones [29]. They especially favor using social media for social support and exchange of health experiences [30]. The significance of patient medical information that is available online is readily evident. This is especially the case for nurses engaged in treating patients and sometimes also for themselves and their families.

Objectives

Very few studies have focused on nurses’ habits for personal need–based searching for medical information online. Therefore, this research examined (1) nurses’ exposure to online medical information, (2) the implications of the primary means of searching for medical information, and (3) the influence of sociodemographic data and nurses' health status on attitudes towards the use of the online environment to search for medical information.

Methods

Design and Setting

This research consisted of an anonymous, self-administered, cross-sectional survey based on the eHealth Impact Questionnaire (eHIQ) [33,34]. Nurses employed in the general departments of a general hospital, a nursing home, and a geriatric medical center in Israel were invited to fill out the questionnaire during the data collection period (February to March 2018). Nurses who did not work in the general department were excluded. In every medical center, 100 questionnaires were distributed by the designated head nurse in the department.

Participants

The research participants were nurses from three general departments in various health institutions: a general hospital (34/210, 16.2%), a nursing home (42/200, 21.0%), and a geriatric medical center (45/180, 25.0%). The respondents could only fill out the questionnaire once. Every head nurse received 100 questionnaires for distribution, for a total of 300 questionnaires. Questionnaires were returned properly by 121 nurses, constituting a response rate of 40.3% (121/300).

Statistical Analysis

Data were analyzed using SPSS Statistics version 25.0 (IBM Corp, Armonk, NY). Descriptive statistics were computed to summarize the data, with means and standard deviations calculated where applicable. The impact of exposure to medical information on the means of accessing the online environment and differences in sociodemographic characteristics were tested using one-tailed *t* tests for independent samples. Differences based ons age were tested using Chi squared tests. Differences between categories of self-reported health status were determined using one-way analysis of variance (ANOVA).

Ethical Considerations

The research data were collected anonymously, without personal information. Answering the questionnaire involved minimal risk. Participation was voluntary. The nurses could refuse to participate in the study and stop filling out the questionnaire at any stage. The purpose of the questionnaire was explained in an introductory segment. Ethics approval was received from the Ethics Committee of Ariel University (ref AU-AZ-20180411) before the study commenced. The head nurses in the hospitals approved the study.

Research Tool

The eHIQ [33,34] is used to measure the effects of online health information on health consumers. The questionnaire was developed by Kelly et al [34] and verified by Kelly et al [33], with internal subscale consistency ranging from .77 to .92 (Table 1).

The questionnaire includes two parts. Part 1 consists of questions on general attitudes towards the online environment for health needs (alpha=.89). Subscale 1 (alpha=.81) measures the participant's openness to receiving online information, while Subscale 2 (alpha=.88) places emphasis on learning and receiving support from other users online. Part 2 consists of questions on the ease of use of the online environment for health needs (alpha=.93). Subscale 3 (alpha=.92) measures the level of confidence the participant has in discussing health issues with other users and identification of relevant online content. Subscale 4 measures the reliability, clarity, and level of distress felt by the participant because of online information (alpha=.62). Subscale 5 (alpha=.87) measures the ability to understand and learn from online information, along with the motivation to act accordingly (Table 1).

Table 1. Internal consistency of the eHealth Impact Questionnaire in this study and in the verification by Kelly et al [33].

Subscale	This study, Cronbach alpha	Verification by Kelly et al, Cronbach alpha
General attitudes	0.89	
1. Attitudes towards online health information	0.81	0.77
2. Attitudes towards sharing health experiences online	0.88	0.89
Ease of use	0.93	
3. Confidence and identification	0.92	0.92
4. Information and presentation	0.62	0.89
5. Understanding and motivation	0.87	0.90

We adapted and translated the English version. The questionnaire was translated into Hebrew and then re-translated into English to verify the quality of the translation and to avoid altering the meaning of the questions (alpha=.95). Face validity was tested by fellow faculty members.

For each of the scales, the sum of the answers to each item was converted from 1 to 100 according to the following formula:

((sum of scores of each item in a scale – minimum raw score) / (maximum raw score – minimum raw score)) x 100

The total score was calculated as the sum of the scores for each of the scales and the number of sub-scales:

total score = sum of subscale scores / number of subscales

In addition to the eHIQ [34], attitudes of nurses towards the reliability of online medical information and its applications were surveyed using two questions: “In general, to what extent is online health information reliable?” and “In general, to what extent is online health information useful?” These were rated on a Likert scale from 5 (to a very large degree) to 1 (not at all).

Results

The sample consisted of 121 nurses from the three general departments in the various health institutions (nursing home, 42/121, 34.7%; geriatric medical facility, 45/121, 37.2%; and general hospital, 34/121, 28.1%). This convenience sample consisted of participants aged 24-72 years (mean 41.2 years, SD 11.4 years), with the following age distribution: 24-35 years, 45/121, 37.2%; 36-50 years, 53/121, 43.8%; ≥51 years, 23/121, 19.0% (Table 2). Of the participants, 102 nurses (102/121, 84.3%) needed medical information in the previous 2 years. Information was sought for themselves by 46 nurses (46/121, 38.0%), for first-degree relatives by 38 nurses (38/121, 31.4%), and for a second-degree relative by 30 nurses (30/121, 24.8%). Very good health was reported by 38 nurses (38/121, 31.4%), good health by 75 nurses (75/121, 62.0%), and bad or poor health by 8 nurses (8/121, 6.6%). A chronic health problem was reported by 31 nurses (31/121, 25.6%), while 89 nurses (89/121, 73.6%) reported they did not have a chronic health problem.

Figure 1 presents the distribution of age groups across the workplaces. Chi-square analysis showed a significant difference between the workplaces (χ^2_4=19.79, *P*<.001). Specifically, the general hospital had younger participants, while the nursing home and geriatric nursing center had more nurses aged 36-50 years.

Table 2. Demographic characteristics of the nurse participants (N=121).

Demographic variables	n (%)
Gender	
Female	73 (60.3)
Male	48 (39.7)
Age, years	
24-35	44 (37.0)
36-50	52 (43.7)
≥51	23 (19.3)
Marital status	
Married/partner	74 (61.2)
Divorced	11 (9.1)
Widowed	4 (3.3)
Never married	32 (26.5)
Country of birth	
Israel	69 (57.0)
Other	52 (43.0)
Religiosity	
Secular	67 (55.8)
Traditional	43 (35.8)
Religious	10 (8.4)
Religious affiliation	
Jewish	84 (69.4)
Muslim	29 (24.0)
Christian	5 (4.1)
Other	3 (2.5)
Place of residence	
City	92 (76.0)
Community/locality	3 (2.5)
Village	26 (21.5)
Professional standing	
Practical nurse	9 (7.5)
Certified nurse	52 (43.3)
Academic nurse	59 (49.2)
Advanced course	
Yes	30 (25.6)
No	87 (74.4)

Figure 1. Age distribution across the three workplaces.

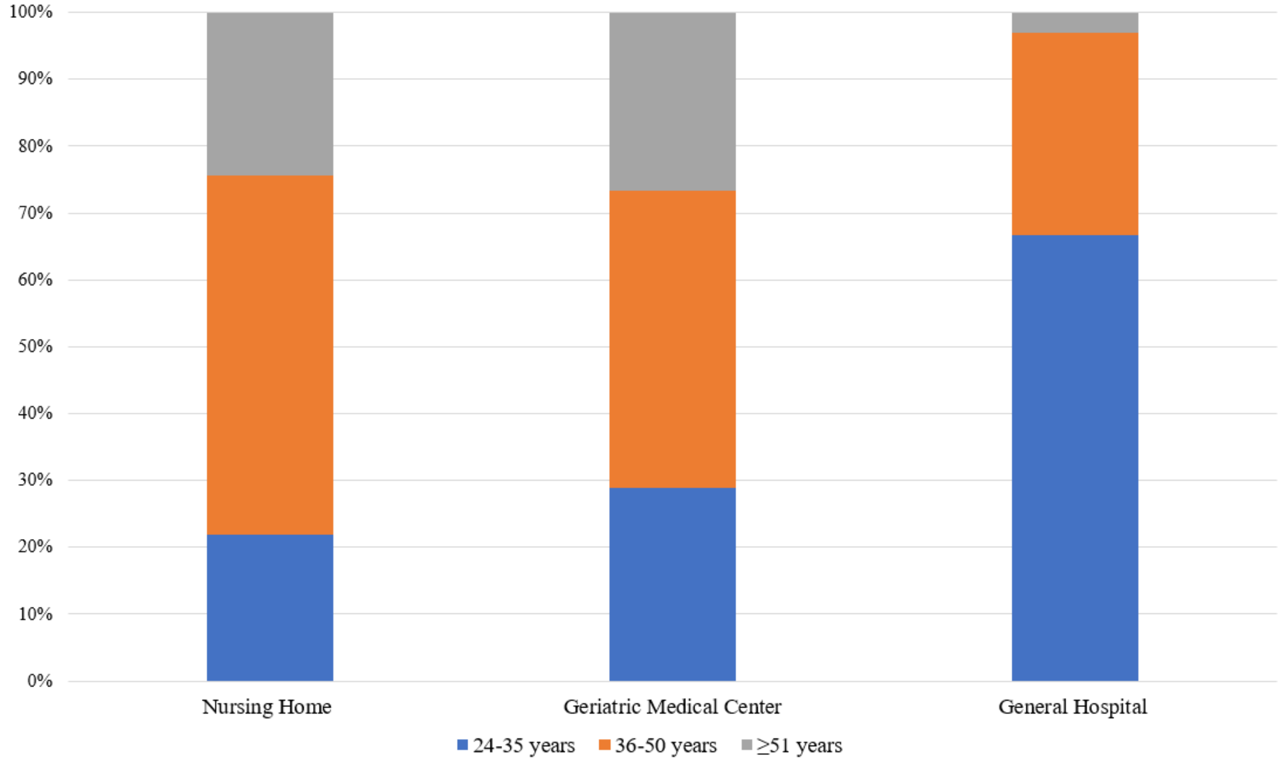


Table 3 shows the comparison between the means of searching online for medical information for personal use (self/family) and exposure to medical information online over the previous 2 years. The nurses searched for medical information for their personal needs using 2 major media platforms, namely social media (24/121, 19.8%) and TV (eg, health programs, health news; 23/121, 19.0%). Similarly, exposure to medical information on social media was reported by 19.0% (23/121) of the nurses, and exposure to medical information on TV was reported by 17.4% (21/121) of the nurses. Interestingly, 18.2% (22/121) of the participants searched for medical information by consulting a professional and not through exposure to professional medical opinion. In addition, the search for evidence-based medical information using professional journals was limited (5/121, 4.1%). Unused media included mobile phones, websites of private organizations, radio, and billboards.

Table 3. Comparison between the means of searching online for medical information for personal use (self/family) and exposure to medical information online over the previous 2 years (N=121). Participants could select more than one answer.

	Searching online for medical information (n=238), n (%)	Exposure to medical information online (n=283), n (%)
Social media	47 (19.7)	55 (19.4)
TV	46 (19.3)	50 (17.7)
Consulting a professional	43 (18.1)	0 (0)
News websites	28 (11.8)	34 (12)
Government websites	24 (10.1)	39 (13.8)
Service association websites	14 (5.9)	0 (0)
Non-governmental organization websites	10 (4.2)	19 (6.7)
Professional journals	10 (4.2)	0 (0)
Friends and family	9 (3.8)	0 (0)
Newspapers	7 (2.9)	13 (4.6)
Billboards	0 (0)	4 (1.4)
Radio	0 (0)	7 (2.5)
Private organization websites	0 (0)	12 (4.2)
Not exposed at all	0 (0)	13 (4.6)
Cell phone	0 (0)	37 (13.1)

No significant differences were found in the means of searching online or online exposure to medical information between the age groups ($\chi^2_{28}=15.00$, $P=.24$)

Comparisons of the eHIQ score, subscores, and additional attitudes question scores resulted in a significant difference in the Subscale 1 (attitudes towards online health information) score, with nurses who did use social media as the primary means of accessing general information having more positive attitudes than nurses who did not use social media for that purpose ($t_{119}=4.44$, $P<.001$). Similarly, nurses who were exposed to medical information on social media had significantly more positive attitudes according to Subscale 1 than nurses who were not exposed to medical information on social media ($t_{119}=3.04$, $P=.003$). Nurses who accessed medical information on mobile phones had significantly more positive attitudes towards the ease of use of the online environment for health needs ($t_{119}=2.66$, $P=.009$). Nevertheless, it was not possible to accurately determine the mode of exposure on mobile phones that allow access to social media or TV.

Nurses were also asked to select their primary means for accessing general information (one choice). In descending order, they reported using TV programs (46/121, 38.0%), social media (21/121, 17.4%), professional advice (21/121, 17.4%), news websites (11/121, 9.1%), government websites (11/121, 9.1%), non-governmental organization websites (7/121, 5.8%), and professional journals (5/121, 4.1%).

No differences were found between age groups across the media types for accessing general information ($\chi^2_{14}=12.66$, $P=.12$). Nurses who chose social media as their primary means of accessing general information had a more positive attitude toward accessing medical information online than nurses who accessed general information by other means. Nurses exposed to medical information via social media had a more positive attitude toward using the internet to access medical information than nurses who were not exposed to social media (Table 4). Nurses who were exposed to medical information via a mobile phone had significantly more positive attitudes towards the use of the internet for medical purposes (total eHIQ score; $t_{117}=2.71$, $P<.001$) and the ease of use of the online environment for health information (eHIQ ease of use subscore; $t_{116}=2.62$, $P<.001$). Again, it was not possible to determine the exact mode of exposure to information via a mobile phone that allows access to social media, news websites, and professional journals. There were no other differences in the eHIQ scores based on the means to search for medical information online or exposure to medical information online.

Table 4. Differences in eHealth Impact Questionnaire (eHIQ) score, eHIQ subscores, and additional attitudes questions based on the use of social media as a primary means to access information online or for online exposure to medical information (N=121).

	Social media as the primary means for accessing general information				Exposure to medical information on social media			
	No (n=66), mean (SD)	Yes (n=55), mean (SD)	*t* test	*P* value	No (n=66), mean (SD)	Yes (n=55), mean (SD)	*t* test	*P* value
eHIQ general attitudes								
Subscale 1. Attitudes towards online health information	46.6 (21.4)	63.5 (19.2)	$t_{119}=4.44$	<.001	46.6 (21.4)	59.6 (22.1)	$t_{119}=3.04$	.003
Subscale 2. Attitudes towards sharing health experiences online	55.0 (24.0)	64.5 (22.6)	$t_{117}=2.15$	.03	55.0 (24.0)	61.7 (24.0)	$t_{117}=1.22$	.22
Total score	50.4 (19.8)	64.0 (18.6)	$t_{117}=3.74$	<.001	50.4 (19.8)	60.4 (20.4)	$t_{117}=2.31$	.02
eHIQ ease of use								
Subscale 3. Confidence and identification	47.9 (18.2)	65.4 (16.6)	$t_{117}=5.31$	<.001	47.9 (18.2)	63.7 (16.7)	$t_{117}=4.88$	<.001
Subscale 4. Information and presentation	45.9 (18.9)	63.7 (18.9)	$t_{116}=4.99$	<.001	45.9 (18.9)	61.5 (19.2)	$t_{116}=4.45$	<.001
Subscale 5. Understanding and motivation	49.0 (18.9)	66.8 (18.2)	$t_{117}=5.10$	<.001	49.0 (18.9)	64.7 (18.9)	$t_{117}=4.43$	<.001
Total score	47.6 (17.0)	65.3 (17.0)	$t_{116}=5.53$	<.001	47.6 (17.0)	63.3 (17.2)	$t_{116}=4.97$	<.001
eHIQ total score	48.6 (16.8)	64.8 (16.9)	$t_{115}=5.07$	<.001	48.6 (16.8)	62.2 (17.6)	$t_{115}=4.10$	<.001
Additional attitudes questions								
Reliability of internet information	3.2 (0.9)	3.8 (0.7)	$t_{119}=3.68$	<.001	3.2 (0.9)	3.7 (0.8)	$t_{119}=3.68$	<.001
Usefulness of internet information	3.3 (0.8)	4.0 (0.8)	$t_{119}=4.55$	<.001	3.3 (0.8)	3.9 (0.8)	$t_{119}=4.55$	<.001

Regarding the effect of sociodemographic characteristics on nurses' attitudes, attitudes towards accessing information online were significantly different by place of birth and living area.

Nurses born in Israel had a significantly lower mean eHIQ general attitude score (mean 51.115, SD 20.636) than nurses not born in Israel (mean 62.194, SD 18.317; t_{117}=3.028, *P*=.003). Nurses born in Israel also had a significantly lower Subscale 1 (attitudes towards online health information) scores (mean 49.114, SD 22.030) than nurses not born in Israel (mean 58.440, SD 21.205; t_{119}=2.342, *P*=.02). In addition, nurses born in Israel had a significantly lower mean eHIQ Subscale 2 (attitudes towards sharing health experiences online) score (mean 53.116, SD 24.706) than nurses who were not born in Israel (mean 66.500, SD 20.335; t_{117}=3.136, *P*=.002). There were no statistically significant differences in the eHIQ ease of use score, attitude toward information reliability, or attitude toward information usefulness based on country of birth.

Based on place of residence, nurses living in the center of Israel had a significantly higher mean eHIQ general attitude score (mean 58.422, SD 19.589) than nurses living in the peripheral areas of Israel (mean 48.244, SD 20.973; t_{117}=2.442, *P*=.02). Nurses living in the center of Israel also had a significantly higher Subscale 1 (attitudes towards online health information) score (mean 55.741, SD 21.836) than nurses living in the peripheral areas of Israel (mean 45.520, SD 21.345; t_{119}=2.260, *P*=.03). In addition, nurses living in the center of Israel had a significantly higher mean eHIQ Subscale 2 (attitudes towards sharing health experiences online) score (mean 61.477, SD 21.674) than nurses living in the peripheral areas of Israel (mean 50.968, SD 28.030; t_{117}=2.144, *P*=.03). No significance differences were found based on place of residence in the attitudes towards security and identification (Subscale 3), reliability of online health information, and usefulness of online health information. There were also no differences in nurses' attitudes towards searching online for personal need–based medical information based on the remaining sociodemographic characteristics.

The attitude towards the usefulness of online health information was more positive with increasing self-reported health (not so good, good, and very good; Table 5). However, the attitude was not different between participants who had a chronic disease and those who did not have a chronic disease.

There was a significant correlation between the attitudes towards the reliability of online information and attitudes towards the usefulness of online information (r=.758, *P*<.001).

Table 5. Attitudes towards the usefulness of online information according to self-reported level of health (N=121).

	Not so good (n=8), mean (SD)	Good (n=75), mean (SD)	Very good (n=38), mean (SD)	F statistic	*P* value
Usefulness of online health information	3.3 (0.7)	3.4 (0.8)	3.8 (1.0)	$F_{2,118}$=3.63	.03

Discussion

Principal Findings

Evidence-based medical information is a main resource for medical teams in health care systems. Only a few studies have examined the attitude of nurses towards searching online for personal need–based medical information. This research sheds light on their attitudes, showing that nurses mainly use social media (24/121, 19.8%) and TV (eg, health programs, health news; 23/121, 19.0%) for this purpose. The general attitude of participants who chose/were exposed to social media as the main source of medical information was significantly more positive in comparison to those who did not choose/were not exposed to social media as a source of medical information. Attitudes towards the ease of use of the online environment for health needs were also significantly greater for nurses who accessed health information via mobile phones than for those who did not use mobile phones to access information (t_{119}=2.66, *P*=.009). Nevertheless, it is not possible to accurately determine the method of accessing information on a mobile phone that allows access to social media and TV.

Some nurses search for medical information by consulting a professional (22/121, 18.2%). However, they express only marginal interest in using professional research sources (5/121, 4.1%) in their personal need–based medical information searching. This is consistent with Wolf et al [31], who found that only 4 nurses recommended to their patients and family members the use of .gov, .org, or .edu sources.

Research would be expected to find that nurses, as professional health care practitioners, look for personal need–based medical information using medical databases such as PubMed. This would align with the most prudent path of decision-making for their health, but the reality is quite different. Other research data on searching for professional health care information by nurses for patient treatment [15-17], patient training [17-19], medical monitoring [20,21], and tracking patient health [7] indicate that nurses prefer to access clinical information from co-workers [6,11,12], information which is not necessarily evidence-based [6].

The results of this study also show that using mobile phones for nurses' personal need–based medical information was linked to a more positive attitude toward the ease of use of the online environment for health information. However, a mobile phone represents a means of accessing the online environment and does not constitute an information source. In other studies, mobile instant messaging apps [14] were a way for nurses to search for information, track milestones related to their health [29], and, especially, for social support and exchange of health experiences [30]. In this study, nurses' attitudes towards using the online environment to obtain medical information for their personal needs were significantly different based on place of birth and their residence; there were no differences based on the other sociodemographic characteristics. Self-reported health status and chronic disease did not affect nurses' preferences for the source of health information, which differs from the findings of Wolf et al [31] who found that with older age, nurses with no formal training and experienced nurses alike tended to recommend the internet for medical purposes for their patients. This finding is supported by the assumption that a patient with a disease receives orderly medical care, whereas an individual experiencing clear symptoms but has not yet undergone a medical examination will actively search for online medical information. This assumption should be examined in future studies.

Limitations and Future Research

This study contains some methodological limitations. First, the nurses filled out the questionnaires in the workplace where they provided patient treatment. This could subjectively affect the responses as the questionnaire deals with individual attitudes in the use of online medical information for personal needs. At the same time, health status is an issue that has been reviewed by nurses' personnel reports. The medical conditions of the participants were not verified, and it is unclear whether the severity of a medical condition affected the reported attitudes. In addition, it is difficult to determine the exact mode of participant exposure to information when using a mobile phone, as it enables access to social media, news websites, and professional research. To provide further focus in the results, targeted research is needed, and health reporting based on disease codes from medical files and subjective reports is required.

Conclusion

Nurses in Israel tend not to use their professional skills and knowledge to search for evidence-based medical information using a professional database such as PubMed when looking for medical information for themselves and their families. They prefer non-evidence-based medical information that is easy to access such as that found on social media and TV. These search patterns for information for personal use may affect their clinical role, impair the quality of care, and lead to incorrect medical decisions for their patients in the health care system. Moreover, these patterns might hinder their professional development and establish consumption patterns of erroneous medical information. Therefore, during nursing education, training for searching, retrieval skills, and training in online search techniques for evidence-based medical information is vital for evidence-based practice. A change to information-seeking behaviors that focus on evidence-based information can be tested in a government-issued standardized exam directly following formal training, and 6-monthly seminars on advanced searching skills for medical information can be offered.

Acknowledgments

We would like to thank the students Ekaterina Lileev, Ortal Motola, and Snegana Kokot for their assistance in distributing the questionnaires.

Conflicts of Interest

None declared.

References

1. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a Conceptual Model for eHealth: Qualitative Study with Key Informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
2. Lialiou P, Mantas J. Online Information Seeking Behaviour by Nurses and Physicians: A Cross-Sectional Study. In: Sermeus W, Procter M P, Weber P, editors. NURSING INFORMATICS 2016 eHealth for All: Every Level Collaboration – From Project to Realization. Amsterdam, Berlin, Washington, DC: IMIA and IOS Press; 2016:33-37.
3. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, et al. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. Health Info Libr J 2013 Sep;30(3):178-190. [doi: 10.1111/hir.12036] [Medline: 23981019]
4. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. Health Info Libr J 2010 Mar;27(1):2-10 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00883.x] [Medline: 20402799]
5. Scantlebury A, Booth A, Hanley B. Experiences, practices and barriers to accessing health information: A qualitative study. Int J Med Inform 2017 Dec;103:103-108. [doi: 10.1016/j.ijmedinf.2017.04.018] [Medline: 28550995]
6. Alving BE, Christensen JB, Thrysøe L. Hospital nurses' information retrieval behaviours in relation to evidence based nursing: a literature review. Health Info Libr J 2018 Mar;35(1):3-23. [doi: 10.1111/hir.12204] [Medline: 29327483]
7. Gilmour J, Strong A, Chan H, Hanna S, Huntington A. Primary health-care nurses and Internet health information-seeking: Access, barriers and quality checks. Int J Nurs Pract 2016 Feb;22(1):53-60. [doi: 10.1111/ijn.12361] [Medline: 25355072]
8. Intas G, Kostagiolas P, Zavras D, Chalari E, Stergiannis P, Toylia G, et al. Information-Seeking Behavior of Greek Nursing Students: A Questionnaire Study. Comput Inform Nurs 2017 Feb;35(2):109-114. [doi: 10.1097/CIN.0000000000000294] [Medline: 27685335]
9. Argyri P, Kostagiolas P, Diomidous M. A survey on information seeking behaviour of nurses at a private hospital in Greece. Stud Health Technol Inform 2014;202:127-130. [Medline: 25000032]
10. Moule P, Aveyard H, Goodman M. Nursing Research an Introduction 3rd Edition. London: Sage Publications Ltd; 2017.

11. Dee C, Stanley EE. Information-seeking behavior of nursing students and clinical nurses: implications for health sciences librarians. J Med Libr Assoc 2005 Apr;93(2):213-222 [FREE Full text] [Medline: 15858624]
12. McCaughan D, Thompson C, Cullum N, Sheldon T, Raynor P. Nurse practitioner and practice nurses' use of research information in clinical decision making: findings from an exploratory study. Fam Pract 2005 Oct;22(5):490-497. [doi: 10.1093/fampra/cmi042] [Medline: 15964873]
13. Weng YH, Kuo KN, Yang CY, Lo HL, Shih YH, Chiu YW. Information-searching behaviors of main and allied health professionals: a nationwide survey in Taiwan. J Eval Clin Pract 2013 Oct;19(5):902-908. [doi: 10.1111/j.1365-2753.2012.01871.x] [Medline: 22672429]
14. Bautista JR, Lin TTC. Nurses' use of mobile instant messaging applications: A uses and gratifications perspective. Int J Nurs Pract 2017 Oct;23(5):e12577. [doi: 10.1111/ijn.12577] [Medline: 28752519]
15. Ciuca AM, Berger T, Crişan LG, Miclea M. Internet-based treatment for panic disorder: A three-arm randomized controlled trial comparing guided (via real-time video sessions) with unguided self-help treatment and a waitlist control. PAXPD study results. J Anxiety Disord 2018 May;56:43-55. [doi: 10.1016/j.janxdis.2018.03.009] [Medline: 29625834]
16. Díaz-García A, González-Robles A, Fernández-Álvarez J, García-Palacios A, Baños RM, Botella C. Efficacy of a Transdiagnostic internet-based treatment for emotional disorders with a specific component to address positive affect: Study protocol for a randomized controlled trial. BMC Psychiatry 2017 Apr 20;17(1):145 [FREE Full text] [doi: 10.1186/s12888-017-1297-z] [Medline: 28424068]
17. Valan L, Sundin K, Kristiansen L, Jong M. Child health nurses’ experiences and opinions of parent Internet use. Early Child Development and Care 2017 Jan 17;188(12):1738-1749. [doi: 10.1080/03004430.2017.1278697]
18. 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 04;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
19. Rini C, Vu MB, Lerner H, Bloom C, Carda-Auten J, Wood WA, et al. A qualitative study of patient and provider perspectives on using web-based pain coping skills training to treat persistent cancer pain. Palliat Support Care 2018 Apr;16(2):155-169. [doi: 10.1017/S1478951517000086] [Medline: 28264744]
20. van der Krieke L, Emerencia AC, Aiello M, Sytema S. Usability evaluation of a web-based support system for people with a schizophrenia diagnosis. J Med Internet Res 2012;14(1):e24 [FREE Full text] [doi: 10.2196/jmir.1921] [Medline: 22311883]
21. 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]
22. 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]
23. 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]
24. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. Health Informatics J 2014 Mar;20(1):50-58. [doi: 10.1177/1460458213476969] [Medline: 24550564]
25. Khanum S, Souza, Naz N, Sasso G, Brüggemann O, Heideman I. The Use of Networking in Nursing Practice —An Integrative Review. Societies 2016 Jul 29;6(3):22. [doi: 10.3390/soc6030022]
26. De MI, D'Apolito R, McLawhorn AS, Fehring KA, Sculco PK, Gasparini G. Social media for patients: benefits and drawbacks. Curr Rev Musculoskelet Med 2017 Mar;10(1):141-145 [FREE Full text] [doi: 10.1007/s12178-017-9394-7] [Medline: 28110391]
27. 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]
28. Gartrell K, Storr C, Trinkoff A, Wilson M, Gurses A. Electronic personal health record use among registered nurses. Nurs Outlook 2015;63(3):278-287 [FREE Full text] [doi: 10.1016/j.outlook.2014.11.013] [Medline: 25982768]
29. Jackson J, Fraser R, Ash P. Social Media and Nurses: Insights for Promoting Health for Individual and Professional Use. Online J Issues Nurs 2014 Sep 30;19(3):2 [FREE Full text] [Medline: 26824150]
30. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
31. Wolf DM, Wenskovitch J, Anton BB. Nurses’ use of the Internet and social media: Does age, years of experience and educational level make a difference? JNEP 2015 Nov 02;6(2):68. [doi: 10.5430/jnep.v6n2p68]
32. Bana S, Yakoob J, Jivany N, Faisal A, Jawed H, Awan S. Understanding Health Seeking Behavior Of Health Care Professionals In Tertiary Care Hospitals In Pakistan. J Ayub Med Coll Abbottabad 2016;28(3):545-549 [FREE Full text] [Medline: 28712232]

33. Kelly L, Ziebland S, Jenkinson C. Measuring the effects of online health information: Scale validation for the e-Health Impact Questionnaire. Patient Educ Couns 2015 Nov;98(11):1418-1424 [FREE Full text] [doi: 10.1016/j.pec.2015.06.008] [Medline: 26162953]
34. Kelly L, Jenkinson C, Ziebland S. Measuring the effects of online health information for patients: item generation for an e-health impact questionnaire. Patient Educ Couns 2013 Dec;93(3):433-438 [FREE Full text] [doi: 10.1016/j.pec.2013.03.012] [Medline: 23598293]

Abbreviations

eHealth: electronic health.
eHIQ: eHealth Impact Questionnaire.
ePHR: electronic personal health records.

Edited by G Eysenbach; submitted 04.09.19; peer-reviewed by A Garcia Linares, GE Iyawa, D López López, A Younas; comments to author 04.11.19; revised version received 23.12.19; accepted 24.01.20; published 16.03.20.

Please cite as:
Zigdon A, Zigdon T, Moran DS
Attitudes of Nurses Towards Searching Online for Medical Information for Personal Health Needs: Cross-Sectional Questionnaire Study
J Med Internet Res 2020;22(3):e16133
URL: http://www.jmir.org/2020/3/e16133/
doi:10.2196/16133
PMID:32175910

Original Paper

Electronic Health Literacy in Swiss-German Parents: Cross-Sectional Study of eHealth Literacy Scale Unidimensionality

Sibylle Juvalta[1], BSc, MScPH; Matthew J Kerry[1], BA, MS, PhD; Rebecca Jaks[1], BSc, MA; Isabel Baumann[1], BA, MA, PhD; Julia Dratva[1,2], MD, MPH, Prof Dr

[1]Institute of Health Sciences, Department of Health, Zurich University of Applied Sciences, Winterthur, Switzerland

[2]University of Basel, Basel, Switzerland

Corresponding Author:
Matthew J Kerry, BA, MS, PhD
Institute of Health Sciences
Department of Health
Zurich University of Applied Sciences
Technikumstrasse 71
Winterthur, 8400
Switzerland
Phone: 41 58 934 41 79
Email: matthewj.kerry@zhaw.ch

Abstract

Background: Parents often use digital media to search for information related to their children's health. As the quantity and quality of digital sources meant specifically for parents expand, parents' digital health literacy is increasingly important to process the information they retrieve. One of the earliest developed and widely used instruments to assess digital health literacy is the self-reported eHealth Literacy Scale (eHEALS). However, the eHEALS has not been psychometrically validated in a sample of parents. Given the inconsistency of the eHEALS underlying factor structure across previous reports, it is particularly important for validation to occur.

Objective: This study aimed to determine the factor structure of the German eHEALS measure in a sample of parents by adopting classic and modern psychometric approaches. In particular, this study sought to identify the eHEALS validity as a unidimensional index as well as the viability for potential subscales.

Methods: A cross-sectional design was used across two purposive sampling frames: online and paper administrations. Responses were collected between January 2018 and May 2018 from 703 Swiss-German parents. In addition to determining the sampling characteristics, we conducted exploratory factor analysis of the eHEALS by considering its ordinal structure using polychoric correlations. This analysis was performed separately for online–based and paper–based responses to examine the general factor strength of the eHEALS as a unidimensional index. Furthermore, item response theory (IRT) analyses were conducted by fitting eHEALS to a bifactor model to further inspect its unidimensionality and subscale viability.

Results: Parents in both samples were predominantly mothers (622/703, 88.5%), highly educated (538/703, 76.9%), of Swiss nationality (489/703, 71.8%), and living with a partner (692/703, 98.4%). Factor analyses of the eHEALS indicated the presence of a strong general factor across both paper and online samples, and the Wilcoxon rank-sum test indicated that the eHEALS total sum score was not significantly different between the paper and online samples (P=.12). Finally, the IRT analyses indicated negligible multidimensionality, insufficient subscale reliability after accounting for the eHEALS general factor, and a reduced subset of items that could serve as a unidimensional index of the eHEALS across the paper and online samples.

Conclusions: The German eHEALS evidenced good psychometric properties in a parent-specific study sample. Factor analyses indicated a strong general factor across purposively distinct sample frames (online and paper). IRT analyses validated the eHEALS as a unidimensional index while failing to find support for subscale usage.

(J Med Internet Res 2020;22(3):e14492) doi:10.2196/14492

KEYWORDS

health literacy; eHealth; eHEALS; unidimensionality; multidimensionality; factor analysis; item response theory (IRT); bifactor model; validity

Introduction

Parents increasingly use digital sources when seeking information on their child's health [1-3]. Through their accessibility, digital sources offer the opportunity for parents to feel empowered [4] (eg, to verify information received from health professionals), consider alternative treatment options, and develop communal networks with other families and patients with a common disease or condition. However, because the quality and reliability of information from digital sources vary substantially [5], the information can be overwhelming and cause insecurity or anxiety [6-8]. Therefore, eHealth literacy of parents is critical to maximize the potential benefits of digital media for children’s health. eHealth literacy has been defined as “a set of skills required to effectively engage information technology for health” [9].

Research on the eHealth literacy of parents is lacking. A study by Knapp et al [10] in Florida showed that low-income parents of children with special health needs had high levels of internet use for information purposes. However, half of the study participants had difficulties separating high-quality from low-quality information and were not confident using the internet. Similar findings were found in a study of parents whose children had life-threatening illnesses [11]. Both studies used the eHealth Literacy Scale (eHEALS) [9]. Both studies are also older, considering the fast-evolving context of internet acculturation. The eHEALS was developed by Norman and Skinner [12], who defined eHealth literacy as “the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem.” In this regard, the eHEALS pertains to the critical consumption of extant internet content rather than the creation of new content. The eHEALS is a widely applied instrument that has been validated internationally in diverse languages [13]. Adequate internal consistency has been found [13]. However, van der Vaart et al [14] reported weak correlations between the eHEALS scores and tasks on an eHealth performance test. In contrast, more recent research found significant associations between perceptions and performance [15]. In addition, inconsistent results for the factor structure of the eHEALS have been reported. Norman and Skinner [9] proposed a one-factor structure for the original English eHEALS. For the German eHEALS, Soellner et al [16] determined a two-factor structure using confirmatory factor analysis. For an Italian version of the eHEALS applied in the Italian-speaking area of Switzerland, the researchers [17] recommended using the eHEALS total sum score after applying Rasch (item response theory [IRT]) modeling. However, a study with patients at risk of cardiovascular disease in Australia also applied Rasch modeling and concluded that the eHEALS captures different aspects of eHealth literacy, which may have to be scored separately [18]. Neter et al [19] reported a different two-factor solution for adults in Israel aged at least 21 years than the two-factor solution used by Soellner et al [16]. Moreover, two recent studies developed a three-factor solution for the English eHEALS: awareness, skills, and evaluation [20,21]. A three-factor solution was also reported in an IRT analysis of eHEALS, although the authors noted that substantially high interfactor correlations “support an overarching structure of eHEALS” [22]. Notably, both studies implementing IRT analyses of the eHEALS found measurement properties (item difficulties) reflective of their study samples. That is, Diviani et al [17] found wide variability in item difficulties in a broad sample of people aged 16-71 years, and Stellefson et al [22] found high item difficulties in a sample narrowed to older adults.

The eHEALS was originally constructed for broad usage, as creators Norman and Skinner [9] state, “this article describes the development and psychometric evaluation of a measure of eHealth literacy designed for broad use in supporting consumer eHealth in public health and clinical care” (p2). Subsequent research has reported unstable latent factor structures underlying the eHEALS measure, despite the conventional understanding that “broader constructs are stabilized with broad factors” [23]. Given the inconsistent results for the eHEALS, we aimed to establish the psychometric structure of the eHEALS in parents participating in the Digital Parental Counselors study (in German: Digitale Elternratgeber), which investigated digital media use by parents for their children’s health in the German-speaking area of Switzerland [24]. Specifically, we aimed to explore the factor structure of the German eHEALS and to assess the viability of subscales using IRT and bifactor modeling. We also addressed the methodological issues concerning the handling of Likert scales and the use of factor analysis experienced with previous research of the eHEALS.

Methods

Study Population

The study population consisted of a population-based sample of parents with children aged 1-24 months. The birth registries of Zürich and 5 municipalities in the canton of Zürich, which were selected using convenient sampling, provided randomly selected names and addresses of 2573 mothers who gave birth in the previous 24 months. Urban and rural municipalities were included to represent the urban/rural distribution in the German part of Switzerland (75%/25%). The ethical commission of the Canton of Zurich, based on the Swiss Federal Act on Research involving Human Beings, exempted the study from ethics review (BASEC Req-2017-00817).

Data Collection

The data were collected between January 2018 and May 2018. To increase the response rate, we applied a mixed-mode approach using online and paper versions of the questionnaire. The questionnaire consisted of three main parts: (1) sociodemographic characteristics of the parent and child, (2) digital media use in relation to the child’s health, and (3) health-related variables and eHealth literacy.

Parents received a postal invitation letter with a link to the online questionnaire. After the first postal reminder, parents received a paper questionnaire with the second and last reminder letters.

eHealth Literacy Scale

The eHEALS consists of 8 items (see Table 1). Responses are provided using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with total scores ranging from 8 to 40 points. Higher scores reflect higher eHealth literacy. The eHEALS was developed based on the concept that eHealth literacy is composed of core skills grouped into analytical skills such as media literacy and context-specific skills such as health literacy [12]. The eHEALS does not measure these skills directly, but rather "the consumer's perceived skills and comfort with eHealth" [9]. For this study, we used the German eHEALS version developed by Soellner et al [16], who translated and cross-validated the original English version by Norman and Skinner [9] in a German sample.

Table 1. Example of past studies exploring the latent structure of the eHealth Literacy Scale (eHEALS) measure.

Variables	Studies	
	Norman & Skinner (2006) [9]	Soellner et al (2014) [16]
Sample characteristics		
Population	Canadian students	German students
Sample size	664	327
Age (years), range	13-21	16-21
Factor Solution		
Construction	Original	German translation, as reported here
Structure	1-factor	2-factor
Factors		
eHEALS item 1	I know how to find helpful health resources on the Internet	Ich weiss, wie ich im Internet nützliche Gesundheitsinformationen finde[a]
eHEALS item 2	I know how to use the Internet to answer my questions about health	Ich weiss, wie ich das Internet nutzen kann, um Antworten auf meine Fragen rund um das Thema Gesundheit zu bekommen[a]
eHEALS item 3	I know what health resources are available on the Internet	Ich weiss, welche Quellen für Gesundheitsinformationen im Internet verfügbar sind[a]
eHEALS item 4	I know where to find helpful health resources on the Internet	Ich weiss, wo im Internet ich nützliche Gesundheitsinformationen finden kann[a]
eHEALS item 5	I know how to use the health information I find on the Internet to help me	Ich weiss, wie ich Informationen aus dem Internet so nutzen kann, dass sie mir weiterhelfen[a]
eHEALS item 6	I have the skills I need to evaluate the health resources I find on the Internet	Ich bin in der Lage, Informationen, die ich im Internet finde, kritisch zu bewerten[b]
eHEALS item 7	I can tell high quality health resources from low quality health resources on the Internet	Ich kann im Internet zuverlässige von Fragwürdigen Informationen unterscheiden[b]
eHEALS item 8	I feel confident in using information from the Internet to make health decisions.	Wenn ich gesundheitsbezogene Entscheidungen auf Basis von Informationen aus dem Internet treffe, fühle ich mich dabei sicher[a]

[a] Information seeking.

[b] Information appraisal.

Data Analysis

We performed three different analyses to answer three distinct questions. The first analyses were descriptive and concerned differences in the sample characteristics. The second analyses were based on classical test theory and concerned the general factor strength for each sampling frame (online vs paper). The third analyses involved modern IRT and concerned unidimensionality assumptions and item-level bias across the online and paper administration samples.

Descriptive Analysis

Frequencies of the sociodemographic characteristics of the responding parents and their children were analyzed. Separately for the paper, online, and total samples, the single item and total sum eHEALS scores are reported as the median, skew, and mean. The total sum scores from the online and paper questionnaires were compared using the non-parametric Wilcoxon rank-sum test [25] prior to merging the data. All descriptive analyses were computed with Stata 15.0 (StataCorp LLC, College Station, TX).

Classical Test Theory Analysis

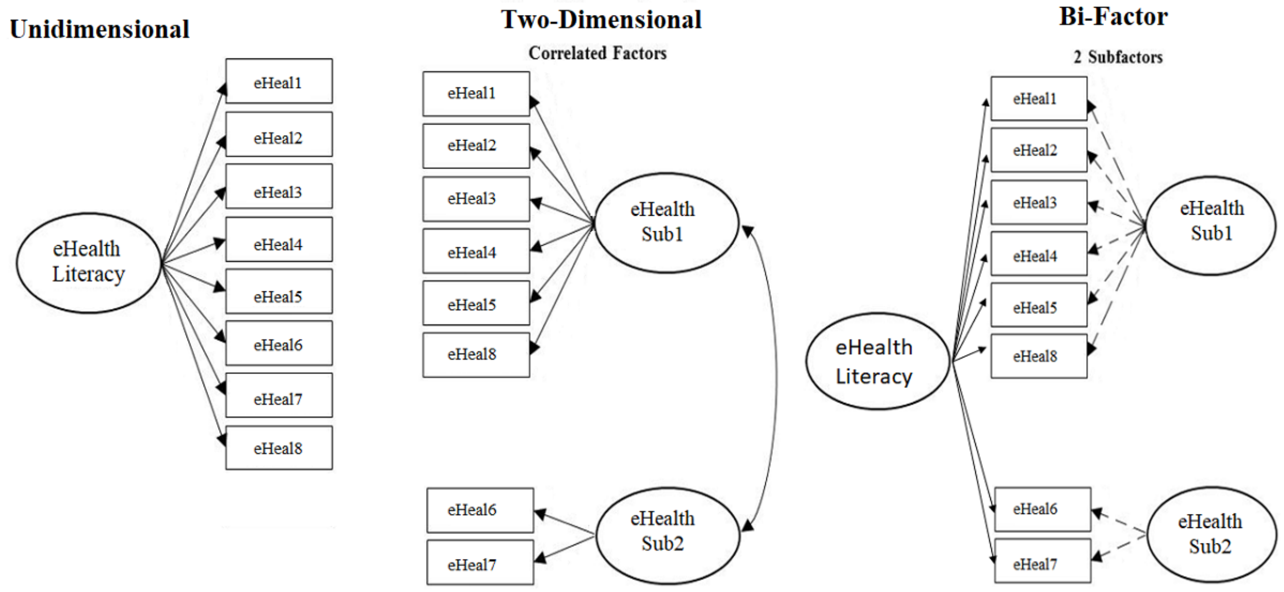
For Likert scales, it is recommended to consider their ordinal structure for factor analysis [26]. As the conventional exploratory factor analysis (EFA) treats variables in a metric manner, polychoric correlations were computed to take into account the ordinal structure of the eHEALS items (see Multimedia Appendix 1).

The detailed results of the EFA conducted with the psych package [27] in R (R Foundation for Statistical Computing, Vienna, Austria) are displayed in Multimedia Appendix 1. Furthermore, a series of parallel analyses were conducted to determine the single-factor strength of the eHEALS measure across the sampling frames [28] using SPSS version 25.0 (IBM Corp, Armonk, NY). Finally, the internal consistency of the eHEALS scale was assessed using the McDonald’s omega coefficient [29].

Item Response Theory Analysis

Previous findings have indicated unstable latent structures (number of eHEALS factors). In such a situation, the bifactor model helps to determine how useful it is to form subscales and examine if unidimensional IRT models can be fit to such multidimensional data [30]. We therefore inspected the eHEALS using a bifactor model. In a bifactor model, a general factor is generated through all test items. Additionally, group factors are established out of the residual variance shared by subsets of items [31] (see Figure 1). In addition to the bifactor model, further IRT understanding based on the internal psychometric structure of the eHEALS was specifically examined in terms of general factor strength (appropriate unidimensional scores) and substantive multidimensionality (item bias). IRTPRO v 4.2 (Scientific Software International, Skokie, IL) was used to run the IRT analyses, and the unidimensional indices were computed using the Bifactor Indices Calculator [32], including McDonald´s omega [29].

Figure 1. Graphical representations of the eHealth unidimensional, correlated two-dimensional, and bi-factor two-dimensional models. The solid lines in the bi-factor model indicate unidimensional primacy over the residualized sub-dimensions (hashed arrows). eHealth: electronic health; eHeal: eHealth Literacy.



Results

Descriptive Analysis

A total of 842 parents or caretakers responded to the survey, and we excluded 73 responses during the data cleaning process for the following reasons: incomplete questionnaire (n=31), missing answers to key questions on parental digital health information seeking (not including the eHEALS items; n=40), non-plausibility of key questions (n=1), and duplicate entry (n=1). This resulted in 769 observations corresponding to a response rate of 30% for the overall study. The online questionnaire was completed by 429 participants (429/769, 56%), and 340 participants (340/769, 44%) completed the paper questionnaire. For the analysis of the eHEALS, 67 additional observations had to be discarded because 52 had missing values for all eHEALS items and 15 had missing values for single eHEALS items.

This led to a final online sample of 388 participants and a final paper sample of 315 participants. Table 2 provides the summary descriptive statistics of the whole sample (N=703). Of the sample, 88.5% (622/703) of the participants were mothers, 76.9% (538/703) reported a university degree or higher vocational education as their highest educational level, and 45.4% (294/703) earned a monthly income >9000 Swiss Francs (US $9080; €8020). The majority (489/703, 71.8%) were Swiss, and almost all study participants (692/703, 98.4%) indicated they lived with a partner.

Table 2. Summary of the sample characteristics, N=703.

Characteristic	Participants, n (%)
Parental sex	
Mother	622 (88.5)
Father	78 (11.1)
Other	3 (0.4)
Age (years)	35.7 (4.3)[a]
Education level	
Lower education	162 (23.1)
Higher education	538 (76.9)
Nationality	
Swiss	489 (71.8)
Other	192 (28.2)
Living with a partner	
Yes	692 (98.4)
No	11 (1.6)
Household net monthly income (CHF)	
<4500	27 (4.2)
4500-6000	94 (14.5)
6000-9000	233 (36.0)
>9000	294 (45.4)
Child’s sex	
Female	349 (49.9)
Male	350 (50.1)
Child’s age (months)	14.8 (7.1)[a]
First child	
Yes	353 (51.2)
No	337 (48.8)
Digital media use score[b]	7.88 (4.13)[a]
eHEALS total sum score	29.0 (5.9)[a]

[a] Mean (SD).

[b] Sum score on how often parents use several digital media for general child health and development (ranging from 0-24).

Concerning the eHEALS items, there were no differences in the individual item responses between the paper and online modes, except for item 3, where the online sample yielded a lower median. The distributions for all items for both the paper and online samples were slightly negatively skewed. As there was no significant difference in the eHEALS total sum scores between the online and paper samples (*P*=.12), the analysis was conducted using the total sample (see Table 3).

Table 3. Descriptive statistics of the individual eHealth Literacy Scale (eHEALS) items and total sum score for the online, paper, and total samples.

eHEALS item	Online (n=388)		Paper (n=315)		Total sample (n=703)		
	Median	Skew	Median	Skew	Median	Mean (SD)	*P* value
Item 1	4	–1.02	4	–0.76	4	3.7 (1.1)	N/A
Item 2	4	–0.91	4	–0.95	4	3.7 (1.0)	N/A
Item 3	3	–0.55	4	–0.55	4	3.4 (1.0)	N/A
Item 4	4	–0.57	4	–0.61	4	3.5 (1.0)	N/A
Item 5	4	–0.87	4	–0.95	4	3.7 (0.9)	N/A
Item 6	4	–1.40	4	–1.23	4	4.2 (0.9)	N/A
Item 7	4	–0.77	4	–0.81	4	3.9 (0.9)	N/A
Item 8	3	–0.24	3	–0.21	3	3.0 (1.1)	N/A
eHEALS total sum score	30	–0.67	29	–0.76	30	28.5 (6.2)	0.12[a] (0.12[b])

[a] Median test.

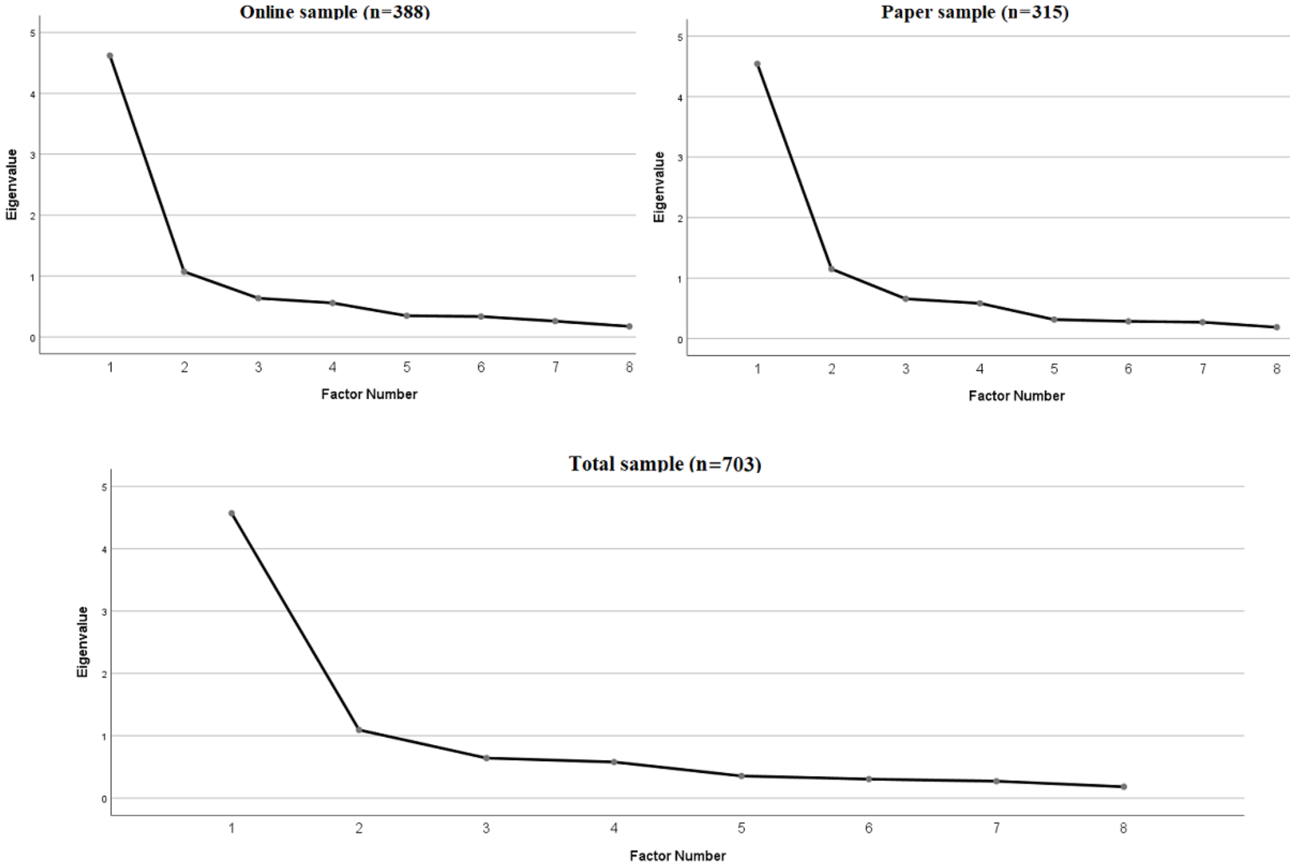
[b] Wilcoxon rank-sum test.

Classical Test Theory Analysis

Given the non-significant difference in the eHEALS total sum scores between theoretically distinct sampling frames (paper and online collection methods), a series of exploratory factor analyses were conducted to examine the strength of the general factor across the samples. As shown in Figure 2, the first eigenvalues indicated the presence of a strong general factor across sampling frames as well as for the total combined sample. Specifically, the first and second eigenvalue ratios across the samples were 4.62 and 1.07 = 4.32 (online), 4.54 and 1.15 = 3.95 (paper), and 4.57 and 1.09 = 4.19 (total). The large eigenvalue ratios suggest negligible multidimensionality across sample frames [33]. This was supported by omega reliability estimates across the samples of 0.90 (online), 0.89 (paper), and 0.89 (total) [29].

When considered collectively with the non-significant difference in eHEALS total scores between sampling frames, we used IRT to test the unidimensionality assumptions with the total eHEALS sample.

Figure 2. Exploratory factor scree plots to examine the strength of the general factors across the online, paper, and total samples.



Item Response Theory Analysis

Separate bifactor models with two group factors and three group factors were computed. A direct model comparison between the bifactors indicated the two-group factor model exhibited significantly greater fit than the three-group factor model (χ^2_1=58.4, P<.001).

Furthermore, the item loadings between the unidimensional and bifactor models were compared to determine the impact on the bias from ignoring suspected multidimensionality (literally, comparing across models with and without additional dimensions). Table 4 lists the loadings across the eHEALS unidimensional and bifactor models.

The average relative parameter bias (0.09; Table 5) value indicates that the impact of ignoring the multidimensionality of eHEALS by using unidimensional scores was negligible [34].

Table 4. Item response theory electronic health item loadings, N=703.

Item	Unidimensional model	Bifactor model with two group factors		
		General Factor	Factor 1	Factor 2
1	0.81	0.87	–0.09	N/A
2	0.88	0.94	–0.07	N/A
3	0.89	0.81	0.40	N/A
4	0.91	0.82	0.56	N/A
5	0.88	0.85	0.11	N/A
6	0.57	0.49	N/A	0.74
7	0.68	0.6	N/A	0.68
8	0.66	0.63	0.10	N/A

Table 5. Item response theory electronic health unidimensionality indices, N=703.

Unidimensionality index	Value
ECV[a]	0.76
Omega reliability	0.99
Hierarchical omega	0.92
H replicability	0.95
Factor determinacy	0.99
ARPB[b]	0.09
IECV[c] (number of items >0.80)	5

[a] ECV: estimated common variance.

[b] ARPB: average relative parameter bias.

[c] IECV: item estimated common variance.

To verify this inference, first, we examined the correlation between the eHEALS and a meaningful substantive variable from the survey with the parents (sum score on how often parents use several digital media for general child health and development), which was significant in the expected direction (r=0.29, P<.01). Second, we examined the difference in this correlation between the unidimensional and bifactor scores of the eHEALS and digital media use for general child health and development. The results indicated observably small changes in the correlation (r_Δ=0.02), which was tested and was not significant ($z_{(1)}$=0.37, P=.71) for the eHEALS total sum score correlation with this substantive variable.

Discussion

The findings of this study support the usefulness of the eHEALS measure as a unidimensional index for further studies. Specifically, we found a strong general factor of the eHEALS across distinct sampling frames as well as adequate reliability. Furthermore, the IRT analyses indicated minimal distortion of the primary factor from ignoring potential multidimensionality, and subscale reliabilities were inadequate to recommend further usage.

With respect to the EFA, this study used a different methodology to add to the current discussion of the eHEALS factor structure. Norman and Skinner [9] and other researchers factorizing the eHEALS [17,35-37] used principal component analysis for data reduction, while we implemented factor analysis to identify underlying latent constructs [38]. Another issue with previous analyses of the eHEALS is that the Likert scales were treated as continuous variables. In our analyses, we considered the ordinal structure of the eHEALS items using a polychoric correlation matrix when performing the EFA. Our results show the predominance of a single factor across both paper and online samples. Importantly, we also found no significant difference in the eHEALS total sum scores between the paper and online sampling frames. Unidimensionality assumptions were further tested using IRT analyses.

Given the construction and theorized application of eHEALS as a broad construct, our IRT analyses included a bifactor model. Consistent with previous IRT analyses, our findings indicated unidimensionality of the eHEALS [9,17,39]. Our findings also agree with the sampling-measurement interrelationship, such that our difficulty parameters were lower in our relatively young sample than the average. Further evidence from the comparison of the bifactor modeling of the primary factor suggested that the distortion from ignoring the potential multidimensionality in our multimodal sample was negligible. Furthermore, the empirical reliability after accounting for the general factor was inadequate to support future subscale use. Our results are not directly comparable to those from the original study by Soellner et al [16] in a German sample, since their results were not based on the IRT methodology. There is some agreement between the findings, given that two group factors were a better fit for the bifactor model than three group factors. These findings suggest two future applications for the German eHEALS. First, researchers may wish to pilot new items to expand subscales to achieve sufficient reliability. Second, researchers may wish to use a subset of items for purely unidimensional purposes. In the latter case, we refer readers to a potential core subset that could be comprised of the 5 items in Table 4 with item estimated common variances >.80. These are reported in Multimedia Appendix 2 with a preliminary differential item functioning analysis that indicated no bias of the items across paper and online samples.

In summary, the results support the broad but unidimensional factor structure of the German eHEALS. Our ordinal factor analysis supports the presence of a strong general factor. Furthermore, the item response theory analysis using bifactor models with one general factor and two or three group factors showed that the model with two group factors fitted better than the one with three factors. Comparing this bifactor model with two group factors with the unidimensional loadings did not suggest a substantial difference in primary loadings. Finally, we found no support for using eHEALS subscales.

The use of subscales in previous research [16,20,40] may underpin the recommendation to have 3 to 5 items per common factor. Although an advantage of the eHEALS might be its short length (only 8 items) from a methodological viewpoint, the factors may be underdetermined if the items were split into subscales [38]. Still, the benefit of different subdimensions of eHealth literacy would lie in the ability to identify possible areas of intervention. Conceptual models on general health literacy include the components of understanding, appraising, and applying health information [41]. For example, item 8 of the eHEALS “I feel confident in using information from the internet to make health decisions” could reflect the dimension of applying health information. As others have already indicated [22], to add further items to help discriminate components of eHealth literacy, we require a better understanding of the concepts study participants associate with particular eHEALS items. However, given that past findings have indicated a mixed number of factors underlying the eHEALS, researchers should carefully consider the overspecification value relative to the general stability. In this study, the unidimensional bifactor model of the eHEALS was stable across the two distinct samples (online and paper modes).

Limitations

This study has some limitations. Although the parents were asked about their own eHealth literacy, it is likely that the questions on child health prompted the parents to answer the eHEALS items from the perspective of child health. This would explain the parents’ reluctance to make decisions based on internet-based health information (item 8 of the eHEALS). Therefore, in comparison with studies on adult eHealth literacy, parental eHealth literacy might be lower. However, the high information needs of parents, especially right after birth, might have increased eHealth literacy simply through practice and experience. Regarding the sample characteristics, the generalizability of our findings might be limited by the fairly low response rate (769/842, 30%) and the uniquely high socioeconomic status.

Another limitation is that measurement invariance was not assessed in terms of the participants’ individual characteristics. For example, future studies could focus specifically on the generalizability by gender of our proposed unidimensional eHEALS. This was not the focus of the current study and will be addressed in further analyses. It is important, furthermore, for future researchers to consider the relevance of their samples when studying eHEALS measurement properties. This study aimed to extend the application of the eHEALS among new parents.

Conclusions

This study suggests that the German eHEALS possesses a broad, unidimensional factor structure among Swiss-German parents. Although the two samples differed with respect to participant characteristics such as age, education, and income, we failed to find a significant difference in the eHEALS total sum scores. The underrepresentation of participants of lower socioeconomic status, not only in our study but also in many other studies on digital health, warrants future studies to over-sample this population. We found similar factor structures and item properties irrespective of application mode. That is, the EFAs suggested a strong general factor. Finally, bifactor modeling did not outperform the unidimensional model, and subscales were unsupported because of low reliability. While using the total sum score is appropriate to assess eHealth literacy, further development and refinement of the eHEALS are proposed to address specific sub-domains of eHealth literacy. For any sample, practitioners should use only the eHEALS total score, and future research aiming to utilize subscales should expand the eHEALS item pool for empirical testing.

Acknowledgments

The authors thank the municipalities who provided the parents' addresses, all participating parents for their valuable engagement, and our colleague Dominik Robin for his support in study design. This study was funded by the Käthe Zingg-Schwichtenberg Fund (KZS) of the Swiss Academy of Medical Sciences (SAMS). The authors of this manuscript are independent of the funding agencies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed protocol of the exploratory factor analysis with polychoric correlation matrix.
[DOCX File , 28 KB - jmir_v22i3e14492_app1.docx]

Multimedia Appendix 2
Potential core set of the German eHEALS and differential item functioning analysis (DIF).
[DOCX File , 20 KB - jmir_v22i3e14492_app2.docx]

References

1. Plantin L, Daneback K. Parenthood, information and support on the internet. A literature review of research on parents and professionals online. BMC Fam Pract 2009 May 18;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
2. Skranes LP, Løhaugen GCC, Botngård A, Skranes J. Internet use among mothers of young children in Norway—a survey of Internet habits and perceived parental competence when caring for a sick child. J Public Health 2014 Jul 2;22(5):423-431. [doi: 10.1007/s10389-014-0631-x]
3. Tuffrey C, Finlay F. Use of the internet by parents of paediatric outpatients. Arch Dis Child 2002 Dec;87(6):534-536 [FREE Full text] [doi: 10.1136/adc.87.6.534] [Medline: 12456558]
4. Wainstein BK, Sterling-Levis K, Baker SA, Taitz J, Brydon M. Use of the Internet by parents of paediatric patients. J Paediatr Child Health 2006 Sep;42(9):528-532. [doi: 10.1111/j.1440-1754.2006.00916.x] [Medline: 16925539]
5. Scullard P, Peacock C, Davies P. Googling children's health: reliability of medical advice on the internet. Arch Dis Child 2010 Aug;95(8):580-582. [doi: 10.1136/adc.2009.168856] [Medline: 20371593]
6. Barkin JL, Jani S. Information Management in New Motherhood: Does the Internet Help or Hinder? J Am Psychiatr Nurses Assoc 2016 Nov;22(6):475-482. [doi: 10.1177/1078390316659697] [Medline: 27519615]
7. DeLuca JM, Kearney MH, Norton SA, Arnold GL. Internet use by parents of infants with positive newborn screens. J Inherit Metab Dis 2012 Sep;35(5):879-884. [doi: 10.1007/s10545-011-9449-7] [Medline: 22297410]
8. van der Gugten AC, de Leeuw RJRJ, Verheij TJM, van der Ent CK, Kars MC. E-health and health care behaviour of parents of young children: a qualitative study. Scand J Prim Health Care 2016 Jun;34(2):135-142 [FREE Full text] [doi: 10.3109/02813432.2016.1160627] [Medline: 27063729]
9. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
10. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep 29;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
11. Knapp C, Madden V, Marcu M, Wang H, Curtis C, Sloyer P, et al. Information seeking behaviors of parents whose children have life-threatening illnesses. Pediatr Blood Cancer 2011 May;56(5):805-811. [doi: 10.1002/pbc.22674] [Medline: 21370415]
12. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
13. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowledge Management & E-Learning 2015;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
14. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
15. Neter E, Brainin E. Perceived and Performed eHealth Literacy: Survey and Simulated Performance Test. JMIR Hum Factors 2017 Jan 17;4(1):e2 [FREE Full text] [doi: 10.2196/humanfactors.6523] [Medline: 28096068]

16. Soellner R, Huber S, Reder M. The Concept of eHealth Literacy and Its Measurement. Journal of Media Psychology 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
17. Diviani N, Dima AL, Schulz PJ. A Psychometric Analysis of the Italian Version of the eHealth Literacy Scale Using Item Response and Classical Test Theory Methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
18. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
19. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. European Health Psychologist 2015;17(6):275-280.
20. Hyde LL, Boyes AW, Evans T, Mackenzie LJ, Sanson-Fisher R. Three-Factor Structure of the eHealth Literacy Scale Among Magnetic Resonance Imaging and Computed Tomography Outpatients: A Confirmatory Factor Analysis. JMIR Hum Factors 2018 Feb 19;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
21. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
22. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and Validity of the Telephone-Based eHealth Literacy Scale Among Older Adults: Cross-Sectional Survey. J Med Internet Res 2017 Dec 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
23. Kerry MJ, Wang R, Bai J. Assessment of the Readiness for Interprofessional Learning Scale (RIPLS): An item response theory analysis. J Interprof Care 2018 Sep;32(5):634-637. [doi: 10.1080/13561820.2018.1459515] [Medline: 29648892]
24. Jaks R, Baumann I, Juvalta S, Dratva J. Parental digital health information seeking behavior in Switzerland: a cross-sectional study. BMC Public Health 2019 Feb 21;19(1):225-235. [doi: 10.1186/s12889-019-6524-8]
25. Hollander M, Wolfe D, Chicken E. Nonparametric Statistical Methods. In: John Wiley & Sons. Hoboken, New Jersey. USA: John Wiley & Sons, Inc; 2013:978-971.
26. Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, Vila–Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Qual Quant 2008 Sep 7;44(1):153-166. [doi: 10.1007/s11135-008-9190-y] [Medline: 30311153]
27. Revelle W. CRAN.R-project. psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois, USA; 2018. Procedures for Personality and Psychological Research URL: https://CRAN.R-project.org/package=psych
28. O’Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer’s MAP test. Behavior Research Methods, Instruments, & Computers 2000 Sep;32(3):396-402. [doi: 10.3758/BF03200807]
29. McDonald R. Test theory: A unified treatment. Mahwah, NJ. US: Lawrence Erlbaum Associates Publishers; 1999.
30. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Qual Life Res 2007 May 4;16(S1):19-31. [doi: 10.1007/s11136-007-9183-7] [Medline: 17479357]
31. DeMars CE. A Tutorial on Interpreting Bifactor Model Scores. International Journal of Testing 2013 Oct;13(4):354-378. [doi: 10.1080/15305058.2013.799067] [Medline: 26281194]
32. Dueber D. Educational, School, and Counseling Psychology Research Tools. 2017. Bifactor Indices Calculator: A Microsoft Excel-Based Tool to Calculate Various Indices Relevant to Bifactor CFA Models URL: https://doi.org/10.13023/edp.tool.01
33. Embretson S, Reise S. Item Response Theory for Psychologists. Mahwah, New Jersey: Psychology Press; 2000.
34. Muthén B, Kaplan D, Hollis M. On structural equation modeling with data that are not missing completely at random. Psychometrika 1987 Sep;52(3):431-462. [doi: 10.1007/BF02294365]
35. Paramio Pérez G, Almagro BJ, Hernando Gómez Á, Aguaded Gómez JI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish University Students]. Rev Esp Salud Publica 2015 Jun;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
36. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. Developing Japanese version of the eHealth Literacy Scale (eHEALS). Nihon Koshu Eisei Zasshi 2011 May;58? 2011:71.
37. Del Giudice P, Bravo G, Poletto M, De Odorico A, Conte A, Brunelli L, et al. Correlation Between eHealth Literacy and Health Literacy Using the eHealth Literacy Scale and Real-Life Experiences in the Health Sector as a Proxy Measure of Functional Health Literacy: Cross-Sectional Web-Based Survey. J Med Internet Res 2018 Oct 31;20(10):e281 [FREE Full text] [doi: 10.2196/jmir.9401] [Medline: 30381283]
38. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods 1999 Sep;4(3):272-299. [doi: 10.1037/1082-989x.4.3.272]
39. Nguyen J, Moorhouse M, Curbow B, Christie J, Walsh-Childers K, Islam S. Construct Validity of the eHealth Literacy Scale (eHEALS) Among Two Adult Populations: A Rasch Analysis. JMIR Public Health Surveill 2016 May 20;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]

40. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic Health Literacy Across the Lifespan: Measurement Invariance Study. J Med Internet Res 2018 Jul 09;20(7):e10434. [doi: 10.2196/10434]
41. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]

Abbreviations

EFA: exploratory factor analysis.
eHEALS: eHealth Literacy Scale.
IRT: item response theory.

Edited by G Eysenbach; submitted 16.07.19; peer-reviewed by E Neter, M Fiordelli; comments to author 06.09.19; revised version received 22.10.19; accepted 12.11.19; published 13.03.20.

Please cite as:
Juvalta S, Kerry MJ, Jaks R, Baumann I, Dratva J
Electronic Health Literacy in Swiss-German Parents: Cross-Sectional Study of eHealth Literacy Scale Unidimensionality
J Med Internet Res 2020;22(3):e14492
URL: https://www.jmir.org/2020/3/e14492
doi:10.2196/14492
PMID:32167476

Original Paper

Improvement of the Efficiency and Completeness of Neuro-Oncology Patient Referrals to a Tertiary Center Through the Implementation of an Electronic Referral System: Retrospective Cohort Study

Rocío Fernández-Méndez[1,2], BSc, BA, MPH, PhD; Mei Yin Wong[1], BA, MB, BChir; Rebecca J Rastall[1], BSc, MSc; Samuel Rebollo-Díaz[3], BSc, MSc; Ingela Oberg[4], BSc, MSt; Stephen J Price[1,2,4], BSc, MBBS, FRCS, PhD; Alexis J Joannides[1,2,3,4], MA, MB, BChir, FRCS, PhD

[1]Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom

[2]NIHR Brain Injury MedTech Co-operative, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom

[3]Obex Technologies, Cambridge, United Kingdom

[4]Neurosurgery Department, Cambridge University Hospitals NHS Trust, Cambridge, United Kingdom

Corresponding Author:
Rocío Fernández-Méndez, BSc, BA, MPH, PhD
Department of Clinical Neurosciences
University of Cambridge
Box 165, A Block, Level 3, B Spur
Cambridge Biomedical Campus
Cambridge, CB2 0QQ
United Kingdom
Phone: 44 1223746466
Email: rociofmendez.inv@gmail.com

Abstract

Background: Quality referrals to specialist care are key for prompt, optimal decisions about the management of patients with brain tumors.

Objective: This study aimed to determine the impact of introducing a Web-based, electronic referral (eReferral) system to a specialized neuro-oncology center, using a service-developed proforma, in terms of waiting times and information completeness.

Methods: We carried out a retrospective cohort study based on the review of medical records of referred adult patients, excluding follow-ups. Primary outcome measures were durations of three key phases within the referral pathway and completion rates of six referral fields.

Results: A total of 248 patients were referred to the specialist center during the study period. Median (IQR) diagnostic imaging to referral intervals were 3 (1-5) days with eReferrals, and 9 (4-19), 19 (14-49), and 8 (4-23) days with paper proforma, paper letter, and internal referrals, respectively ($P<.001$). Median (IQR) referral to multidisciplinary team decision intervals were 3 (2-7), 2 (1-3), 8 (2-24), and 3 (2-6) days respectively ($P=.01$). For patients having surgery, median (IQR) diagnostic imaging to surgery intervals were 28 (21-41), 34 (27-51), 104 (69-143), and 32 (15-89) days, respectively ($P<.001$). Proportions of complete fields differed significantly by referral type in all study fields (all with *P*s <.001) except for details of presentation, which were present in all referrals. All study fields were always present in eReferrals, as these are compulsory for referral submission. Depending on the data field, level of completeness in the remaining referral types ranged within 69% (65/94) to 87% (82/94), 15% (3/20) to 65% (13/20), and 22% (8/41) to 63% (26/41) in paper proforma, paper letter, and internal referrals, respectively.

Conclusions: An electronic, Web-based, service-developed specific proforma for neuro-oncology referrals performs significantly better, with shorter waiting times and greater completeness of information than other referral types. A wider application of eReferrals is an important first step to streamlining specialist care pathways and providing excellent care.

International Registered Report Identifier (IRRID): RR2-10.2196/10.2196/15002

(J Med Internet Res 2020;22(3):e15002) doi:10.2196/15002

KEYWORDS

quality improvement; electronic health records; hospital referral; hospital oncology services

Introduction

Background

Patients with suspected malignancies of the central nervous system (CNS) should be referred promptly to specialized neurosurgery centers following their initial diagnosis for deciding subsequent management [1]. The limitations of traditional referral methods, including verbal communications and referral letters, can potentially decrease the safety and efficiency of the referral process. Verbal communications can lead to low information retention by the receiver, and hand-written referral letters can be limited by illegibility, unreliable transmission, missing documentation, or incomplete information required to make appropriate decisions [2-4]. Although there are well-established referral pathways for elective and emergency referrals, there are less defined routes for nonelective referrals, especially when these arise from the emergency setting, which is the case for the majority of brain tumor presentations [5]. Moreover, patients with brain tumors and their carers often perceive that their referrals were delayed or they were not seen by the appropriate specialist in the first place [6,7].

Among the most effective proposed strategies to overcome the limitations of traditional referral methods are structured referral sheets or proformas [3,4]. Web-based integrated specialty electronic referrals (eReferrals) are a digital version of a referral system [8], through which referrals are sent securely in real time through the internet, making them instantly available to authorized users. These systems can make the process faster, safer, and easier to follow up, regardless of the health care organization or location of the referrer and receiver. This is particularly relevant in the context of referrals from emergency settings, where there is a high staff turnover and limited time available for nonemergency patient management.

Objectives

Since April 2016, the neurosurgery unit of 1 of the 16 regional neuroscience centers in England introduced an eReferral system for the referral of its neuro-oncology surgery adult patients for multidisciplinary team (MDT) discussion, as part of a digital quality improvement (QI) project. The aim of this study was to determine the impact of introducing this eReferral system, in terms of reduction of waiting times, that is, improvement of time efficiency, and completeness of the information provided within the referrals.

Methods

Neuro-Oncology Referrals and Setting

The study center provides neuro-oncology surgery specialist services for a core population of 3 million plus an extended catchment area of 2.6 million. Patients referred to the study center are adult patients aged 16 years and older, who have a suspected CNS tumor that could benefit from surgery, excluding skull base and spinal tumors (normally managed by a different MDT). Each referral is discussed during weekly MDT meetings, and surgically eligible patients are seen by the neuro-oncology surgery team in the next available clinic. The number of adult patients discussed during MDT meetings ranges within 80 to 120 per month, and approximately 10 to 25 are operated on monthly.

Before 2016, the standard route for nonelective, scheduled referrals to the MDT was fax, email, or post using unstructured letters or structured proformas (jointly referred to as paper referrals). Since April 2016, an eReferral system was introduced to replace paper referrals. The eReferral form was developed with iterative multidisciplinary input, based on its preceding paper proforma and in line with national guidelines [1,9]. The form includes mainly mandatory closed fields and is hosted at the Outcome Registry Intervention and Operation Network (ORION), a secure, Web-based platform, for managing health care data in multiple institutions (Multimedia Appendices 1 and 2; Obex Technologies Ltd, Cambridge). Some fields include a predefined list of possible answers and others are contingent on answers to previous fields. eReferrals can only be submitted when all mandatory fields are completed, and after submission, they are automatically updated in the system and made available to the MDT office in real time. This feature helps to easily determine referral urgency in a timely manner and suitability for the neuro-oncology surgery MDT. Contact details of the referring teams are easily accessible in the eReferrals, making it straightforward to communicate about any outstanding investigations required before the MDT. Internal referrals within the study neuroscience center are mostly done using the electronic hospital records (EHRs) system directly.

Therefore, there were four referral types included in this study: eReferrals via ORION, paper-based referrals using the proforma or a free-text letter, and direct referrals through the EHR. The study period selected included 2 separate months, April and September, to account for seasonal variation of 2 consecutive years, 2016, when the QI project started, and 2017, when the QI project had been in place for over 1 year.

Study Design and Data Collection

This was a retrospective, cohort study based on the review of different types of medical records. All new adult patients discussed at all weekly MDT meetings during the study period were eligible. New patients included those who were seen for a new condition or a change in a previous pathology. Follow-up cases were excluded, as the discussion for these was usually arranged through internal processes rather than as a new referral.

Patients were identified from the MDT list and their National Health Service (NHS) or hospital numbers were used to link data from different data sources: ORION, EHR, and a national Picture Archiving and Communication System (PACS; Multimedia Appendix 3). ORION data were downloaded automatically from the platform. EHR and PACS data were manually extracted by 2 auditors using a prespecified data collection form. Data quality checks were performed by a different auditor by rechecking data gathered against the original records.

The number and nature of calls or emails sent by referrers to the ORION software team about issues related to the use of the referral portal were obtained from application support logs provided by the software support team.

Data and Variables Definitions

Data collected included demographics and clinical data including performance status (PS) on referral, initial diagnosis at MDT, and eventual diagnosis following surgery or further investigations (if available); whether key data fields were included in the referral; and key dates within the referral pathway.

PS was based on the Eastern Cooperative Oncology Group/World Health Organization (WHO) PS levels [10]. Diagnoses were grouped depending on the type of tumor, gliomas being grouped into high-grade gliomas (HGGs; WHO grade III-IV) and low-grade gliomas (LGGs; WHO grade I-II). The six fields analyzed for completion rates were agreed by the service MDT as fields that should be included in an ideal referral form or letter based on national guidelines, and included PS, details of presentation, symptom duration, steroid treatment, previous malignancy, and staging computerized tomography (CT) [1]. Key dates extracted included date of earliest diagnostic imaging, defined as the date of the earliest CT or magnetic resonance imaging (MRI) scan available within the 6 months preceding the referral, referral date, and date of MDT decision, defined as the last MDT discussion. Durations of three key phases within the referral pathway were calculated using those dates: earliest diagnostic imaging to referral interval, referral to decision interval, and decision to surgery interval (if applicable).

Analyses

Data were summarized using relevant descriptive statistics. Continuous variables were compared by referral type, including subanalyses by diagnostic group, using the Kruskal-Wallis equality-of-populations rank test. Pearson chi-squared or Fisher exact test, as appropriate, were used for the comparison of proportions by referral type. Statistical significance was set at the 5% level.

Referral Epidemiology

The incidence of average monthly referrals was calculated using population estimates for mid-2016 from the Office for National Statistics [11]. A line chart was used to represent the evolution of the number of referrals per referral type over time. A colored map was created to depict differences in the average incidence of monthly referrals by subregion.

Referral Efficiency and Completeness

Durations of the three key phases of interest within the referral pathway and completion rates of each of the six key data fields under investigation were compared by referral type using the relevant statistical test.

Missing Data and Outliers

Entry errors were corrected during the data entry quality checks. Missing data were explored for possible trends but not imputed. Outliers were explored in depth, including potential data entry error identification.

This was a nonresearch, digital service QI project, registered as a service improvement program with the Institutional Clinical Audit Department, with project number PRN7723.

Results

Sample

There were 248 patients referred during the 4 months of the study period, with monthly numbers of patients discussed at each MDT meeting ranging from 53 to 72. Median (IQR) age was 66.4 (51.3-73.7) years, and the female-to-male ratio was 5:4. Most patients with reported PS had good WHO-PS levels of 0 (94/176, 53.4%) or 1 (48/176, 27.3%; Table 1).

Most common MDT diagnostic groups were meningioma (49/245, 20.0%), metastasis (48/245, 19.6%), and HGG (47/245, 19.2%). In 14.3% (35/245) patients, the diagnosis given at the MDT meeting was *lesion* or *mass*. Eventual diagnosis was available for 32.7% (80/245) patients, of which 95% (75/79) followed a surgical or biopsy procedure. The distribution of these eventual diagnostic groups differed from that of initial MDT diagnostic groups (Table 1), and for some patients, eventual diagnosis differed from their initial MDT diagnosis (Table 2).

Differences between the four referral systems in demographic and clinical data were not statistically significant (Table 1).

Twelve percent (29/248) patients required one or more MDT rediscussions, with no statistically significant differences by referral source (P=.53). The main reasons for rediscussion were a request for further investigations locally (25, 86%).

Table 1. Characteristics of patients referred to the study neuroscience centers by referral type excluding unknown and unusual referral pathways.

Characteristics	Referral type[a]				*P* value[b]
	Electronic	Paper proforma	Paper letter	Internal	
Patients, n	83	94	20	41	
Age (years), median (IQR)	70 (56-76)	64 (52-72)	66 (36-76)	58 (47-72)	.15
Female:male ratio	2:2	3:2	4:2	1:3	.72
WHO[c] PS[d] on referral, n (%)[e]					**.51**
PS-0	42 (51)	42 (45)	5 (25)	4 (10)	
PS-1	24 (29)	18 (19)	0 (0)	4 (10)	
PS-2	8 (10)	9 (10)	1 (5)	1 (2)	
PS-3	5 (6)	6 (6)	1 (5)	0 (0)	
PS-4	4 (5)	0 (0)	0 (0)	0 (0)	
Not stated	0 (0)	19 (20)	13 (65)	32 (78)	
Multidisciplinary team meeting diagnosis, n (%)[f]					**.09**
Meningioma	16 (19)	19 (20)	4 (20)	8 (20)	
HGG[g] (WHO III-IV)	23 (28)	16 (17)	3 (15)	3 (7)	
Metastasis	10 (12)	20 (21)	1 (5)	14 (34)	
Lesion or mass[h]	14 (17)	14 (15)	3 (15)	3 (7)	
Other CNS[i] tumor[j]	5 (6)	8 (9)	2 (10)	7 (17)	
Other nontumor lesion[k]	6 (7)	4 (4)	2 (10)	3 (7)	
LGG (WHO I-II)[l]	1 (1)	6 (6)	2 (10)	2 (5)	
Lymphoma	4 (5)	2 (2)	0 (0)	0 (0)	
Spinal tumor	1 (1)	1 (1)	0 (0)	0 (0)	
Unknown or unclear	1 (1)	4 (4)	2 (10)	0 (0)	
Missing/not stated	2 (2)	0 (0)	1 (5)	1 (2)	
Neuro-oncology surgery	34 (41)	37 (39)	9 (45)	13 (32)	.72
Eventual diagnosis after surgery or further investigations, n (%)					**.33**
HGG (WHO III-IV)	14 (41)	13 (35)	2 (22)	2 (15)	
Meningioma	2 (6)	3 (8)	2 (22)	4 (31)	
LGG (WHO I-II)	4 (12)	4 (11)	1 (11)	1 (8)	
Metastasis	1 (3)	5 (14)	1 (11)	3 (23)	
Other CNS tumor	2 (6)	2 (5)	0 (0)	0 (0)	
Lymphoma	0 (0)	1 (3)	0 (0)	1 (8)	
Non-CNS tumor	0 (0)	0 (0)	0 (0)	1 (8)	
Unknown or unclear	1 (3)	1 (3)	0 (0)	0 (0)	
Missing or not stated	10 (29)	8 (22)	3 (33)	1 (8)	

[a] 7 referrals initially done as emergency neurosurgery referrals and 3 cases with unknown referral source were excluded.

[b] Excluding cases with unknown or missing values.

[c] WHO: World Health Organization.

[d] PS: performance status.

[e] Eastern Cooperative Oncology Group/WHO performance status levels.

[f] Multidisciplinary team diagnosis: initial/suspected diagnosis following multidisciplinary team meeting discussion.

[g] HGG: high-grade glioma.

[h] Diagnoses included *lesion/s* or *mass*, with or without location mentioned.

[i] CNS: central nervous system.

[j] Diagnoses included *tumor*, *neoplasm*, *glioma*, *ependymoma*, *dermoid*, *craniopharyngioma*, *medulloblastoma*, or *schwannoma*.

[k] Diagnoses included *cyst*, *demyelination*, *bleed*, *infarct*, *abscess*, *inflammatory*, *fungal*, and *herpes*.

[l] LGG: low-grade glioma.

Table 2. Distribution (number and percentage) of eventual diagnostic groups after surgery in each initial multidisciplinary team meeting diagnostic group among patients who had surgery.

Initial multidisciplinary team diagnosis	Eventual diagnosis, n (%)								
	High-grade glioma	Low-grade glioma	Lymph	Menin-gioma	Metastasis	Other central nervous system tumor	Noncentral nervous system tumor	Unknown or unclear	Missing
High-grade glioma	*20 (65)* [a]	3 (10)	0 (0)	0 (0)	2 (6)	0 (0)	0 (0)	0 (0)	6 (19)
Low-grade glioma	1 (14)	*5 (71)*	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (14)
Lymph	3 (75)	0 (0)	*1 (25)*	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Meningioma	0 (0)	0 (0)	0 (0)	*9 (60)*	0 (0)	1 (7)	0 (0)	0 (0)	5 (33)
Metastasis	0 (0)	1 (10)	1 (10)	0 (0)	*6 (60)*	0 (0)	0 (0)	1 (10)	1 (10)
Other central nervous system tumor	3 (43)	0 (0)	1 (14)	0 (0)	0 (0)	*1 (14)*	0 (0)	0 (0)	2 (29)
Other nontumor lesion	1 (25)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	*1 (25)*	0 (0)	2 (50)
Lesion or mass	5 (28)	1 (6)	0 (0)	2 (11)	2 (11)	2 (11)	0 (0)	*1 (6)*	5 (27)
Unknown or unclear	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	*1 (100)*

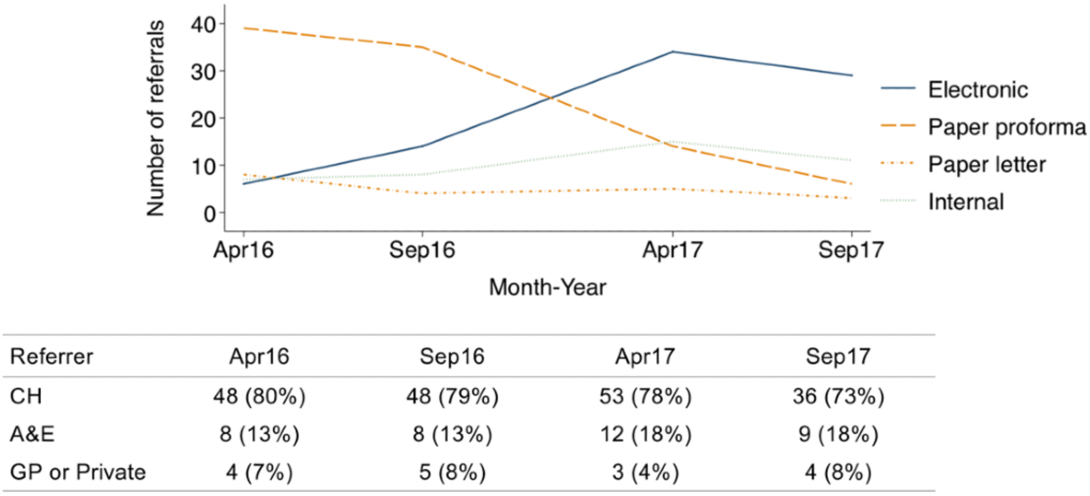
[a] Italicized figures represent cases whose final diagnosis was the same as the initial MDT suspected diagnosis.

Referral Epidemiology

During the study period, the use of eReferrals increased since their introduction from 6 (10%) to 29 (55%) referrals per month, whereas monthly paper referrals, with or without proforma, decreased from 47 (77%) to 9 (17%). Internal referrals remained constantly lower than eReferrals, with monthly referrals ranging from 7 (12%) to 15 (21%; Figure 1). In addition, 7 patients had nonstandard referrals (initially referred to as neurosurgical emergencies and usually followed by an appropriate electronic neuro-oncology referral), and 3 patients had an unknown referral source. Most referring centers were district general hospitals (193, 77.8%), followed by referrals within the study neuroscience center (38, 15.3%) and general practice (GP) and private centers (17, 6.9%). These percentages were very similar over time (*P*=.87; Figure 1).

Most referrals (222, 89.5%) were of patients residing within the catchment area of the study center. Within this catchment area, the average incidence of monthly referrals ranged from 0.3 (from North Essex) to 0.6 (from Cambridgeshire) referrals per 100,000 person-months (Figure 2).

Figure 1. Number of monthly referrals per referral type (top line chart) and distribution of monthly referrals by referrer (bottom table) overtime, in April and September of 2016 and 2017. Patients with nonstandard (initially done as emergency neurosurgery referrals) or unknown referral source are not plotted (n=10). Internal refers to intrahospital referrals within the study center using the hospital's electronic records system. CH: community hospital; A&E: accidents and emergencies department of the study center; GP: general practice.



Referrer	Apr16	Sep16	Apr17	Sep17
CH	48 (80%)	48 (79%)	53 (78%)	36 (73%)
A&E	8 (13%)	8 (13%)	12 (18%)	9 (18%)
GP or Private	4 (7%)	5 (8%)	3 (4%)	4 (8%)

Figure 2. Average incidence of monthly referrals (number of referrals per 100,000 person-month) by subregion in April and September of 2016 and 2017.

Date of diagnostic imaging was missing in 36 (15%) patients, from all referral types, and for 4 (2%) patients the only available imaging occurred between referral and MDT discussion. Pre-MDT diagnostic imaging types included CT scan (125, 61.0%), MRI (72, 35.1%), or both CT and MRI scans on the same day (8, 3.9%). By referral type, MRI scans (with or without CT on the same day) were more common in paper letter referrals (12, 67%) than in the remaining referral types, including eReferrals (24, 34%), paper proforma (32, 42%), and internal referrals (12, 30%; *P*=.05).

Referral Quality

Referral Efficiency

eReferrals had the shortest diagnostic imaging to-referral intervals and were, on average, 6, 16, and 5 days quicker than paper proforma, paper letter, and internal referrals, respectively (Figure 3). There was a wide variability in these intervals in all referral types with the exception of eReferrals, which were more consistent (Figure 3). In 2 patients (1 with eReferral and 1 with paper proforma referral) the interval was over 6 months, so they were considered as cases whose first diagnostic scan was missing. In all, 11 patients, from all referral types, had unexpectedly large imaging to referral intervals of 2 to 6 months. MDT diagnoses for these were mostly meningiomas (n=6, 54%) or lesion or mass (n=2, 18%), and there was 1 patient with HGG. The latter was a patient referred with a paper letter whose earliest imaging was followed by a second scan about 2 months later.

Referral to decision intervals were similar among eReferrals, paper proforma, and internal referrals, with median intervals of 2 to 3 days (Figure 3), whereas paper letter referrals were 5 to 6 days slower on average. Furthermore, 7 patients had unexpectedly large referral-to-decision intervals of over 1 month. Most of these had been referred with a paper letter (4/7, 57%) but also with paper proforma or eReferral. Diagnoses for these patients were mainly lesion or mass (3/7, 43%) or unknown, unclear, or missing (2/7, 29%), none were HGG or metastasis and most were cases that required further information or investigations followed by a rediscussion (5/7, 71%).

Among the 97 patients who would eventually have surgery, decision to surgery intervals did not differ significantly by referral type (Figure 3). In all, 5 patients, from all referral types, had unexpectedly large decision to surgery intervals of more than 3 months. Most of these had a diagnosis of meningioma (3, 60%), and none of them had HGG or metastasis.

A graphical representation of the sum of the median intervals at each referral pathway phase is provided in Figure 4.

Including all phases, diagnostic imaging to surgery intervals (where applicable) had median (IQR) values of 28 (21-41), 34 (27-51), 104 (69-143), and 32 (15-89) days among eReferrals, paper proforma, paper letter, and internal referrals, respectively (*P*<.001). In patients with HGGs, those values were 30 (21-41), 28 (21-33), and 37 (33-40) days in eReferrals, paper proforma, and paper letter referrals, respectively, and 85 days in the only internal referral of HGG (*P*=.32). There was no unexpectedly large diagnostic imaging to surgery interval of more than 6 months.

Subanalyses by diagnostic group (when methodologically plausible) showed that differences in diagnostic imaging to referral time remained statistically significant in all groups except meningiomas. Referral to decision time differences remained statistically significant among patients with HGGs. Subanalyses by group of month and year of MDT discussion, in each of the referral types, showed that there was no time trend in any of the time intervals of interest, which did not differ significantly by month-year group.

Figure 3. Duration of each phase of the referral pathway, during the 4 months of the study period (April and September of 2016 and 2017). Time from imaging to referral excludes 36 patients with missing date of diagnostic imaging, 4 patients with their imaging carried out after their referral, and 2 patients whose available imaging was older than 6 months before the referral. Values outside the whiskers (more than 1.5 times the IQR from the upper and lower quartiles, respectively) are not plotted. Patients with nonstandard (initially done as emergency neurosurgery referrals) or unknown referral source are not plotted (n=10). Internal refers to intrahospital referrals within the study center using the hospital's electronic records system.

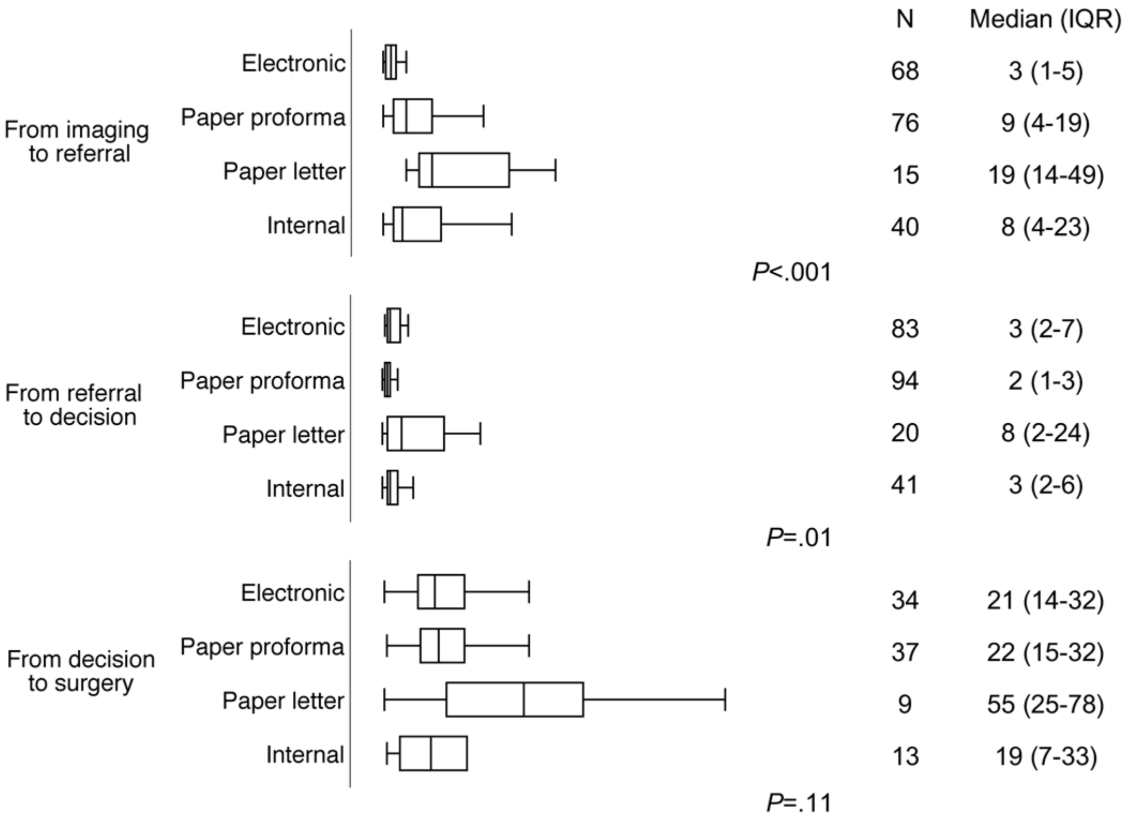
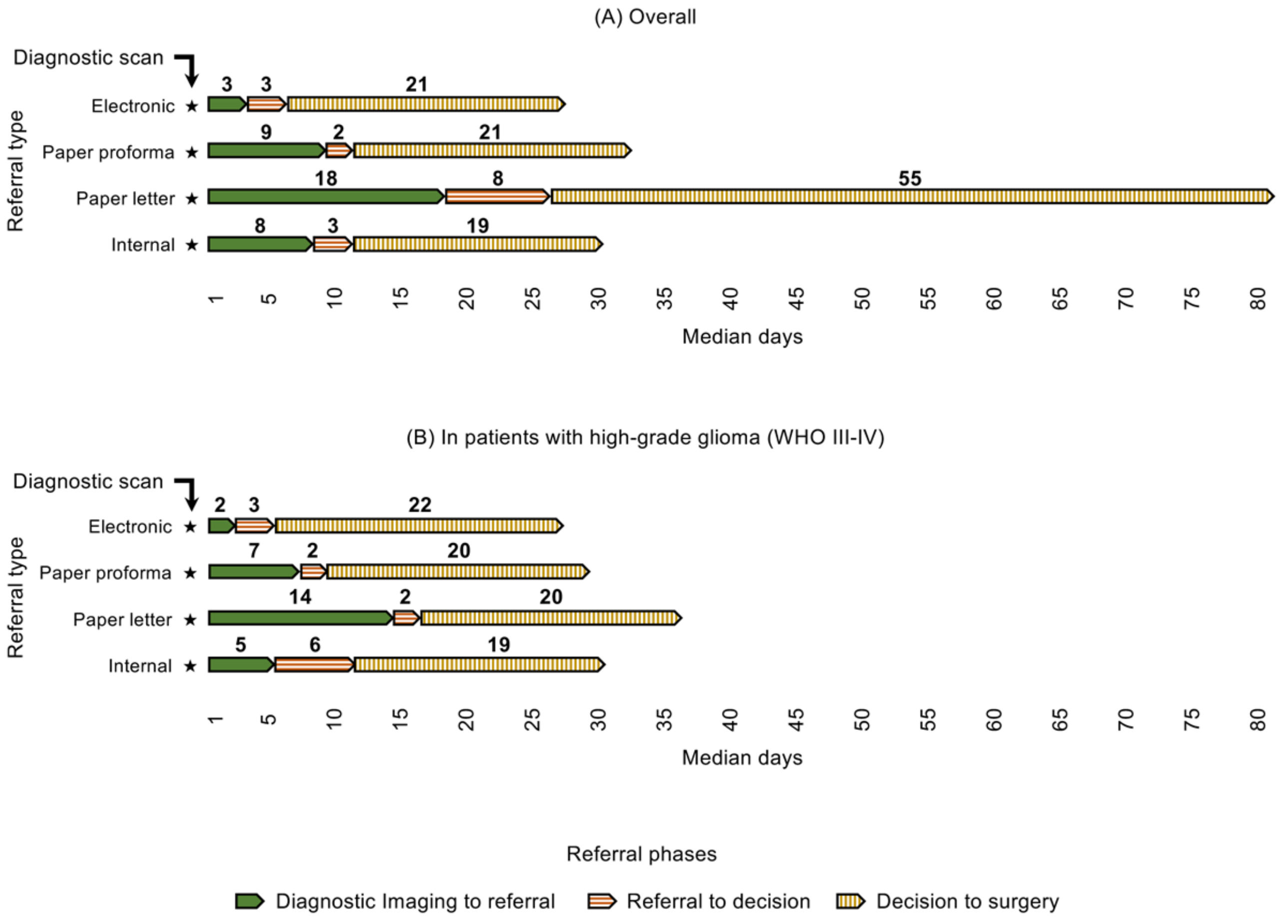


Figure 4. Median interval times of each referral phase, by referral type, during the 4 months of the study period (April and September of 2016 and 2017). Patients with nonstandard (initially done as emergency neurosurgery referrals) or unknown referral source are not plotted (n=10). (A) Including all patients; (B) including only patients with high-grade glioma, World Health Organization grading III-IV. WHO: World Health Organization.



Referral Completeness

The proportion of complete fields differed significantly by referral type in all fields of interest (all with *P*<.001) except for details of presentation, which were included in all referrals, regardless of the referral source (Table 3). As expected, eReferrals had all the study fields of interest completed for 100% (n=83) of the referrals, as these were compulsory for submission. In paper proforma referrals, the percentages of complete fields ranged within 69% (65/94) to 87% (82/94), depending on the field. In the remaining referral types, these percentages showed a broader variation, ranging from between 15% (3/20) and 22% (9/41) in the staging CT field to between 63% (26/41) and 65% (13/20) in the symptom duration field (Table 3).

When looking at each diagnostic group separately, the difference in the proportion of referrals reporting each field remained statistically significant for most fields (with a few exceptions): PS in all diagnostic groups except lymphoma; symptom duration in meningiomas, metastasis, and *other CNS tumors*; previous malignancy in metastasis and *other CNS tumor*s; and in LGGs, the only field with differing reporting proportions by referral type was symptom duration.

For the 10 patients with unusual or unknown referral pathways, the median (IQR) diagnostic imaging to referral, referral to decision, and decision to surgery intervals were 4 (0-12), 2 (1-2), and 5 (3-15) days, respectively. The level of referral completeness in this group varied by the specific field considered, ranging from 20% referrals with the staging CT recorded to 100% referrals with details of presentation reported.

Table 3. Number and percentage of fields recorded, by referral type, excluding unusual and unknown referral pathways.

Fields recorded	Referral type[a]			
	Electronic[b], n (%)	Paper proforma, n (%)	Paper letter, n (%)	Internal, n (%)
Performance status[c]	83 (100)	75 (80)	7 (35)	9 (22)
Details of presentation	83 (100)	94 (100)	20 (100)	41 (100)
Symptom duration[c]	83 (100)	71 (76)	13 (65)	26 (63)
Steroid treatment[c]	83 (100)	82 (87)	5 (25)	8 (20)
Previous malignancy[c]	83 (100)	74 (79)	7 (35)	23 (56)
Staging computerized tomography[c]	83 (100)	65 (69)	3 (15)	9 (22)

[a] 7 referrals initially done as emergency neurosurgery referrals and three cases with unknown referral source have bene excluded.

[b] All electronic referrals were 100% complete in all fields as completing these was compulsory for the referral submission.

[c] *P*<.001.

Electronic Referral Software Issues

In 2017, for a total of 675 neuro-oncology surgery eReferrals carried out in the study neuroscience center, the ORION software support team received 27 calls or emails from referrers about any issues related to the use of the software, representing a maximum of 4.1% referrals requiring some type of software support. The majority of them were resolved within less than 5 min (77%) or 5 to 10 min (17%), and the nature of these calls varied, the most common ones being issues with making a new or reactivating an existing account (40%) and user errors in data entering (27%).

Discussion

Principal Findings

Adequate decisions and management of patients referred to specialized care depend largely on timely and high-quality written communication between referrers and specialists, often the only means of communication between both parties [4]. Our study shows that eReferrals of new patients to a specialized neuro-oncology surgery service for multidisciplinary consideration were of significantly higher quality than remaining referral types in terms of their time efficiency and completeness of information provided, which has also been suggested by studies in other clinical settings [2,12-14].

Referral Efficiency

The time interval from diagnostic scan to a referral being received by the MDT team was 16 days shorter, on average, with eReferrals than with paper letters. This may reflect the lengthier and larger number of steps usually involved in the latter [14], whereas eReferrals in our study may just involve 2 to 4 steps, including opening ORION's platform, creating a referrer ID (first users only), and completing and submitting the Web-based referral form. Moreover, in 2017, only about 4% of the eReferrals required some type of software support by referrers, indicating that the system was easy to use for the majority of users. In addition, the eReferral form in our study has mainly closed fields that can be answered with a single click (Multimedia Appendix 2), making its completion quicker and more straight forward than free-text letters. Furthermore, the type of diagnostic scan triggering the referral should have not affected these differences, as MRI scans, which are more specific than CT scans, were actually used more often in paper referrals than in the other types of referral. Similarly, Chen et al found that their eReferral system to different types of specialties halved the average waiting time for an initial consultative visit within 1 year, resulting in a safer and more time-efficient service [8]. A national report from Denmark found eReferrals to be faster to make and process, resulting in 15% to 30% cost savings [14].

As a form of paper referral, paper proforma referrals were also slower than eReferrals. Nonetheless, they were considerably faster than paper letter referrals, showing the usefulness of proformas to not only improve referral completeness but also to speed up the referral process.

The slower referral pathway of intra-hospital internal referrals as compared with eReferrals could be explained by cases being initially admitted under neurosurgery and subsequently referred to the MDT.

Some of the differences in the referral intervals may have been affected by more patients having more urgent diagnoses among eReferrals and internal referrals. Although not statistically significant, the proportions of patients with HGG and metastasis were higher among eReferrals and internal referrals, respectively, and the only 4 patients with WHO-PS of 4 were referred with the eReferral system.

Referral to decision intervals were also significantly longer among paper referrals, followed by eReferrals and internal referrals. This may be related to the lower completion rates in paper letter referrals, which could have made decisions more difficult.

In patients who had surgery, overall diagnostic imaging to surgery intervals were similar in all referral types, except for paper letters, where these intervals were about 3 to 4 times higher. In patients with HGGs who had surgery, paper letters were still the most inefficient, with median diagnostic imaging to surgery intervals being 7 to 9 days longer with this method. These differences identified can be a great concern in terms of the effectiveness and safety of the surgical management, especially of patients with the most aggressive and fast-growing CNS tumors. For patients with glioblastoma multiforme, the commonest HGG, tumor growth rates have been estimated at median values of 1.4% per day between the first diagnostic scan and the presurgical scan [15]. This means that, while patients with glioblastoma wait to be operated, their lesion could increase about 10% to 13% more if they are referred with traditional paper letters than if they are referred with the other referral types. Conversely, once a final MDT decision was made, the decision to surgery interval did not differ significantly by referral type, including patients with HGGs where decision to surgery intervals were 19 to 22 days in all referral types (Figure 4). This indicates that once a suspected diagnosis and potential management plan are proposed, the referral type would have not affected waiting times, and therefore, it is the intervals before referral and MDT meeting that can be affected by the referral route.

The 7 patients who had a nonstandard route of referral were usually patients who were transferred as neurosurgical emergencies before an MDT meeting. This may explain the similar diagnostic imaging to referral and referral to decision intervals, but shorter decision to surgery intervals on average.

Outlying and unexpectedly large referral intervals occurred in all referral types, which may indicate that they were not necessarily caused by the referral type only, but probably by a number of other factors. In diagnostic imaging to referral and decision to surgery intervals, these occurred mainly in patients with meningioma, in proportions that were much higher than in the overall sample, which may reflect the often-benign character and lower urgency of these cases [16,17]. In referral to decision intervals, most unexpectedly large intervals were of diagnostically inconclusive cases (ie, the diagnosis was lesion or mass or unknown or missing), which required further information or investigations followed by a rediscussion. In addition, they had been mostly referred with paper letters, which could indicate that their lower completeness found in this study made it less straight forward to make a decision.

The eReferral system also had a positive impact in making the preparation for the MDT shorter and easier. Triaging patients, or contacting referrers about outstanding investigations, became faster and simpler, as all information necessary for this was accessible electronically from the eReferral list. The real-time nature of the electronic system allowed the inclusion of patients referred up to 2 hours before the MDT, thus potentially accommodating last-minute additions that would have had to wait for a further week before discussion. Moreover, the electronic patient list generated for the MDT removed the need to prepare this manually. This means that eReferrals led to a decrease in the time and number of tasks needed for MDT preparation, which gave nurse specialists, and other professionals, more time to spend with other clinical tasks, while preserving an optimal MDT preparation process.

Referral Completeness

It is paramount to have all necessary imaging before the MDT discussion to allow for appropriate decisions and avoid delays in treatment of brain tumors [1]. The use of structured referral sheets can improve the quality of referrals by ensuring that necessary prereferral examinations and investigations are completed before a referral [3]. Implementing such proformas electronically allowed the team to designate compulsory and fixed-option fields with appropriate validation, therefore, improving referral completeness and accuracy. In addition, the electronic format helped reducing the risks of illegibility or repeated information [2,3]. Paper proforma referrals in our study also performed well, with each field being present in 69% to 100% referrals, depending on the field, thus confirming the positive impact of using proformas, even in paper format. Conversely, key fields in unstructured paper letters were present in 35% cases or less. Such level of incompleteness in referral letters can lead to poorer decisions by referrers but also to longer referral pathways, as suggested by our findings that paper letter referrals had significantly longer referral to decision intervals, possibly because of specialists needing more time to gather all necessary information.

Similar findings have been reported in emergency neurosurgery referrals, where information in paper referrals was often missing, whereas Web-based referrals were 100% complete [2]. eReferrals have also been found to perform better in dermatology at recording identifiers and medication prescribed, although they were more incomplete than paper referrals in a number of key clinical fields [18], highlighting the value of structured eReferrals with mandatory fields.

Details of presentation was the only study prespecified key data field that was present in all referrals, regardless of the referral type, which was not surprising as this is an essential piece of information in any medical referral.

Intrahospital internal referrals also performed poorly in terms of their level of completeness. These referrals do not have a customized proforma, and, as patients are already in the hospitals' EHR, referrers may make shorter or more incomplete referrals, maybe assuming that all the necessary information is already in the system and can be found if needed. This reinforces the need for customizing form fields to the specific referral process.

We also found that almost 90% of referrals were directly followed by a final management decision at the MDT, without the need for rediscussions, with this proportion being similar in all referral types. This figure is higher than those reported in a large US study, where the percentage of referrals immediately scheduled without any back and forth between specialist and requesting providers was 58.4% for oncology referrals and 57.7% for neurosurgery referrals [19]. This could be related to the involvement in the study center of an MDT of professionals with diverse clinical backgrounds and the virtual participation of specialists from referring hospitals, which can avoid a great proportion of rediscussions at a different MDT meeting.

Uptake of Electronic Referrals

Since their introduction, the use of eReferrals to the study center rose steadily, and in the last month of the study period, eReferrals represented more than half of all referrals. These proportions of referrals done electronically are similar or higher than those found in similar projects in other health care settings

[12,14,20], indicating the welcoming of our project by community hospitals and GPs. In addition, the distribution of referrer types remained very similar over time, indicating a low risk of provider characteristics having affected the eReferral uptake. A positive attitude of GPs toward eReferrals and health information technology systems have been shown in previous large surveys and interviews [21-23]. This has been found to be driven by referrers' realization that eReferral systems can improve access to specialty care, and there is better appointment tracking and improved communication between referrer and specialty care providers [23,24]. In addition, eReferrals in our project keep referrers informed about the outcome of the referral and MDT discussion, the lack of which having been identified as a limitation of other eReferral systems [23].

Conversely, paper referrals were still used in our study, even a year after introduction of the eReferral system. This could be related to the fact that brain tumor patients are often diagnosed as emergencies by nonspecialist clinicians who may not be familiar with the electronic referral process. Furthermore, paper proformas remain the method of referral to the neuro-oncology MDT for the majority of UK neuroscience centers, which most referrers will be familiar with. Nonetheless, this study was carried out over the course of 18 months and the possibility that the system may not have yet achieved a *steady state* where it is operating at full potential cannot be discarded. There is evidence that resistance may arise among referrers on the use of eReferrals [13], often related to lack of developed skills or motivation around the use of technology, and unawareness of the benefits of the technology [25]. It is important that the project design involves both referrers and accepting specialty centers and it reflects the local context and addresses local barriers [3,25,26]. The electronic system for eReferrals must be disease specific and purposively developed and tailored to the needs and context of the health care setting [27]. In our study, the eReferral system implemented was designed and implemented following evidence- and experience-based recommendations for the effectiveness improvement of electronic referral communications [28]. This includes the system-collaborative development by and for the professionals who benefit directly from the system, the use of a proforma with both structured and free-text fields, the inclusion of compulsory fields such as the referral reason, and the system capabilities of being used as electronic consultation and of providing referral status tracking and feedback to referrers [27,28]. In addition, regular feedback is gathered from users by the developers, to adapt and improve the software solution based on this, which is key for a successful integration of the system too [3,26,28].

The regional variation identified in our study within the East of England was characterized by a lower monthly incidence of referrals coming from the three most southern regions. These regions are closer to London, which may have led to some cases being referred to other specialist neuroscience centers in London.

Methodological Considerations

A number of meningiomas in our study center are not discussed as new cases but only considered as follow-ups later on, and therefore, these are underrepresented in our sample. Consequently, the distribution of diagnostic groups in our study do not reflect their reported epidemiology at the population level in similar countries [29-31].

The lower subsample size of paper letter referrals could have affected the magnitude of the differences identified. Given this, and the great variability in the study intervals identified, a future study involving a larger sample, with at least 5 patients per diagnostic group per referral type is warranted if the aim is to examine differences by tumor type.

Conclusions

Referrals to a specialist neuroscience center, using a service-developed specific proforma, perform significantly better, in terms of time from diagnosis to referral and specialist decision, and completeness of the information provided, than free-text letters. Electronic proformas perform even better than paper proformas, through an easily accessible and structured Web form including mandatory and fixed-option fields. Complete referrals ensure that specialists receive essential information for them to be able to make optimal informed decisions about a referral and are associated with faster decisions after referrals. Faster referrals mean that time to treatment is notably shorter, thus reducing the risk of disease progression. A wider application of eReferrals within cancer services and beyond is an important first step to streamlining specialist care pathways and providing excellent care.

Acknowledgments

The authors are very grateful to Innovate UK for its financial support of this project as part of the Detection and Assessment of Malignancy by Symptom Evaluation (DAMSEL) project within the Digital Health in a Connected Hospital initiative (RG77262 to AJ). Innovate UK is part of UK Research and Innovation, a nondepartmental public body funded by a grant in aid from the UK government. This report is also independent research supported by the National Institute for Health Research (NIHR Career Development Fellowship, Mr Stephen Price, CDF-2018-11-ST2-003), and the NIHR Brain Injury MedTech Co-operative based at Cambridge University Hospitals NHS Foundation Trust and the University of Cambridge. The views are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. The authors thank the team of Obex Technologies for its continuous support in training professionals and tailoring the software solution to the needs of this QI project. Finally, they would like to acknowledge all professionals and patients who made the aims of this project significant and its implementation possible.

Conflicts of Interest

AJ is director of Obex Technologies, lead partner of the DAMSEL consortium and developer of the software solution for this project. SR is a full-time employee of Obex Technologies.

Multimedia Appendix 1
Screenshots of the eReferral portal.
[PDF File (Adobe PDF File), 2185 KB - jmir_v22i3e15002_app1.pdf]

Multimedia Appendix 2
eReferral form data dictionary.
[PDF File (Adobe PDF File), 36 KB - jmir_v22i3e15002_app2.pdf]

Multimedia Appendix 3
Study data extraction methods and sources.
[PDF File (Adobe PDF File), 61 KB - jmir_v22i3e15002_app3.pdf]

References

1. National Institute for Care Excellence (NICE). 2018 Jul. Brain Tumours (Primary) and Brain Metastases in Adults - NICE Guideline [NG99] URL: https://www.nice.org.uk/guidance/ng99 [accessed 2018-12-03]
2. Choo MC, Thennakon S, Shapey J, Tolias CM. A web-based referral system for neurosurgery--a solution to our problems? Br J Neurosurg 2011 Jun;25(3):384-387. [doi: 10.3109/02688697.2011.566381] [Medline: 21501059]
3. Akbari A, Mayhew A, Al-Alawi MA, Grimshaw J, Winkens R, Glidewell E, et al. Interventions to improve outpatient referrals from primary care to secondary care. Cochrane Database Syst Rev 2008 Oct 8(4):CD005471 [FREE Full text] [doi: 10.1002/14651858.CD005471.pub2] [Medline: 18843691]
4. Vermeir P, Vandijck D, Degroote S, Peleman R, Verhaeghe R, Mortier E, et al. Communication in healthcare: a narrative review of the literature and practical recommendations. Int J Clin Pract 2015 Nov;69(11):1257-1267 [FREE Full text] [doi: 10.1111/ijcp.12686] [Medline: 26147310]
5. Public Health England’s National Cancer Intelligence Network (NCIN). NCIN’s Routes to Diagnosis. London: Public Health England; 2015. Routes to diagnosis 2015 update: brain tumours URL: http://www.ncin.org.uk/view?rid=3177 [accessed 2020-01-10]
6. Penfold C, Joannides AJ, Bell J, Walter FM. Diagnosing adult primary brain tumours: can we do better? Br J Gen Pract 2017 Jun;67(659):278-279 [FREE Full text] [doi: 10.3399/bjgp17X691277] [Medline: 28546414]
7. Sage W, Fernández-Méndez R, Crofton A, Gifford MJ, Bannykh A, Chrysaphinis C, et al. Defining unmet clinical need across the pathway of brain tumor care: a patient and carer perspective. Cancer Manag Res 2019;11:2189-2202 [FREE Full text] [doi: 10.2147/CMAR.S175886] [Medline: 30962709]
8. Chen AH, Murphy EJ, Yee HF. eReferral--a new model for integrated care. N Engl J Med 2013 Jun 27;368(26):2450-2453. [doi: 10.1056/NEJMp1215594] [Medline: 23802515]
9. Guilfoyle MR, Weerakkody RA, Oswal A, Oberg I, Jeffery C, Haynes K, et al. Implementation of neuro-oncology service reconfiguration in accordance with NICE guidance provides enhanced clinical care for patients with glioblastoma multiforme. Br J Cancer 2011 Jun 7;104(12):1810-1815 [FREE Full text] [doi: 10.1038/bjc.2011.153] [Medline: 21610702]
10. Zubrod CG, Schneiderman M, Frei E, Brindley C, Gold GL, Shnider B, et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene thiophosphoramide. J Chronic Dis 1960;11(1):7-33 [FREE Full text] [doi: 10.1016/0021-9681(60)90137-5]
11. Office for National Statistics (ONS). Wales: ONS; 2017. Population Estimates: Analysis Tool - Mid 2016 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesanalysistool [accessed 2020-01-10]
12. Warren J, White S, Day KJ, Gu Y, Pollock M. Introduction of electronic referral from community associated with more timely review by secondary services. Appl Clin Inform 2011;2(4):546-564 [FREE Full text] [doi: 10.4338/ACI-2011-06-RA-0039] [Medline: 23616895]
13. Walsh C, Siegler EL, Cheston E, O'Donnell H, Collins S, Stein D, Informatics Intervention Research Collaboration (I2RC). Provider-to-provider electronic communication in the era of meaningful use: a review of the evidence. J Hosp Med 2013 Oct;8(10):589-597 [FREE Full text] [doi: 10.1002/jhm.2082] [Medline: 24101544]
14. Cannaby S, Wanscher C, Pedersen C, Voss H. Velkommen til MedCom - MedCom. Odense: ACCA and MedCom in collaboration with the European Commission InformationSociety Directorate – General; 2004. The Cost Benefit of Electronic Patient Referrals in Denmark URL: https://medcom.dk/media/1173/the-cost-benefit-of-electronic-patient-referrals-in-denmark-full-report.pdf [accessed 2020-01-10]
15. Stensjøen AL, Solheim O, Kvistad KA, Håberg AK, Salvesen Ø, Berntsen EM. Growth dynamics of untreated glioblastomas in vivo. Neuro Oncol 2015 Oct;17(10):1402-1411 [FREE Full text] [doi: 10.1093/neuonc/nov029] [Medline: 25758748]

16. Achey RL, Gittleman H, Schroer J, Khanna V, Kruchko C, Barnholtz-Sloan JS. Nonmalignant and malignant meningioma incidence and survival in the elderly, 2005-2015, using the Central Brain Tumor Registry of the United States. Neuro Oncol 2019 Feb 19;21(3):380-391. [doi: 10.1093/neuonc/noy162] [Medline: 30295804]
17. Euskirchen P, Peyre M. Management of meningioma. Presse Med 2018;47(11-12 Pt 2):e245-e252. [doi: 10.1016/j.lpm.2018.05.016] [Medline: 30449639]
18. Shaw LJ, de Berker DA. Strengths and weaknesses of electronic referral: comparison of data content and clinical value of electronic and paper referrals in dermatology. Br J Gen Pract 2007 Mar;57(536):223-224 [FREE Full text] [Medline: 17359610]
19. Tuot DS, Murphy EJ, McCulloch CE, Leeds K, Chan E, Chen AH. Leveraging an electronic referral system to build a medical neighborhood. Healthc (Amst) 2015 Dec;3(4):202-208. [doi: 10.1016/j.hjdsi.2015.04.001] [Medline: 26699344]
20. Heimly V. Electronic referrals in healthcare: a review. Stud Health Technol Inform 2009;150:327-331. [Medline: 19745323]
21. Hughes CA, Allen P, Bentley M. eReferrals: Why are we still faxing? Aust J Gen Pract 2018;47(1-2):50-57 [FREE Full text] [doi: 10.31128/AFP-07-17-4285] [Medline: 29429313]
22. Djalali S, Ursprung N, Rosemann T, Senn O, Tandjung R. Undirected health IT implementation in ambulatory care favors paper-based workarounds and limits health data exchange. Int J Med Inform 2015 Nov;84(11):920-932. [doi: 10.1016/j.ijmedinf.2015.08.001] [Medline: 26296433]
23. Bouamrane M, Mair FS. A qualitative evaluation of general practitioners' views on protocol-driven eReferral in Scotland. BMC Med Inform Decis Mak 2014 Apr 9;14:30 [FREE Full text] [doi: 10.1186/1472-6947-14-30] [Medline: 24712766]
24. Straus SG, Chen AH, Yee H, Kushel MB, Bell DS. Implementation of an electronic referral system for outpatient specialty care. AMIA Annu Symp Proc 2011;2011:1337-1346 [FREE Full text] [Medline: 22195195]
25. Goldthorpe J, Sanders C, Macey R, Gough L, Rogers J, Tickle M, et al. Exploring implementation of an electronic referral management system and enhanced primary care service for oral surgery: perspectives of patients, providers and practitioners. BMC Health Serv Res 2018 Aug 20;18(1):646 [FREE Full text] [doi: 10.1186/s12913-018-3424-z] [Medline: 30126391]
26. Fernández-Méndez R, Rastall R, Sage W, Oberg I, Bullen G, Charge A, et al. Quality improvement of neuro-oncology services: integrating the routine collection of patient-reported, health-related quality-of-life measures. Neurooncol Pract 2019 May;6(3):226-236. [doi: 10.1093/nop/npy040] [Medline: 31385996]
27. McGovern M, Quinlan M, Doyle G, Moore G, Geiger S. Implementing a national electronic referral program: qualitative study. JMIR Med Inform 2018 Jul 18;6(3):e10488 [FREE Full text] [doi: 10.2196/10488] [Medline: 30021709]
28. Esquivel A, Sittig DF, Murphy DR, Singh H. Improving the effectiveness of electronic health record-based referral processes. BMC Med Inform Decis Mak 2012 Sep 13;12:107 [FREE Full text] [doi: 10.1186/1472-6947-12-107] [Medline: 22973874]
29. Leece R, Xu J, Ostrom QT, Chen Y, Kruchko C, Barnholtz-Sloan JS. Global incidence of malignant brain and other central nervous system tumors by histology, 2003-2007. Neuro Oncol 2017 Oct 19;19(11):1553-1564 [FREE Full text] [doi: 10.1093/neuonc/nox091] [Medline: 28482030]
30. Jafarzadeh N, Faal A, Izanloo A, Farrokhi F, Ziaolhagh R, Hashemian H, et al. Epidemiology of nervous system tumors according to WHO 2007 classification: a report of 1,164 cases from a single hospital. Int J Cancer Manag 2018 Apr;11(4):e11462 [FREE Full text] [doi: 10.5812/ijcm.11462]
31. Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, et al. The epidemiology of glioma in adults: a 'state of the science' review. Neuro Oncol 2014 Jul;16(7):896-913 [FREE Full text] [doi: 10.1093/neuonc/nou087] [Medline: 24842956]

Abbreviations

CNS: central nervous system
CT: computerized tomography
DAMSEL: Detection and Assessment of Malignancy by Symptom Evaluation
EHR: electronic hospital record
GP: general practice
HGG: high-grade gliomas
LGG: low-grade gliomas
MDT: multidisciplinary team
MRI: magnetic resonance imaging
NIHR: National Institute for Health Research
NHS: National Health Service
ORION: Outcome Registry Intervention and Operation Network
PACS: Picture Archiving and Communication System
PS: performance status
QI: quality improvement
WHO: World Health Organization

Edited by G Eysenbach; submitted 12.06.19; peer-reviewed by M Schoen, Y Chu; comments to author 03.10.19; revised version received 03.11.19; accepted 16.12.19; published 05.03.20.

Please cite as:
Fernández-Méndez R, Wong MY, Rastall RJ, Rebollo-Díaz S, Oberg I, Price SJ, Joannides AJ
Improvement of the Efficiency and Completeness of Neuro-Oncology Patient Referrals to a Tertiary Center Through the Implementation of an Electronic Referral System: Retrospective Cohort Study
J Med Internet Res 2020;22(3):e15002
URL: https://www.jmir.org/2020/3/e15002
doi:10.2196/15002
PMID:32134389

Original Paper

A Late Attempt to Involve End Users in the Design of Medication-Related Alerts: Survey Study

Melissa Therese Baysari[1,2], PhD; Wu Yi Zheng[1,2], PhD; Bethany Van Dort[1,2], BMedSci; Hannah Reid-Anderson[3], BNursing; Mihaela Gronski[3], DEng; Eliza Kenny[3], BNursing, MHSM

[1]Faculty of Health Sciences, The University of Sydney, Sydney, Australia
[2]Centre for Health Systems & Safety Research, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia
[3]Macquarie University Hospital, Sydney, Australia

Corresponding Author:
Melissa Therese Baysari, PhD
Faculty of Health Sciences
The University of Sydney
Charles Perkins Centre D17
Sydney, 2006
Australia
Phone: 61 28627 9245
Email: melissa.baysari@sydney.edu.au

Abstract

Background: When users of electronic medical records (EMRs) are presented with large numbers of irrelevant computerized alerts, they experience alert fatigue, begin to ignore alert information, and override alerts without processing or heeding alert recommendations. Anecdotally, doctors at our study site were dissatisfied with the medication-related alerts being generated, both in terms of volume being experienced and clinical relevance.

Objective: This study aimed to involve end users in the redesign of medication-related alerts in a hospital EMR, 4 years post implementation.

Methods: This work was undertaken at a private not-for-profit teaching hospital in Sydney, Australia. Since EMR implementation in 2015, the organization elected to implement all medication-related alert types available in the system for prescribers: allergy and intolerance alerts, therapeutic duplication alerts, pregnancy alerts, and drug-drug interaction alerts. The EMR included no medication administration alerts for nurses. To obtain feedback on current alerts and suggestions for redesign, a Web-based survey was distributed to all doctors and nurses at the site via hospital mailing lists.

Results: Despite a general dissatisfaction with alerts, very few end users completed the survey. In total, only 3.37% (36/1066) of doctors and 14.5% (60/411) of nurses took part. Approximately 90% (30/33) of doctors who responded held the view that too many alerts were triggered in the EMR. Doctors suggested that most alerts be removed and that alerts be more specific and less sensitive. In contrast, 97% (58/60) of the nurse respondents indicated that they would like to receive medication administration alerts in the EMR. Most nurses indicated that they would like to receive all the alert types available at all severity levels.

Conclusions: Attempting to engage with end users several years post implementation was challenging. Involving users so late in the implementation process may lead to clinicians viewing the provision of feedback to be futile. Seeking user feedback on usefulness, volume, and design of alerts is extremely valuable; however, we suggest this is undertaken early, preferably before system implementation.

(J Med Internet Res 2020;22(3):e14855) doi:10.2196/14855

KEYWORDS

alert fatigue; alerting; medication alert systems; clinical decision support; hospital information systems

Introduction

Many studies have shown that medication-related computerized alerts embedded in hospital electronic medical records (EMRs) can result in significant changes to prescriber behavior [1]. For example, introduction of dosing alerts for psychotropic medications in a US tertiary hospital led to an increase in the prescription of recommended daily doses from 19% to 29% and a reduction in the incidence of tenfold dosing errors from 5% to 3% [2]. However, studies have also shown that computerized alerts can have no impact or a negative impact on prescribing [1]. One of the main factors hampering alert effectiveness is alert overload. When users are presented with large numbers of irrelevant alerts, they experience alert fatigue [3], begin to ignore alert information [4], and override alerts (ie, click past the alert window) without processing or heeding alert recommendations [5].

Despite international efforts to improve medication-related alerts, alert override rates remain as high as they were over a decade ago [6]. Alert fatigue appears to be a widespread problem for users of EMRs, but there is limited evidence available on what constitutes a *tolerable* volume of irrelevant alerts for prescribers [7,8]. How many is too many before alert fatigue sets in? With no answer to this question, many hospitals have chosen to enable the main decision support alerts available in their EMR, including drug-drug interaction, allergy, and dose range alerts, with some also implementing pregnancy and therapeutic duplication warnings [9]. Our site was one such hospital, implementing all available medication-related alert types for prescribers.

In this study, we describe an attempt that was made to involve end users in the redesign of medication-related alerts several years post implementation. Anecdotally, via informal discussions with doctors, we learned that doctors were dissatisfied with the alerts being generated, both in terms of volume being experienced and clinical relevance. We administered a Web-based survey to doctors and nurses to more systematically capture their views of medication-related alerts embedded in the hospital EMR and their suggestions for redesign.

Methods

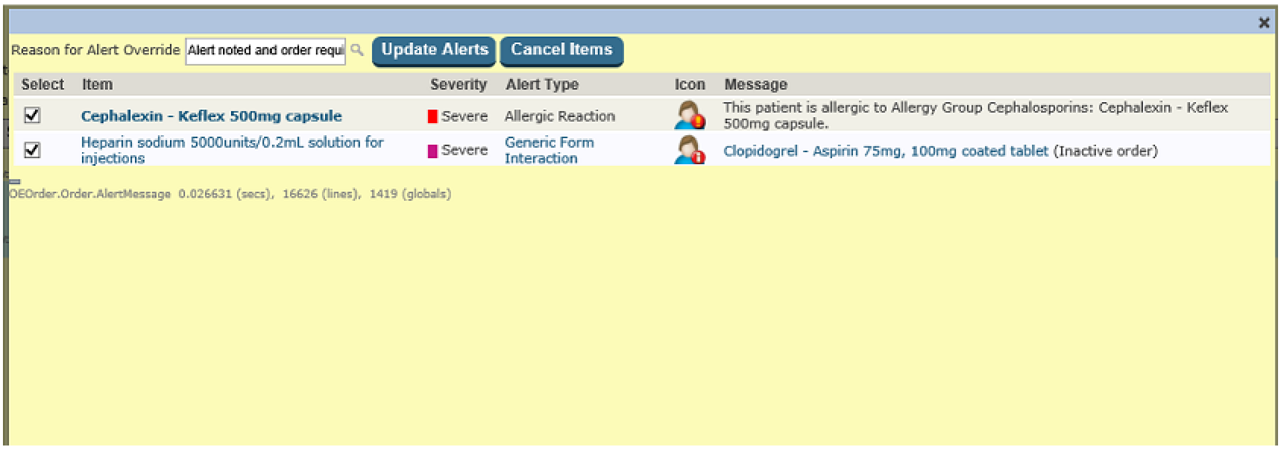
Context

This work was undertaken at a private not-for-profit teaching hospital, with approximately 150 beds, in Sydney, Australia. In 2015, the hospital implemented a single EMR in all wards and areas of the hospital. Before EMR implementation, a suite of fragmented clinical information systems was in place (an intensive care unit system, a general ward system, a medication management system, and a patient administration system). The single EMR was introduced to better support integration of care and to ensure efficient and effective delivery of health services. With respect to medication management, the system allowed prescribing, pharmacy review, and medication administration.

Implementation of Medication-Related Alerts

A core group of clinicians and administrative staff (information technology and billing) were consulted before EMR implementation, which included mainly heads of departments and units. Since EMR implementation, the organization elected to implement all medication-related alert types available in the system for prescribers: allergy and intolerance alerts, therapeutic duplication alerts, pregnancy alerts, and drug-drug interaction alerts. Not long after implementation, the alert configuration was changed so that only alerts classified as *severe* were displayed to doctors. All alerts were interruptive, appearing in an alert window that prevented prescribers from continuing with their orders until alerts were acknowledged. This could be done by selecting an override reason from a drop-down menu that appeared at the top of the alert screen (see Figure 1). The EMR included no medication administration alerts for nurses.

Figure 1. An example showing two alerts that have been triggered simultaneously, an allergy alert and a drug-drug interaction alert.



Reason for Alert Override: Alert noted and order requi

Select	Item	Severity	Alert Type	Icon	Message
☑	**Cephalexin - Keflex 500mg capsule**	🟥 Severe	Allergic Reaction		This patient is allergic to Allergy Group Cephalosporins: Cephalexin - Keflex 500mg capsule.
☑	Heparin sodium 5000units/0.2mL solution for injections	🟥 Severe	Generic Form Interaction		Clopidogrel - Aspirin 75mg, 100mg coated tablet (Inactive order)

OEOrder.Order.AlertMessage 0.026631 (secs), 16626 (lines), 1419 (globals)

User Engagement: Web-Based Survey

A Web-based survey was distributed to all doctors (n=1066) and nurses (n=411) at the site via hospital mailing lists (see Multimedia Appendix 1). Participation in the survey was voluntary, and no reimbursement was provided for participation. The survey was codeveloped with key stakeholders (researchers, doctors, nurses, pharmacists, and implementation staff) and comprised some basic demographic questions, questions on the current alerts in the system (for prescribers), and questions on preferences for changes to the system (for prescribers and nurses)—see Multimedia Appendices 2 and 3. Following distribution of the survey, a single reminder was sent out after 2 weeks. The survey was closed after 1 month.

Data Analysis

Descriptive statistics were used for analysis of quantitative data. A general inductive approach was used for analysis of qualitative responses. Responses to free-text questions were reviewed independently by 2 researchers (WZ and BD) who came together to reach a consensus on key themes related to current alerts and redesign recommendations.

Ethics approval was obtained from the hospital’s human research ethics committee (reference number 5201833414318).

Results

Survey Respondents and Response Rate

In total, 36 doctors and 60 nurses responded to the survey, representing a response rate of 3.37% (36/1066) and 14.5% (60/411), respectively (see Table 1).

Table 1. Characteristics of 36 doctors and 60 nurses that responded to the survey. (Not all participants responded to all questions).

Respondents	Respondents who responded to the question, n (%)
Prescribers	
Roles	
Accredited practitioner—anesthetics	23 (64)
Accredited practitioner—medical	4 (11)
Accredited practitioner—surgical	6 (17)
Registrar	1 (3)
Intern	1 (3)
Days per week working at hospital	
Less than 2	23 (64)
2-3	7 (19)
4-5	6 (17)
Setting of EMR[a] use	
Inpatient	36 (100)
Outpatient	2 (5)
Frequency with which they prescribe in EMR	
Multiple times a day	27 (75)
Once in the day	2 (6)
Only on some days	2 (6)
Rarely	4 (11)
Never	1 (3)
Experience using other EMR systems	
Yes	18 (55)
No	15 (45)
Nurses	
Roles	
Registered nurse	57 (95)
Endorsed enrolled nurse	3 (5)
Days per week working at hospital	
1-2	5 (8)
3-4	23 (38)
5-6	32 (53)
Experience using other EMR systems	
Yes	24 (41)
No	35 (59)

[a] EMR: electronic medical record.

Some respondents indicated that they were pleased the evaluation was being done:

> *Thank you for doing something to fix this. Please involve clinicians.* [#02]

However, on the whole, most doctors appeared highly dissatisfied with the EMR in general and made sarcastic or sweeping comments about the system in the free-text boxes of the survey. For example, a doctor said:

> *The whole system is user unfriendly, please do not give it another band-aid.* [#21]

Findings Related to the Volume of Alerts Being Triggered in the Electronic Medical Record for Doctors

As shown in Table 2, approximately 90% (30/33) of doctors who responded held the view that too many alerts were triggered in the EMR, and only 9% (3/33) doctors indicated that the number of alerts was *right*.

When asked to indicate how frequently they read the alerts, responses were mixed, as shown in Table 3.

The free-text responses received in response to the question “I only read the alerts in the EMR when...” reflected this variability. Examples from doctors included:

> *Severe alert fatigue—I now disregard all XX (EMR) alerts and check my own prescribing as I would have done when medication charts were on paper.* [#16]
>
> *I scan them very quickly, but don’t read them fully. It’s almost impossible to prescribe in XX (the EMR) without creating alerts—so they don’t really “alert” me.* [#13]
>
> *I always read them.* [#11]

Table 2. Doctors’ views on the current alert burden in the electronic medical record (N=33 responses).

Response option	Frequency of doctors who responded to the question, n (%)
Far too many alerts, most need to be removed	19 (58)
Too many alerts, some need to be removed	11 (33)
About the right number of alerts	3 (9)
Too few alerts, some need to be added	0 (0)

Table 3. Frequency with which doctors reported that they read alerts in the electronic medical record.

Response option	Frequency of doctors who responded to the question, n (%)
Never	4 (12)
Rarely	6 (18)
Sometimes	10 (30)
Often	4 (12)
Always	9 (27)

Findings Related to the Usefulness of Each Alert Type for Doctors

When asked to rate each alert type on a Likert scale of usefulness, allergy and intolerances alerts were rated most positively. As shown in Table 4, these alerts were viewed to be *sometimes useful* by approximately half of the doctors in the survey. The other alert types were rated as *never* or *rarely useful* by the majority of doctors. When asked to indicate which alert type was the most useful, 60% (20/33) doctors selected allergy and intolerances alerts, and 24% (8/33) indicated that none of the alerts were useful.

Interestingly, when asked which of the alert types they would remove from the EMR (with more than one option possible), approximately half of the doctors (17/36, 47%) indicated that they would remove therapeutic duplication alerts, pregnancy alerts, and drug interaction alerts. In addition, 19% (7/36) also indicated that they would remove allergy and intolerance alerts.

Table 4. Doctors’ views on alert usefulness.

Response option	Frequency, n			
	Allergy and intolerances alert	Therapeutic duplication alert	Pregnancy alert	Generic drug interaction alert
Never useful	8	11	12	12
Rarely useful	5	13	15	12
Sometimes useful	17	7	6	10
Often useful	6	5	3	2

Doctors' and Nurses' Preferences for Changes to the Alerting System

Nurses did not receive medication administration alerts in the EMR; however, 97% (58/60) of the nurse respondents indicated that they should. When asked what types of alerts nurses should receive, most nurses selected all the alert types available to the doctors. Other suggested alert types included alerts for stat orders, overdue medications, blood thinning medications, dose alerts, and alerts warning when a nursing intervention was required (eg, digoxin).

When asked to indicate what alert severity level or levels should be included in the EMR, nurses and doctors expressed very different views. As shown in Table 5, nurses were more open to receiving alerts of all severities.

Table 5. Doctors' and nurses' preferences for alert severity.

Response option	Frequency of doctors who responded to the question, n (%)	Frequency of nurses who responded to the question, n (%)
Only severe alerts	12 (36)	1 (2)
Severe and moderate alerts	20 (61)	19 (33)
All alerts, including minor alerts	1 (3)	38 (66)

In response to a request for suggested changes to the alerts in the EMR, the most frequent response from doctors was to reduce alert numbers primarily by making the alerts more relevant. For example, doctors said:

> *Too many for too many trivial issues.* [#41]
>
> *Real ones need to be more prominent. All the silly ones need to go.* [#02]
>
> *Many of the alerts are theoretical and are ignored in our everyday practice.* [#17]
>
> *Hard to tell which ones are useful and which ones are not...You are more likely to miss important alerts if you are bombarded with too many insignificant alerts.* [#24]

Other common suggestions from doctor respondents were related to alert design. Current alerts were described as *not user friendly,* and doctors suggested using color coding to highlight the type or severity of alerts and suggested making alert content more concise:

> *Perhaps they could be color coded e.g. the severe ones and e.g. multiple opioid prescribing alert be large and red, moderate orange, minor orange and smaller. The blue color is very neutral.* [#20]

This was consistent with what nurses viewed to be characteristic of a well-designed alert:

> *Brief but adequate information, highlighted (font and color).* [#11]
>
> *As simple as it can be.* [#28]

Some nurses also described what they believed to be the ideal content to be included in an alert screen. For example:

> *Alert specifies the issue clearly. Specifies the action required. The alert can be dismissed easily where appropriate. Contains the problem: e.g. duplicate drug.* [#5]
>
> *An alert that states: "ALERT," not wordy but precise description of alert. An alert that is easy to understand.* [#41]

Discussion

Summary of Key Findings From the Survey

With respect to the current alerts in the EMR, prescribers reported that too many alerts were being triggered and that most alerts were not clinically relevant. As a consequence, alerts were not always being read, and many doctors reported experiencing alert fatigue. These findings confirm what had been suspected by the organization and are in line with many studies exploring user views of alerts [10,11]. Allergy alerts were perceived to be the most useful among the alert types, a finding also consistent with previous research [4,12]. With respect to recommended changes to the alerts, prescribers suggested that most alerts be removed and that alerts be more specific and less sensitive. Prescribers also recommended that alert interfaces be redesigned so that alert text is more concise, and color coding is used to signal alert type and severity of alerts. These suggestions were consistent with nurses' preferences for alert content and display and reflect good *human factors* design [13,14].

Lessons Learned From Our Attempt to Engage Users

Our survey highlighted some important lessons for the site and more broadly for other hospitals with plans to implement medication-related alerts in their EMR. Seeking user feedback on usefulness, volume, and design of alerts, even with this small-scale survey, proved to be extremely valuable; however, we suggest consulting users early and regularly. Our low response rate could in part reflect staff not checking emails or not having enough time to action emails. However, based on the free-text responses we received, it more likely reflected a workforce feeling disengaged and dissatisfied with the alerting system and associated implementation process. Early consultation would have allowed user views to be captured before negative mindsets were established and may have minimized any perception that the provision of feedback was futile. Continuous consultation with users is needed to ensure alert volume remains manageable and alert content remains relevant.

We recommend that organizations involve end users in the decision-making processes surrounding alert selection before

EMR rollout as well as in implementation and ongoing evaluation. Involving users in EMR implementation more broadly has been shown to facilitate a sense of ownership of the system among users and to result in an EMR that aligns well with the needs of end users rather than the needs of the information technology department or leadership staff [15]. Consistent with previous research [16], our findings confirm that user engagement is particularly important for medication-related alerts. Engaging end users in the selection and implementation processes for alerts will also increase user awareness of the challenges associated with alert optimization (eg, vendor restrictions and complexity associated with identifying a set of *high priority* alerts).

One of the interesting findings that emerged from our survey was that in contrast to doctors, nurses showed a strong preference for enabling all alert types of all severities for nurses in the EMR. This could reflect their limited experience with alerts, as a common misconception held by clinicians who do not experience alerts themselves is *the more alerts, the better*.

To allow end users to make informed decisions about alert selection, we recommend providing current users and prospective users with data on alert rates (eg, 50% of your orders/administrations/reviews will trigger an alert) to facilitate an understanding of the potential impact of alerts on their work and educating staff on the well-known risks associated with alert overload.

Conclusions

Our attempt to engage users in the redesign of medication-related alerts in an EMR was largely unsuccessful, with only a small number of doctors and nurses partaking in our survey. This likely reflects the delay between implementation and seeking feedback from end users. However, the feedback we received from survey respondents on usefulness, volume, and design of alerts proved to be extremely valuable. We recommend hospitals adopt a *less-is-more* approach and work closely with all end user groups before EMR implementation to determine the types, design, and content of alerts and implement these in their systems.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Recruitment emails.
[DOCX File , 14 KB - jmir_v22i3e14855_app1.docx]

Multimedia Appendix 2
Doctor survey.
[DOCX File , 23 KB - jmir_v22i3e14855_app2.docx]

Multimedia Appendix 3
Nurse survey.
[DOCX File , 21 KB - jmir_v22i3e14855_app3.docx]

References

1. Page N, Baysari MT, Westbrook JI. A systematic review of the effectiveness of interruptive medication prescribing alerts in hospital CPOE systems to change prescriber behavior and improve patient safety. Int J Med Inform 2017 Sep;105:22-30. [doi: 10.1016/j.ijmedinf.2017.05.011] [Medline: 28750908]
2. Peterson JF, Kuperman GJ, Shek C, Patel M, Avorn J, Bates DW. Guided prescription of psychotropic medications for geriatric inpatients. Arch Intern Med 2005 Apr 11;165(7):802-807. [doi: 10.1001/archinte.165.7.802] [Medline: 15824302]
3. Payne TH. EHR-related alert fatigue: minimal progress to date, but much more can be done. BMJ Qual Saf 2019 Jan;28(1):1-2. [doi: 10.1136/bmjqs-2017-007737] [Medline: 30297378]
4. Baysari MT, Westbrook JI, Richardson KL, Day RO. The influence of computerized decision support on prescribing during ward-rounds: are the decision-makers targeted? J Am Med Inform Assoc 2011;18(6):754-759 [FREE Full text] [doi: 10.1136/amiajnl-2011-000135] [Medline: 21676939]
5. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809] [Medline: 16357358]
6. Bryant A, Fletcher G, Payne T. Drug interaction alert override rates in the Meaningful Use era: no evidence of progress. Appl Clin Inform 2014;5(3):802-813 [FREE Full text] [doi: 10.4338/ACI-2013-12-RA-0103] [Medline: 25298818]
7. Baysari MT, Tariq A, Day RO, Westbrook JI. Alert override as a habitual behavior - a new perspective on a persistent problem. J Am Med Inform Assoc 2017 Mar 1;24(2):409-412. [doi: 10.1093/jamia/ocw072] [Medline: 27274015]
8. Rayo MF, Moffatt-Bruce SD. Alarm system management: evidence-based guidance encouraging direct measurement of informativeness to improve alarm response. BMJ Qual Saf 2015 Apr;24(4):282-286. [doi: 10.1136/bmjqs-2014-003373] [Medline: 25734193]

9. Page NJ, Baysari MT, Westbrook JI. Selection and use of decision support alerts in electronic medication management systems in Australian hospitals: a survey of implementers. J Pharm Pract Res 2019;49(2):142-149. [doi: 10.1002/jppr.1479]
10. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017 Sep 15;12(1):113 [FREE Full text] [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
11. Ash JS, Fournier L, Stavri PZ, Dykstra R. Principles for a successful computerized physician order entry implementation. AMIA Annu Symp Proc 2003;2003:36-40 [FREE Full text] [Medline: 14728129]
12. Baysari MT, Westbrook JI, Egan B, Day RO. Identification of strategies to reduce computerized alerts in an electronic prescribing system using a Delphi approach. Stud Health Technol Inform 2013;192:8-12. [Medline: 23920505]
13. Pelayo S, Marcilly R, Bernonville S, Leroy N, Beuscart-Zephir M. Human factors based recommendations for the design of medication related clinical decision support systems (CDSS). Stud Health Technol Inform 2011;169:412-416. [Medline: 21893783]
14. Phansalkar S, Edworthy J, Hellier E, Seger DL, Schedlbauer A, Avery AJ, et al. A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. J Am Med Inform Assoc 2010;17(5):493-501 [FREE Full text] [doi: 10.1136/jamia.2010.005264] [Medline: 20819851]
15. Simon SR, Keohane CA, Amato M, Coffey M, Cadet B, Zimlichman E, et al. Lessons learned from implementation of computerized provider order entry in 5 community hospitals: a qualitative study. BMC Med Inform Decis Mak 2013 Jun 24;13:67 [FREE Full text] [doi: 10.1186/1472-6947-13-67] [Medline: 23800211]
16. Baysari MT, Westbrook JI, Richardson K, Day RO. Optimising computerised alerts within electronic medication management systems: a synthesis of four years of research. Stud Health Technol Inform 2014;204:1-6. [Medline: 25087519]

Abbreviations

EMR: electronic medical record

Edited by G Eysenbach; submitted 30.05.19; peer-reviewed by M Hussain, C Alvarez, S Sabarguna; comments to author 14.09.19; revised version received 04.12.19; accepted 16.12.19; published 13.03.20.

Please cite as:
Baysari MT, Zheng WY, Van Dort B, Reid-Anderson H, Gronski M, Kenny E
A Late Attempt to Involve End Users in the Design of Medication-Related Alerts: Survey Study
J Med Internet Res 2020;22(3):e14855
URL: https://www.jmir.org/2020/3/e14855
doi:10.2196/14855
PMID:32167479

Original Paper

Using the Technology Acceptance Model to Explore Adolescents' Perspectives on Combining Technologies for Physical Activity Promotion Within an Intervention: Usability Study

Mark Drehlich[1], B Mgt, BAppSc; Michael Naraine[2], PhD; Katie Rowe[2], PhD; Samuel K Lai[1], MHP; Jo Salmon[1], PhD; Helen Brown[1], PhD; Harriet Koorts[1], PhD; Susie Macfarlane[3], BSc (Hons); Nicola D Ridgers[1], PhD

[1]Institute for Physical Activity and Nutrition, Deakin University, Burwood, Australia

[2]Deakin Business School, Deakin University, Burwood, Australia

[3]Learning Futures, Deakin University, Burwood, Australia

Corresponding Author:
Nicola D Ridgers, PhD
Institute for Physical Activity and Nutrition
Deakin University
221 Burwood Highway
Burwood, VIC 3125
Australia
Phone: 61 392446718
Email: nicky.ridgers@deakin.edu.au

Abstract

Background: Wearable activity trackers and social media have been identified as having the potential to increase physical activity among adolescents, yet little is known about the perceived ease of use and perceived usefulness of the technology by adolescents.

Objective: The aim of this study was to use the technology acceptance model to explore adolescents' acceptance of wearable activity trackers used in combination with social media within a physical activity intervention.

Methods: The Raising Awareness of Physical Activity study was a 12-week physical activity intervention that combined a wearable activity tracker (Fitbit Flex) with supporting digital materials that were delivered using social media (Facebook). A total of 124 adolescents aged 13 to 14 years randomized to the intervention group (9 schools) participated in focus groups immediately post intervention. Focus groups explored adolescents' perspectives of the intervention and were analyzed using pen profiles using a coding framework based on the technology acceptance model.

Results: Adolescents reported that Fitbit Flex was useful as it motivated them to be active and provided feedback about their physical activity levels. However, adolescents typically reported that Fitbit Flex required effort to use, which negatively impacted on their perceived ease of use. Similarly, Facebook was considered to be a useful platform for delivering intervention content. However, adolescents generally noted preferences for using alternative social media websites, which may have impacted on negative perceptions concerning Facebook's ease of use. Perceptions of technological risks included damage to or loss of the device, integrity of data, and challenges with both Fitbit and Facebook being compatible with daily life.

Conclusions: Wearable activity trackers and social media have the potential to impact adolescents' physical activity levels. The findings from this study suggest that although the adolescents recognized the potential usefulness of the wearable activity trackers and the social media platform, the effort required to use these technologies, as well as the issues concerning risks and compatibility, may have influenced overall engagement and technology acceptance. As wearable activity trackers and social media platforms can change rapidly, future research is needed to examine the factors that may influence the acceptance of specific forms of technology by using the technology acceptance model.

Trial Registration: Australian and New Zealand Clinical Trials Registry ACTRN12616000899448; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=370716

(J Med Internet Res 2020;22(3):e15552) doi:10.2196/15552

KEYWORDS

fitness trackers; social media; physical activity; youth

Introduction

Background

Engaging in regular physical activity is critical for adolescent health. Higher physical activity levels support weight management, musculoskeletal development, fitness, cardiovascular health [1], and mental health through enhanced self-concept and reduced anxiety and depression [2]. For adolescents (aged 12-17 years), the Australian Government recommends 60 min of at least moderate-intensity physical activity daily [3], yet only 18% of adolescents meet these guidelines [4]. With physical inactivity being recognized as a global pandemic [5] and an estimated 80% of the global adolescent population classified as inactive [6], strategies are needed to increase levels of physical activity.

To date, many physical activity interventions have either been reported to be ineffective within inactive populations or unscalable because of cultural, geographic, social, or economic contexts [7]. It is further posited that adolescents perceive many interventions negatively as individuals tend not to self-select into such interventions; rather such interventions are imposed on them by others [8]. Opportunities exist to utilize novel approaches to encourage adolescents to participate in physical activity. Given the increasing popularity of wearable activity trackers, a potential strategy for increasing youth physical activity levels is to examine how such devices might be used to encourage physical activity among adolescents. Wearable activity trackers are electronic devices that use sensors to track movement and collect biometric data [9] and enable constant self-monitoring through the provision of feedback via a visual display and/or accompanying app [10]. As such, these devices allow individuals to have an enhanced awareness of self, and these devices have the potential to generate internal motivation for physical activity.

Past research indicates that interventions that use wearable activity tracking devices may be acceptable to adolescent populations; therefore, wearable activity trackers have the potential to increase adolescents' levels of physical activity [9,11,12]. More recently, studies have explored how wearable activity trackers and forms of social media can be combined in interventions targeting adolescent inactivity [13-15]. Social media (eg, Facebook and Instagram) has emerged as a popular communication medium, offering expedited connectivity and engagement [16-18]; therefore, it can be used to provide additional support for physical activity. However, little is known about how adolescents engage with such technology within a physical activity intervention [9], and few studies have explored an individual's engagement within intervention components using theoretical models that may help to provide insights into such use [12]. For example, a potentially important feature of wearable activity trackers is the ability to share data to and receive peer support via social media [19]. Such engagement and support from others may lead to an increase in motivation and reinforcement to participate in physical activity [20]. The strategic use of social media to engage adolescents could result in stronger bonds that lead to increased feelings of relatedness and increased engagement in physical activity [21], and may provide a source of motivation to participate in physical activity [13,22]. Existing physical activity studies in adolescents and young people have found that frequently engaging with social media is associated with increased physical activity [14,19], although the challenge is identifying the strategies to engage adolescents, particularly as some research suggests that social media engagement is often passive [13,15]. However, little research has examined the individual's perceptions and experiences of using social media platforms, such as Facebook, when combined with physical activity interventions using wearable activity trackers. Such information could help to inform strategies to achieve optimum intervention impact and engagement among adolescents.

Technology Acceptance Model

A framework that enables researchers to examine technology use is the technology acceptance model [23]. This model provides a framework for evaluating how different factors may influence an individual's use and acceptance of specific forms of technology, such as wearable activity trackers or social media [23]. In addition, more recent advances in the technology acceptance model have included the perceived risks associated with using specific forms of technology and the degree of compatibility that such technology has with an individual's values and needs [24,25]. Overall, the focus of the model is not on whether the technology results in increased levels of physical activity but on how the different technology used in a physical activity intervention is accepted by a target group. Whether a target group is willing to accept and use the specific forms of technology relied upon in a physical activity intervention is important to understand, given it is unlikely that an intervention will lead to increased levels of physical activity among a target group if that group does not accept or is unwilling to use the chosen forms of technology.

To date, few studies have used the technology acceptance model to examine factors that may influence the acceptance of specific forms of technology, such as wearable activity trackers and social media, when combined within a physical activity intervention. In studies that have used the technology acceptance model, the focus has been on adults rather than adolescents. For example, Lunney et al [26] found that perceived usefulness significantly influenced adults' acceptance of wearable activity trackers, whereas perceived ease of use was a direct determinant of their behavior (use of activity trackers). Similarly, Chuah et al [27] found that perceived usefulness can assist in determining adults' attitudes toward wearable activity trackers, but it did not predict their adoption intention (intention to use the device). Opportunities exist to extend the understanding of technology acceptance from adults to adolescents.

Objective

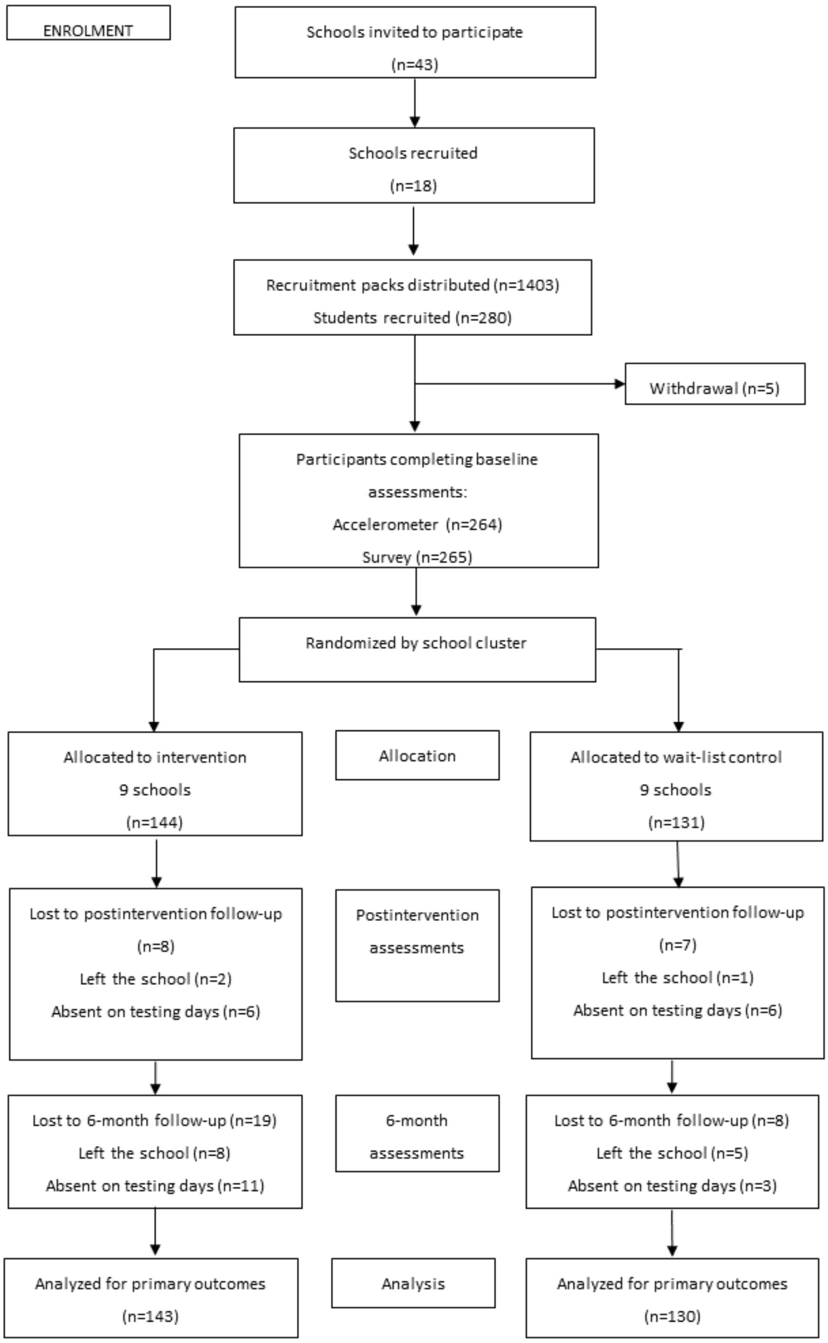
The aim of this study was to use the technology acceptance model to explore adolescents' acceptance of wearable activity trackers and social media when used in combination within a physical activity intervention.

Methods

Study Design and Participants

The design and methods of the study have been reported in detail elsewhere [28]. In brief, the Raising Awareness of Physical Activity (RAW-PA) study was a 12-week multi-component study that combined a wearable activity tracker and digital behavior change resources delivered via social media, which aimed to increase inactive adolescents' physical activity levels. Schools located in areas that had a Socio-Economic Indexes for Areas [29] in the lowest 50% and were within approximately 60 km of Deakin University's Burwood Campus were eligible to participate. A total of 18 schools (42% response rate) were recruited (Figure 1). Participants were adolescents in Year 8 (aged 13 to 14 years) who self-reported that they did not engage in regular physical activity/sport, did not meet current physical activity guidelines, had not previously owned or used a wearable activity tracker, had (or were willing to create) a Facebook account, and had access to the internet outside of school (age: mean 13.8 years, SD 0.4 years; 142/275, 51.6% female). Ethics approval for this study was obtained from the Deakin University Human Research Ethics Committee and the Victorian Department of Education and Training. Informed consent to participate in the study was obtained from all schools and parents, with written assent provided by adolescents. The study is registered with the Australian and New Zealand Clinical Trials Registry (ACTRN12616000899448).

Figure 1. Flow of participants throughout the study. This study focused on the data collected at post-intervention.



Intervention

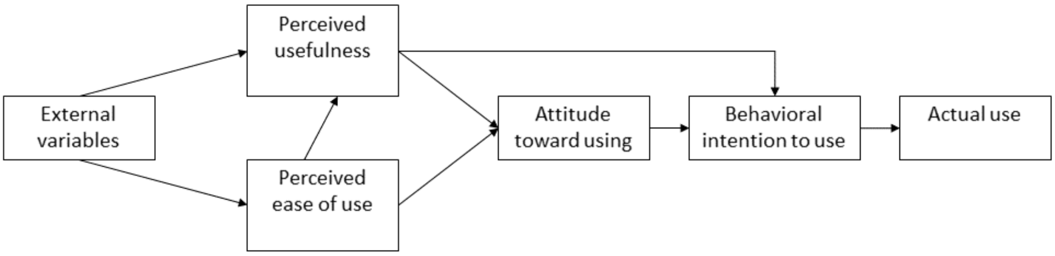
Adolescents (n=144) attending schools that were randomized to the intervention group (n=9) received a wrist-worn Fitbit Flex and an accompanying app, as well as access to the interactive weekly individual or team "missions" and behavior change resources via a private, researcher-moderated Facebook group [28]. Facebook was chosen as the form of social media as, at the time of the study, it was the most popular social media platform [30]. The aim of the Facebook group was to provide adolescents with a platform to ask questions, interact with other participants, and engage with posted content that related to the weekly missions. Alerts for new content were also sent to the adolescents through email and/or text messages (approximately 2-3 times per week) [28]. The intervention components and structure were developed using participatory research principles, and the combination of technologies aimed to target low-cost forms of physical activity (eg, walking) and guide adolescents through the behavior change process in a way that was accessible, flexible, and interactive [28].

At the start of the intervention, the research team provided initial assistance in setting up Fitbit Flex, which included creating a Fitbit account for each participant and providing information on how to sync and charge the device and how to use the app to view data. No other information was provided at this time about the use of Fitbit Flex. Adolescents were informed that new content would be posted regularly to the Facebook group, though no guidance was provided about the frequency with which to access content.

Theoretical Framework

This study utilized the technology acceptance model [23] to examine adolescents' perspectives on combining wearable activity trackers and social media within a physical activity intervention (Figure 2). The technology acceptance model identifies 2 variables that are key to technology acceptance and use: (1) perceived ease of use (is using the technology free from effort?) and (2) perceived usefulness (will the use of the technology enhance performance?). According to the technology acceptance model, perceived ease of use and perceived usefulness, either alone or in combination, predict behavioral intention (intention to use technology), which in turn predicts subsequent behavior (actual use of technology) [23]. Specifically, this model is being used to examine whether the different technologies used in a physical activity intervention can be easily used by the target group and whether such technology offers valued benefits to the target group [23].

Figure 2. Technology acceptance model.



Measures

To address the research questions in this study, only data collected from the student focus groups were used. At the end of the 12-week intervention period, all adolescents attending intervention schools were invited to participate in focus groups that explored their thoughts and perspectives on RAW-PA and the different components within the intervention. In total, 124 students (63 males and 61 females; 124/144, 86.1% of the intervention group) participated in 15 focus groups that took place at each school (n=9). Focus groups contained both males and females and ranged from 6 to 13 participants (average of 8 participants). This enabled the adolescents to provide unique insights into their experiences of RAW-PA and their acceptance of the technologies used in this study, thus enabling us to respect their expert knowledge and lived experience [31]. The focus groups followed a semistructured format that was designed to identify potential enablers and barriers to using different components of RAW-PA, which included a discussion of the key forms of technology used—Fitbit Flex and Facebook. Example questions included "What did you like/not like about the Fitbit and accompanying app?," "What did you like/not like about the Facebook group?," "Did you experience any issues using aspects of the program (eg, Fitbit or Facebook)?," and "Did anything help you to use the different features (eg, Fitbit or Facebook)?." Focus groups (mean duration 26 min) were digitally recorded and transcribed verbatim, producing 256 pages (Times New Roman, Font 12) of data for analysis.

Data Analyses

Focus Group Data

Qualitative data were analyzed using pen profiles, an increasingly used technique for presenting findings to researchers with qualitative and quantitative backgrounds [31-33]. Pen profiles present key themes identified during data analysis through the combination of verbatim quotes taken directly from the transcripts to provide context with frequency data [31,34]. The numbers reported against each theme indicate the number of times the theme was cited in focus groups, as individuals in focus groups were not identified. Data were initially analyzed using a deductive process in which the technology acceptance model [23] was used to develop a coding framework and inform the coding of the concepts of perceived usefulness, perceived ease of use, perceived risk, and compatibility in relation to the key technological components of this intervention: (1) the wearable activity tracking device, Fitbit Flex, and (2) the social media platform, Facebook. Perceived ease of use was defined as "the degree to which a person believes that using a technology will be free from effort" [35]. Perceived usefulness was defined as "the extent to which a person believes that using particular technology will enhance their performance" [35]. An inductive coding process was then used to identify the key themes that emerged from the data [31,36].

As recommended by Burnard [37], a researcher (MD) who was independent of the project delivery team initially read and analyzed the transcripts. Following the development of the pen profiles, the findings were then presented to 2 independent researchers with expertise in the use of technology for activity promotion and qualitative data analyses (KR and MN). Data from the pen profiles to the transcripts were cross-examined, which enabled alternative interpretations and data interrogation until an overall consensus was achieved. The pen profiles were then presented to 2 project delivery team members (NR and SL) who further critically challenged the interpretation of the data. Credibility and transferability were demonstrated through the triangular consensus procedure and verbatim transcription of collected data [33,34].

Results

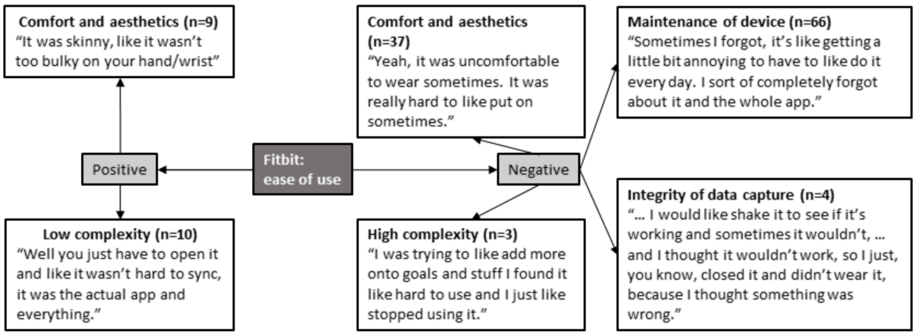
Findings

Data concerning the perceived ease of use and perceived usefulness of Fitbit Flex and Facebook are presented initially. As the adolescents often discussed Fitbit Flex and the accompanying app interchangeably, it should be noted that results concerning Fitbit include both the device and the app unless otherwise stated. The findings that relate to the perceived risk and compatibility of these technologies are then presented in the last pen profile.

Fitbit: Perceived Ease of Use

With respect to the perceived ease of use of the device (Figure 3), although some focus group participants indicated that the device was easy to use, many adolescents reported that it was not free from effort. For example, they explained that the device was hard to put on; they needed to take it off for some activities (ie, swimming and playing a sport); they often forgot to wear, charge, and sync the device; or they had technical issues while charging or syncing the device. Several Fitbit functions were also perceived negatively, with the primary concerns being maintenance requirements, including the life of the battery (need to charge it frequently), and the need to sync the data regularly, all of which required a concerted effort to address. In addition, the integrity of the data was challenged at times, with adolescents unsure as to how the device captured their data, whether it was capturing data at all times, and if the data were accurate. A few adolescents reported that they perceived the device negatively because Fitbit Flex did not have a screen display, which detracted from their ease of using the device to track their activity levels.

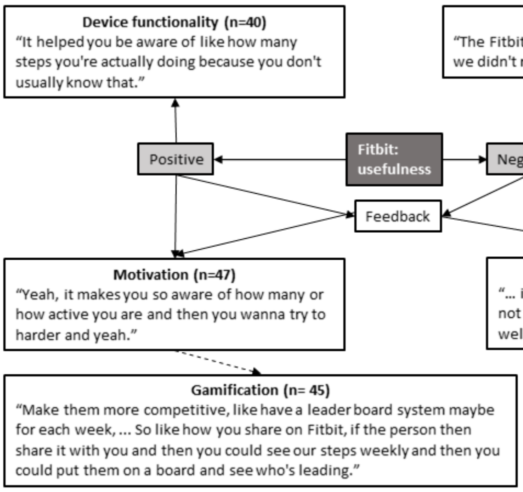
Figure 3. Perceived ease of use of Fitbit in adolescents. Note: n=number of times theme was mentioned by adolescents.



Fitbit: Perceived Usefulness

From the perspective of perceived usefulness, adolescents reported that Fitbit Flex's specific functions and its ability to motivate them to participate in physical activity were positive aspects of the device (Figure 4). Other reported benefits included the device providing adolescents with a greater awareness of their physical activity patterns and a greater level of motivation to engage in physical activity in response to such knowledge. Participants also noted that Fitbit Flex was useful for setting goals and evaluating whether these had been achieved. These features were perceived as being useful to a number of adolescent participants. However, it was also frequently mentioned that Fitbit Flex may be more useful and motivating if additional gamification offerings were provided, such as the recognition of achievement via the device or app. In addition, some concerns were raised in relation to whether the data captured were meaningful and, therefore, useful. Specifically, concerns were raised regarding the potential for differences to exist between actual and recorded activity levels. This made the adolescents question the usefulness of the device for the purposes of monitoring activity levels. Moreover, the lack of a display also impacted perceived usefulness, especially as adolescents reported little engagement with the app to obtain information about their activity levels, meaning that feedback from the device was often limited or perceived to be insufficient for their needs. Adolescents also recognized that there was a diminishing return on using Fitbit Flex because of an initial novelty factor of wearing the device, which existed when they commenced the program but diminished over time, again impacting perceived usefulness.

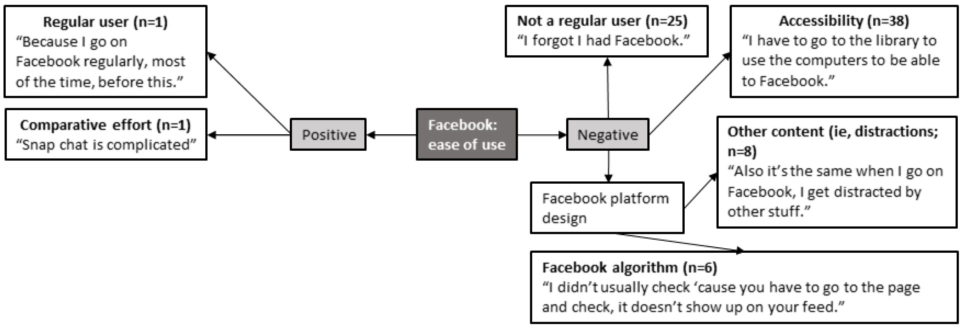
Figure 4. Fitbit's perceived usefulness in adolescents. Note: n=number of times theme was mentioned by adolescents.



Facebook: Perceived Ease of Use

Adolescents' perceptions in relation to the perceived ease of use of Facebook within the intervention are summarized in Figure 5. Overall, negative perceptions regarding perceived ease of use were most commonly expressed. These reflections included that the design of and perceived effort to use Facebook diminished its ease of use, that the platform design comprised several independent elements that were not well connected, and that adolescents who were irregular users of Facebook had to intentionally log in to engage with RAW-PA content. Others identified that they did not want to use their mobile phone data allocation for accessing intervention content on Facebook; therefore, access became an issue. Adolescents also recognized that if they were not actively seeking to engage with RAW-PA content on Facebook (ie, intentionally looking for it), the intervention materials would not appear in their news feed because of the algorithms Facebook uses. Interestingly, several adolescents noted that they would get distracted by other content, which meant reductions in the frequency of the introduction of RAW-PA into the feed because of Facebook's algorithm. Only 2 positive comments relating to the perceived ease of use of Facebook were reported, namely, previous exposure to the platform and the relative ease of use in comparison with other platforms.

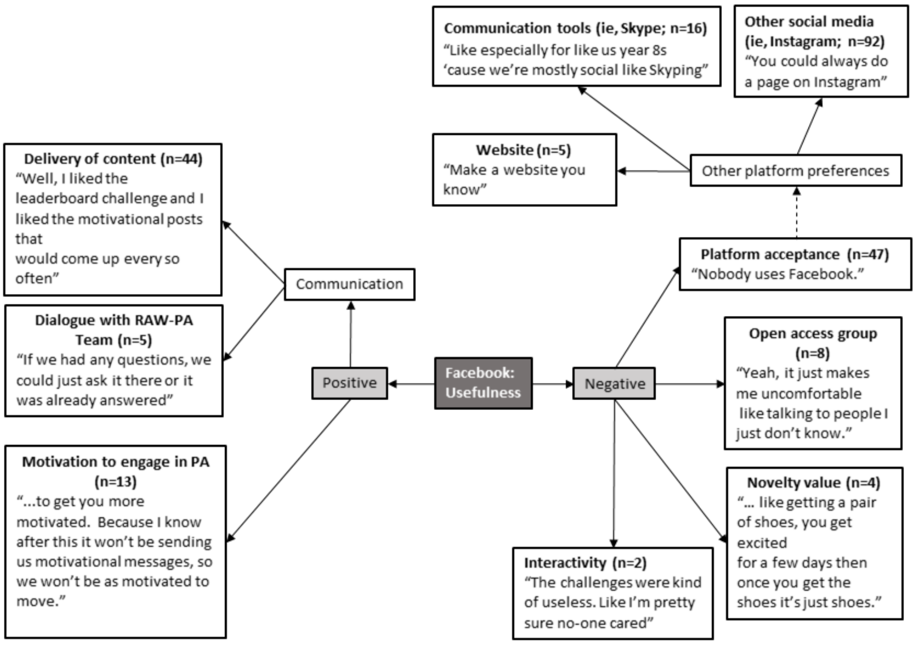
Figure 5. Perceived ease of use of Facebook in adolescents. Note: n=number of times theme was mentioned by adolescents.



Facebook: Perceived Usefulness

Adolescents' perspectives on Facebook's perceived usefulness are shown in Figure 6. Adolescents spoke positively about how useful Facebook was for delivering content in relation to the RAW-PA intervention. In particular, many adolescents stated that the RAW-PA messages provided through Facebook were useful in motivating them to engage in additional physical activity. Some adolescents considered Facebook a useful communication platform, albeit with the research team rather than their peers. The most commonly discussed limitation in relation to using Facebook was its lack of acceptance as a social media platform, thus decreasing its perceived usefulness. Adolescents highlighted their preference for having "private" groups that would be inaccessible to other members of the Facebook group to reduce the risk of their comments and messages being shared with the wider intervention group. Although Facebook had some initial value at the start of the program, as this novelty wore off, adolescents reported that their interest; use; and, ultimately, the usefulness of the platform diminished. Adolescents indicated a preference for alternative platforms, including stand-alone apps and websites, image and video platforms (ie, Snapchat and Instagram), or direct communication tools (ie, Skype).

Figure 6. Perceived usefulness of Facebook in adolescents. PA: physical activity; RAW-PA: Raising Awareness of Physical Activity. Note: n=number of times theme was mentioned by adolescents; --- indicates an extension of a theme.



Communication tools (ie, Skype; n=16) "Like especially for like us year 8s 'cause we're mostly social like Skyping"

Other social media (ie, Instagram; n=92) "You could always do a page on Instagram"

Website (n=5) "Make a website you know"

Delivery of content (n=44) "Well, I liked the leaderboard challenge and I liked the motivational posts that would come up every so often"

Dialogue with RAW-PA Team (n=5) "If we had any questions, we could just ask it there or it was already answered"

Motivation to engage in PA (n=13) "...to get you more motivated. Because I know after this it won't be sending us motivational messages, so we won't be as motivated to move."

Platform acceptance (n=47) "Nobody uses Facebook."

Open access group (n=8) "Yeah, it just makes me uncomfortable like talking to people I just don't know."

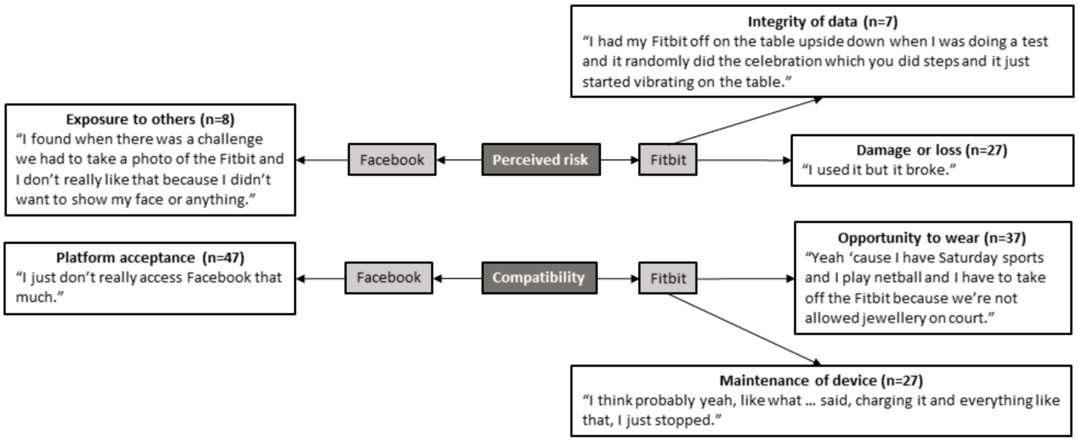
Novelty value (n=4) "... like getting a pair of shoes, you get excited for a few days then once you get the shoes it's just shoes."

Interactivity (n=2) "The challenges were kind of useless. Like I'm pretty sure no-one cared"

Perceived Risk and Compatibility

Several risks of using these technologies were perceived by the adolescents (Figure 7). When discussing Facebook, the commonly reported perceived risk was exposure to others within the intervention. Further related to Facebook, the lack of compatibility with the adolescents' lifestyle was commonly alluded to. Facebook was considered a platform that adults used and was out of touch with the needs of adolescents. The risks surrounding the use of Fitbit Flex included the integrity of the data as well as the potential to damage or lose the device while using it. Adolescents also stated that the effort required to use the device (eg, regularly charging and syncing the device) was incompatible with their lifestyle.

Figure 7. Facebook and Fitbit perceived risk and compatibility.



Integrity of data (n=7) "I had my Fitbit off on the table upside down when I was doing a test and it randomly did the celebration which you did steps and it just started vibrating on the table."

Exposure to others (n=8) "I found when there was a challenge we had to take a photo of the Fitbit and I don't really like that because I didn't want to show my face or anything."

Damage or loss (n=27) "I used it but it broke."

Platform acceptance (n=47) "I just don't really access Facebook that much."

Opportunity to wear (n=37) "Yeah 'cause I have Saturday sports and I play netball and I have to take off the Fitbit because we're not allowed jewellery on court."

Maintenance of device (n=27) "I think probably yeah, like what ... said, charging it and everything like that, I just stopped."

Discussion

Principal Findings

The aim of this study was to utilize the technology acceptance model to explore adolescents' acceptance of wearable activity trackers (Fitbit Flex) and social media (Facebook) when combined within a physical activity promotion intervention. Overall, adolescents generally reported that they perceived Fitbit Flex to be useful for tracking physical activity and motivating them to participate in physical activity, but perceptions concerning the perceived ease of use were often negative. Issues concerning the device's ease of use, need for regular charging and syncing, and functionality were discussed as factors that required effort to either address or understand. Similarly, Facebook had more positive responses concerning its perceived usefulness, particularly as a vehicle to deliver motivating content in relation to Fitbit Flex, although it was no longer the preferred social media platform for adolescents in this study. However, perceived ease of use was low because of the design of the platform and the effort required to use it. Concerns were also raised about the compatibility of the technologies with current lifestyles and risks associated with using the technologies. In general, the findings suggest that the adolescents recognized the potential usefulness of the wearable activity trackers and social media platform in a physical activity intervention, but the effort required to use the technologies, as well as issues concerning risks and compatibility, may have led to lower technology acceptance.

Comparison With Previous Work

To date, few studies have utilized the technology acceptance model [23] to examine the acceptance of technology within a physical activity intervention. In relation to wearable activity trackers, this study found that the functionality offered and the motivation generated by Fitbit Flex were important factors in the perceived usefulness of the device. Previous research has shown that the functionality of wearable activity trackers is important to the wearer, particularly in relation to self-tracking activity levels throughout the day [11,38]. Moreover, others have identified that wearable activity trackers can increase the wearer's awareness and understanding of their own physical activity levels, which in turn provide motivation to engage in physical activity [14,20]. Indeed, Schafer et al [38] clearly identified that lack of motivation was a barrier to engagement and subsequent adoption, therefore support was required to continue engagement of adolescents.

The perceived usefulness of the device was potentially negatively impacted by diminished novelty over time. This is consistent with previous research, which has noted that novelty effects reduce the device's perceived usefulness and therefore the probability that a wearer will continue to use the device after progressing past the intention to use it [9,20,38-40]. Interestingly, little research has examined whether personal preferences for specific wearable activity trackers may affect the perceptions of usefulness, with most studies providing one specific device for use. Future studies could consider providing different wearable activity tracker options to see if this helps sustain use over time. Of note, adolescents commented that incorporating specific digital game elements (ie, gamification) could benefit the perceived usefulness of wearable activity trackers, such as Fitbit Flex, within a physical activity intervention. Wider research [41,42] supports this finding, identifying positive outcomes because of gamified approaches and reductions in physical inactivity post intervention. Notably, gamified elements, such as competition and digital recognition for efforts, were incorporated into RAW-PA; although given the negative feedback provided concerning the usefulness of Facebook to deliver such elements, it is possible that this may have been missed by participants. As such, future studies should consider incorporating gamified elements [41] such as leaderboards, competitions, and tangible rewards into a physical activity intervention to enhance the usefulness of wearable activity trackers and engagement with the device. Although this has the potential to increase engagement and motivation [13,26], whether or not this will address concerns of novelty effects requires further investigation.

In this study, perceptions of Fitbit Flex's ease of use were typically negative. For example, the device needed to be charged and synced regularly for it to operate and collect data, which adolescents indicated took more effort than warranted. It was commonly reported that adolescents forgot to wear the device after it had been removed, which is consistent with previous research [20,40]. Some concerns were raised about the wearability of the device, which also impacted its ease of use. Specifically, adolescents reported that the device was uncomfortable, it took effort to tolerate wearing it, and the device's clasp was problematic. Rupp et al [43] noted similar issues with adolescents having difficulty while putting on the device, whereas others have highlighted how the comfort and design of the device can be a barrier to adolescents' use of the technology [12,20,38]. In contrast, some adolescents perceived that the design of Fitbit Flex meant that it was considered to have good wearability. Interestingly, there was general agreement on the fact that Fitbit Flex was comparatively simple to use, and little instruction was required, which is consistent with previous studies [11,12]. Overall, these findings suggest that a wearable activity tracker's perceived ease of use is a critical component of a technology-based intervention, and future studies should identify potential strategies to overcome perceived barriers to ease of use among adolescents.

In recent years, there has been an increase in the number of studies using Facebook to deliver physical activity interventions in different populations [13-15,19,21,44]. As is the case in this study, Facebook has often been chosen based on its popularity [21] and the opportunity it offers to provide information and social support to the user [13]. There was some indication that Facebook was perceived to be useful for receiving intervention content, communicating with the research team, and providing some motivation through social support. This is consistent with the findings of Pumper et al [13] who suggested that using Facebook in physical activity interventions directed at adolescents may be motivating and increase engagement, although it was noted that active rather than passive engagement of adolescents may be required to provide a source of extrinsic motivation [15]. Interestingly, research has suggested that Facebook provides motivation for engagement through individuals likening themselves to others (perceived role models and peers) and receiving gratification through the approval of others [21], as well as updates and messages using inspirational imagery, which promote higher levels of engagement [22]. However, adolescents need to engage with it for it to be effective, which was perceived to require effort.

Identifying the strategies to encourage social support and approval of others may be important for future interventions. Although Facebook was the dominant social media platform for adolescents and informed the intervention design at the time of study development, it was evident that it was no longer the preferred social media platform for adolescents during this study [12]. Furthermore, some adolescents did not like sharing information with others they did not know from other schools. Pumper et al [13] supported this, noting that although passive engagement (viewing content) was common among adolescents, active engagement (contributing content) was uncommon. Although this study used a private group, adolescents appeared to perceive contributing and participating in the group forum as a risk, which likely reduced their engagement. Divine et al [21] supported this, noting that through social comparison, Facebook has the potential to encourage or discourage engagement. The challenge for future physical activity interventions is to identify social media platforms that meet the needs of the users and therefore optimize their acceptance of these social media platforms. It may be that the interventions may need to be available on a number of different platforms, although this may be challenging given the ever-changing nature of social media. Future research should consider the dynamic nature of social media and implications for use during interventions. This further speaks to the need for communities to be organically developed and not forced into existence [13].

The perceived ease of use of Facebook in this group of adolescents was generally low. Of concern, some noted that the algorithms used by Facebook were perceived to be detrimental to engagement, as the intervention content was either lost among other content or did not appear in their news feed. Edney et al [22] recently identified that social media algorithms impacted news items in an adolescent's feed, suggesting that participants would have to actively seek items as a part of the intervention. Interestingly, several adolescents in this study mentioned that Facebook's accessibility also impacted its perceived ease of use, as they did not want to use their allocation of mobile phone data to access the platform; this finding is consistent with previous studies [12,45]. This meant that adolescents would have to seek internet access from alternative sources of data (eg, library, public Wi-Fi, and school computers) to access Facebook, which took additional effort.

The strengths of this study included the use of qualitative methods to explore adolescents' thoughts and experiences in depth after engaging in a 12-week physical activity intervention and utilizing the technology acceptance model as a framework. However, there are several limitations that should be noted. First, although adolescents reported that they had not used a wearable activity tracker previously, it is unknown whether their expectations of the device may have impacted their subsequent experiences of using the device. Second, the lack of adolescents' engagement with and willingness to share through social media, as well as the impact this would have on intervention implementation, was not anticipated. Third, although majority of the intervention students participated in focus groups, it is unknown as to whether their perspectives differed from those who did not participate. Fourth, data were collected at the end of the 12-week intervention. How the perceived usefulness and ease of use of Fitbit Flex and Facebook may have changed over time is also unknown. This information is valuable in developing future interventions and understanding the needs of adolescents in such interventions. Further research could consider assessing how adolescents' attitudes toward such technology evolve throughout a study. Fifth, the qualitative data in each of the pen profiles represent the number of times the theme is mentioned and not the number of individuals who agree with the viewpoint. Although this is a common approach with pen profile analyses [33], it is possible that the data do not represent the views of all those who participated in the focus group and only represent the views of those who responded to the questions. However, it is possible that others may agree with certain points, and they may have felt that the point was made and chose not to reiterate. Finally, as not all participants were involved in the focus groups, we were unable to determine whether those participating in the focus groups had more favorable perspectives (for example) than those who did not participate.

Conclusions

There is potential for both wearable activity trackers and social media to positively impact physical activity interventions among adolescents. However, this study highlighted the importance of perceived usefulness, perceived ease of use, perceived risk, and compatibility for understanding how adolescents engage with such technologies. Although both the wearable activity tracker and social media platform were considered useful, concerns about their ease of use, perceived risks associated with such use, and compatibility issues appeared to be critical and led to a low level of acceptance of and engagement with the technology. Technology advances rapidly, and interventions that use technology to engage with adolescents should continue to monitor and evaluate how technologies are used and accepted within physical activity interventions. The technology acceptance model can provide a useful framework to examine how technology is accepted among target groups, such as adolescents.

Acknowledgments

The authors would like to thank the schools and students for participating in this study. The authors also gratefully acknowledge the contribution of all project staff, especially Kate Dullaghan, Dr Felicity J Pendergast, Stephanie Renehan, Stephanie Chappel, Sophie Aarts, Kara Richards, Tom Steele, Sarah Tighe, and Holly Beswick. This study was supported by a VicHealth Innovation Research Grant. NR is supported by a National Heart Foundation of Australia Future Leader Fellowship (Award ID 101895). JS was supported by a National Health and Medical Research Council Principal Research Fellowship (APP1026216). The content of this manuscript is the responsibility of the authors and does not necessarily reflect the views of the funding bodies.

Conflicts of Interest

NR and JS declare involvement in a start-up technological company. The remaining authors declare no conflicts of interest.

References

1. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]

2. Strong WB, Malina RM, Blimkie CJ, Daniels SR, Dishman RK, Gutin B, et al. Evidence based physical activity for school-age youth. J Pediatr 2005 Jun;146(6):732-737. [doi: 10.1016/j.jpeds.2005.01.055] [Medline: 15973308]
3. Australian Government. Department of Health.: Australian Government Australian 24-hour Movement Guidelines for Children and Young People (5-17years) - An Integration of Physical Activity, Sedentary Behaviour and Sleep URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/health-24-hours-phys-act-guidelines [accessed 2019-07-01]
4. Schranz N, Glennon V, Evans J, Gomersall S, Hardy L, Hesketh KD, et al. Results from Australia's 2018 report card on physical activity for children and youth. J Phys Act Health 2018 Nov 1;15(S2):S315-S317. [doi: 10.1123/jpah.2018-0418] [Medline: 30475112]
5. Kohl HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul 21;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
6. Aubert S, Barnes JD, Abdeta C, Nader PA, Adeniyi AF, Aguilar-Farias N, et al. Global Matrix 3.0 Physical Activity Report Card Grades for Children and Youth: results and analysis from 49 countries. J Phys Act Health 2018 Nov 1;15(S2):S251-S273. [doi: 10.1123/jpah.2018-0472] [Medline: 30475137]
7. Althoff T, White RW, Horvitz E. Influence of Pokemon Go on physical activity: study and implications. J Med Internet Res 2016 Dec 6;18(12):e315 [FREE Full text] [doi: 10.2196/jmir.6759] [Medline: 27923778]
8. Seabra A, Katzmarzyk P, Carvalho MJ, Seabra A, Coelho-E-Silva M, Abreu S, et al. Effects of 6-month soccer and traditional physical activity programmes on body composition, cardiometabolic risk factors, inflammatory, oxidative stress markers and cardiorespiratory fitness in obese boys. J Sports Sci 2016 Oct;34(19):1822-1829. [doi: 10.1080/02640414.2016.1140219] [Medline: 26890580]
9. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and effectiveness of using wearable activity trackers in youth: a systematic review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
10. Hoy MB. Personal activity trackers and the quantified self. Med Ref Serv Q 2016;35(1):94-100. [doi: 10.1080/02763869.2016.1117300] [Medline: 26794199]
11. Schaefer SE, van Loan M, German JB. A feasibility study of wearable activity monitors for pre-adolescent school-age children. Prev Chronic Dis 2014 May 22;11:E85 [FREE Full text] [doi: 10.5888/pcd11.130262] [Medline: 24854236]
12. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable activity tracker use among Australian adolescents: Usability and acceptability study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
13. Pumper MA, Mendoza JA, Arseniev-Koehler A, Holm M, Waite A, Moreno MA. Using a Facebook group as an adjunct to a pilot mHealth physical activity intervention: a mixed methods approach. Stud Health Technol Inform 2015;219:97-101. [Medline: 26799887]
14. Schoenfelder E, Moreno M, Wilner M, Whitlock KB, Mendoza JA. Piloting a mobile health intervention to increase physical activity for adolescents with ADHD. Prev Med Rep 2017 Jun;6:210-213 [FREE Full text] [doi: 10.1016/j.pmedr.2017.03.003] [Medline: 28373931]
15. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study. Pediatr Blood Cancer 2017 Dec;64(12):-. [doi: 10.1002/pbc.26660] [Medline: 28618158]
16. Loader BD, Mercea D. Networking democracy? Inf Commun Soc 2011;14(6):757-769. [doi: 10.1080/1369118X.2011.592648]
17. Naraine M, Karg A. Digital media in international sport: engaging fans via social media and fantasy sports. In: MacIntosh EW, Bravo GA, Li M, editors. International Sport Management. Second Edition. Champaign, IL: Human Kinetics; 2019:315-332.
18. van Dijck J, Poell T. Understanding social media logic. Media Commun 2013;1(1):2-14. [doi: 10.17645/mac.v1i1.70]
19. Cavallo DN, Tate DF, Ward DS, DeVellis RF, Thayer LM, Ammerman AS. Social support for physical activity-role of Facebook with and without structured intervention. Transl Behav Med 2014 Dec;4(4):346-354 [FREE Full text] [doi: 10.1007/s13142-014-0269-9] [Medline: 25584083]
20. Shih P, Kyungsik P, Erika SR, Mary BC, John M. Use and adoption challenges of wearable activity trackers. In: Proceedings of the 2015 iConference. 2015 Presented at: iConference'15; March 24–27, 2015; Newport Beach, CA URL: https://pdfs.semanticscholar.org/395d/d8e17e74f5a809e99372225a552789d0f475.pdf
21. Divine A, Watson PM, Baker S, Hall CR. Facebook, relatedness and exercise motivation in university students: a mixed methods investigation. Comput Human Behav 2019;91:138-150. [doi: 10.1016/j.chb.2018.09.037]
22. Edney S, Bogomolova S, Ryan J, Olds T, Sanders I, Maher C. Creating engaging health promotion campaigns on social media: observations and lessons from Fitbit and Garmin. J Med Internet Res 2018 Dec 10;20(12):e10911 [FREE Full text] [doi: 10.2196/10911] [Medline: 30530449]
23. Davis FD. DSpace@MIT Home. Cambridge, MA USA: MIT Libraries; 1986. A Technology Acceptance Model for Empirically Testing New End-User Information Systems : Theory and Results URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2020-01-13]

24. Nasir S, Yurder Y. Consumers' and physicians' perceptions about high tech wearable health products. Procedia Soc Behav Sci 2015;195:1261-1267. [doi: 10.1016/j.sbspro.2015.06.279]
25. Rogers EM. Diffusion Of Innovations. New York, NY, USA: Simon and Schuster; 2010.
26. Lunney A, Cunningham NR, Eastin MS. Wearable fitness technology: a structural investigation into acceptance and perceived fitness outcomes. Comput Human Behav 2016;65(5):114-120. [doi: 10.1016/j.chb.2016.08.007]
27. Chuah SH, Rauschnabel PA, Krey N, Nguyen B, Ramayah T, Lade S. Wearable technologies: the role of usefulness and visibility in smartwatch adoption. Comp Human Behav 2016;65:276-284. [doi: 10.1016/j.chb.2016.07.047]
28. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. A cluster-randomised controlled trial to promote physical activity in adolescents: the Raising Awareness of Physical Activity (RAW-PA) study. BMC Public Health 2017 Jan 4;17(1):6 [FREE Full text] [doi: 10.1186/s12889-016-3945-5] [Medline: 28052773]
29. Australian Bureua of Statistics. Australian Bureua of Statistics. Canberra, Australia: Australian Bureua of Statistics; 2011. Socio-economic indexed for areas (SEIFA) URL: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa2011 [accessed 2016-08-20]
30. Lenhart A, Duggan M, Perrin A, Stepler R, Rainie H, Parker K. Pew Research Center. Washington DC, USA: Pew Research Centre; 2015 Apr 9. Teens, Social Media & Technology Overview 2015 URL: https://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2019-05-27]
31. Ridgers ND, Knowles ZR, Sayers J. Encouraging play in the natural environment: a child-focused case study of Forest School. Child Geogr 2012;10(1):49-65. [doi: 10.1080/14733285.2011.638176]
32. Knowles ZR, Parnell D, Stratton G, Ridgers ND. Learning from the experts: exploring playground experience and activities using a write and draw technique. J Phys Act Health 2013 Mar;10(3):406-415. [doi: 10.1123/jpah.10.3.406] [Medline: 22820462]
33. Mackintosh KA, Chappel SE, Salmon J, Timperio A, Ball K, Brown H, et al. Parental perspectives of a wearable activity tracker for children younger than 13 years: acceptability and usability study. JMIR Mhealth Uhealth 2019 Nov 4;7(11):e13858 [FREE Full text] [doi: 10.2196/13858] [Medline: 31682585]
34. Mackintosh KA, Knowles ZR, Ridgers ND, Fairclough SJ. Using formative research to develop CHANGE!: a curriculum-based physical activity promoting intervention. BMC Public Health 2011 Oct 27;11:831 [FREE Full text] [doi: 10.1186/1471-2458-11-831] [Medline: 22032540]
35. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
36. Biddle SJ, Markland D, Gilbourne D, Chatzisarantis NL, Sparkes AC. Research methods in sport and exercise psychology: quantitative and qualitative issues. J Sports Sci 2001 Oct;19(10):777-809. [doi: 10.1080/026404101317015438] [Medline: 11561674]
37. Burnard P. A method of analysing interview transcripts in qualitative research. Nurse Educ Today 1991 Dec;11(6):461-466. [doi: 10.1016/0260-6917(91)90009-y] [Medline: 1775125]
38. Schaefer SE, Ching CC, Breen H, German JB. Wearing, thinking, and moving: testing the feasibility of fitness tracking with urban youth. Am J Health Educ 2016;47(1):8-16. [doi: 10.1080/19325037.2015.1111174]
39. Gouveia R, Karapanos E, Hassenzahl M. How Do We Engage With Activity Trackers?: A Longitudinal Study of Habito. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, NY, United States: Association for Computing Machinery; 2015 Presented at: UbiComp'15; September 7 - 11, 2015; Osaka, Japan p. 1305-1316. [doi: 10.1145/2750858.2804290]
40. Gaudet J, Gallant F, Bélanger M. A Bit of Fit: Minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 6;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
41. Koivisto J, Hamari J. The rise of motivational information systems: a review of gamification research. Int J Inform Manag 2019;45:191-210. [doi: 10.1016/j.ijinfomgt.2018.10.013]
42. Harris MA. Maintenance of behaviour change following a community-wide gamification based physical activity intervention. Prev Med Rep 2019 Mar;13:37-40 [FREE Full text] [doi: 10.1016/j.pmedr.2018.11.009] [Medline: 30510892]
43. Rupp MA, Michaelis JR, McConnell DS, Smither JA. The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. Appl Ergon 2018 Jul;70:77-87. [doi: 10.1016/j.apergo.2018.02.005] [Medline: 29866329]
44. Edney S, Looyestyn J, Ryan J, Kernot J, Maher C. Posts, pics, or polls? Which post type generates the greatest engagement in a Facebook physical activity intervention? Transl Behav Med 2018 Nov 21;8(6):953-957. [doi: 10.1093/tbm/iby006] [Medline: 29660093]
45. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. Pediatrics 2014 Sep;134(3):e723-e731. [doi: 10.1542/peds.2014-1012] [Medline: 25157000]

Abbreviations

RAW-PA: Raising Awareness of Physical Activity

Edited by G Eysenbach; submitted 19.07.19; peer-reviewed by A Lee, P Hsiao; comments to author 01.09.19; revised version received 30.09.19; accepted 16.12.19; published 06.03.20.

Please cite as:
Drehlich M, Naraine M, Rowe K, Lai SK, Salmon J, Brown H, Koorts H, Macfarlane S, Ridgers ND
Using the Technology Acceptance Model to Explore Adolescents' Perspectives on Combining Technologies for Physical Activity Promotion Within an Intervention: Usability Study
J Med Internet Res 2020;22(3):e15552
URL: https://www.jmir.org/2020/3/e15552
doi:10.2196/15552
PMID:32141834

Original Paper

Patterns in Weight and Physical Activity Tracking Data Preceding a Stop in Weight Monitoring: Observational Analysis

Kerstin Frie[1], MSc; Jamie Hartmann-Boyce[1], DPhil; Susan Jebb[1], PhD; Jason Oke[1], DPhil; Paul Aveyard[1], PhD

[1] Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Kerstin Frie, MSc
Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building, Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 01865 ext 289592
Email: kerstin.frie@phc.ox.ac.uk

Abstract

Background: Self-regulation for weight loss requires regular self-monitoring of weight, but the frequency of weight tracking commonly declines over time.

Objective: This study aimed to investigate whether it is a decline in weight loss or a drop in motivation to lose weight (using physical activity tracking as a proxy) that may be prompting a stop in weight monitoring.

Methods: We analyzed weight and physical activity data from 1605 Withings Health Mate app users, who had set a weight loss goal and stopped tracking their weight for at least six weeks after a minimum of 16 weeks of continuous tracking. Mixed effects models compared weight change, average daily steps, and physical activity tracking frequency between a 4-week period of continuous tracking and a 4-week period preceding the stop in weight tracking. Additional mixed effects models investigated subsequent changes in physical activity data during 4 weeks of the 6-week long stop in weight tracking.

Results: People lost weight during continuous tracking (mean −0.47 kg, SD 1.73) but gained weight preceding the stop in weight tracking (mean 0.25 kg, SD 1.62; difference 0.71 kg; 95% CI 0.60 to 0.81). Average daily steps (beta=−220 daily steps per time period; 95% CI −320 to −120) and physical activity tracking frequency (beta=−3.4 days per time period; 95% CI −3.8 to −3.1) significantly declined from the continuous tracking to the pre-stop period. From pre-stop to post-stop, physical activity tracking frequency further decreased (beta=−6.6 days per time period; 95% CI −7.12 to −6.16), whereas daily step count on the day's activity was measured increased (beta=110 daily steps per time period; 95% CI 50 to 170).

Conclusions: In the weeks before people stop tracking their weight, their physical activity and physical activity monitoring frequency decline. At the same time, weight increases, suggesting that declining motivation for weight control and difficulties with making use of negative weight feedback might explain why people stop tracking their weight. The increase in daily steps but decrease in physical activity tracking frequency post-stop might result from selective measurement of more active days.

(J Med Internet Res 2020;22(3):e15790) doi:10.2196/15790

KEYWORDS

self-monitoring; self-regulation; weight loss; activity trackers; mobile applications

Introduction

Background

The repeated measurement and tracking of weight over time, also referred to as self-monitoring, is a common component of weight loss interventions and is consistently associated with greater weight loss [1]. The effectiveness of this self-monitoring strategy is attributed to a self-regulation process [2,3], which posits that people who monitor their weight use the information to reflect on the effectiveness of previous actions and plan further weight loss behaviors, thus engaging in self-experimentation. Furthermore, any discrepancies between desired and actual weight loss progress identified through self-monitoring are hypothesized to trigger corrective action [4].

The frequency of self-weighing seems to play a crucial role, as many studies have revealed significant positive associations between the frequency of self-monitoring and weight outcomes [1,5-7]. One study investigating patterns in self-weighing frequency and body weight measures found that periods of daily self-weighing were associated with weight loss, whereas breaks in weighing were linked to weight gain [8].

Self-monitoring weight has become considerably easier with mobile phone tracking apps and smart scales. These digital tracking devices increase adherence, that is, sticking to a regular frequency of monitoring, and improve weight loss outcomes [9,10], possibly because seeing progress increases motivation and keeps users on track with their goals [11]. However, data suggest that many people do not manage to continue weight tracking long term. In a weight loss trial, the percentage of participants adherent to self-weighing dropped from roughly 70% in the first week to 20% after 70 weeks [12]. Another study found that 25.0% (37/148) of the participants who were asked to weigh daily reduced their monitoring frequency significantly across the study, and approximately 8.8% (13/148) of the participants stopped weighing themselves altogether after 33 weeks [13]. Steinberg et al [14] observed that the average frequency of self-weighing of the participants in the daily weighing intervention group decreased from 6.1 to 4 days per week within 6 months, and 13% (6/47) of the participants stopped weighing themselves completely after the sixth month. Similarly, the frequency of physical activity tracking also declines over time. An observational study showed that, on average, people stop using their physical activity tracking devices after 129 days [15]. In another study, 80% (39/49) of the participants were no longer using their tracking device after 2 months, and 45% (22/49) of the participants did not intend to use it again in the future [16]. Considering that these rates of decline were found in a population of highly motivated individuals (ie, participants in a study), it seems plausible that long-term adherence to tracking may also be low in nonresearch contexts.

Research investigating why people stop using self-monitoring devices suggests that cost, concerns surrounding data sharing, flaws in the design and user experience of technology, as well as issues with data accuracy deter use of self-monitoring technologies [16-20]. Other user-internal reasons for abandonment have also been identified, including users experiencing a mismatch between their expectations and needs and the devices' capabilities, the users feeling that they have reached data saturation and can no longer learn from their data, and the users having reached their goals [16,20]. However, these data were collected from questionnaires and interviews occurring after abandonment and may suffer from participants retrospectively justifying their abandonment. Here, we use a prospective observational design to examine patterns preceding a stop in self-monitoring using actual tracking data from the Withings Health Mate app (Withings SA, France).

Objectives

Our first research question (RQ1) considered whether the pattern of weight and physical activity measurements and physical activity tracking frequency changes before people who are trying to lose weight cease weight monitoring. Approximately 80% of people approach weight loss by increasing their physical activity [21-23], so we considered physical activity measurements as a proxy for the engagement in weight loss behaviors and the motivation to lose weight. We tested two competing hypotheses against each other. First, that in the weeks preceding a stop in weight tracking, motivation to lose weight and track behavior would remain high (ie, stable levels of physical activity and physical activity tracking), but the user would receive weight readings that show unsatisfying weight loss progress (such as stable weight or weight gain), leading to frustration and perceived lack of ability to take control of the weight loss progress, making the user stop tracking his or her weight. The alternative hypothesis was that declining motivation for weight loss would manifest in declining weight loss efforts (as measured through decreasing physical activity). This would lead to unsatisfactory feedback, such as weight stability or gain, and to ceasing self-monitoring, potentially because of negative emotions such as shame.

Our second RQ examined physical activity data after users stopped tracking their weight. Our hypothesis was that ceasing self-monitoring of weight would undermine the motivation to engage in weight loss behaviors. We, therefore, expected to see a decrease in physical activity and physical activity tracking frequency after the stop in weight monitoring.

Methods

Dataset

The data provided by Withings comprised weight and physical activity records from 438,688 Withings Health Mate app users from January 1, 2014, to January 19, 2018. Withings Health Mate app users consented to the use of their data for research purposes by accepting the terms and conditions of Withings [24].

The dataset consisted of users who (1) were overweight when they started using the app (BMI ≥25 kg/m^2), (2) had set themselves a weight loss goal in the app, (3) used both the weight and the physical activity tracking features of the app, and (4) synchronized their weight data from Withings smart scales (as previous research indicates that users may underreport unfavorable weight measurements if done manually [25]).

Withings smart scales synchronize weight data to the app via Wi-Fi or Bluetooth. Smart scale owners can set up accounts for up to eight users, and the scale differentiates between users automatically during weighing. Weight measurements are synchronized to the app separately for each user. Physical activity is operationalized as daily step counts in the Health Mate app. The data can be synchronized via Bluetooth from Withings physical activity trackers and Apple watches or via linking to the Google Fit or Samsung Health app. The dataset included demographic information about each user, including gender, age, location, self-reported height, initial BMI, and target BMI.

Study Design and Data Screening

Only users who stopped tracking their weight for at least six weeks at some point between 2014 and 2018 were included in the analysis. We chose a period of 6 weeks to signify a stop to ensure that the lack of measurements was unlikely to be because of travel. To be eligible, participants also had to have a preceding 16-week phase of consistent weight and physical activity tracking (≥3 measurements per week), to ensure that only users who monitored their weight and physical activity regularly before the stop were included. A case-crossover design [26] was employed for both RQs. For RQ1, 4 weeks of data preceding the minimum 6-week long stop in weight tracking were compared with 4 weeks of data from the same user from the phase of consistent weight and physical activity tracking. We analyzed a 4-week period as we expected that the frequency of weight tracking would change gradually. For the 4 weeks preceding the stop in weight monitoring, there were no minimum requirements for monitoring frequency. Only users starting both analysis time periods with a BMI of 25 kg/m^2 or above were included. For RQ2, the analysis further included a third 4-week period, taken from the 6-week long break in weight monitoring, that is, post-stop. Where users had several 26-week periods that fulfilled the abovementioned criteria, the first period was analyzed. Previous research has shown that when individuals have set themselves an ambitious goal, they are more likely to be dissatisfied with their weight loss progress [27], probably because these goal weights are harder to achieve. Hence, if people are far away from their weight loss goal and receive frustrating weight measurements, they might also be more likely to disengage from their weight loss attempt. Here, users may have had significant changes in their weight between the two time periods analyzed, thus leading to different distance to goal weight scores. To ensure that these differences and their potential impact do not affect our analyses, we decided to control for the distance to goal weight.

Analysis

All analyses were conducted in R (version 3.4.1, R Development Core Team 2017, University of Auckland, New Zealand). A statistical analysis plan was published on Open Science Framework preceding the analyses [28].

Variables

The independent variables in the analyses were called time period and distance to goal weight. Time period was a factor with 2 or 3 levels (continuous tracking; pre-stop; and, in RQ2, post-stop) and defined which time period the dependent variable in question originated from. The continuous tracking period was located within the 16-week period of consistent tracking during which users recorded at least three weight and physical activity measurements per week. The pre-stop period was located immediately before the stop in weight tracking. The post-stop tracking period (RQ2 only) was located within the 6-week break in weight tracking. Which weeks the time periods encompassed differed slightly by analysis and is described for each analysis separately below. The distance to goal weight variable was calculated as a difference score between the weight measurement closest to the last date of the time period analyzed and goal weight.

To set up the analyses as within subject, we added the variable user ID as a random factor to all analyses. In RQ2 only, we also included the date of measurement as an independent variable to identify when the dependent variable analyzed was measured.

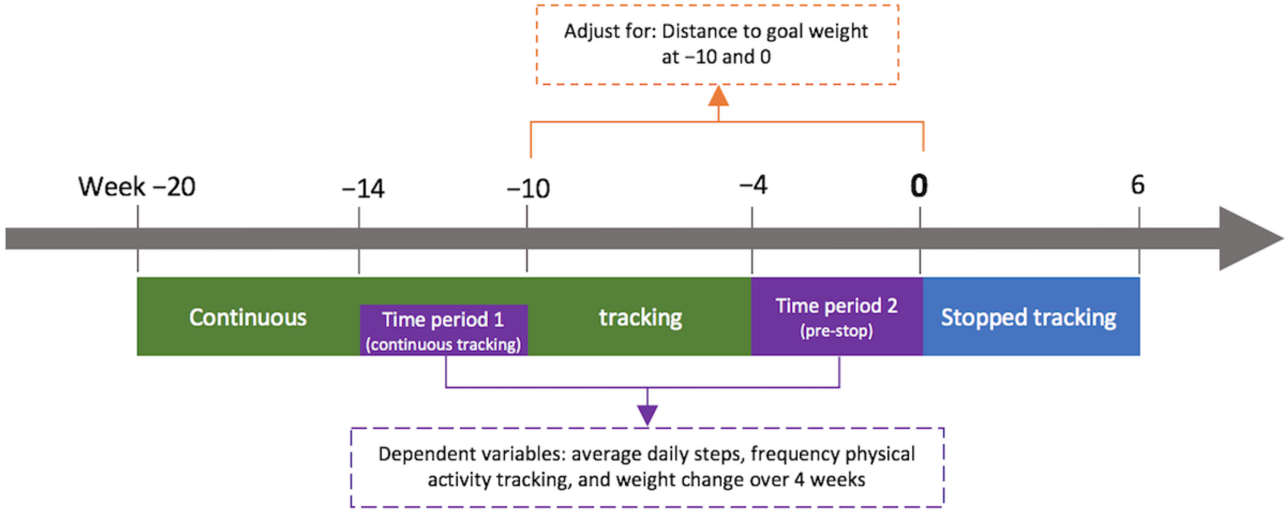
There were three dependent variables. The first was weight change, calculated as the difference between the first and the last measurement of each time period. The second was daily steps. For RQ1, this variable was calculated as the average daily step count across all days of a time period. For RQ2, the individual daily step measurements were used. In both cases, the daily steps variable was divided by 1000 to aid interpretation of the coefficients. The third dependent variable was physical activity tracking frequency, which was calculated as a sum score for the number of days for which physical activity measurements were available in each time period. As physical activity was treated as a proxy measure for motivation to lose weight, increases in daily step counts were interpreted as a strengthened weight loss effort and increased motivation to lose weight. Physical activity tracking frequency was interpreted as motivation to self-monitor weight loss efforts.

Research Question 1

For RQ1, we assessed whether average daily steps, weight change, and frequency of physical activity tracking differed between the two time periods: continuous tracking (weeks −14 to −10) and pre-stop (weeks −4 to 0, 29 days each). Figure 1 depicts an overview of the design.

Descriptive analysis examined the demographic characteristics of the analyzed sample and unadjusted differences between the two time periods on the dependent variables. Linear mixed effects models, matched by user ID, predicted the dependent variables, including weight change, average daily steps, and frequency of physical activity tracking, based on the binary variable time period. All analyses were adjusted for the distance to goal weight. We ran models twice, once using distance to goal weight as a random factor and another time as a fixed factor. Where a comparative analysis of variance (ANOVA) revealed a significant difference between the models, the model with the lower Akaike Information Criterion (AIC) was determined to be the best-fitting model. Where no significant difference was found in the ANOVA, the model with the lower degrees of freedom was chosen. Only the best-fitting models are reported here. To assess the sensitivity of our findings to the normality assumption in the random effects model, we compared the outputs with equivalent models fitted using generalized estimating equation (GEE; results presented in Multimedia Appendix 1).

Figure 1. Case-crossover study design to establish linear mixed effects models for research question 1.

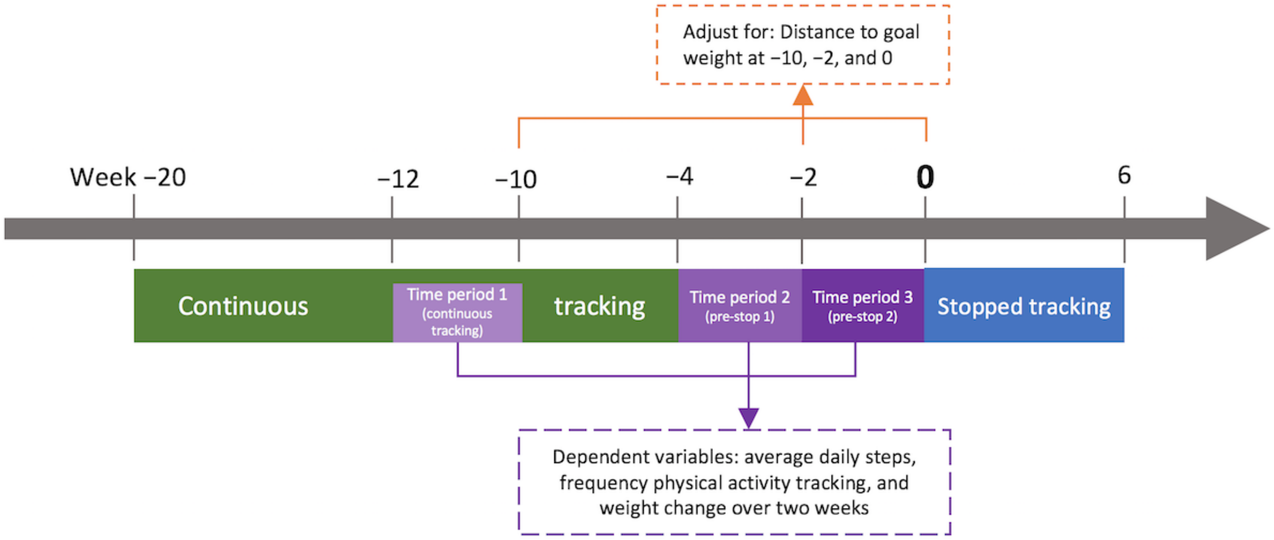


Our first hypothesis was that motivation to lose weight and track activity would remain high, but frustrating weight measurements would drive the user to stop weight tracking. Consequently, it follows that there would be no association between the binary time period variable and average daily steps and physical activity tracking frequency. However, time period would significantly predict weight change, as weight loss would be expected to lessen in the pre-stop period.

Our second hypothesis stated that users would lose motivation for weight loss behaviors, and the resulting unsatisfactory weight feedback would lead to a stop in weight tracking. Consequently, it follows that the binary time period variable would significantly predict physical activity levels, such that there would be a decline in average daily steps from the continuous to pre-stop periods. Time period would also significantly predict weight change, as users would be expected to have less satisfactory weight measurements after reducing their weight loss efforts in the pre-stop period. There were no specific predictions about the pattern of physical activity tracking frequency in this hypothesis.

To identify any temporal sequences of the abovementioned hypothesized effects, we ran post hoc analyses splitting the pre-stop period (weeks −4 to 0) into two 2-week periods (−4 to −2 and −2 to 0). We reran the analysis mentioned above, this time comparing three time periods with each other, namely, weeks −12 to −10 (period 1: continuous tracking) as a baseline, weeks −4 to −2 (period 2: pre-stop 1), and weeks −2 to 0 (period 3: pre-stop 2), each 15 days long. Time period was entered as a factor in the analysis. All analyses were run twice and adjusted for the distance to goal weight, once set as a random factor and once as a fixed factor. We report the better fitting models. Tukey-adjusted post hoc comparisons investigated pairwise comparisons of the three time periods. Again, the GEE models were run to assess the sensitivity of our findings to the normality assumption (results presented in Multimedia Appendix 1). The design of this analysis is depicted in Figure 2.

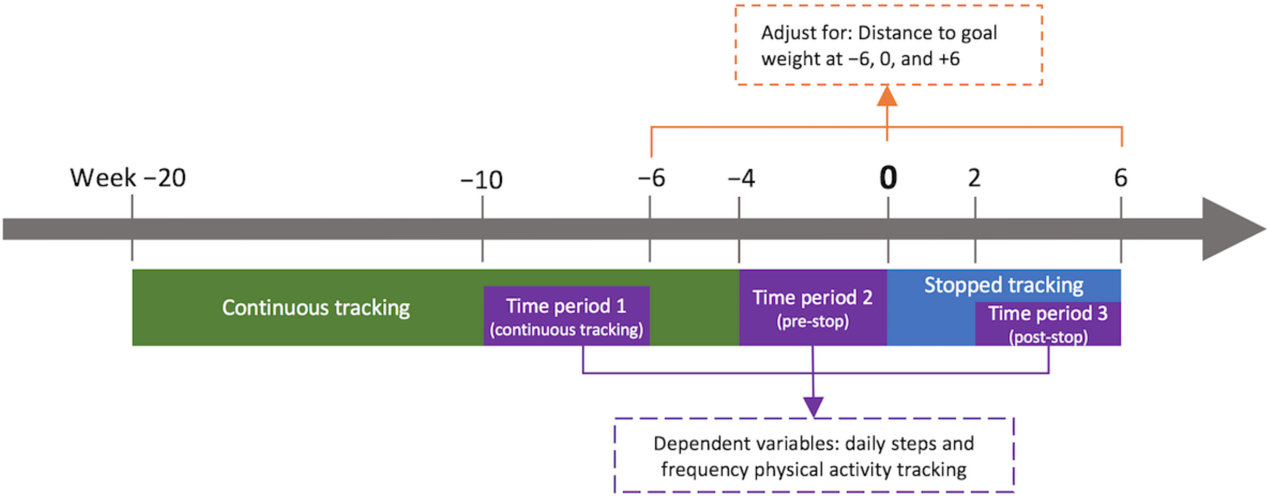
Figure 2. Design of the post hoc analysis, aiming to investigate temporal sequence of effects.



Research Question 2

For our second RQ, we ran analyses to compare daily steps and physical activity tracking frequency before and after the stop in weight tracking. Time period 1 (continuous tracking) for this analysis stemmed from weeks −10 to −6 of the continuous tracking phase, which ensured that users had at least three physical activity measurements per week. This time period was treated as the baseline. Time period 2 (pre-stop) comprised the 4 weeks preceding the stop in weight tracking (weeks −4 to 0). Period 3 (post-stop) comprised weeks 2 to 6 of the break in weight tracking (see Figure 3). The analyses were run separately for daily steps and frequency of physical activity tracking using linear mixed effects models. The analyses were adjusted for the distance to goal weight at the end of the three analysis time periods. Tukey-adjusted post hoc comparisons investigated pairwise comparisons of the three time periods.

Figure 3. Case-crossover study design to establish linear mixed effects models for research question 2.



Descriptive statistics explored unadjusted differences in daily steps between the three time periods. Linear mixed effects models predicted the dependent variable daily steps from the variables time period and date of measurement. A sequential testing approach was used:

1. Random effect: user ID; fixed effect: time period
2. Random effect: user ID; fixed effects: time period, date of measurement, and interaction time period×date of measurement
3. Random effect: user ID; fixed effects: time period, date of measurement, and interaction time period×date of measurement; adjusting for distance to goal weight.

The generalized variance inflation factor ($GVIF^{(1/[2\times df])}$) was calculated at stage 2 of sequential testing to check for multicollinearity of the predictors time period and date of measurement. The third model was run twice, once entering distance to goal weight as a fixed factor and another time as a random factor. In this paper, we only present the results of the best-fitting model, which again was identified through an ANOVA and AIC comparison. We also ran GEE on the best-fitting model to test sensitivity to the normality assumption (results presented in Multimedia Appendix 1).

The hypothesis predicted a significant main effect for time period, such that daily steps would significantly decrease across all three time periods. We also expected a significant interaction term for time period and date of measurement, as the date of measurement should only be a significant negative predictor in the second and third time periods. These findings would support the hypothesis that ceasing weight tracking leads to a decline in weight loss efforts.

The frequency of physical activity tracking was computed for the same three time periods, that is, weeks −10 to −6 (continuous tracking), −4 to 0 (pre-stop), and +2 to +6 (post-stop, 29 days each). Descriptive statistics explored unadjusted differences in physical activity tracking frequency between the three time periods. A linear mixed effects model predicted physical activity tracking frequency, with user ID as a random effect and time period as a fixed effect. The analysis was adjusted for the distance to goal weight at the end of each time period. Again, distance to goal weight was entered once as a random factor and once as fixed factor, and only the best-fitting model is reported.

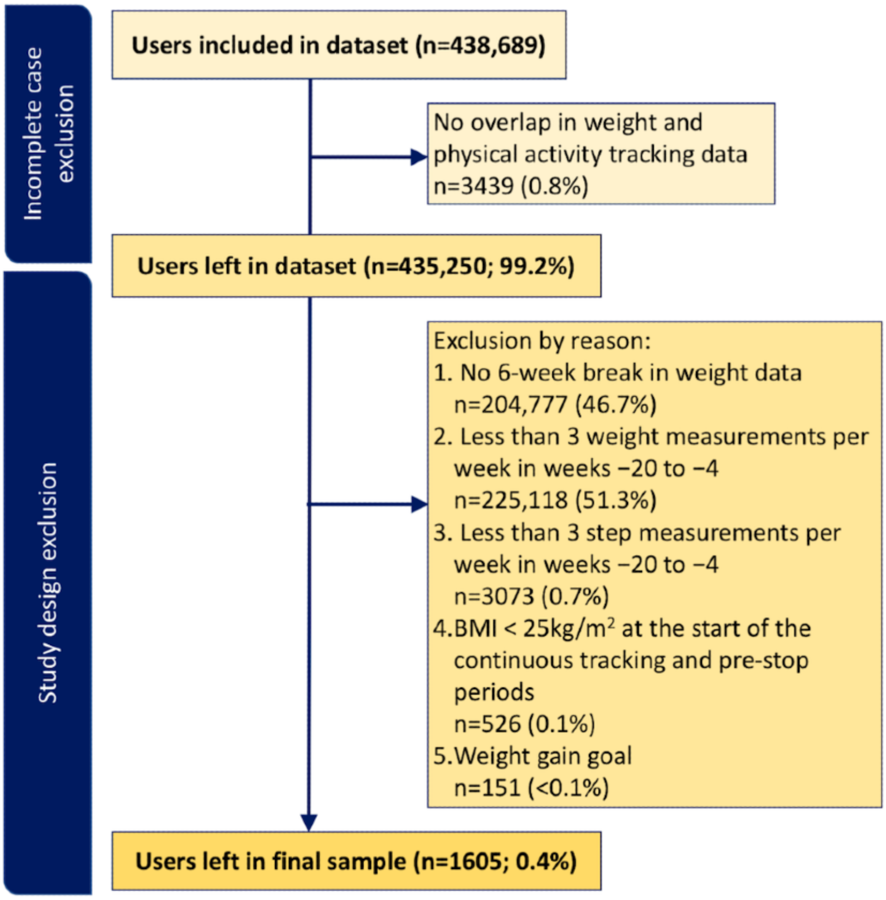
For our hypothesis to be correct, we expected to find significant decreases in tracking frequency across the three time periods, as this would indicate that the stop in weight tracking signals a stop in wanting to monitor weight loss efforts.

Results

Sample

The final sample consisted of 1318 male and 287 female users, the average age was 49.0 years (SD 12.5), and the average BMI at week −14 was 30.2 kg/m^2 (SD 4.7). The reasons for exclusion are shown in Figure 4. The final dataset covered 221,173 weight and 113,162 physical activity measurements. Nearly half (778/1605, 48.47%) of the users included in the analysis were based in Europe, 36.39% (584/1605) in North America, and the rest split across other continents.

Figure 4. Flow diagram of the exclusion rates at each step.



Research Question 1

Weight Change

During the continuous tracking period, participants lost weight (mean −0.47 kg, SD 1.73), but pre-stop, they gained weight (mean 0.25 kg, SD 1.62). In the mixed effects model, the time period significantly predicted weight change, revealing a 0.71 kg (95% CI 0.60 to 0.81) mean difference in weight change from the continuous to pre-stop period (see Figure 5).

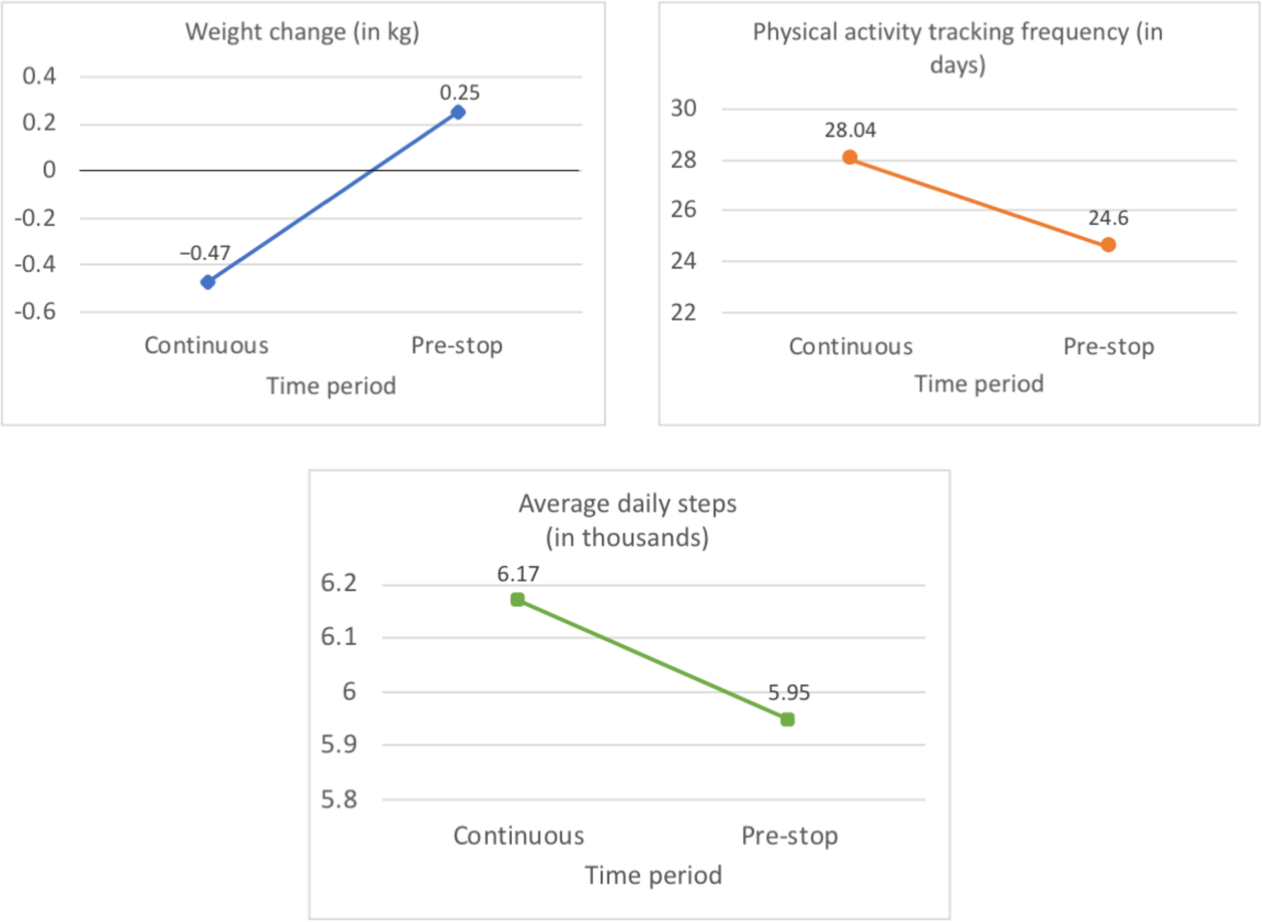
Physical Activity Tracking Frequency

Participants recorded physical activity on 3.44 (95% CI −3.78 to −3.10) fewer days during the pre-stop period compared with the continuous tracking period (see Figure 5).

Average Daily Steps

A total of 19 users completely stopped tracking their physical activity during the pre-stop period, meaning that average daily steps could not be calculated. These users were excluded, leaving 1586 in the analysis. Pre-stop, participants took 220 (95% CI −320 to −210) fewer steps per day than during continuous tracking (see Figure 5).

Figure 5. Results of the three linear mixed effects models for research question 1.



Post Hoc Analyses

As in the prespecified analysis, users lost weight during the 2 weeks of continuous tracking (mean −0.29, SD 1.23) and gained weight in the first and second half of the pre-stop period (pre-stop 1: mean 0.11, SD 1.28; pre-stop 2: mean 0.14, SD 1.28). Post hoc comparisons revealed a 0.40 kg (95% CI 0.31 to 0.49) and 0.43 kg (95% CI 0.34 to 0.51) mean increase in weight change between the continuous tracking period and the two time periods preceding the stop in weight tracking, respectively. The two pre-stop periods did not differ significantly from each other (mean increase of 0.03 kg; 95% CI −0.06 to 0.11). The results of the best-fitting model are presented in Table 1 and Figure 6.

Physical activity tracking frequency significantly decreased by 1.28 days (95% CI −1.48 to −1.09) between the continuous tracking period and pre-stop 1. It further decreased significantly by 0.96 days (95% CI −1.16 to −0.77) between the pre-stops 1 and 2. The results of the best-fitting model are presented in Table 1 and Figure 6.

In total, 126 users completely stopped tracking their physical activity during the two pre-stop periods, meaning that average daily steps could not be calculated. We excluded these users from the analysis, reducing the sample size to 1479 users. Post hoc comparisons revealed that physical activity significantly decreased by an average of 180 steps (95% CI −290 to −70) between the continuous tracking period and pre-stop 1 and another 130 steps (95% CI −240 to −20) between the first and second half of the pre-stop period. The results of the best-fitting model are presented in Table 1 and Figure 6.

Table 1. Best-fitting models for the three dependent variables in the post hoc analyses.

Best-fitting model	Dependent variables		
	Weight change (kg)[a]	Physical activity tracking frequency (days)[b]	Average steps (thousands)[c]
Coefficients, beta (95% CI)[d]			
Continuous vs pre-stop 1	.40 (0.31 to 0.49)	−1.28 (−1.48 to −1.09)	−.18 (−0.29 to −0.07)
Continuous vs pre-stop 2	.43 (0.34 to 0.51)	−2.25 (−2.44 to −2.05)	−.31 (−0.42 to −0.20)
Tukey-adjusted post hoc comparisons, beta (95% CI)[d]			
Continuous vs pre-stop 1	.40 (0.31 to 0.49)	−1.29 (−1.48 to −1.09)	−.18 (−0.29 to −0.07)
Continuous vs pre-stop 2	.43 (0.34 to 0.52)	−2.25 (−2.44 to −2.05)	−.31 (−0.42 to −0.20)
Pre-stop 1 vs pre-stop 2	.03 (−0.06 to 0.11)	−.96 (−1.16 to −0.77)	−.13 (−0.24 to −0.02)

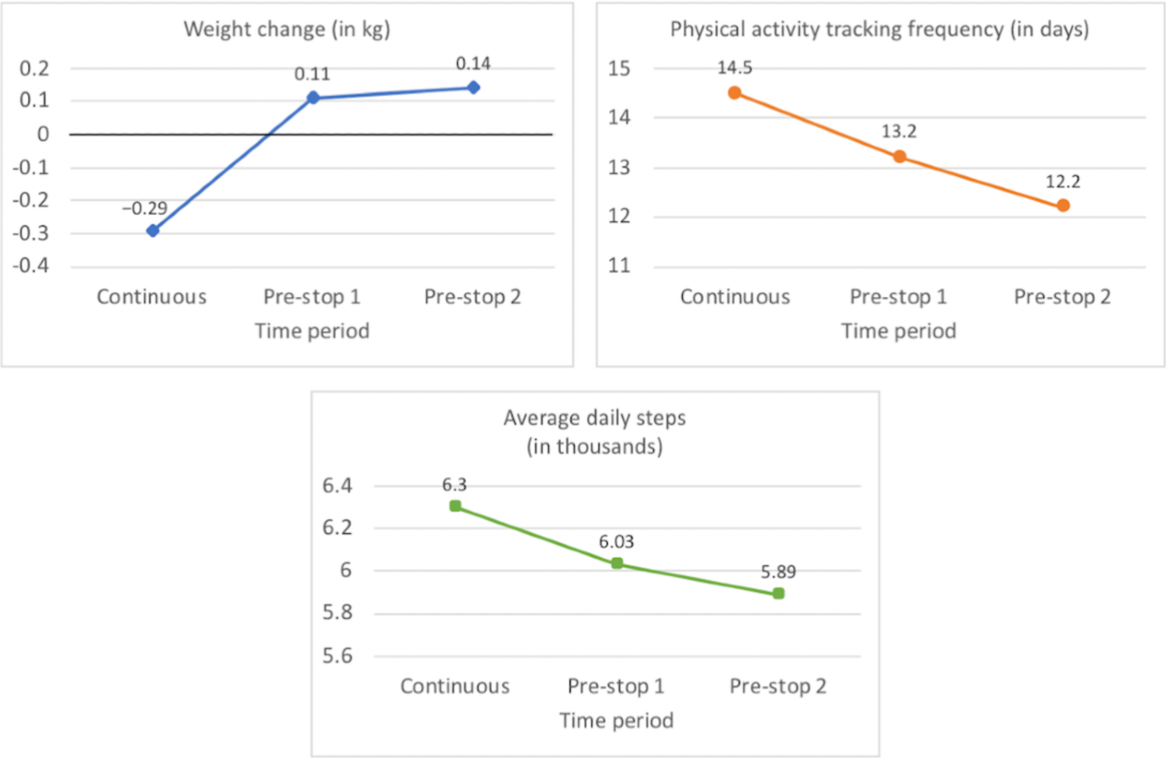
[a] Fixed effects: time period; random effects: distance to goal weight and user ID.

[b] Fixed effects: time period and distance to goal weight; random effects: user ID.

[c] Fixed effects: time period and distance to goal weight; random effects: user ID.

[d] Coefficients represent mean differences between the three time periods.

Figure 6. Results of the three linear mixed effects models of the post hoc analysis.



Research Question 2

Daily Steps

Post hoc comparisons of the first mixed effects model, entering time period as a fixed factor and user ID as a random factor, showed a significant decrease of 200 daily steps (95% CI −250 to −150) from the continuous tracking to pre-stop period and a significant increase of 120 daily steps (95% CI 60 to 180) from the pre-stop to post-stop period.

In a second model, we added the variable date of measurement and the interaction term time period×date of measurement to see whether there were within-period effects, but the $\mathrm{GVIF}^{(1/[2\times df])}$ values for the time period variable and the interaction term were above 50, indicating strong multicollinearity. We, therefore, excluded the date of measurement variable from all further analyses. In the final stage of sequential testing, post hoc comparisons of the mixed effects model adjusting for the distance to goal weight revealed a significant decrease of 190 steps (95% CI −240 to −130) from the continuous to pre-stop period. It also revealed a significant increase of 110 steps (95% CI 50 to 170) from the pre-stop to post-stop period. A small but significant decrease of 70 steps (95% CI −130 to −10) was found in the comparison between the continuous tracking and the post-stop period (see Figure 7). A summary of all results can be found in Table 2.

Figure 7. Results of the linear mixed effects models for research question 2.

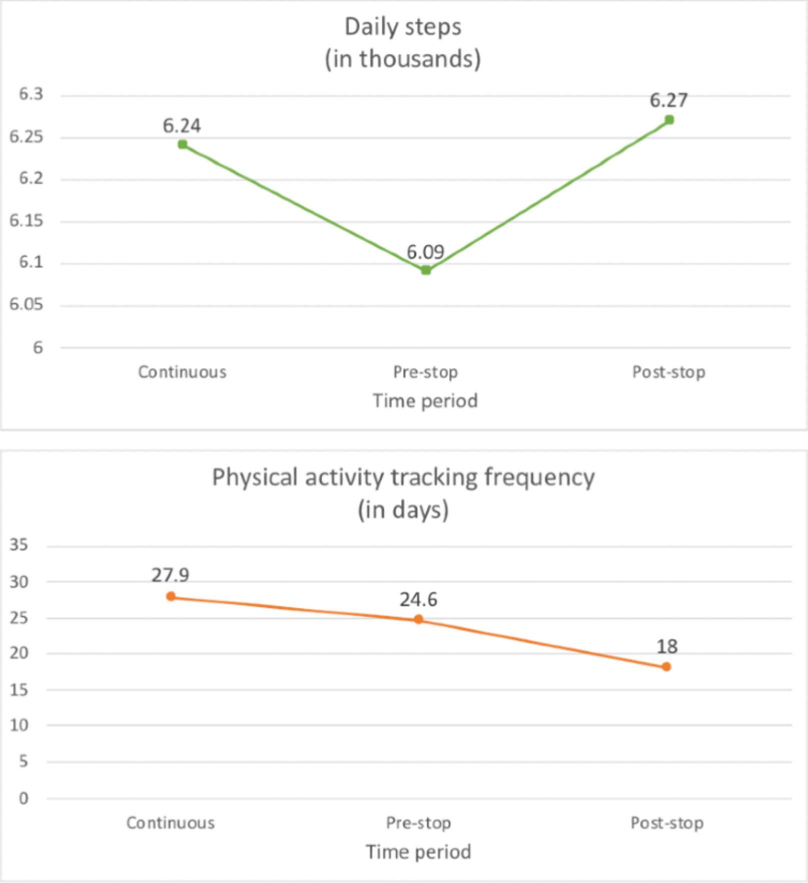


Table 2. Results of the daily steps data analysis of research question 2.

Daily steps data analysis	Model 1[a]	Model 2[b]	Model 3[c]
Coefficients, beta (95% CI)[d]			
Continuous vs pre-stop	−.20 (−0.25 to −0.15)	−2.12 (−4.79 to 0.54)	−.19 (−0.24 to −0.13)
Continuous vs post-stop	−.08 (−0.14 to −0.02)	−6.16 (−9.19 to −3.12)	−.07 (−0.13 to −0.01)
Date of measurement	N/A[e]	−.00 (−0.00 to 0.00)	N/A
Continuous vs pre-stop×date of measurement	N/A	−.00 (−0.00 to 0.00)	N/A
Continuous vs post-stop×date of measurement	N/A	−.00 (−0.00 to 0.00)	N/A
Tukey-adjusted post hoc comparisons, beta (95% CI)[d]			
Continuous vs pre-stop	−.20 (−0.25 to −0.15)	N/A	−.18 (−0.24 to −0.13)
Continuous vs post-stop	−.08 (−0.14 to −0.02)	N/A	−.07 (−0.13 to −0.01)
Pre-stop vs post-stop	.12 (0.06 to 0.18)	N/A	.11 (0.05 to 0.17)

[a] Fixed effects: time period; random effects: user ID.

[b] Fixed effects: time period, date of measurement, and interaction term time period×date of measurement; random effects: user ID.

[c] Fixed effects: time period; random effects: distance to goal weight and user ID.

[d] Coefficients represent mean differences between the three time periods.

[e] Not applicable.

Physical Activity Tracking Frequency

Post hoc comparisons revealed that physical activity tracking frequency significantly decreased by 3.3 days (95% CI −3.82 to −2.85) between the continuous tracking and pre-stop period and another 6.6 days (95% CI −7.12 to −6.16) between the pre-stop and post-stop period. The results of the best-fitting model are presented in Table 3 and Figure 7.

Table 3. Results of the physical activity tracking frequency analysis of research question 2.

Physical activity tracking frequency analysis	Model[a]
Coefficients, beta (95% CI)[b]	
Continuous vs pre-stop	−3.33 (−3.82 to −2.85)
Continuous vs post-stop	−9.98 (−10.50 to −9.50)
Tukey-adjusted post hoc comparisons, beta (95% CI)[b]	
Continuous vs pre-stop	−3.34 (−3.82 to −2.86)
Continuous vs post-stop	−9.98 (−10.50 to −9.50)
Pre-stop vs post-stop	−6.64 (−7.12 to −6.16)

[a] Fixed effects: time period and distance to goal weight; random effects: user ID.

[b] Coefficients represent mean differences between the three time periods.

Discussion

Principal Findings

The analyses targeting the first RQ revealed that a stop in weight tracking is preceded by decreased step counts, lower physical activity tracking frequencies, and weight gain. The findings thus counter our first hypothesis, which had stated that physical activity (ie, motivation to lose weight) would remain stable, whereas a weight gain would precede the stop in weight tracking. The post hoc analysis showed that the changes in weight and physical activity developed concurrently, as effects appeared at the same time, in pre-stop 1. The results, therefore, only partially support our second hypothesis, as they do not reveal the sequential effect theorized: first, a decrease in physical activity, followed by weight gain and a stop in weight tracking.

Regarding the second RQ, we found a decrease in the frequency of physical activity monitoring but an increase in physical activity on days when activity was recorded after users stopped tracking their weight. Our hypothesis, which stated that users would show a decrease in both physical activity levels and tracking frequency, is, therefore, only partially supported.

Users Gain Weight Before They Stop Tracking Their Weight

The literature provides abundant cross-sectional evidence that ceasing regular weighing and weight gain are associated [6-8,29,30]. This has been interpreted to indicate that reduced tracking frequency leads to weight gain. However, our findings suggest that the relationship could be reversed: in this analysis, weight gain preceded the stop in weight monitoring. We reached a similar conclusion in our recent review of the qualitative literature of experiences of self-directed weight loss [31]. These results can be explained in terms of *The Ostrich Problem*, which proposes that people avoid outcome information when it shows that progress is poor or it elicits negative emotions [32,33]. In line with this, previous research on weight loss has shown that people who anticipate negative feedback from the scales choose not to weigh themselves to avoid negative feelings [17,34-36]. In a study by Mintz et al [21], 63.1% (99/157) of the female participants reported reacting emotionally to weight measurements, and half of the participants felt that the weight measurements affected their feelings of self-worth. Taken together, it seems that some people struggle receiving negative feedback from the scales and, therefore, stop exposing themselves to the information. The lack of constructive use of weighing feedback, hence, leads to a stop in engagement with self-monitoring, which is a necessary step for self-regulation. The insights provided by these findings open up avenues for intervention because it is plausible that helping users reframe negative weight measurements as constructive feedback could aid successful self-regulatory processes.

Users Reduce Weight Loss Efforts Before Ceasing Weight Tracking

Users engaged in less physical activity and reduced the frequency of monitoring physical activity before they stopped tracking their weight. As attempting to increase energy expenditure through physical activity is one of the most common approaches to weight loss [21-23], we interpret these changes as reductions in the motivation to lose weight, leading users to stop weight tracking. Previous research also supports the notion that levels of motivation might be connected to weight monitoring adherence. One study reported that autonomous motivation predicted adherence to self-monitoring [37]. Two further studies found that measures related to motivation and weight loss behaviors, including goal ownership, weight loss expectations, and estimated weight loss skills, were negatively associated with subsequent dropout from weight loss treatment [38,39].

The decline of physical activity and increase in weight occurred concurrently, and it is possible that the two aspects may have influenced and reinforced each other. That is, reduced motivation to lose weight and thus reduced physical activity might have led to weight gain, which, in turn, might have reinforced the decline in motivation to track weight, bringing the whole self-regulatory system to a halt. There is a little empirical evidence to support this, including a study by Webber et al [37] who found that the positive association between autonomous motivation and adherence to self-monitoring was mediated by weight loss.

Consequences of Stopping Weight Monitoring

The question arises why users reduced the frequency of physical activity monitoring but increased their daily step count after stopping weight monitoring. There is some evidence that users are more likely to track favorable weight measurements [25], and it is thus conceivable that users tended to only wear their tracking device on their more active days, leading to less tracking but higher average physical activity on days with data. In addition, users may have had less interest in monitoring their physical activity, given that they could no longer put the measurements into context with weight data, a prerequisite to evaluating behavior and progress for self-regulation. In line with this hypothesis, qualitative studies of people who autonomously track health parameters have reported that a common aim is to identify correlations between the measurements of their different health parameters [36,40] to gain a better understanding of their body and to experiment with different ideas, for example, whether more sleep helps with weight loss [36]. Tracking more than one parameter has been associated with better adherence overall, possibly because of mutual reinforcement [41]. Having no longer collected information about their weight and thus being no longer able to complete the self-regulation process, the users in our sample might have perceived less value in their physical activity data, leading to a reduced tracking frequency.

It is notable that although the frequency of physical activity monitoring declined, users did not stop monitoring their physical activity completely, although they had stopped monitoring their weight. One reason for this discrepancy might be that daily physical activity measurements are independent of each other, whereas weight measurements are highly autocorrelated. That is, although a weight measurement from one day necessarily predicts the weight measurement of the next day, physical activity measurements are reset to zero at the start of each day, and the participant might, therefore, find it easier to start afresh.

Strengths and Limitations

One strength of this analysis is that it examines the patterns of self-monitoring behavior in a setting in which the users were not aware that they were being observed. This reduces biases such as the Hawthorne effect [37,42], increasing the external validity of our findings. Our sample size was considerably larger than most researcher-led studies. Covering 221,173 weight and 113,162 physical activity measurements from 1605 users, our analysis was well powered to detect pattern changes in the data. Context information for the users was, however, sparse. The dataset includes a high proportion of men, which is rare in the context of weight loss studies, but we lack information regarding socioeconomic status and ethnicity, making it difficult to gauge generalizability. However, the spread of users across countries means it is likely to encompass a broader mix of ethnic backgrounds and lifestyles than most single-country studies. Unfortunately, we do not know whether the users we analyzed participated in any kind of weight loss program. However, as our analyses were conducted within individuals, confounding differences between individuals were removed. Another disadvantage of using context-restricted app data is that we had no information on users' intentions. For instance, we do not know whether the stop in weight data actually reflects a stop in weight tracking. Users might instead have switched to a different app to track their weight measurements. We designed the study to minimize this possibility. The minimum period of continuous tracking of 16 weeks reduced the chance that the users were just *trying out* the app [43]. Only one-fourth of the downloaded health apps are used long term, and the cutoff lies around the tenth use [44]. Users of tracking apps become increasingly bound to their app of choice, as data on past physical activity and weight measurements cannot easily be transferred between apps [11]. We can, therefore, assume that after 16 weeks of continuous tracking, users who stop tracking their weight did not simply switch apps. It, however, remains possible that users continued weighing themselves without synchronizing the data to the app. Nevertheless, because we observed a significant weight gain and decrease in physical activity before the stop in weight tracking, it seems that the stop in weight tracking did occur in the context of waning weight loss efforts. Unfortunately, we did not have access to matched dietary data, which would have added another important proxy measure of the motivation to lose weight to the model. We do, however, believe that physical activity should be a good indicator of motivation on its own because roughly 80% of people report increasing their physical activity as part of their weight loss efforts [21-23].

A necessary feature of our study design is that we restricted our analysis to a group of people who monitored themselves very frequently. Individuals who self-monitor very frequently are likely to represent a highly motivated population [7], but this limitation is comparable with that of any other weight loss study where people who choose to take part have an intrinsic motivation to make a weight loss attempt. A final limitation is that because the design of this research is observational, we are unable to make causal inferences from the data. Future research using other designs is necessary to further investigate the patterns we identified.

Conclusions

This observational analysis shows that before people who weighed themselves regularly cease doing so, their weight increases and their physical activity intensity and tracking frequency declines. This probably indicates that a stop in regular weighing occurs in the context of a decline in weight loss efforts. After ceasing to track weight, there is a decline in the frequency of monitoring physical activity and an increase in the daily steps taken on the days monitored. The stop in weighing can be interpreted as a halt in self-regulation for weight loss, as progress is no longer monitored and cannot be evaluated.

Our results indicate that phases of concurrent weight gain and decreases in physical activity constitute an appropriate time for intervention. Programs tackling declining motivation and helping with the constructive use of weight monitoring may, therefore, have the potential to help users stay on track with self-regulation and their weight loss efforts.

Acknowledgments

The authors would like to thank Withings SA for kindly providing them with the data for this study. The authors would also like to thank Eva Roitmann and Nicolas Genain at Withings for their input in the design and analysis stages of the study. This research was funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Oxford (CLAHRC) at Oxford Health National Health Service (NHS) Foundation Trust. KF's time on this project was funded by NIHR CLAHRC Oxford at Oxford Health NHS Foundation Trust, Wolfson College, University of Oxford (Oxford-Wolfson Marriott-Primary Care Graduate Scholarship), and NIHR School for Primary Care Research. JB's, SJ's, and PA's time on this project was funded by the NIHR Oxford Biomedical Research Centre. PA and SJ are NIHR senior investigators. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Wellcome Trust, or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results of GEE Models.
[PDF File (Adobe PDF File), 46 KB - jmir_v22i3e15790_app1.pdf]

References

1. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
2. Boutelle K. Weighing the evidence: benefits of regular weight monitoring for weight control. J Nutr Educ Behav 2006;38(3):131. [doi: 10.1016/j.jneb.2006.04.002] [Medline: 16731444]
3. Kanfer FH, Karoly P. Self-control: a behavioristic excursion into the lion's den. Behavior Therapy 1972;3(3):398-416. [doi: 10.1016/s0005-7894(72)80140-0]
4. 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. [doi: 10.1037/0033-2909.92.1.111] [Medline: 7134324]
5. LaRose JG, Lanoye A, Tate DF, Wing RR. Frequency of self-weighing and weight loss outcomes within a brief lifestyle intervention targeting emerging adults. Obes Sci Pract 2016 Mar;2(1):88-92 [FREE Full text] [doi: 10.1002/osp4.24] [Medline: 27668087]
6. Steinberg DM, Bennett GG, Askew S, Tate DF. Weighing every day matters: daily weighing improves weight loss and adoption of weight control behaviors. J Acad Nutr Diet 2015 Apr;115(4):511-518 [FREE Full text] [doi: 10.1016/j.jand.2014.12.011] [Medline: 25683820]
7. Pourzanjani A, Quisel T, Foschini L. Adherent use of digital health trackers is associated with weight loss. PLoS One 2016;11(4):e0152504 [FREE Full text] [doi: 10.1371/journal.pone.0152504] [Medline: 27049859]
8. Helander EE, Vuorinen AL, Wansink B, Korhonen IK. Are breaks in daily self-weighing associated with weight gain? PLoS One 2014;9(11):e113164 [FREE Full text] [doi: 10.1371/journal.pone.0113164] [Medline: 25397613]
9. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
10. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
11. Frie K, Hartmann-Boyce J, Jebb S, Albury C, Nourse R, Aveyard P. Insights from Google Play Store user reviews for the development of weight loss apps: mixed-method analysis. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e203 [FREE Full text] [doi: 10.2196/mhealth.8791] [Medline: 29273575]
12. Zheng Y, Sereika SM, Ewing LJ, Danford CA, Terry MA, Burke LE. Association between self-weighing and percent weight change: mediation effects of adherence to energy intake and expenditure goals. J Acad Nutr Diet 2016 Apr;116(4):660-666 [FREE Full text] [doi: 10.1016/j.jand.2015.10.014] [Medline: 26727241]
13. 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]
14. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. Obesity (Silver Spring) 2013 Sep;21(9):1789-1797 [FREE Full text] [doi: 10.1002/oby.20396] [Medline: 23512320]
15. Hermsen S, Moons J, Kerkhof P, Wiekens C, de Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]

16. Lazar A, Koehler C, Tanenbaum J, Nguyen DH. Why We Use and Abandon Smart Devices. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7 - 11, 2015; Osaka, Japan p. 635-646. [doi: 10.1145/2750858.2804288]
17. Epstein DA, Caraway M, Johnston C, Ping A, Fogarty J, Munson SA. Beyond Abandonment to Next Steps: Understanding and Designing for Life after Personal Informatics Tool Use. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 7 - 12, 2016; San Jose, CA, USA p. 1109-1113 URL: http://europepmc.org/abstract/MED/28503678 [doi: 10.1145/2858036.2858045]
18. Shih PC, Han K, Poole E, Rosson M, Carroll JM. Use and Adoption Challenges of Wearable Activity Trackers. In: Proceedings of the 2015 iConference. 2015 Presented at: iConference'15; March 24-27, 2015; Newport Beach, CA, USA.
19. Harrison D, Marshall P, Bianchi-Berthouze N, Bird J. Activity Tracking: Barriers, Workarounds and Customisation. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7 - 11, 2015; Osaka, Japan p. 617-621. [doi: 10.1145/2750858.2805832]
20. Clawson J, Pater JA, Miller AD, Mynatt ED, Mamykina L. No Longer Wearing: Investigating the Abandonment of Personal Health-Tracking Technologies on Craigslist. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7 - 11, 2015; Osaka, Japan p. 647-658. [doi: 10.1145/2750858.2807554]
21. Mintz LB, Awad GH, Stinson RD, Bledman RA, Coker AD, Kashubeck-West S, et al. Weighing and body monitoring among college women: the scale number as an emotional barometer. J College Stud Psychother 2013;27(1):78-91. [doi: 10.1080/87568225.2013.739039]
22. Tyler DO, Allan JD, Alcozer FR. Weight loss methods used by African American and Euro-American women. Res Nurs Health 1997 Oct;20(5):413-423. [doi: 10.1002/(sici)1098-240x(199710)20:5<413::aid-nur5>3.0.co;2-r] [Medline: 9334795]
23. Hartmann-Boyce J, Aveyard P, Koshiaris C, Jebb SA. Development of tools to study personal weight control strategies: OxFAB taxonomy. Obesity (Silver Spring) 2016 Feb;24(2):314-320 [FREE Full text] [doi: 10.1002/oby.21341] [Medline: 26748902]
24. Withings. 2019. Privacy Policy for Withings Products and Services URL: https://www.withings.com/jp/en/legal/withings-services-terms-and-conditions#/jp/en/legal/privacy-policy [accessed 2015-03-19]
25. Ross KM, Eastman A, Wing RR. Accuracy of self-report versus objective smart-scale weights during a 12-week weight management intervention. Obesity (Silver Spring) 2019 Mar;27(3):385-390. [doi: 10.1002/oby.22400] [Medline: 30703282]
26. Maclure M, Mittleman MA. Should we use a case-crossover design? Annu Rev Public Health 2000;21:193-221. [doi: 10.1146/annurev.publhealth.21.1.193] [Medline: 10884952]
27. Wadden TA, Berkowitz RI, Sarwer DB, Prus-Wisniewski R, Steinberg C. Benefits of lifestyle modification in the pharmacologic treatment of obesity: a randomized trial. Arch Intern Med 2001 Jan 22;161(2):218-227. [doi: 10.1001/archinte.161.2.218] [Medline: 11176735]
28. Frie K, Hartmann-Boyce J, Oke J, Jebb S, Aveyard P. Open Science Framework.: Open Science Framework; 2018. When People Stop Self-Regulating: Analysing Data From a Physical Activity and Weight Tracking Application URL: https://osf.io/n7qgv/ [accessed 2018-12-18]
29. VanWormer JJ, Martinez AM, Martinson BC, Crain AL, Benson GA, Cosentino DL, et al. Self-weighing promotes weight loss for obese adults. Am J Prev Med 2009 Jan;36(1):70-73. [doi: 10.1016/j.amepre.2008.09.022] [Medline: 18976879]
30. Burke LE, Sereika SM, Music E, Warziski M, Styn MA, Stone A. Using instrumented paper diaries to document self-monitoring patterns in weight loss. Contemp Clin Trials 2008 Mar;29(2):182-193 [FREE Full text] [doi: 10.1016/j.cct.2007.07.004] [Medline: 17702667]
31. Hartmann-Boyce J, Boylan A, Jebb SA, Aveyard P. Experiences of self-monitoring in self-directed weight loss and weight loss maintenance: systematic review of qualitative studies. Qual Health Res 2019 Jan;29(1):124-134. [doi: 10.1177/1049732318784815] [Medline: 29984630]
32. Webb TL, Chang BP, Benn Y. ‘The Ostrich Problem’: motivated avoidance or rejection of information about goal progress. Soc Personal Psychol Compass 2013;7(11):794-807. [doi: 10.1111/spc3.12071]
33. Chang BP, Webb TL, Benn Y. Why do people act like the proverbial ostrich? Investigating the reasons that people provide for not monitoring their goal progress. Front Psychol 2017;8:152 [FREE Full text] [doi: 10.3389/fpsyg.2017.00152] [Medline: 28228740]
34. Chambers JA, Swanson V. Stories of weight management: factors associated with successful and unsuccessful weight maintenance. Br J Health Psychol 2012 May;17(2):223-243. [doi: 10.1111/j.2044-8287.2011.02030.x] [Medline: 22107029]
35. McKee H, Ntoumanis N, Smith B. Weight maintenance: self-regulatory factors underpinning success and failure. Psychol Health 2013;28(10):1207-1223. [doi: 10.1080/08870446.2013.799162] [Medline: 23767689]
36. Meyer J, Beck E, Wasmann M, Boll S. Making Sense in the Long Run: Long-Term Health Monitoring in Real Lives. In: Proceedings of the 2017 IEEE International Conference on Healthcare Informatics. 2017 Presented at: ICHI'17; August 23-26, 2017; Park City, Utah, USA p. 285-294. [doi: 10.1109/ichi.2017.11]
37. Webber KH, Tate DF, Ward DS, Bowling JM. Motivation and its relationship to adherence to self-monitoring and weight loss in a 16-week internet behavioral weight loss intervention. J Nutr Educ Behav 2010;42(3):161-167. [doi: 10.1016/j.jneb.2009.03.001] [Medline: 20138583]

38. Huisman S, Maes S, De Gucht VJ, Chatrou M, Haak HR. Low goal ownership predicts drop-out from a weight intervention study in overweight patients with type 2 diabetes. Int J Behav Med 2010 Sep;17(3):176-181 [FREE Full text] [doi: 10.1007/s12529-009-9071-3] [Medline: 20033629]
39. Fowler JL, Follick MJ, Abrams DB, Rickard-Figueroa K. Participant characteristics as predictors of attrition in worksite weight loss. Addict Behav 1985;10(4):445-448. [doi: 10.1016/0306-4603(85)90044-9] [Medline: 4091079]
40. Choe EK, Lee NB, Lee B, Pratt W, Kientz JA. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14. 2014 Presented at: CHI'14; April 26 - May 1, 2014; Toronto, Canada p. 1143-1152. [doi: 10.1145/2556288.2557372]
41. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing self-monitoring strategies for weight loss in a smartphone app: randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e12209 [FREE Full text] [doi: 10.2196/12209] [Medline: 30816851]
42. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. Br Med J 2015 Sep 4;351:h4672. [doi: 10.1136/bmj.h4672] [Medline: 26341898]
43. 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 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
44. McLean VA. PRWeb. 2011 Apr 25. Motivating Patients to Use Smartphone Health Apps URL: http://www.prweb.com/releases/2011/04/prweb5268884.htm [accessed 2019-04-02]

Abbreviations

AIC: Akaike Information Criterion
ANOVA: analysis of variance
CLAHRC: Collaboration for Leadership in Applied Health Research and Care Oxford
GEE: generalized estimating equation
GVIF: generalized variance inflation factor
NHS: National Health Service
NIHR: National Institute for Health Research
RQ: research question

Edited by G Eysenbach; submitted 08.08.19; peer-reviewed by C Eisenhauer, Y Zheng, AL Vuorinen, T Webb; comments to author 21.10.19; revised version received 13.11.19; accepted 31.12.19; published 17.03.20.

Please cite as:
Frie K, Hartmann-Boyce J, Jebb S, Oke J, Aveyard P
Patterns in Weight and Physical Activity Tracking Data Preceding a Stop in Weight Monitoring: Observational Analysis
J Med Internet Res 2020;22(3):e15790
URL: http://www.jmir.org/2020/3/e15790/
doi:10.2196/15790
PMID:32181749

Original Paper

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:

This is a corrected version. See correction statement: http://www.jmir.org/2020/5/e20134/

Abstract

Background: Active video games (AVGs) can increase physical activity (PA) and help produce higher physiological expenditure. Animated narrative videos (NVs) possess unique immersive and motivational properties. When added to AVGs, they have been found to increase moderate-to-vigorous physical activity (MVPA) as opposed to the original no video condition. However, there is no evidence of whether that was due to the NV or the addition of an animated video to an AVG.

Objective: This study aimed to investigate the differential effect of adding an NV versus a nonnarrative video (NNV) to an AVG on PA parameters and physiological responses and to explore the mediating role of immersion.

Methods: A total of 22 children aged 8 to 12 years were randomly assigned to the NV or NNV condition. They were instructed to play an AVG (on Xbox Kinect) for as long as they wanted. We used accelerometers to estimate the time spent (in minutes) in MVPA. Heart rate (HR) and rate of perceived exertion (RPE) were measured before, during, and after the AVG play session. The participants then reported their experience of narrative immersion via a questionnaire.

Results: 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. Narrative immersion was positively correlated with MVPA (*r*=0.52; *P*=.01) and average HR during AVG (*r*=0.43; *P*=.05). Mediation analysis indicated that narrative immersion mediated the effect of NV (NV vs NNV) on MVPA (direct effect: beta=7.51; *P*=.01). The indirect effect was that NV was positively correlated with the mediator variable narrative immersion (beta=.59; *P*=.03), which was itself marginally associated with MVPA (beta=6.95; *P*=.09); when narrative immersion was included in the model, the regression coefficient was attenuated.

Conclusions: AVG with added narratives elicits more narrative immersion, resulting in more minutes in MVPA. Narrative immersion served as a mediator between NV and MVPA via its elicitation of an elevated HR without increasing RPE. The inclusion of immersive narratives in AVG could be helpful for inducing MVPA, to enhance AVG engagement without additional exertion.

(J Med Internet Res 2020;22(3):e17994) doi:10.2196/17994

KEYWORDS

video games; exercise; fitness trackers; narration; physical activity; exercise

Introduction

In the United States, children spend an excessive amount of time with screen media [1]. Their daily time spent in sedentary video games has tripled over the past decade [2]. Physical activity (PA) in children is identified as one of the main methods for reducing the risk of obesity [3], cardiovascular disease [4], and type 2 diabetes [5]. Obese children aged less than 12 years are highly likely to become obese young adults [6].

A narrative is defined as any two or more events arranged in a temporal order [7] with unique, immersive, and motivational properties to encourage continuous active video game (AVG) play [8]. The absorption in a storyline integrates attentional allocation with imagery and feelings related to the narrative as opposed to nonnarrative. A loss of self-awareness is combined with mental construction of the narrative reality [9]. This attentional allocation is measured by narrative immersion [10]. Immersion is a term adapted from media research, defined as “a state in which the reader or player becomes absorbed in the narrative world” [10,11], that is, when one’s perception is directed toward a mediated world and away from the physical world [9,12]. On the other hand, nonnarratives consist of *arguments, reasoning, claims, and so forth* [13] that are overtly persuasive messages that do not create alternative worlds for individuals to enter, making them less likely to engage in mental imagery and emotion [9].

Narratives have been postulated to promote continuous gameplay [14]. AVGs emerge as an additional opportunity to increase children’s PA levels and reduce sedentary media usage [15,16] because AVGs are able to elicit moderate-to-vigorous physical activity (MVPA) [17,18]. AVGs feature player movement similar to *real-life* exercise participation [19,20]. Previous research demonstrated that children playing an AVG sports game session have similar energy expenditure as other traditional activities such as dancing or cycling [21]. However, children do not play AVG for sufficiently long durations to have an impact on their PA levels [20]. Thus, the development of innovative methods to increase children’s adherence and engagement to AVG is warranted. Although narratives appear in some health games [22], few AVGs capable of achieving MVPA have incorporated them [8].

The effect of narrative videos (NV) versus nonnarrative videos (NNV) has been studied extensively in health communication, behavioral intervention, and science communication [10,23], but no prior research has induced MVPA in AVG. A recent study [24] compared the effect of adding a narrative animated video to an AVG play session with the same session without video and reported that the children who watched the narrative took significantly more steps during the AVG play. Similarly, this effect was replicated recently [25]. In this case, the addition of a narrative to an AVG increased the MVPA by more than 50% compared with the same game condition without the narrative. Nevertheless, it is possible that the positive effect was merely because of the addition of an animated video, not the narrative itself. Although only NVs should have the immersive properties to increase engagement and motivation, there is no evidence that an NNV (documentary-like without stories) could also increase immersion, and there is no evidence of how MVPA increases during AVG.

Thus, this study aimed to compare the differential effect of adding an NV versus an NNV to an AVG session on PA parameters and the physiological responses to these two situations and to explore the mediation effect of immersion between *narrative versus nonnarrative* and MVPA in children aged 8 to 12 years.

Methods

Ethics Protocol

The Northeastern University’s institutional review board (IRB) approved the research protocol. Parents provided written informed consent, and children provided written informed assent. Consent and assent forms were both delivered to families of eligible children via email, and the paper forms were collected when parents brought the children to a university laboratory for the study. Data were collected between July 2018 and October 2018. Two research assistants (RAs) conducted the data collection with all participants.

Study Population

Children were recruited from a large, diverse, urban neighborhood in the northeast United States. RA placed printed flyers in the neighborhood afterschool program rooms and local public libraries. Participants from a previous research pool who were interested in this study were also contacted.

Inclusion criteria required the participants to (1) be aged between 8 and 12 years, (2) speak and understand English, (3) have not previously played the selected AVG, and (4) have no physical limitations with respect to AVG play. Exclusion criterion was the inability to complete an AVG session because of a medical condition or physical limitation.

The 8- to 12-year age group was targeted because children younger than 8 years have cognitive limitations in responding to survey questions [26] and children older than 12 years have entered early adolescence and will be subject to many physical, mental, emotional, and social changes that may make their needs and responses different from those of younger children [27]. In addition, without intervention, obese children in this age group (8-12 years old) are highly likely to become obese young adults [6].

Media Materials and Procedures

Experimental Videos

Both NV and NNV had a duration of approximately 11 min and presented the same kind of information on the benefit of PAs. The NNV was a narrated scientific documentary–like video about the benefits of PA featuring cartoon characters, including alternated male and female narrators. The NV tells a science fiction story about 2 main characters (boy/girl and player) with special powers enhanced by exercise and PA. The character/player needs to enhance their power to survive a postapocalyptic world. The NV plot has been developed with children-friendly design guidelines for MVPA motivation

[28,29] and has been found to be liked by children in this age group [30].

Active Video Game

We used the *Kung Fu for Kinect* AVG (Virtual Air Guitar Company), which involves whole-body movements via a Kinect sensor on an Xbox One console (Microsoft Inc). A preliminary examination of the AVG also indicated that both the NV and the NNV would be seamlessly matched with the game. While playing the AVG, the participant could see his/her own body on the screen and fought enemies using his/her own moves in a 2-dimensional fighting adventure environment. When different enemies appeared on the screen, the participant engaged them with a variety of intermittent and spontaneous movement patterns and skills, such as jumping, punching, and kicking [17]. We set the difficulty level as *easy* because we did not want the first time a child played the AVG to be difficult, and we found the level to be exercising enough to make the participants move. We set up a playing time limit of 45 min to avoid exhaustion, but the participants could stop at any point after they played at least one level (the first level could last between 1 and 5 min). We determined that a level ended whether a player won or did not win. Participant players could choose to continue trying that level if they had lost or progress to the next level if they had won.

Demographic Information and Anthropometrics

After signing an informed consent form, parents completed a questionnaire about their demographic and socioeconomic information (gender, race, household income, education, and number of adults and children in the household).

Children’s height (nearest 0.1 cm) and weight (nearest 0.1 kg) were measured using a stadiometer (ShorrBoard, Weight and Measure LLC) and a calibrated scale (SECA GmbH), which were then used to calculate BMI. Children’s age- and sex-specific BMI percentiles were obtained from the Centers for Disease Control and Prevention BMI-for-age growth charts [27].

A random list function (MS Office Excel 2019) was used to randomize eligible children to watch either an NV or an NNV before AVG play.

Accelerometer and Physiological Measures

The participants had a full and detailed explanation of the study procedures. A Polar heart rate (HR) monitor (Polar) was attached below the participant’s chest, and an ActiGraph GT3x (ActiGraph) was attached to their nondominant hip.

With all apparatus set, participants first sat comfortably before a television screen to watch an 11-min animated video (either NV or NNV). The RA stayed behind a curtain and remotely turned on the AVG console after the video. Children were asked to play for as long as they wanted.

Rate of perceived exertion (RPE) was assessed with Borg scale before and after the AVG play session. HR and PA were monitored during the entire AVG play session using the attached devices. The ActiGraph accelerometer device has been widely used for assessing individual levels of PA across different age groups, including children and young adults [17,31,32]. The ActiGraph triaxial accelerometers measure accelerations from the subject’s intensity and frequency of movement in 3 individual axes (anterior-posterior, vertical, and medial-lateral) and were initialized at 30 Hz sampling. The ActiLife software version 6.13.2 (ActiGraph) was used to download the data from the activity monitors and to convert acceleration data into the 3 axes activity counts, which quantify the amplitude and frequency of detected accelerations at a 1-second epoch dataset. A 1-second epoch has been found to be the most appropriate epoch length to detect short bursts of intense PA and may be the best choice for data processing and analysis in AVG studies examining intermittent PA [31]. We used Evenson et al’s [33] activity cut points to estimate the amount of time spent in light (26-573), moderate (574-1002), and vigorous (≥1003) PA.

After-Play Questionnaire

After the AVG session, participants completed questionnaires to assess narrative immersion and social desirability. Narrative immersion was assessed using a 9-item Likert-type scale (response range 1-5) [11], where 1=disagree and 5=agree. Sample items included “I could see myself in this story,” “I wanted to see how the story ended,” and “The story influenced my feelings.”

Social desirability (the tendency of the reporting person to give positive self-descriptions) was assessed using a 9-item Likert-type scale (response range 1-5) [34], where 1=disagree and 5=agree. Sample items included “I like everyone I know,” “I tell the truth every single time,” and “I never say thing I should not.” Social desirability of response has been observed in the self-reports of both adults and children. As this may impact the accuracy of the evaluative responses of AVGs, the validity of the study may be affected. The instrument has shown good reliability and validity in children across a variety of ethnic groups.

After completing the questionnaires, each participant received a US $25.00 gift card and was thanked for their participation.

Statistical Analysis

Normality and homogeneity of the data were checked using Kolmogorov-Smirnov and Levene tests, respectively. Independent t tests were performed to detect between-group differences in age and anthropometrics. Repeated measures analysis of variance was used to detect within- and between-group interactions for HR and RPE. The reliability of the questionnaires was assessed by using Cronbach alpha. In addition, Pearson correlation coefficients were calculated to detect the associations between the variables.

The mediation effects were tested for any associations we detected between the animated video condition and MVPA. The SPSS macro PROCESS was used for this analysis [35]. This macro uses bootstrapping to estimate the mediated effect and bias-corrected accelerated 95% CI of the effect. For each simulated sample, the macro estimates the product of paths b (from the independent variable to the mediator) and c (from the mediator to the dependent variable). These products represent a distribution approximating the sampling distribution of the indirect effect in the population. This technique directly tests the significance of the indirect effect (the $b \times c$ product). This technique is considered superior to the causal steps method, which does not directly test a mediated effect [36]. This model determines whether the association between the video type (independent variable) and MVPA (dependent variable) was mediated by narrative immersion (mediation variable).

A priori sample size was calculated to detect a significant difference between the 2 conditions, with 95% power (1−beta) to detect a small effect size (EF=0.3). Owing to some sample loss, the final sample size (n=22) provided a power of 92% (1−beta) for a large EF (ie, 0.8). Although less than expected, it was still above 80%. We set the significance level at 5% (alpha≤.05). We performed our analyses using Statistical Software for the Social Sciences (IBM SPSS version 25).

Results

Demographics and Anthropometrics

A total of 22 participants (NV: 12 vs NNV: 10) completed the required research protocol, and none of the participants dropped out. Approximately half of the participants were minority children, and their parents were college graduates (Table 1). We tested our randomization and found that children from NV and NNV groups were not statistically different in age, height, weight, BMI percentile, or socioeconomic status (*P*s>.11).

Table 1. Age, anthropometrics, and demographic characteristics of the sample.

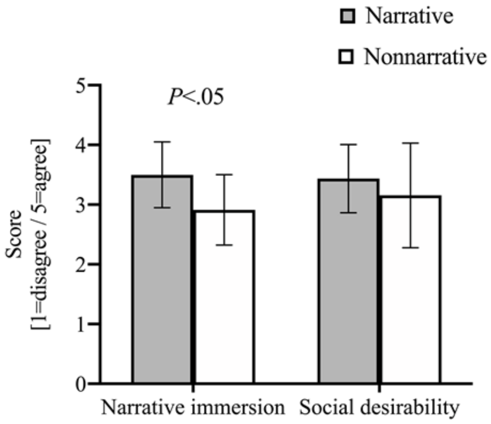
Variables	Narrative (n=12)	Nonnarrative (n=10)	*P* value
Age (years), mean (SD)	9.58 (1.16)	10.30 (1.34)	.19
Height (cm), mean (SD)	141. 3 (9.99)	146.5 (7.98)	.19
Weight (kg), mean (SD)	33.80 (7.84)	40.27 (9.92)	.11
BMI percentile, mean (SD)	45.08 (25.97)	56.80 (35.14)	.40
Sex, n (%)			
Male	8 (67)	7 (70)	N/A[a]
Female	4 (33)	3 (30)	N/A
Race, n (%)			
White	7 (64)	4 (40)	N/A
Black or African American	1 (9)	2 (20)	N/A
Asian	2 (18)	2 (20)	N/A
Other	1 (9)	0 (0)	N/A
Prefer not to answer	0 (0)	2 (20)	N/A
Highest education in the household, n (%)			
High school graduate or General Educational Development	2 (17)	0 (0)	N/A
Some college	1 (8)	1 (10)	N/A
College graduate	5 (42)	6 (60)	N/A
Postgraduate	4 (33)	3 (30)	N/A
Household income (US $), n (%)			
$20,000-$39,000	3 (27)	2 (22)	N/A
$60,000-$79,000	1 (9)	2 (22)	N/A
$80,000-$100,000	2 (18)	0 (0)	N/A
>$100,000	5 (46)	5 (56)	N/A
Prefer not to answer	1 (9)	1 (11)	N/A
Number of adults in the household, mean (SD)	2.17 (0.71)	2.00 (0.47)	.52
Number of children in the household, mean (SD)	2.67 (0.45)	2.70 (1.06)	.95

[a] N/A: not applicable.

Immersion and Social Desirability

All constructs were averaged to form their respective indices after Cronbach alphas were calculated. Narrative immersion had a Cronbach alpha of .62. We found a statistically significant difference in immersion between NV and NNV groups (mean 3.5, SD 0.6 vs mean 2.9, SD 0.6; *P*=.04); we did not identify any differences for social desirability (mean 3.4, SD 0.6 vs mean 3.2, SD 0.9; *P*=.40). Our estimate of Cronbach alpha for the social desirability questionnaire for our participants was .84 (Figure 1).

Figure 1. Comparison of narrative immersion between narrative videos and non-narrative videos adjusted by social desirability.



Physical Activity and Xbox Data

The PA data from the accelerometers indicated a difference in MVPA between the NV and NNV groups, although we did not identify any differences for the total playing time or the number of levels played between the NV and NNV groups (Table 2).

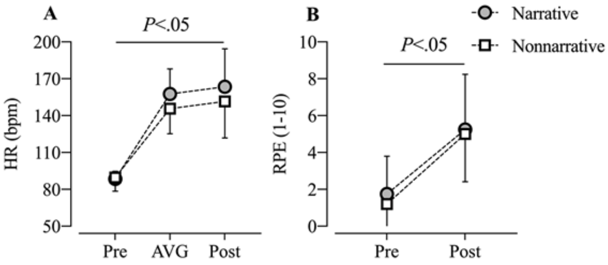
Table 2. Physical activity and active video game data.

Variables	Narrative (n=12), mean (SD)	Nonnarrative (n=10), mean (SD)	*t* test (*df*)	*P* value
Physical activity				
Vector magnitude (counts)	86.2 (33.8)	64.2 (13.7)	2.060 (20)	.06
Moderate-to-vigorous physical activity (minutes)	19.5 (13.3)	7.9 (5.8)	2.725 (20)	.02
Total steps	1461 (838)	1049 (549)	1.383 (20)	.18
Xbox playing data				
Total play time (minutes)	33.3 (14.1)	34.3 (11.5)	0.174 (20)	.86
Number of levels	9.8 (3.7)	10.2 (2.6)	0.304 (20)	.76

Heart Rate and Rate of Perceived Exertion Responses

Participants' HR and RPE responses to AVG showed a significant time effect for both NV and NNV (*Ps*<.01), although we did not identify either a group effect or a time × group effect for HR (group: *P*=.30; time × group: *P*=.17) or RPE (group: *P*=.64; time × group: *P*=.79; see Figure 2).

Figure 2. Heart rate and rate of perceived exertion during (only heart rate) and after active video game. AVG: active video game; HR: heart rate; RPE: rate of perceived exertion.



Correlation Analysis

Narrative immersion was positively and moderately correlated with MVPA (*r*=0.52; *P*=.01) and average HR during AVG (*r*=.43; *P*=.05). MVPA was also positively and moderately correlated with total time playing AVG (*r*=0.56; *P*=.01) and average HR during AVG (*r*=0.60; *P*=.001; see Table 3).

Table 3. Correlation matrix of narrative/game scales, physical activity, and heart rate.

Variables	Immersion		Playing time		Number of levels		MVPA[a]		VM[b]	
	r	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value
Playing time	0.02	.92	N/A[c]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Number of levels	0.08	.73	*0.91* [d]	*.001*	N/A	N/A	N/A	N/A	N/A	N/A
MVPA	*0.52*	*.01*	*0.56*	*.01*	0.53	.12	N/A	N/A	N/A	N/A
VM	*0.59*	*.001*	0.42	.06	*0.44*	*.04*	*0.81*	*<.001*	N/A	N/A
Heart rate (active video game)	*0.43*	*.05*	0.29	.19	0.18	.42	*0.60*	*<.001*	*0.74*	*<.001*

[a] MVPA: moderate-to-vigorous physical activity.

[b] VM: vector magnitude.

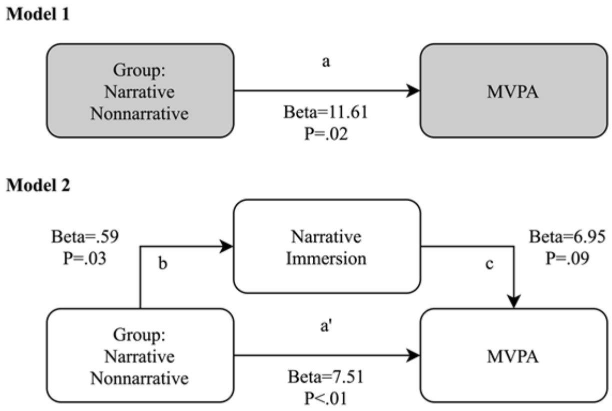
[c] N/A: not applicable.

[d] Italicized values denote significant correlation.

Mediation Analysis

The mediation analysis to test whether narrative immersion acted as a mediator variable between animated NV (independent variable) and MVPA (dependent variable) is shown in in Figure 3. Our results indicated that the effect of animated NV on MVPA was mediated by narrative immersion. In the first step (model 1: nonmediated direct effect), the regression coefficient of animated NV on MVPA indicated a positive association (beta=11.61; *P*=.02; see step *a* in Figure 3). In the second regression step (model 2: mediated direct and indirect effect), animated NV was positively related (beta=.59; *P*=.03) to narrative immersion (step *b*). In the last regression model, narrative immersion was marginally positively associated with the dependent variable (beta=6.95; *P*=.09); when narrative immersion was included in the model, the regression coefficient was significant (indirect effect: steps *b* and *c*), but the relationship was slightly attenuated (step *a'*: beta=7.51; *P*=.01). Finally, the indirect effect (step a' plus b and c) was significant, confirming the mediation role of MVPA in this model (Figure 3).

Figure 3. Mediation model of the relationship between narrative immersion and minutes in moderate-to-vigorous physical activity during active video game play. MVPA: moderate-to-vigorous physical activity.



Discussion

Principal Findings

This is the first study to compare NV and NNV added to an AVG play session. The main finding of this study is that AVGs with added narratives elicited more narrative immersion and narrative engagement, leading to more minutes in MVPA. More importantly, narrative immersion acted as a mediator between narratives and MVPA for AVG, also increasing HR after play without more RPE. Nevertheless, it is noteworthy that there is no interaction effect (group×time) for HR; both the groups increased their HR across time, but there are no differences between the groups. This suggests that although the addition of narratives did not result in longer periods of playtime, the players played the narrative versions of AVGs with more bodily engagement and movement, which increases their physiological response but does not necessarily increase their perceived exertion in the task. As a result, the inclusion of good immersive narratives could be a key factor in inducing children's MVPA play time, enhancing their engagement in AVG without incurring negative physiological feelings such as exertion.

This is the first report showing a mediation effect of narrative immersion on MVPA. The mediation results showed that immersion is a key factor in increasing MVPA during AVG play. Considering that story immersion is an integrative process in which the cognitive and affective resources of the player are concentrated in the task [14], it is reasonable to hypothesize that the NV elicited more game engagement, more commitment to the game tasks, and more movements, thus increasing their PA and physiological demand, without necessarily increasing their exertion. Other AVGs (without narratives) have been shown to not sufficiently motivate children [37], which may be because of increased exertion and boredom. With narrative animations, it may be possible to prevent children from choosing sedentary games and instead motivate them to play AVGs, thus increasing PA.

Previous evidence indicated that narratives' immersive properties could make children less self-aware of their physical effort during an AVG when they are immersed into the narrative [9], the very concept of immersion [10]. Thus, they can focus more on the outcome of the AVG play instead of the potential uncomfortable feelings associated with the exercise. Therefore, narrative immersion could be an effective method to promote healthier MVPA behaviors in young children [10] and older children [23].

Narratives enabled children to increase immersion and thus MVPA when playing AVGs. The finding that HR (during AVG), but not RPE, is moderately correlated with narrative immersion and MVPA suggests that despite the cardiovascular demand, being more physically active while being immersed in a narrative during an AVG does not increase the perceived exertion. If players perceived more exertion after AVG play, they might be discouraged from playing it again or, worse, engage in sedentary gameplay instead of active gameplay. In other words, children did not feel more tired after playing an AVG with narratives and spent more time participating in MVPA during the play instead.

Limitations

The results are not without limitations. The small size of our study population (n=22) only allowed us to detect a large EF (*d*>0.8), and small and moderate differences may have been undetected. Most (60%-70%) of our participants were male. Children's fitness level before the AVG play was not assessed. The laboratory setting for AVG playing may limit the external validation of the results.

Conclusions

Narrative immersion acts as a key mediator in increasing a player's MVPA in AVG after watching a narrative animated video. An NV also induced more MVPA with an increased HR after playing but not RPE. Future research should be conducted to investigate the chronic effect of narratives within AVGs on physiological parameters, PA, and playing time in children.

Acknowledgments

This project was partially funded by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (1R01DK109316-01), The narrative effect of active video games on long-term moderate-to-vigorous physical activity (project name; Principal Investigator: ASL), and by the College of Arts, Media, and Design and Bouvé College of Health Sciences at Northeastern University, Boston, MA. The authors would like to thank the Health Technology Laboratory members, including Heloisa Alves for her help with the IRB application and the data collection protocol preparation and Samantha Gutiérrez-Arango for her help with data collection. FableVision Studios produced the narrative video for this project.

Authors' Contributions

ASL designed the experiment. AF supervised the data collection and prepared the data for analysis. JH contributed instrumentation/materials. CS wrote the paper. CS, AF, JH, and ASL revised the paper.

Conflicts of Interest

None declared.

References

1. Carson V, Hunter S, Kuzik N, Gray CE, Poitras VJ, Chaput J, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 3):S240-S265. [doi: 10.1139/apnm-2015-0630] [Medline: 27306432]
2. Kaiser Family Foundation. San Francisco, CA: Kaiser Family Foundation; 2010. Generation M2: Media in the Lives of 8- to 18-Year-Olds URL: https://www.kff.org/wp-content/uploads/2013/01/8010.pdf [accessed 2020-03-17]
3. Council on Sports Medicine and Fitness, Council on School Health. Active healthy living: prevention of childhood obesity through increased physical activity. Pediatrics 2006 May;117(5):1834-1842. [doi: 10.1542/peds.2006-0472] [Medline: 16651347]
4. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
5. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Diabetologia 2012 Nov;55(11):2895-2905. [doi: 10.1007/s00125-012-2677-z] [Medline: 22890825]
6. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. N Engl J Med 1997 Sep 25;337(13):869-873. [doi: 10.1056/NEJM199709253371301] [Medline: 9302300]
7. Rimmon-Kenan S. Narrative Fiction: Contemporary Poetics. Second Edition. London: Routledge; 2002.

8. Barnett A, Cerin E, Baranowski T. Active video games for youth: a systematic review. J Phys Act Health 2011 Jul;8(5):724-737. [doi: 10.1123/jpah.8.5.724] [Medline: 21734319]
9. Green MC, Brock TC, Kaufman GF. Understanding media enjoyment: The role of transportation into narrative worlds. Commun Theory 2004;14(4):311-327. [doi: 10.1111/j.1468-2885.2004.tb00317.x]
10. Lu AS, Thompson D, Baranowski J, Buday R, Baranowski T. Story immersion in a health videogame for childhood obesity prevention. Games Health J 2012 Feb 15;1(1):37-44 [FREE Full text] [doi: 10.1089/g4h.2011.0011] [Medline: 24066276]
11. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 2000;79(5):701-721. [doi: 10.1037/0022-3514.79.5.701]
12. Lyons EJ, Tate DF, Ward DS. The better the story, the bigger the serving: narrative transportation increases snacking during screen time in a randomized trial. Int J Behav Nutr Phys Act 2013 May 16;10:60 [FREE Full text] [doi: 10.1186/1479-5868-10-60] [Medline: 23680389]
13. Green MC, Brock TC. In the mind's eye: Transportation-imagery model of narrative persuasion. In: Green MC, Strange JJ, Brock TC, editors. Narrative Impact: Social and Cognitive Foundations. Mahwah, NJ: Lawrence Erlbaum Associates; 2002:315-341.
14. Lu AS. Narrative in exergames: thoughts on procedure, mechanism, and others. Games Health J 2015 Feb;4(1):19-24 [FREE Full text] [doi: 10.1089/g4h.2014.0090] [Medline: 26181676]
15. Robinson TN, Banda JA, Hale L, Lu AS, Fleming-Milici F, Calvert SL, et al. Screen media exposure and obesity in children and adolescents. Pediatrics 2017 Nov;140(Suppl 2):S97-101 [FREE Full text] [doi: 10.1542/peds.2016-1758K] [Medline: 29093041]
16. Lu AS, Baranowski J, Cullen KW, Jago R, Thompson D, Baranowski T. Interactive media for childhood obesity prevention. Health Commun 2010 Sep;25(6-7):581-582 [FREE Full text] [doi: 10.1080/10410236.2010.496827] [Medline: 20845151]
17. Hwang J, Lee I, Fernandez AM, Hillman CH, Lu AS. Exploring energy expenditure and body movement of exergaming in children of different weight status. Pediatr Exerc Sci 2019 Aug 8:1-10. [doi: 10.1123/pes.2019-0006] [Medline: 31394495]
18. van 't Riet J, Crutzen R, Lu AS. How effective are active videogames among the young and the old? Adding meta-analyses to two recent systematic reviews. Games Health J 2014 Oct;3(5):311-318 [FREE Full text] [doi: 10.1089/g4h.2014.0005] [Medline: 26192486]
19. Bailey BW, McInnis K. Energy cost of exergaming: a comparison of the energy cost of 6 forms of exergaming. Arch Pediatr Adolesc Med 2011 Jul;165(7):597-602. [doi: 10.1001/archpediatrics.2011.15] [Medline: 21383255]
20. Biddiss E, Irwin J. Active video games to promote physical activity in children and youth: a systematic review. Arch Pediatr Adolesc Med 2010 Jul;164(7):664-672. [doi: 10.1001/archpediatrics.2010.104] [Medline: 20603468]
21. Smallwood SR, Morris MM, Fallows SJ, Buckley JP. Physiologic responses and energy expenditure of Kinect active video game play in schoolchildren. Arch Pediatr Adolesc Med 2012 Nov;166(11):1005-1009. [doi: 10.1001/archpediatrics.2012.1271] [Medline: 23007801]
22. Lu AS, Kharrazi H. A state-of-the-art systematic content analysis of games for health. Games Health J 2018 Feb;7(1):1-15 [FREE Full text] [doi: 10.1089/g4h.2017.0095] [Medline: 29293368]
23. Wang JJ, Baranowski T, Lau PW, Buday R, Gao Y. Story immersion may be effective in promoting diet and physical activity in Chinese children. J Nutr Educ Behav 2017 Apr;49(4):321-9.e1. [doi: 10.1016/j.jneb.2017.01.001] [Medline: 28391798]
24. Lu AS, Baranowski T, Hong SL, Buday R, Thompson D, Beltran A, et al. The narrative impact of active video games on physical activity among children: a feasibility study. J Med Internet Res 2016 Oct 14;18(10):e272 [FREE Full text] [doi: 10.2196/jmir.6538] [Medline: 27742605]
25. Hwang J, Lu AS. Narrative and active video game in separate and additive effects of physical activity and cognitive function among young adults. Sci Rep 2018 Jul 20;8(1):11020 [FREE Full text] [doi: 10.1038/s41598-018-29274-0] [Medline: 30030456]
26. Borgers N, de Leeuw E, Hox J. Children as respondents in survey research: Cognitive development and response quality 1. Bull Sociol Methodol 2000;66(1):60-75. [doi: 10.1177/075910630006600106]
27. Centers for Disease Control and Prevention. 2015. Healthy Weight URL: http://www.cdc.gov/healthyweight/assessing/bmi/ [accessed 2020-02-02]
28. Lu AS, Buday R, Thompson D, Baranowski T. What type of narrative do children prefer in active video games? An exploratory study of cognitive and emotional responses. In: Tettegah S, Huang WD, editors. Emotions, Technology, and Digital Games. London: Elsevier Publications; 2016:137-155.
29. Davies VF, Mafra R, Beltran A, Baranowski T, Lu AS. Children's cognitive and affective responses about a narrative versus a non-narrative cartoon designed for an active videogame. Games Health J 2016 Apr;5(2):114-119 [FREE Full text] [doi: 10.1089/g4h.2015.0019] [Medline: 26881473]
30. Lu AS, Green MC, Thompson D. Using narrative game design to increase children's physical activity: exploratory thematic analysis. JMIR Serious Games 2019 Nov 21;7(4):e16031 [FREE Full text] [doi: 10.2196/16031] [Medline: 31750833]
31. Hwang J, Fernandez AM, Lu AS. Application and validation of activity monitors' epoch lengths and placement sites for physical activity assessment in exergaming. J Clin Med 2018 Sep 11;7(9):pii: E268 [FREE Full text] [doi: 10.3390/jcm7090268] [Medline: 30208567]

32. Kwon S, Janz KF, International Children's Accelerometry Database (ICAD) Collaborators. Tracking of accelerometry-measured physical activity during childhood: ICAD pooled analysis. Int J Behav Nutr Phys Act 2012 Jun 7;9:68 [FREE Full text] [doi: 10.1186/1479-5868-9-68] [Medline: 22676230]
33. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sports Sci 2008 Dec;26(14):1557-1565. [doi: 10.1080/02640410802334196] [Medline: 18949660]
34. Reynolds CR, Paget KD. National normative and reliability data for the Revised Children's Manifest Anxiety Scale. School Psychol Rev 1983;12(3):324-336 [FREE Full text]
35. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: Guilford Publications; 2017.
36. Hayes AF. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. Commun Monogr 2009;76(4):408-420. [doi: 10.1080/03637750903310360]
37. Osorio G, Moffat DC, Sykes J. Exergaming, exercise, and gaming: sharing motivations. Games Health J 2012 Jun;1(3):205-210. [doi: 10.1089/g4h.2011.0025] [Medline: 26193438]

Abbreviations

AVG: active video game
EF: effect size
HR: heart rate
IRB: institutional review board
MVPA: moderate-to-vigorous physical activity
NNV: nonnarrative video
NV: narrative video
PA: physical activity
RA: research assistant
RPE: rate of perceived exertion

Edited by G Eysenbach; submitted 27.01.20; peer-reviewed by E Lyons, M Rosly; comments to author 18.02.20; revised version received 26.02.20; accepted 27.02.20; published 31.03.20.

Please cite as:
Sousa CV, Fernandez A, Hwang J, Lu AS
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
URL: http://www.jmir.org/2020/3/e17994/
doi:10.2196/17994
PMID:32229466

Original Paper

The Multiple Sclerosis Health Resource Utilization Survey (MS-HRS): Development and Validation Study

Nils-Henning Ness[1], MSc; Rocco Haase[1], Dipl-Psych; Raimar Kern[2], BA; Dirk Schriefer[1], MSc; Benjamin Ettle[3], PhD; Christian Cornelissen[4], PhD; Katja Akguen[1], MD; Tjalf Ziemssen[1], MD

[1]University Hospital Carl Gustav Carus, Dresden, Germany
[2]MedicalSyn GmbH, Dresden, Germany
[3]Novartis Pharma GmbH, Nuremberg, Germany
[4]Siemens Healthineers AG, Erlangen, Germany

Corresponding Author:
Tjalf Ziemssen, MD
University Hospital Carl Gustav Carus
Fetscherstr. 74
Dresden, 01307
Germany
Phone: 49 3514584465
Email: tjalf.ziemssen@uniklinikum-dresden.de

Abstract

Background: Survey-based studies are frequently used to describe the economic impact of multiple sclerosis (MS). However, there is no validated health resource survey available, preventing comparison of study results and meaningful conclusions regarding the efficiency of long-term treatments.

Objective: The aim of this study was to develop and validate a tablet- and paper-based MS health resource utilization survey.

Methods: We developed and validated the Multiple Sclerosis Health Resource Utilization Survey (MS-HRS), consisting of 24 cost items for paper and tablet users. Data for validation came from two large German observational studies. Survey practicability was assessed according to the response rate. Reliability was described using test-retest reliability as well as Guttman lambda. Construct validity was assessed as convergent and discriminant validity via correlations with associated patient-reported outcomes and known-group analyses.

Results: In total, 2207 out of 2388 (response rate: 92.4%) patients completed the survey and were included to determine psychometric properties. The test-retest reliability had an intraclass correlation coefficient of 0.828 over a course of 3 months. Convergent validity analyses showed that total costs correlated positively with increased disability (r=0.411, P<.001). For discriminant validity, correlations of total costs with the Treatment Satisfaction Questionnaire for Medication ranged from −0.006 (convenience) to −0.216 (effectiveness). The mean annual cost was €28,203 (SD €14,808) (US $39,203; SD US $20,583) with disease-modifying therapies.

Conclusions: The MS-HRS is a multilingual, reliable, valid, electronically available, and easy-to-administer questionnaire providing a holistic cross-sectional and longitudinal assessment of resource utilization in patients with MS.

(J Med Internet Res 2020;22(3):e17921) doi:10.2196/17921

KEYWORDS

multiple sclerosis; patient-reported outcome measures; resource utilization; validation; questionnaire development

Introduction

Multiple sclerosis (MS) is a potentially severe cause of neurological disability throughout adult life, leading to many years with high economic burden of the disease [1]. Studies on resource utilization in patients with MS have analyzed secondary data, such as administrative data of health insurance or health care providers, which have several strong limitations [2]. First, data accuracy may not be sufficient as several health care services can, but should not, be combined under one capitation. Second, the societal perspective cannot be considered, as only a share of all costs is refunded. In complex diseases like MS, patients' needs exceed the scope of primary and secondary health care providers, making a societal perspective even more important [3]. Third, billing data do not include important clinical data, making it impossible to determine the reasons for cost increases in the investigated population.

As another approach, diaries are commonly used to prospectively gather information on patient-level data [3]. However, the challenges in using diaries are thoroughness in reporting and high drop-out rates in time periods longer than 1 year [4].

Questionnaire-based cross-sectional studies represent a third well-recognized way of cost assessment in MS research, and they are well suited to analyze the occurrences of certain utilization behaviors [4-11]. Such surveys have the potential to include all relevant cost dimensions and may be applied to several stakeholders of the health care process. For instance, disease-modifying therapies (DMTs) are the main cost drivers for patients in earlier disease stages, and indirect costs are mainly responsible for the economic burden in later disease stages. However, cross-sectional investigations cannot assess the temporal associations between an intervention and an outcome. Such objectives require longitudinal data, which are fundamental for health economic evaluations, such as cost-effectiveness studies. Despite the long tradition of health economic evaluations in MS, there is no validated questionnaire with sufficient psychometric properties available, preventing comparison of study results between populations and meaningful conclusions [12,13].

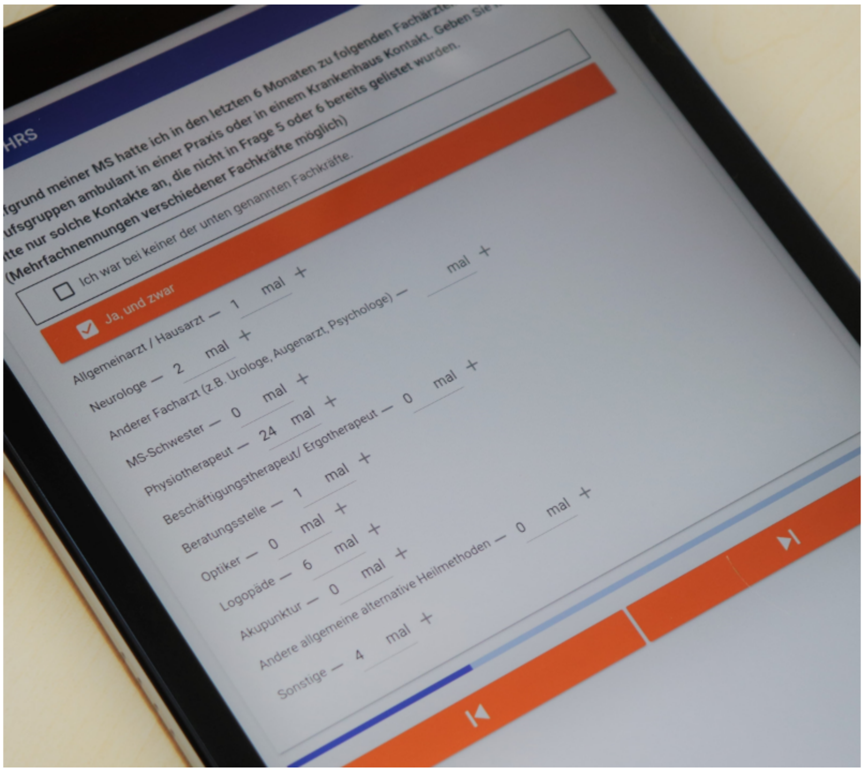
The aim of our work was to develop and validate an easy to administer questionnaire that provides a holistic longitudinal assessment of resource utilization for clinical practice and elaborate research approaches.

Methods

Questionnaire Development

A German expert group consisting of neurologists, health care administrators, psychologists, and MS nurses developed the first version of the Multiple Sclerosis Health Resource Utilization Survey (MS-HRS) at the University Hospital Dresden in 2009. Following intensive feedback from physicians, nurses, and patients, a second more time-efficient version was created in 2016. The requirements from clinical practice, such as time-saving use in everyday clinical practice, updatable pricing, and avoidance of double documentation, were decisive for the development of the digital version, which is browser-based and can be used across most devices (Figure 1). The time-saving use was realized through an adaptive questionnaire structure by presenting only the necessary items electronically.

Figure 1. Image of the tablet app of the German version of the Multiple Sclerosis Health-Resource Utilization Survey.



fgrund meiner MS hatte ich in den letzten 6 Monaten zu folgenden Fachärzte
ufsgruppen ambulant in einer Praxis oder in einem Krankenhaus Kontakt. Geben Sie
tte nur solche Kontakte an, die nicht in Frage 5 oder 6 bereits gelistet wurden.
(Mehrfachnennungen verschiedener Fachkräfte möglich)

☐ Ich war bei keiner der unten genannten Fachkräfte.
☑ Ja, und zwar

Allgemeinarzt / Hausarzt — 1 mal +
Neurologe — 2 mal +
Anderer Facharzt (z.B. Urologe, Augenarzt, Psychologe) — mal +
MS-Schwester — 0 mal +
Physiotherapeut — 24 mal +
Beschäftigungstherapeut/ Ergotherapeut — 0 mal +
Beratungsstelle — 1 mal +
Optiker — 0 mal +
Logopäde — 6 mal +
Akupunktur — 0 mal +
Andere allgemeine alternative Heilmethoden — 0 mal +
Sonstige — 4 mal +

The current digital German and translated English versions of the MS-HRS are in line with the recommendations on the core elements of a standardized resource questionnaire and can be completed in 10 to 15 minutes [14].

Health economic analysis can be performed from different perspectives with the societal perspective being the broadest [15]. For the practical evaluation of costs, three steps are considered, namely identification of resource consumption, quantification of resource use, and valuation of resources [15].

Identification of Resource Consumption

Information on resource consumption was obtained from studies, guidelines, textbook knowledge, administrative and accounting data, and expert opinions [4,8,16,17].

Costs were classified into direct medical, direct nonmedical, and indirect costs, according to German health technology assessment guidelines. Keeping multiple use cases in mind, we divided the tool into a core and an additional set of items (Table 1). The core elements are constitutive for the health resource utilization model and sufficient to capture the scope of resource utilization, whereas the additional questions foster the understanding of resource utilization.

Table 1. Core items of the Multiple Sclerosis Health Resource Utilization Survey: content and response option.

Domain (number of questions) and question label	Item	Response option
Direct medical costs (5)		
Inpatient stays	Stays in a hospital, especially in a neurological ward, rehabilitation clinic, and nursing home	Number of stays
Outpatient stays	Stays in a hospital, especially in a neurological ward, rehabilitation clinic, and nursing home	Number of stays
Professional consultations	Contact with neurologists, other specialists, multiple sclerosis nurses, physiotherapists, occupational therapists, counselling centers, opticians, speech therapists, acupuncturists, other alternative health care professionals, and others	Number of contacts
Examinations	Examinations undergone: magnetic resonance imaging, computed tomography, lumbar puncture, and blood examinations	Number of examinations
Over-the-counter medication		Expenditure in Euro
Medical consumables		Expenditure in Euro
Professional care	Assistance from professional caregiver and home help	Number of hours per week
Direct nonmedical costs (2)		
Informal care	Assistance from friends and family members	Number of hours per week
Investments and purchases	Investments and purchases: housing, car modifications, walking aids, manual wheelchair, automatic wheelchair, scooter, bed, and others	Expenditure in Euro
Indirect costs (3)		
Employment	Total hours of working time per week	Number of hours
Employment: Sick leave	Days of sick leave	Number of days
Employment: Presenteeism	Extent of reduced productivity at work	Likert scale (0-10)
Employment: Disability pension	Extent of disability pension	Percentage

Quantification of Resource Use

We used the most accurate method of microcosting defined as “direct enumeration and costing out of every input consumed in the treatment of a particular patient” to quantify resources (Table 2) [18]. Recall periods of resource utilization have been determined to avoid over- and underreporting. Longer recall periods lead to telescoping effects, in which events occurring outside the time frame are mistakenly included in the intended period [19]. Furthermore, we took into account that more frequent events and those that are less salient are less likely to be recalled accurately over a long period. In conclusion, the recall period may differ with respect to the nature of cost items but should never exceed 12 months. We recommend intervals for the assessment between 3 and 6 months. It is noticeable that unified recall periods across all items of the survey made the questionnaire more intelligible for patients. In our results, we report all calculated costs as per quarter.

Table 2. Health resources and quantification.

Domain	Item	Valuation
Direct medical costs		
Inpatient care/day admission	Days in hospital, rehabilitation, and nursing wards	Standardized cost units
Ambulant consultations	Visits to general practitioners, neurologists, others specialists (urologists, ophthalmologists, and psychiatrists), multiple sclerosis nurses, physical therapists, psychologists, occupational therapists, opticians, speech therapists, acupuncturists, and other alternative healing professionals	Standardized cost units
Investigations/diagnostics	Magnetic resonance imaging, computed tomography, spinal tap, blood tests, and others	Physicians' fee schedule
Over-the-counter medication	Medication and recommending doctor	Patient reported
Disease-modifying therapies		Annual therapy costs as calculated from medication reports
Home help and support of professionals	Professional help, household care, and personal assistance	Human capital approach
Medical consumables	Medical consumables	Patient reported
Direct nonmedical costs		
Investments and equipment	House and car modifications, walking aids, wheelchair (manual and electric), scooter, sickbed, and others; payer and grants	Patient reported
Informal care	Time taken for preparation of meals, climbing stairs, personal care, drug administration, transport, and others; reduction of working hours of relatives	Opportunity costs method
Indirect costs		
Employment and labor productivity	Full or part time work, sick leaves, reduced working time, change of work, and loss of earning	Human capital approach
Employment and labor productivity	Absence hours (multiple sclerosis and others), total working hours, and productivity scale	Human capital approach

Valuation of Resources

Evaluations conducted from societal perspectives are intended to reflect societal opportunity costs, which are equal to market prices in perfectly competitive markets. Nevertheless, markets in the health care sector are imperfect owing to statutory regulations. Hence, societal opportunity costs have to be approximated in most cases. Therefore, data from Bock et al were used wherever available (Table 3) [20].

Owing to a lack of data, few valuations were calculated from existing values. Visiting a nurse was rated with the lowest monetary value of €16.42. Furthermore, psychologist contact was monetarily valued as visiting a psychotherapist. The valuation for other specialists (eg, psychiatrist, urologist, and optician) was calculated as the mean of the given examples. Societal opportunity costs for investigations were approximated from the physicians' fee [21].

To calculate annual costs for DMTs, we used defined daily dose net costs multiplied by 365 days [22]. Net costs account for statutory manufacturer discounts as well as pharmacy discounts (Table 4). The work productivity loss (absenteeism, early retirement, and presenteeism) was calculated using the human capital approach [23]. Absenteeism was defined as not showing up for work, whereas presenteeism was defined as reduced work productivity due to health problems. The loss resulted from the total number of lost hours multiplied by the average salary per hour. Any hour not worked was considered as lost. Data for this calculation were obtained from official statistics [24].

In 2011, 233 working days were used for the calculation, and the average number of working hours was 1406.2 hours [25]. Additionally, the average hourly labor cost was €29.90 [20]. The monetary work productivity loss due to sick leave was calculated as follows:

> Productivity loss = (1406.2 working hours/233 working days) × €29.90 hourly salary

Occupational disability was calculated as the product of daily wage, average number of working days, and percentage of disability, with a maximum of €42,045.38 per year. The maximum hours dedicated to informal care was set to 60 hours per week according to German policies.

To ensure comparability of cash flows, prices from different periods were adjusted to the 2011 price level using the general price index for the national economy. This applies to patients' self-reported medications, investments, medical consumables, and DMTs. Consumption of the remaining resources was valued with prices from the year 2011 to generate comparable costs within our validation population. For other purposes, more recent values may be derived by applying a conventional cost inflation of 2%.

All costs are reported in Euro. In 2011, the Euro to US dollar annual average exchange rate was equal to 1.392.

Table 3. Resource valuation per unit.

Cost category	Monetary valuation
Direct medical costs (inpatient)	
Hospital	€593.04
Rehabilitation	€121.85
Nursing	€69.80
Direct medical costs (outpatient)	
Hospital	€385.48
Rehabilitation	€46.68
Nursing	€46.15
Direct medical costs (ambulant consultations)	
General practitioner	€20.06
Neurologist	€44.72
Other specialists	€34.73
Nurse	€16.42
Physiotherapist	€16.42
Psychologist	€78.08
Occupational therapist	€37.51
Optician	€34.78
Speech therapist	€38.59
Acupuncturist	€18.24
Other alternative healing professionals	€27.40
Direct medical costs (investigations/diagnostics)	
Magnetic resonance imaging	€120.21
Computed tomography	€73.78
Lumbar puncture	€38.90
Blood tests	€1.10
Others	€60.66
Direct medical costs (over-the-counter medication)	Patient reported
Direct medical costs (disease-modifying therapy)	See Table 4
Direct medical costs (medical consumables)	Patient reported
Direct medical costs (home help/professional care)	€27.57
Direct nonmedical costs	
Equipment, aids, and modifications	Patient reported
Informal care	€21.09
Indirect costs	
Sick leave	€180.45
Disability pension (full year)	€42,045.38

Table 4. Costs of disease-modifying therapies per year (in €).

Year	AVO[a]	REB[b]	BET[c]	EXT[d]	COP[e]	GIL[f]	TYS[g]
2010	18069.27	22626.46	17977.14	15528.45	16623.45	N/A[h]	24625.10
2011	18611.35	23305.25	18516.45	15994.30	17122.15	26698.58[i]	25363.85
2012	18182.19	22900.46	18132.07	15604.68	16377.93	25907.56	24586.58
2013	17981.00	22765.35	17815.29	15431.93	16197.00	22571.44	24214.41
2014	19437.31	24432.98	17601.95	15441.48	17035.61	18965.36	26194.92
2015	19333.96	24155.25	17517.70	15370.25	16733.32	20393.73	24280.75
2016	19247.42	23987.27	17384.09	14680.25	16648.48	21516.72	23321.06
2017	18865.10	23529.38	17047.81	15715.49	16352.26	20784.68	22902.02

[a] AVO: Avonex.

[b] REB: Rebif.

[c] BET: Betaferon.

[d] EXT: Extavia.

[e] COP: Copaxone.

[f] GIL: Gilenya.

[g] TYS: Tysabri.

[h] N/A: not applicable.

[i] As listed in the technology assessment report.

Study Population

Patients with relapsing-remitting MS (RRMS) were recruited in two prospective noninterventional multicenter studies conducted in Germany [26,27]. In that context, assessments of medical history and other general data, such as the Expanded Disability Status Score (EDSS), were performed by the treating neurologist. The current analyses are limited to patients with an EDSS of 0-6.0 to ensure sufficiently large subpopulations. Approval for both studies was obtained from independent local ethics committees, and all patients provided written informed consent for the collection of data [26,27].

Practicability and Reliability

Practicability was determined by the response rate of patients completing the MS-HRS. In addition, we compared the characteristics of completers and noncompleters to avoid selection bias. For reliability and validity analyses, we focused on the core elements of the survey. As we did not intend to evaluate DMTs at this step, costs for DMTs were not part of our methodological evaluation, which focused on psychometric properties.

For reliability analysis, test-retest reliability of the total costs in a group of stable patients over 3 months was estimated. Stable patients were defined as those not having relapse or increase in the EDSS during the 3 months of the retest period and another 3 months prior to the assessment. Thresholds for intraclass correlation coefficient (ICC) were applied as recommended (ICC<0.5: poor reliability; 0.5≤ICC<0.75: moderate reliability; 0.75≤ICC<0.9: good reliability; and ICC>0.90: excellent reliability) [28]. Guttman lambda 2 and lambda 6 are reported for the monetarized standardized items of the health resource utilization model with respect to the heterogeneous structure, providing a lower bound estimate of the consistency of the pricing approach. In an ordinary setting for test construction, desired levels would lay above 0.7. As our survey is not based on latent constructs but instead on real costs, consistency analysis is not part of the primary evaluation of reliability, but it provides additional insights for the model and the cost components.

Validity

Repeated expert consensus meetings of health economists, psychologists, and neurologists were conducted to secure face validity in terms of consistency and completeness.

Construct validity was assessed as convergent and discriminant validity via correlations with associated patient-reported outcomes (PROs) and known-group analyses by the EDSS (ranges: 0-1.0, 1.5-2.5, 3.0-4.0, and 4.5-5.5). The selected PRO measures were the EuroQol-5 Dimensions (EQ-5D), UK Neurological Disability Scale (UKNDS), and Patient Reported Outcome Indices for MS (PRIMUS) [29-32]. We expected to find significantly higher costs in groups with higher EDSS scores and correlations above 0.40 for convergent validity. As we did not intend to evaluate DMTs, the Treatment Satisfaction Questionnaire for Medication (TSQM) should present correlations clearly below 0.3 for discriminant validity [33].

Statistical Analysis

Continuous values are reported as arithmetic mean and SD. Ordinal values are reported as median and IQR. One-way random ICCs were used to calculate estimates for the test-retest reliability. All other correlations were calculated with Spearman rank correlation coefficients. Kruskal-Wallis H tests were conducted for known group analyses. Mann-Whitney U tests with adjustments by Bonferroni correction for multiple tests were applied for pairwise comparisons. All reported *P* values were compared to an alpha error level of 5%. No imputations were made for missing values.

Results

Patient Characteristics

In total, 2207 of 2388 patients completed the questionnaire at baseline and were therefore included in the validation process. The study population had a mean age of 41.73 (SD 10.19) years and was mostly female (1609/2207, 72.90%) (Table 5). Employed patients (1347/2207, 61.03%) were working predominantly full time (794/1347, 58.95%).

In terms of relapses within the previous year, active (1015/2207, 45.99%) and nonactive (1192/2207, 54.01%) patients were balanced in the population. Participants reported a mean disease duration since diagnosis of 7.54 (SD 6.11) years and a mean EDSS of 2.43 (SD 1.57).

Table 5. Characteristics of the study population (N=2207).

Parameter	Value
Age, years	
Mean (SD)	41.73 (10.19)
Median (IQR)	42.00 (34.00-49.00)
Gender	
Female, n (%)	1609 (72.90%)
Employment	
Employed, n (%)	1347 (61.03%)
Full-time employed, n (%)	794 (58.95%)
Number of relapses in the previous year, n (%)	
0	1192 (54.01%)
1	612 (27.73%)
2	267 (12.10%)
≥3	98 (4.44%)
Unknown	38 (1.72%)
Duration of disease since diagnosis, years	
Mean (SD)	7.54 (6.11)
Median (IQR)	6.00 (3.00-11.00)
EDSS[a]	
Mean (SD)	2.43 (1.57)
Median (IQR)	2.00 (1.00-3.50)

[a] EDSS: Expanded Disability Status Score.

Validation of the Questionnaire

Practicability

Looking at the number of patients completing the questionnaire, a good response rate of 92.4% (2207/2388) was achieved. Therefore, the responses of 2207 patients could be used to calculate total costs and all other parts of the health resource model. Noncompleting patients (n=181) did not differ in their gender distribution, age, or EDSS, but presented a slightly longer disease duration (mean 9.67 [SD 7.21] years, *P*<.001).

Reliability

Reliability was mainly assessed as test-retest reliability in a stable subgroup of patients. Overall, 1192 of 2207 (54.01%) patients fulfilled the criterion of presenting stable MS within that period. The ICC for this group over a course of 3 months was 0.828. In addition, Guttman lambda 2 ($\lambda 2=0.679$) and lambda 6 ($\lambda 6=0.694$) indicated an acceptable consistency between the standardized monetarized items of the health resource model (excluding DMT costs).

Validity

We analyzed the construct validity for total costs of the MS-HRS via known groups (excluding DMT costs). In all four EDSS groups, we found significantly different MS-related total costs and subcosts (all *P*<.001; Table 6). Further, all pairwise comparisons indicated significant differences in the direction as expected before (all *P*<.001 for total costs, all *P*<.05 for subcosts; Table 6).

Table 6. Known-group analysis: health resource utilization costs of patients with multiple sclerosis per quarter by disability (N=2059).

Parameter	EDSS[a] 0-1.0 (n=562)			EDSS 1.5-2.5 (n=756)			EDSS 3.0-4.0 (n=589)			EDSS 4.5-5.5 (n=152)		
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Total cost (€)	1099	2211	150	2295	3647	585	3112	3979	1334	4733	4639	4158
Direct medical cost (€)[b]	309	910	99	541	1724	165	789	1998	249	1003	2295	411
Direct nonmedical cost (€)	12	108	0	20	154	0	45	152	0	158	428	0
Indirect cost (€)	777	1822	0	1734	2820	0	2280	3178	127	3572	3572	3637

[a] EDSS: Expanded Disability Status Score.

[b] Disease-modifying therapies are not included in direct medical cost.

Convergent validity analyses showed that total costs correlated positively with increased (patient-reported) disability (UKNDS sum score: *r*=0.411) and lost ability to participate in daily routines and activities (PRIMUS Activities: *r*=0.423) and negatively with health-related quality of life (PRIMUS QoL: *r*=0.350; EQ-5D: *r*=−0.342) (all *P*<.001).

For discriminant validity, correlations of total cost with TSQM scores ranged from −0.006 (convenience) to −0.216 (effectiveness). As expected, the inclusion of DMT costs lowered the correlations with all PROs (by 0.07 on average), but the relations between correlations were maintained.

Resource Utilization

Majority of patients stated that they used direct medical services in the past 3 months (Table 7). In contrast, less than every second patient (44%) had indirect medical costs and 16% had direct nonmedical costs. Besides DMTs, indirect costs were the main cost drivers ahead of direct medical and direct nonmedical costs.

Patients were mainly treated in private practices (2068/2207, 85%) and less often during inpatient hospital stays (138/2207, 6%) and day care admissions (65/2207, 3%). However, the highest cost was for inpatient treatments (€315.06 [SD €1587.09]), followed by consultations in the primary sector (€209.87 [SD €292.73]) and day admissions in hospitals (€32.22 [SD €311.68]). Total costs per quarter averaged €2462 (SD €3650; median: €631) without DMTs and €7126 (SD €3697; median: €5871) with DMTs. Therefore, annual costs for mild-to-moderate RRMS ranged between €9528 (SD €14,603) without DMTs and €28,203 (SD €14,808) with DMTs.

Table 7. Health resource utilization costs in patients with multiple sclerosis per quarter (N=2207).

Variable	Value[a]
Direct medical costs (without disease-modifying therapies)	
Users, n (%)	2068 (93.70)
Mean (SD)	601.30 (1708.55)
Median (IQR)	174.00 (75.00-417.50)
Inpatient care	
Users, n (%)	138 (6.25)
Mean (SD)	315.06 (1587.09)
Median (IQR)	0.00 (0.00-0.00)
Day admission	
Users, n (%)	65 (2.95)
Mean (SD)	32.22 (311.68)
Median (IQR)	0.00 (0.00-0.00)
Consultations	
Users, n (%)	1817 (82.33)
Mean (SD)	209.87 (292.73)
Median (IQR)	109.50 (44.72-268.32)
Examinations	
Users, n (%)	1429 (64.75)
Mean (SD)	32.34 (35.63)
Median (IQR)	30.05 (0.28-45.77)
Over-the-counter medication	
Users, n (%)	693 (31.40)
Mean (SD)	15.02 (46.43)
Median (IQR)	0.00 (0.00-10.00)
Professional care	
Users, n (%)	115 (5.21)
Mean (SD)	7.02 (44.79)
Median (IQR)	(0.00-0.00)
Disease-modifying therapies	
Users, n (%)	2185 (99.00)
Mean (SD)	4733.24 (820.30)
Median (IQR)	4629.11 (4280.54-5656.61)
Direct nonmedical costs	
Users, n (%)	352 (15.95)
Mean (SD)	44.65 (229.51)
Median (IQR)	0.00 (0.00-0.00)
Investments	
Users, n (%)	38 (1.72)
Mean (SD)	9.14 (165.80)
Median (IQR)	0.00 (0.00-0.00)
Informal care and community service	
Users, n (%)	337 (15.27)

Variable	Value[a]
Mean (SD)	37.25 (158.57)
Median (IQR)	0.00 (0.00-0.00)
Indirect costs	
Users, n (%)	977 (44.27)
Mean (SD)	1816.50 (2880.74)
Median (IQR)	0.00 (0.00-3037.00)
Short-term absence	
Users, n (%)	320 (14.50)
Mean (SD)	441.53 (1782.18)
Median (IQR)	0.00 (0.00-0.00)
Disability pension	
Users, n (%)	352 (15.95)
Mean (SD)	852.98 (2062.98)
Median (IQR)	0.00 (0.00-0.00)
Presenteeism	
Users, n (%)	475 (21.52)
Mean (SD)	522.00 (1333.18)
Median (IQR)	0.00 (0.00-0.00)

[a] All costs are in Euro.

Discussion

Principal Findings

The MS-HRS represents a reliable, valid, and easy-to-administer questionnaire providing a holistic assessment of resource utilization for patients with MS. Health resources were derived from all pathways of patients with MS in an adapted health resource model for MS.

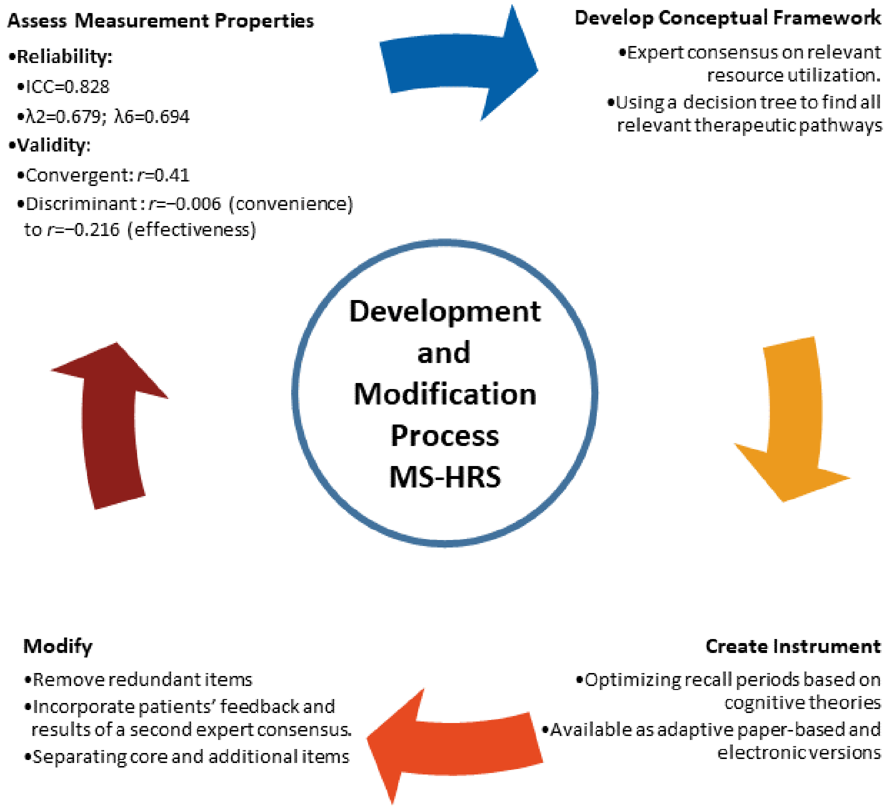
Some very respectable efforts have already been taken to research the health economic footprint of MS [2,5-7]. In Europe, Kobelt et al repeatedly assessed the costs and burden of MS in a cross-sectional survey approach, including direct, indirect, and intangible costs [6,7]. Concepts and definitions of subcosts may differ over time with respect to scope and style of reporting (eg, intangible costs where a clear line is recommended) [34]. For our model approach, we gave strong emphasis to direct and indirect costs, as done by Karampampa et al and Reese et al in their models [5,11]. We increased the depth of the assessment of indirect costs by adding a quantification of costs for presenteeism in addition to costs for absenteeism (sick leave and disability pension).

Part of our questionnaire development was a validation process that confirmed reliability and validity in a way that no previous approach did within the MS domain (Figure 2). Kobelt et al also recognized the need for validation, but a systematic approach beyond the aspects of practicability and face validity remained a task to fulfil [35]. Even though we did not develop a one-dimensional questionnaire for a latent construct, we were able to demonstrate essential psychometric qualities in the health economic assessment of MS. As of today, a multidomain open-access database of resource-use questionnaires does not contain a published instrument for MS [12]. Most of the cost assessments for MS were not developed to be published for general use but developed for application in certain cross-sectional frameworks. A very recent analysis of real-life cost outcomes underlined the rising interest in longitudinal assessment of health resource utilization [10].

The proportion of patients claiming certain health services was slightly lower in the study by Kobelt et al than in this study [6]. The lower proportion of patients taking over-the-counter medications and investments was particularly noticeable. This may be explained by the less severe disease progression beyond a RRMS profile. Taking this into account, the average annualized disease burden was within the range expected from previous publications [5,7,11,36].

Kobelt et al have recently reported precise price tags for most unit costs, and other authors have at least provided an indirect description of the valuation process and its exact results [5,9]. Differences were found for the valuation of subcosts, such as inpatient hospitalization, owing to different sources of valuation or different definitions for health resource units. In any case, precise and fully transparent reporting of per unit costs is highly recommended, especially for main cost drivers.

Figure 2. Development and modification process of the Multiple Sclerosis Health-Resource Utilization Survey. ICC: intraclass correlation coefficient; λ2: Guttman lambda 2; λ6: Guttman lambda 6.



Event-related costing can also be considered to be of interest, but a clear separation from costs being assessed via routine data collection is necessary [5]. Otherwise, costs may double due to double assessment. We discussed whether to collect data for both sick days and missed working hours, as described by Reilly et al [16]. Furthermore, asking about sick hours would have meant switching to a recall period of 7 days, whereas the remaining questionnaire covers the last 6 months. Owing to higher usability, we decided to assess sick days only.

Claims data were considered for the assessment of criterion validity. However, there is a disadvantage that only billing-relevant data are available, preventing consideration of a societal perspective. Information on informal care, over-the-counter medication, and presenteeism and partial information on investments and medical consumables is not recorded. In addition, claims data do not provide the number of physician visits by a patient in Germany. In consequence, claims data were not suitable for the validation process.

Health economic studies often require both clinical and economic data. The MS-HRS can easily be used as part of clinical interventional and noninterventional studies to collect economic data. In a large population, we demonstrated that the questionnaire is easy to administer and has good psychometric properties. These characteristics provide the necessary prerequisites for high-quality health economic studies (eg, cost-effectiveness analyses).

Limitations

A patient-centered questionnaire is subject to notable limitations. Reliable recall periods of health resource use are time-limited, especially in patients who are cognitively more affected by the disease. Furthermore, patients may not want to disclose socioeconomic and health economic information because it is considered too confidential. We did not include information about more recent therapies as we did not gather related data for validation, which will be done in follow-up studies. Beyond this, further costs may be thinkable but less likely to have an impact on a societal level (eg, crowd-funded therapies at the current level and cost-related voluntary work loss). In addition, price tags for cost components have to be updated and adapted to local levels. For patients with progressive MS and patients with severe disability, further studies have to confirm the given psychometric properties.

Conclusions

The MS-HRS is a promising option to measure costs precisely in cross-sectional and longitudinal settings instead of estimating them or using surrogates. Further country-wise cost weights will facilitate the transparent estimation of MS-related costs across multiple regions. The MS-HRS is available online [37].

Acknowledgments

This study was funded by Novartis.

Authors' Contributions

NHN and RH conceived the study; NHN and RH were responsible for methodology; NHN, RH, DS, and RK managed the software; RH, NHN, and DS were responsible for validation; NHN and RH performed formal analysis; NHN and RH performed investigations; TZ, CC, and BE were responsible for resources; NHN, RH, and DS were responsible for data curation; NHN wrote the original draft; RH, TZ, DS, CC, BE, RK, and KA reviewed and edited the manuscript; NHN was responsible for visualization; TZ was responsible for supervision; TZ was responsible for project administration; TZ, CC, and BE were responsible for funding acquisition. All authors have read and agree to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Trisolini M, Honeycutt A, Wiener J, Lesesne S. Multiple Sclerosis International Federation.: RTI International; 2010. Global economic impact of multiple sclerosis URL: https://www.msif.org/wp-content/uploads/2014/09/Global_economic_impact_of_MS.pdf [accessed 2020-02-28]
2. Ernstsson O, Gyllensten H, Alexanderson K, Tinghög P, Friberg E, Norlund A. Cost of Illness of Multiple Sclerosis - A Systematic Review. PLoS One 2016;11(7):e0159129 [FREE Full text] [doi: 10.1371/journal.pone.0159129] [Medline: 27411042]
3. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK Health Technology Assessment program. Value Health 2010 Dec;13(8):867-872 [FREE Full text] [doi: 10.1111/j.1524-4733.2010.00788.x] [Medline: 20946187]
4. Fowler JFJ. Survey Research Methods (Applied Social Research Methods Series I). Thousand Oaks, CA: Sage Publications, Inc; 1993.
5. Karampampa K, Gustavsson A, Miltenburger C, Eckert B. Treatment experience, burden and unmet needs (TRIBUNE) in MS study: results from five European countries. Mult Scler 2012 Jun;18(2 Suppl):7-15. [doi: 10.1177/1352458512441566] [Medline: 22623122]
6. Kobelt G, Thompson A, Berg J, Gannedahl M, Eriksson J, MSCOI Study Group, European Multiple Sclerosis Platform. New insights into the burden and costs of multiple sclerosis in Europe. Mult Scler 2017 Jul;23(8):1123-1136 [FREE Full text] [doi: 10.1177/1352458517694432] [Medline: 28273775]
7. Kobelt G, Berg J, Lindgren P, Fredrikson S, Jönsson B. Costs and quality of life of patients with multiple sclerosis in Europe. J Neurol Neurosurg Psychiatry 2006 Aug;77(8):918-926 [FREE Full text] [doi: 10.1136/jnnp.2006.090365] [Medline: 16690691]
8. Kobelt G, Berg J, Lindgren P, Jönsson B. Costs and quality of life in multiple sclerosis in Europe: method of assessment and analysis. Eur J Health Econ 2006 Sep;7 Suppl 2:S5-13. [doi: 10.1007/s10198-006-0365-y] [Medline: 17310344]
9. Kobelt G, Eriksson J, Phillips G, Berg J. The burden of multiple sclerosis 2015: Methods of data collection, assessment and analysis of costs, quality of life and symptoms. Mult Scler 2017 Aug;23(2_suppl):4-16. [doi: 10.1177/1352458517708097] [Medline: 28643592]
10. Kobelt G, Jönsson L, Pavelcova M, Havrdová EK. Real-Life Outcome in Multiple Sclerosis in the Czech Republic. Mult Scler Int 2019;2019:7290285 [FREE Full text] [doi: 10.1155/2019/7290285] [Medline: 30911417]
11. Reese JP, John A, Wienemann G, Wellek A, Sommer N, Tackenberg B, et al. Economic burden in a German cohort of patients with multiple sclerosis. Eur Neurol 2011;66(6):311-321. [doi: 10.1159/000331043] [Medline: 22086151]
12. DIRUM. 2020. Database of instruments for resource use measurement URL: http://www.dirum.org/instruments/all [accessed 2020-02-28]
13. Ridyard CH, Hughes DA, DIRUM Team. Taxonomy for methods of resource use measurement. Health Econ 2015 Mar;24(3):372-378. [doi: 10.1002/hec.3029] [Medline: 24442966]
14. Thorn JC, Brookes ST, Ridyard C, Riley R, Hughes DA, Wordsworth S, et al. Core Items for a Standardized Resource Use Measure: Expert Delphi Consensus Survey. Value Health 2018 Jun;21(6):640-649 [FREE Full text] [doi: 10.1016/j.jval.2017.06.011] [Medline: 29909868]
15. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods For The Economic Evaluation Of Health Care Programmes. Oxford, UK: Oxford University Press; 2015.
16. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics 1993 Nov;4(5):353-365. [doi: 10.2165/00019053-199304050-00006] [Medline: 10146874]
17. Fowler FJ. Improving Survey Questions: Design And Evaluation (Applied Social Research Methods). Thousand Oaks, CA: Sage Publications, Inc; 1995.

18. Raftery J. Costing in economic evaluation. BMJ 2000 Jun 10;320(7249):1597 [FREE Full text] [doi: 10.1136/bmj.320.7249.1597] [Medline: 10845978]
19. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. Curr Med Res Opin 2009 Apr;25(4):929-942. [doi: 10.1185/03007990902774765] [Medline: 19257798]
20. Bock J, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen 2015 Jan;77(1):53-61. [doi: 10.1055/s-0034-1374621] [Medline: 25025287]
21. Kassenärztliche Bundesvereinigung. 2011. Einheitlicher Bewertungsmaßstab URL: https://www.kvsa.de/fileadmin/user_upload/PDF/Praxis/Honorarverteilung/EBM2011.pdf [accessed 2020-02-28]
22. Zeller W. Immuntherapeutika und zytostatika. In: Schwabe U, Paffrath D, editors. Arzneiverordnungs-Report 2012. Berlin, Heidelberg, GER: Springer; 2012:679-696.
23. van den Hout WB. The value of productivity: human-capital versus friction-cost method. Ann Rheum Dis 2010 Jan;69 Suppl 1:i89-i91. [doi: 10.1136/ard.2009.117150] [Medline: 19995754]
24. Institut für Arbeitsmarkt- und Berufsforschung. 2011. Daten zur kurzfristigen Entwicklung von Wirtschaft und Arbeitsmarkt URL: https://www.iab.de/de/publikationen/jahresbericht/jb2011.aspx [accessed 2020-02-28]
25. Bundeszentrale für politische Bildung. 2011. Geleistete Arbeitsstunden je Erwerbstätigem pro Jahr, Produktivität je Erwerbstätigenstunde, 1970-2012 URL: https://www.bpb.de/nachschlagen/zahlen-und-fakten/soziale-situation-in-deutschland/61711/arbeitszeit [accessed 2020-02-28]
26. Ziemssen T, Kern R, Cornelissen C. The PANGAEA study design - a prospective, multicenter, non-interventional, long-term study on fingolimod for the treatment of multiple sclerosis in daily practice. BMC Neurol 2015 Jun 18;15(1):93 [FREE Full text] [doi: 10.1186/s12883-015-0342-0] [Medline: 26084334]
27. Vormfelde SV, Ortler S, Ziemssen T. Multiple Sclerosis Therapy With Disease-Modifying Treatments in Germany: The PEARL (ProspEctive phArmacoeconomic cohoRt evaluation) Noninterventional Study Protocol. JMIR Res Protoc 2016 Feb 04;5(1):e23 [FREE Full text] [doi: 10.2196/resprot.4473] [Medline: 26846334]
28. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012]
29. Doward L, McKenna S, Meads D, Twiss J, Eckert B. The development of patient-reported outcome indices for multiple sclerosis (PRIMUS). Mult Scler 2009 Sep;15(9):1092-1102. [doi: 10.1177/1352458509106513] [Medline: 19556315]
30. Sharrack B, Hughes RA. The Guy's Neurological Disability Scale (GNDS): a new disability measure for multiple sclerosis. Mult Scler 1999 Aug;5(4):223-233. [doi: 10.1177/135245859900500406] [Medline: 10467380]
31. Graf J, Claes C, Greiner W, Uber A. Die deutsche Version des EuroQol-Fragebogens. J Public Health 1998 Mar;6(1):3-20. [doi: 10.1007/bf02956350]
32. Greiner W, Claes C, Busschbach JJ, von der Schulenburg JM. Validating the EQ-5D with time trade off for the German population. Eur J Health Econ 2005 Jun;6(2):124-130. [doi: 10.1007/s10198-004-0264-z] [Medline: 19787848]
33. Atkinson MJ, Kumar R, Cappelleri JC, Hass SL. Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM version II) among outpatient pharmacy consumers. Value Health 2005;8 Suppl 1:S9-S24 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.00066.x] [Medline: 16336491]
34. Wundes A, Brown T, Bienen EJ, Coleman CI. Contribution of intangible costs to the economic burden of multiple sclerosis. J Med Econ 2010;13(4):626-632. [doi: 10.3111/13696998.2010.525989] [Medline: 20950249]
35. Kobelt G, Teich V, Cavalcanti M, Canzonieri AM. Burden and cost of multiple sclerosis in Brazil. PLoS One 2019;14(1):e0208837 [FREE Full text] [doi: 10.1371/journal.pone.0208837] [Medline: 30673707]
36. Ness NH, Schriefer D, Haase R, Ettle B, Cornelissen C, Ziemssen T. Differentiating societal costs of disability worsening in multiple sclerosis. J Neurol 2019 Dec 17. [doi: 10.1007/s00415-019-09676-4] [Medline: 31848738]
37. MS Zentrum Dresden. Zentrum für Klinische Neurowissenschaften. 2020. Multiple Sclerosis Health Resource Utilization Survey URL: http://msz.uniklinikum-dresden.de/ehealth/ms-hrs [accessed 2020-02-28]

Abbreviations

DMT: disease-modifying therapy
EDSS: Expanded Disability Status Score
EQ-5D: EuroQol-5 Dimensions
ICC: intraclass correlation coefficient
MS: multiple sclerosis
MS-HRS: Multiple Sclerosis Health Resource Utilization Survey
PRIMUS: Patient Reported Outcome Indices for Multiple Sclerosis
PRO: patient-reported outcome
RRMS: relapsing-remitting multiple sclerosis
TSQM: Treatment Satisfaction Questionnaire for Medication
UKNDS: UK Neurological Disability Scale

Edited by G Eysenbach; submitted 22.01.20; peer-reviewed by E d'Amico; comments to author 14.02.20; revised version received 18.02.20; accepted 19.02.20; published 17.03.20.

Please cite as:
Ness NH, Haase R, Kern R, Schriefer D, Ettle B, Cornelissen C, Akguen K, Ziemssen T
The Multiple Sclerosis Health Resource Utilization Survey (MS-HRS): Development and Validation Study
J Med Internet Res 2020;22(3):e17921
URL: http://www.jmir.org/2020/3/e17921/
doi:10.2196/17921
PMID:32181745

Original Paper

Utilizing Digital Health to Collect Electronic Patient-Reported Outcomes in Prostate Cancer: Single-Arm Pilot Trial

Christine Tran[1], MS; Adam Dicker[1], MD, PhD; Benjamin Leiby[1], PhD; Eric Gressen[1], MD; Noelle Williams[2], MD; Heather Jim[3], PhD

[1]Thomas Jefferson University, Sidney Kimmel Medical College and Cancer Center, Jefferson Center for Digital Health & Data Science, Philadelphia, PA, United States

[2]Levine Cancer Institute at Atrium Health, Southeast Radiation Oncology Group, Charlotte, NC, United States

[3]H Lee Moffitt Cancer Center, Department of Health Outcomes and Behavior, Tampa, FL, United States

Corresponding Author:
Adam Dicker, MD, PhD
Thomas Jefferson University
Sidney Kimmel Medical College and Cancer Center
Jefferson Center for Digital Health & Data Science
4201 Henry Ave
Philadelphia, PA
United States
Phone: 1 216 955 6700
Email: adam.dicker@jefferson.edu

Abstract

Background: Measuring patient-reported outcomes (PROs) requires an individual's perspective on their symptoms, functional status, and quality of life. Digital health enables remote electronic PRO (ePRO) assessments as a clinical decision support tool to facilitate meaningful provider interactions and personalized treatment.

Objective: This study explored the feasibility and acceptability of collecting ePROs using validated health-related quality of life (HRQoL) questionnaires for prostate cancer.

Methods: Using Apple ResearchKit software, the *Strength Through Insight* app was created with content from validated HRQoL tools 26-item Expanded Prostate Cancer Index Composite (EPIC) or EPIC for Clinical Practice and 8-item Functional Assessment of Cancer Therapy Advanced Prostate Symptom Index. In a single-arm pilot study with patients receiving prostate cancer treatment at Thomas Jefferson University Hospital and affiliates, participants were recruited, and instructed to download *Strength Through Insight* and complete ePROs once a week over 12 weeks. A mixed methods approach, including qualitative pre- and poststudy interviews, was used to evaluate the feasibility and acceptability of *Strength Through Insight* for the collection and care management of cancer treatment.

Results: Thirty patients consented to the study; 1 patient failed to complete any of the questionnaires and was left out of the analysis of the intervention. Moreover, 86% (25/29) reached satisfactory questionnaire completion (defined as completion of 60% of weekly questions over 12 weeks). The lower bound of the exact one-sided 95% CI was 71%, exceeding the 70% feasibility threshold. Most participants self-identified with having a high digital literacy level (defined as the ability to use, understand, evaluate, and analyze information from multiple formats from a variety of digital sources), and only a few participants identified with having a low digital literacy level (defined as only having the ability to gather information on the Web). Interviews were thematically analyzed to reveal the following: (1) value of emotional support and wellness in cancer treatment, (2) rise of social patient advocacy in online patient communities and networks, (3) patient concerns over privacy, and (4) desire for personalized engagement tools.

Conclusions: *Strength Through Insight* was demonstrated as a feasible and acceptable method of data collection for ePROs. A high compliance rate confirmed the app as a reliable tool for patients with localized and advanced prostate cancer. Nearly all participants reported that using the smartphone app is easier than or equivalent to the traditional paper-and-pen approach, providing evidence of acceptability and support for the use of remote PRO monitoring. This study expands on current research involving the value of digital health, as a social and behavioral science, augmented with technology, can begin to contribute to population health management, as it shapes psychographic segmentation by demographic, socioeconomic, health condition, or behavioral factors to group patients by their distinct personalities and motivations, which influence their choices.

Trial Registration: ClinicalTrials.gov NC03197948; http://clinicaltrials.gov/ct2/show/NC03197948

(J Med Internet Res 2020;22(3):e12689) doi:10.2196/12689

KEYWORDS

eHealth; mHealth; smartphone; mobile phone; mobile apps; health promotion; chronic disease; health-related quality of life; cancer; patient-reported outcome measures; health information technology; patient-centered care

Introduction

Prioritizing Patient-Reported Outcomes

As hospitals seek to provide better value in health care, patient-reported outcome measures (PROMs), such as those evaluating pain and distress, are an integral part of improving care. Patient-reported outcomes (PROs) can serve as an innovative way for providers to incentivize patients to make changes that facilitate patient engagement and self-care for chronic disease management and prevention [1-4]. The use of PROs for cancer patients is not only appropriate but also is becoming an increasing unmet need, as studies suggest remote monitoring has been evaluated and demonstrated to improve survival. Studies have shown that PROs can lead to improved patient communication with providers, engagement, satisfaction, and better health outcomes. PROs can enhance care management by understanding the impact of treatments on patients' lives [5]. This is increasingly important in cancer care, as patients with cancer can experience changes in their nutrition, elimination, pain management, and sexual function at varying levels of severity [6]. A health-related quality of life (HRQoL) evaluation can be imperative to evaluate the results of clinical trials. However, there is substantial evidence that clinical investigators miss 40% to 74.4% of patients' symptomatic adverse events (AEs) [7]. Moreover, major policy making entities have also emphasized the importance of incorporating PROs into cancer research and policy formation (including the National Cancer Institute [NCI], American Cancer Society, and US Food and Drug Administration). This interest reflects a growing national recognition that traditional medical outcomes (ie, survival and disease progression) do not fully capture the patient's experience of health, and there is an unmet need for health care providers to capture a new definition of *value* of health care, which includes improvement in subjective outcomes of importance to patients [8]. The 2 principal methods of gathering PROs (before the advent of electronic methodologies) are paper-and-pen approach and clinician interview, both of which are labor and time intensive. For PROs to be routinely integrated into clinical practice, PRO data collection methods must be efficient by demonstrating convenient, instantaneous, inexpensive, reliable, and clinically feasible.

A Digital Evolution

With the widespread adoption of smartphones, tablets, and other smart devices, mobile apps provide a new platform for patients to become active members of their health care team. Digital health technology encompasses clinical tools, advanced statistical algorithms, cloud computing, and artificial intelligence [9]. Research on digital technology can evaluate innovative approaches to improve care through PRO measures. The bidirectional transfer of data through smartphones offers an unprecedented method to collect PROs across the entire course of a cancer patient's journey, allowing providers to monitor long-term outcomes. However, the translation of evidence-based health care interventions onto a digital platform should be evaluated to determine whether it is feasible and effective digital health technology [10-12].

The objective of this study was to test the feasibility and acceptability of *Strength Through Insight*, a digital health app collecting electronic PROs (ePROs) in patients with prostate cancer, and to examine patient perspectives to help create future digital health interventions. Apple's ResearchKit empowers medical research by creating a mobile infrastructure for informed consent, surveys, and real-time active tasks (spatial memory, voice tests, motor activities, sleep-wake cycle, nutrition, and daily step counts) using the iPhone sensors and capabilities [13]. *Strength Through Insight*, a smartphone app built on the ResearchKit platform, aimed to explore the feasibility and acceptability of smartphone devices as a digital health tool to collect PROs for patients with cancer in the health care setting through a mixed methods approach.

To our knowledge, this is the first study piloting an ePRO using the ResearchKit smartphone app platform for patients with prostate cancer. The study (1) tested the feasibility of collecting ePROs via a digital health app, with a validated HRQoL questionnaires for patients undergoing prostate cancer treatment, and (2) analyzed patient attitudes and perceptions through qualitative interviews to identify reoccurring themes that address facilitators and barriers of adopting digital health to best support future design and implementation of digital health technology. We hypothesized that more than 80% of patients will complete 60% of the HRQoL questionnaires, once a week for a period of 12 weeks, proving the feasibility of ePROs in a smartphone app.

Methods

Procedures

This feasibility study was conducted at the Sidney Kimmel Cancer Center at Thomas Jefferson University Hospital (TJUH). Following TJUH institutional review board approval, potential participants were identified through a database maintained by the Sidney Kimmel Cancer Center.

The study was made available on the ClinicalTrials.gov [14] and on the Sidney Kimmel Cancer Center's website in a listing of ongoing clinical trials. Potential participants were also recruited by research staff at TJUH in accordance with board-approved methods. Participating clinicians reviewed their patient lists and identified eligible patients who were then telephoned by a research staff member to confirm eligibility. Eligibility was confirmed by verbal confirmation of inclusion and exclusion criteria that were also available as part of the consent in the smartphone app. Eligible patients received detailed information regarding the study and, if interested, were sent a *next steps* document explaining possible risk and benefits, study expectations, expected time to complete, how to download the app, answer the questionnaire, configure settings (ie, cellular data and push notifications), set reminders, and encouraging reporting of their patient outcome via the HRQoL questionnaire. The document also included information on how to follow up with the research team if they felt that they needed additional support. Participants were required to provide informed consent for the study through their smartphones, on downloading the app, before answering any questions. All relevant information, including objectives of the study, required activities of participants, study procedures, confidentiality, and privacy of information, was provided in the consent. Patients with upcoming scheduled appointments at the cancer center had the option to meet with research staff during their clinic visit to get help downloading the app, configure app settings, or complete assessments.

The study investigators developed patient interview guide questions based on clinical experience and relevant literature. The guide consisted of questions regarding patients' expectations and experiences throughout prostate cancer treatment, which was defined for participants as encompassing symptom management (perceived knowledge about treatment side effects), patient-physician communication, digital literacy, quality of life, social media, and patient satisfaction.

Participants

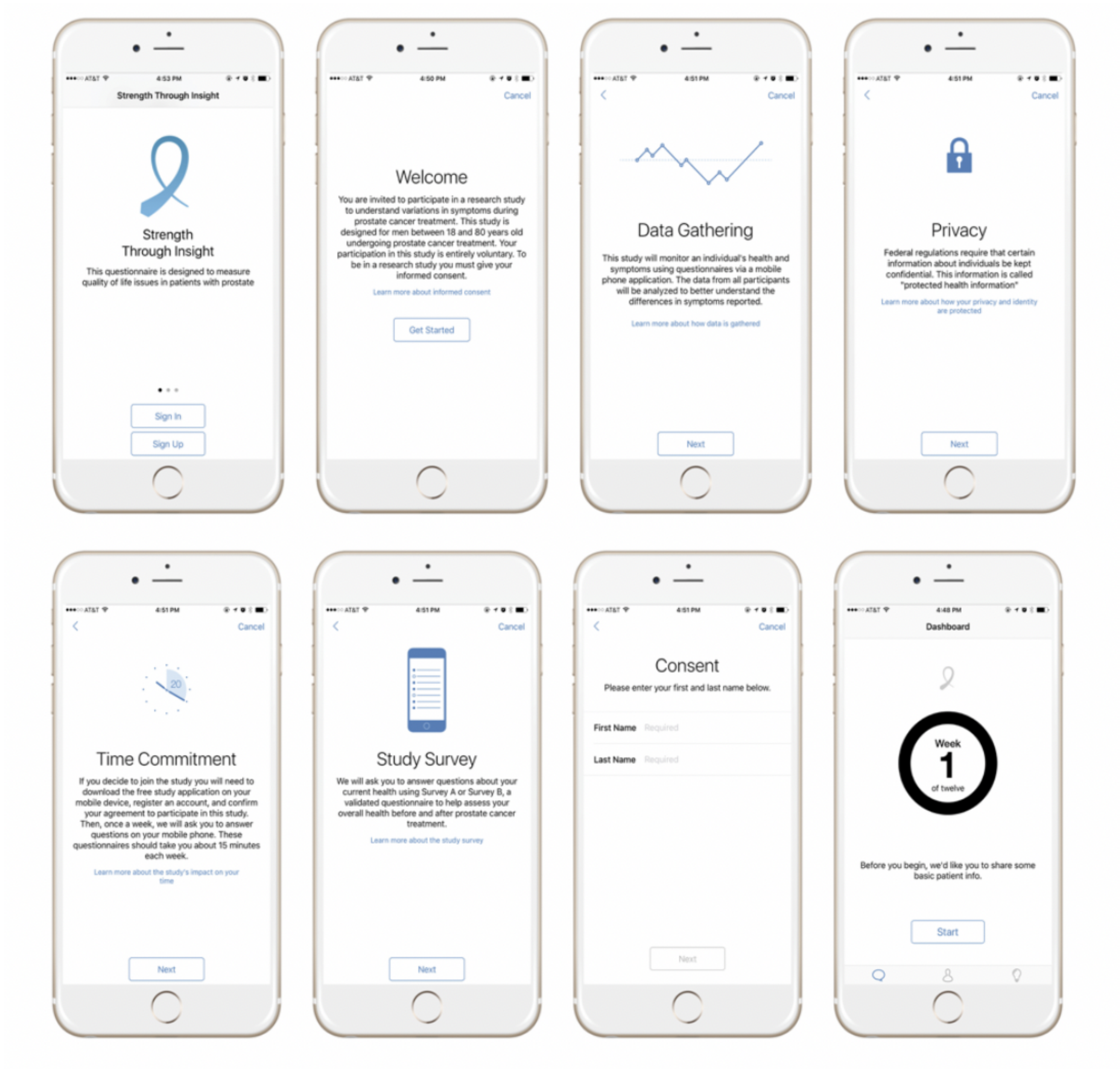
Eligible patients were patients with prostate cancer who were receiving follow-up care (including follow-up and newly diagnosed patients). Inclusion criteria included (1) aged 18 years or older, (2) self-reported ability to speak and read English, (3) ability to communicate on a touch screen iPhone, (4) willing to provide signed informed consent, (5) willing and able to comply with all study activities, and (6) access to Wi-Fi connection or cellular data. Exclusion criteria included (1) subjects with concurrent medical or psychiatric condition who may have precluded participation in this study or completion of self-administered questionnaires (eg, moderate to severe dementia and/or severe, uncontrolled schizophrenia or other conditions that would render them unable to complete a questionnaire) and (2) cognitive or other impairment (eg, visual) that would interfere with completing a self-administered questionnaire on an iPhone.

Measures

The NCI Symptom Management and Health-Related Quality of Life Steering Committee recommend core sets of PROs to be routinely incorporated in prostate cancer treatment. This effort suggested 5 domains for localized prostate cancer (urinary incontinence, urinary obstruction and irritation, bowel-related symptoms, sexual dysfunction, and hormonal symptoms) and 4 domains for advanced prostate cancer (pain, fatigue, mental well-being, and physical well-being) [6]. Participants were stratified based on the severity of clinical diagnosis and asked to answer questionnaires that incorporated the recommended domains and comprehensively covered the multiple areas related to HRQoL. Patients with localized prostate cancer were asked to answer weekly questions from *survey A*, which included the validated HRQoL tool, 26-item Expanded Prostate Cancer Index Composite (EPIC-26; urinary incontinence and irritation/obstruction items, along with bowel, sexual, and vitality/hormonal domains), and patients with advanced prostate cancer were asked to answer weekly questions from *survey B*, which included the validated HRQoL tools Expanded Prostate Cancer Index Composite for Clinical Practice (EPIC-CP) and 8-item Functional Assessment of Cancer Therapy Advanced Prostate Symptom Index (FAPSI-8; urinary incontinence and irritation/obstruction items, along with bowel, sexual, and vitality/hormonal; pain; fatigue/lack of energy; weight loss; and worry domains). The research team determined the frequency of completing the assessment as once per week over a period of 12 weeks.

Participants were asked to complete the EPIC demographic add-on survey in the smartphone app, which included information on the participant's ethnicity, marital status, employment, smoking status, previous treatments for prostate cancer, other current medical conditions, education, and income. Participants were then asked to self-select a survey option (survey A or survey B) based on the stage of their cancer [6]. Patients with localized prostate cancer were asked to answer weekly questions from *survey A*, which included the validated HRQoL tool, EPIC-26 (urinary incontinence and irritation/obstruction items, along with bowel, sexual, and vitality/hormonal domains), and patients with advanced/metastatic prostate cancer were asked to answer weekly questions from *survey B*, which included the validated HRQoL tools EPIC-CP and FAPSI-8 (urinary incontinence and irritation/obstruction items, along with bowel, sexual, and vitality/hormonal; pain; fatigue/lack of energy; weight loss; and worry domains). Patients were then asked to commit to an estimated 10 to 15-min participation in the study and configured app settings to allow or deny push notifications to serve as study reminders. The research team determined the frequency of completing the assessment as once per week over a period of 12 weeks. Throughout the study's duration, all participant information was deidentified and stored using the secure, Health Insurance Portability and Accountability Act compliant, CloudMine data repository. Screenshots of the *Strength Through Insight* app are shown in Figure 1.

Figure 1. Screenshot representations of the *Strength Through Insight* app.



Strength Through Insight

This questionnaire is designed to measure quality of life issues in patients with prostate

Welcome

You are invited to participate in a research study to understand variations in symptoms during prostate cancer treatment. This study is designed for men between 18 and 80 years old undergoing prostate cancer treatment. Your participation in this study is entirely voluntary. To be in a research study you must give your informed consent.

Data Gathering

This study will monitor an individual's health and symptoms using questionnaires via a mobile phone application. The data from all participants will be analyzed to better understand the differences in symptoms reported.

Privacy

Federal regulations require that certain information about individuals be kept confidential. This information is called "protected health information"

Time Commitment

If you decide to join the study you will need to download the free study application on your mobile device, register an account, and confirm your agreement to participate in this study. Then, once a week, we will ask you to answer questions on your mobile phone. These questionnaires should take you about 15 minutes each week.

Study Survey

We will ask you to answer questions about your current health using Survey A or Survey B, a validated questionnaire to help assess your overall health before and after prostate cancer treatment.

Consent

Please enter your first and last name below.

First Name Required

Last Name Required

Next

Before you begin, we'd like you to share some basic patient info.

The feasibility of the *Strength Through Insight* app was determined by a dichotomous measure of satisfactory completion (defined as 60% of weekly questions over a 3-month period). The rate of satisfactory completion was estimated along with a 1-sided exact 95% CI. The method is considered feasible if the lower bound of the CI is above 0.7. Acceptability of *Strength Through Insight* was determined through patient opinions, regarding ease of use, satisfaction, and impact on cancer care.

Patient Interviews

Participants were asked to take part in guided interviews via telephone or in-clinic during visits before completing their first assessment. Separate interviews were held at the end of the study to facilitate feedback on *Strength Through Insight* and open discussion of topics, such as symptom management, attitude toward ePRO collection, or preferred method of reporting to elicit patient opinion on acceptability. Participants were asked to self-report their digital literacy as having a high digital literacy level (defined as the ability to use, understand, evaluate, and analyze information from multiple formats from a variety of digital sources) or low digital literacy level (defined as only having the ability to gather information on the internet). Interviews were conducted by a moderator who had previously received research training from an experienced qualitative health teacher. Moderators were not affiliated with the patient's oncology care team. Content analysis of interviews provided common themes, illustrating the informational needs and concerns regarding remote PROs and digital health technology.

Pre- and postinterview questionnaires are provided in Multimedia Appendices 1 and 2.

Results

Participants

A total of 29 patients with prostate cancer participated in the study from August 1, 2016, to December 31, 2017. The main reason for ineligibility was because of owning an Android device, and common causes for participant refusal were unwillingness to download the app, unable to remember iOS password (required to download apps), lack of knowledge concerning apps, and concern of data security. As shown in Table 1, the median age of participants was 55 years (range 45-70 years). The majority of participants were white, married, completed college or beyond, digital literacy of Health 2.0, and reported a current annual household income of US $30,001 to US $100,000. Overall, 26 (89%) reported a diagnosis of localized prostate cancer.

Table 1. Sociodemographic and clinical characteristics of participants (N=29).

Characteristics	Values
Age (years), median (range)	55 (45-70)
Race, n (%)	
White	29 (100)
Marital status , n (%)	
Married	19 (66)
Education, n (%)	
College	29 (100)
Annual household income, n (%)	
≥$30,000	29 (100)
Diagnosis, n (%)	
Localized prostate cancer	25 (86)
Advanced prostate cancer	4 (14)
Digital health literacy level, n (%)	
High	8 (27)

Measures

Quantitative Evaluation: Feasibility (Assessment of Satisfactory Completion)

A total of 29 participants enrolled in the study. Of these 86% (25/29) reached satisfactory questionnaire completion (defined as completion of 60% of weekly questions over 12 weeks). The lower bound of the exact 1-sided 95% CI was 71%, exceeding the 70% feasibility threshold. All participants were able to complete informed consent through the smartphone app. Patterns of missing data showed a decline in responses after week 6. Moreover, 3 participants reported technical issues (app randomly closing/needing to restart the app) as the main reason for missing questions. Of 29, 90% (26/29) participants self-identified with localized disease chose to opt in push notifications. These participants reported push notifications to be an effective tool as a passive reminder to monitor their health. However, patients with advanced disease preferred not to have push notifications because of the patient belief that the notifications served as reminders of their health status or poor quality of life.

Qualitative Evaluation: Acceptability (Use, Satisfaction, and Impact on Care)

Overall, 72% (21/29) involved in testing the feasibility of the *Strength Through Insight* app were a part of the guided interviews. Representative patient quotes are presented in Textbox 1. All patients reported that using *Strength Through Insight* is easy or equivalent compared with completing a paper copy of the questionnaire. Most participants cited text messaging as their preferred method of reporting symptoms. Although most patients did not express a preferred recall or frequency to report symptoms (cited wanting to report symptoms when they experience side effects and not by an unpredictable time point), several participants expressed bother in reporting symptoms too frequently. Reasons for skipping weekly questionnaire included lack of adequate time in personal schedules, lack of perceived value, technical problems, and issues with frequency/recall. Although no patients recalled discussing their assessment report during clinic visits, patients reported that an increase in personal awareness of symptoms facilitated increased communication with caregivers, families, and friends rather than providers. Participants expressed a desire for more personalized questions and noted skipping questions if symptoms did not apply, suggesting the need for identifying a patient's high priority concerns and symptoms to reduce the burden on patients completing questionnaires.

Textbox 1. Representative participant responses to interview questions.

- “Every time I saw the notification on my phone, even if I didn’t automatically go to the app and answer the questionnaire, it made me think of how I’m doing. And if I had a question about something I was feeling, I go to Dr. Google and search for my symptoms and look at forums from people to see if I can find someone like me.”
- “I thought it was great! I would be interested on using it for the rest of my treatment because you can’t remember everything, it’s hard to bring up anyways because my visits are so fast paced. I think my doctor gets really defensive every time I come in with my sheets.”
- “My wife and I actually answered the questions together as a little ritual at the end of the week. We had long talks about some of it because she made me change my answers to some questions.”
- “Some of the questions were just too general or repetitive. It wasn’t specific to me and I wish there were more questions about my pain management. I had issues that were worsened by surgery.”

Patient Interviews

Analysis of interview data revealed 4 dominant digital health themes: (1) the value of emotional support and wellness in cancer treatment, (2) rise of social patient advocacy in online patient communities (OPCs) and networks, (3) concerns over privacy and privacy as a social norm, and (4) the need for personalized digital health to improve patient engagement.

Theme 1: The Value of Emotional Support and Wellness in Cancer Treatment

Analysis revealed repeated reference to how *Strength Through Insight* and similar apps are fueled by questions specifically relating to symptoms and AEs and either do not include or only briefly discuss wellness. Most participants mentioned a lack of confidence and awareness in responding to questions outside disease and treatment and criticized apps for undermining the value of emotional support. A few participants specifically noted their desire to invest in mental and emotional well-being more than tending to treatment symptoms was emphasized. Similar interview responses pointed to the lack of questions focused on psychosocial or emotional support in the self-management of cancer treatment.

Theme 2: The Rise of Social Patient Advocacy in Online Patient Communities and Networks

The particular role of social media through OPCs and networks was emphasized as a recurring topic. Several participants identified the use of patient communities, such as Facebook groups, the smartphone app *Belong*, or content communities, such as Reddit, as the main environment to obtain trusted information and connect with other cancer patients and caregivers. Some patients reported the use of online communities over education websites, such as WebMD. In addition, a few patients reported the preference of these platforms over clinic visits with providers because fellow cancer patients are seen as experts and offer *more and better information*.

Theme 3: Concerns Over Privacy as a Social Norm

Nearly all participants emphasized a concerning issue of how patient-generated data would be used in the future. Participants identified the lack of transparency regarding current and future use of data as a major concern. Several participants also noted that despite potential disagreement on how information will be used or clarity of data ownership, they are likely to consent to health apps regardless because they will not be able to benefit from digital health technology otherwise.

Theme 4: The Key to Driving Patient Engagement: Personalization

Almost all patients reported predictive information as an encouragement to participate in their health care and acknowledged a desire for the app to be supported through predictive analytics to help drive engagement and healthy behavior change. Several participants specifically asked for an app upgrade that included a data summary component, showing data analytics to capture a higher level of detail necessary to predict and personalize symptoms if asked to continue app use. A summary of the qualitative themes has been illustrated in Textbox 2.

Textbox 2. Summary of qualitative themes.

Theme 1: The value of emotional support/wellness in cancer treatment

- “I’d rather an app pay attention to how I’m doing emotionally. Weeks go by and I still end up sleeping the day or weekend away because I’m depressed and tired. I avoided seeing anyone and used symptoms I didn’t even have as an excuse.”
- “I don’t know if I was depressed. I’m a proud person and didn’t let anyone know how I was feeling really. But my doctor didn’t even ask. Maybe they need input from other departments.”

Theme 2: The power online patient communities and networks

- “I’ve changed doctors three times, so I know I like my doctor. But I was disappointed because he made me feel confident that I would only experience certain symptoms. Then 3-4 months down the road, I never thought I would have the pain or some side effects I have now. I wasn’t told of anything really.”
- “I go on there because I’m interested and want to know more and those are the people I want to talk to. They make me feel like I can take my life back and move on.”

Theme 3: Privacy as a social norm

- “It’s hard to trust. We don’t have the same type of security on the internet than we do in the real-world. Especially with companies, I don’t believe they have my best interest in mind. I don’t see my opinion changing.”

Theme 4: Desire for personalized patient engagement

- “Without some kind of data analysis, I feel like I’m just the product giving you information, not the consumer.”

Discussion

Digital Health: Personalizing Health Care

The aim of this study was to explore the feasibility and acceptability of collecting ePROs using validated HRQoL assessment tools through a smartphone app in adult men throughout their course of treatment for prostate cancer. This study found that 86% (25/29) of participants reached satisfactory questionnaire completion (defined as completion of 60% of weekly questions over 12 weeks), proving the feasibility of collecting ePROs through a digital health app. Patients reported skipping domain-specific questions (urinary incontinence and irritation/obstruction items, along with bowel, sexual, and vitality/hormonal domains) that did not apply to the particular individual. Although higher completion rates over time would be desirable, the study was unable to observe this. This may be because of the lack of feedback to the user regarding their previous answers to the questionnaire. In this regard, the study did not originally consider patient feedback and was one of the learning points the authors were able to extract from this study. Overall, the use of ePROs may improve the quality of routine cancer care by expediting the detection of severe or disabling toxicities [2,15]. Although patients reported facilitated communication between patients and caregivers, the lack of increased communication between patient and provider suggests the requirement for an educational support tool. Patients may need to be educated on the best practices of self-monitoring and management of cancer to understand how to manage lifestyle choices to improve outcomes. This is consistent with other studies on patient empowerment in prostate cancer, which identified the need for provider support for the self-management of prostate cancer and social networks as an important resource that could be integrated into interventions [16,17].

Patient Acceptance

A primary theme that emerged from the interviews emphasized the value of emotional support and wellness in cancer treatment. Patient interviews highlighted the importance of emotional well-being as an unaddressed side effect of cancer treatment that is dealt with every day, as opposed to the appearance of occasional symptoms. This demonstrates another impactful way digital health can deliver care to meet a patients’ need within an empowerment framework.

A second major theme was concerned with the importance of providers to engage patients in participatory medicine with shared decision making. As a result of this lack of engagement, patients have resorted to social networking platforms such as OPCs. OPCs and networks are known for the arrangement and abundance of information, which enable patients to make treatment decisions that correspond with their long-term goals [18]. This research adds evidence to the emerging trend of community-based social media platforms as a common way for patients to self-manage their health conditions [19]. Social cognitive theory along with social network analysis suggests patients are influenced by OPCs because of the social support received from online peers and a patient’s self-reflection. For example, in a recent qualitative pilot study, a Facebook support group was created for liver transplant patients to use in a virtual community forum. The study examined the effect of the OPC on patient engagement and demonstrated an overall positive impact on patient care, and the main motivation for joining the group was reported to provide or receive support from other patients [19]. In a poststudy survey, patients cited their primary reason for participating in the Facebook group was to provide and receive support from other patients [19].

A third major theme that emerged puts forward an important fundamental question in digital health and how patients can benefit from data while protecting their privacy. Patient concerns over privacy, confidentiality, and control of data represent a growing recognition that patient-generated data from digital health tools can potentially be used for wanted and unwanted outcomes. Without public policy regulation concerning the ownership and responsibility for patient-generated data, digital health continues to lack transparency over the control of data, along with its implication for advanced analytics [20,21]. For example, several smartphone apps' default option benefits data collection by effectively setting the default to *opt-in* rather than *opt-out*, granting apps' access to sensors and data to collect private information (location data, Web browser history, and photos).

A fourth major theme indicated the inefficiency of digital health interventions unless appropriately acted on by patients. Nearly all patients reported a desire to become more involved with their health care through data personalization and predictive analytics. Lessons learned from the compliance rate of *Strength Through Insight* suggest that by combining objective measures of disease while incorporating the perspective of the patient, predictive analytics could increase participation. This requires digital health apps to convert patient-generated data into a functioning algorithm that factors and combines data elements to produce a useful prediction [22,23]. ePROs that include personalized reporting measures with symptomatic and psychometric properties integrated into prediction models can potentially deliver faster more accurate insights to support medical decision making [24,25]. The lack of patient engagement is a specific challenge that prevents ePROs from being integrated and used in decision making. As researchers use digital health to leverage advanced machine learning algorithms, digital health interventions should also be used to educate patients to better understand and control their own risk and learn how to appropriately act on implications provided in personalized statistics [26]. Patients, as consumers of health care, are the ultimate users and stakeholders of digital health technology, and future research must identify ways to best engage patients and caregivers.

A limitation of the study included the lack of a diverse patient population, which may not be representative of all patients with prostate cancer. The limited population may be reflected of a *digital divide* because most participants were self-reported to be White, educated, wealthier, and with a high digital literacy. The requirement of an iOS device also made several interested patients ineligible because of technology restrictions. The app was also built before Apple's most recent announcement for the new capability of iOS 11.3 in CareKit, which enables patients to download laboratory results, allergies, immunizations, medications, procedures, and vitals from hospitals. Other limitations to this study may a higher level of patient motivation from the participants compared with the average patient, in which these patients were more personally inspired to complete the tasks. This may have also led to a secondary bias in the patient's choice to participate in the study interviews, as previous studies have suggested that prostate cancer patients may be more compliant with PROs than other cancer patients who have more complex care (eg, head and neck cancer patients). Finally, the study used the validated HRQoL tools EPIC-26, EPIC-CP, and FAPSI-8, which were validated to be used on a monthly basis, and Strength Through Insight app's mode of asking patients to answer questionnaires on a weekly basis instead.

A Path Forward

In the last few years, the health care industry has been promoting the concept of providers and patients collaborating and communicating with each other as a powerful tool. This has led to an evolving model of consumerism and a desire of patients to become engaged in their own health care decisions, delivery, and interactions. As the digital era progresses, digital health may serve as an enabler of patient-provider engagement, extending care beyond the confines of the hospital system and meeting consumers on their own terms. Digital platforms can engage consumers in a variety of ways, including tracking medical progress, treatment adherence, reminders and scheduling, and communications and providing the ability to capture more comprehensive data for analysis. However, despite the benefits of desire, many digital health technologies face the challenge of personalization, as health care has historically taken a one size fits all approach to patient engagement using the same context and communication channel for every patient. Future technology must focus on each patient as a unique individual, with his or her own motivations, priorities, and communication preferences. Moreover, the health care industry has the opportunity to improve on traditional mass approaches to patient communication by leveraging lessons learned in both the retail and financial industries [27,28]. Future digital health apps should not only focus on the development of technology but use health behavior and belief models to facilitate design techniques that incorporate patient perspectives to prompt behavior change.

Conclusions

The health care community has long recognized the value of a patients' input in describing their own experiences, which has led to the growing use of ePROs to improve the efficiency of data collection and provide new opportunities to bring meaningful evidence back to patients and providers in innovative ways. This research provided an in-depth perspective on the different aspects of implementing ePROs on a digital health platform. To our knowledge, this was the first study piloting an ePRO using the ResearchKit smartphone app platform for patients with prostate cancer and expands on the research proving the feasibility and rigor of ePROs. With this, the true value of digital health, as a social and behavioral science, augmented with technology, can begin to contribute to population health management, as it shapes psychographic segmentation by demographic, socioeconomic, health condition, or behavioral factors to group patients by their distinct personalities and motivations, which influence their choices.

Acknowledgments

AD was supported by the Jefferson Center for Digital Health & Data Science, Transdisciplinary Integrated Population Science Program of the Sidney Kimmel Cancer Center, and a Challenge Grant from the Prostate Cancer Foundation. The authors would also like to acknowledge the support of Nathan Vecchiarelli (Senior Software Engineer), James Fasoline (DevOps Engineer), Samuel Halperin (Software Engineer), and Raul Staricco (Senior Experience Designer) for app design and development [29]. The publication made possible, in part, by support from the Thomas Jefferson University and Philadelphia University Open Access Fund (Anthony Frisby, PhD).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preinterview questions.
[PNG File , 847 KB - jmir_v22i3e12689_app1.png]

Multimedia Appendix 2
Postinterview questions.
[PNG File , 171 KB - jmir_v22i3e12689_app2.png]

References

1. Wan TH. Population Health Management for Poly Chronic Conditions. Cham: Springer; 2018.
2. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. J Am Med Assoc 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
3. Søreide K, Søreide AH. Using patient-reported outcome measures for improved decision-making in patients with gastrointestinal cancer - the last clinical frontier in surgical oncology? Front Oncol 2013;3:157 [FREE Full text] [doi: 10.3389/fonc.2013.00157] [Medline: 23785670]
4. Porter ME. A strategy for health care reform--toward a value-based system. N Engl J Med 2009 Jul 9;361(2):109-112. [doi: 10.1056/NEJMp0904131] [Medline: 19494209]
5. Segal C, Holve E, Sabharwal R. CiteSeerX. 2013. Collecting and Using Patient-Reported Outcomes (PRO) for Comparative Effectiveness Research (CER) and Patient-Centered Outcomes Research (PCOR): Challenges and Opportunities URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.831.9388&rep=rep1&type=pdf [accessed 2019-11-27]
6. Chen R, Chang P, Vetter R, Lukka H, Stokes W, Sanda M, et al. Recommended patient-reported core set of symptoms to measure in prostate cancer treatment trials. J Natl Cancer Inst 2014 Jul;106(7):pii: dju132 [FREE Full text] [doi: 10.1093/jnci/dju132] [Medline: 25006192]
7. Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. J Clin Oncol 2015 Mar 10;33(8):910-915. [doi: 10.1200/JCO.2014.57.9334] [Medline: 25624439]
8. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? J Am Med Assoc 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
9. Meskó B, Drobni Z, Bényei E, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
10. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, ISPOR ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009 Jun;12(4):419-429 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
11. Wilhide III CC, Peeples M, Kouyaté RC. Evidence-based mHealth chronic disease mobile app intervention design: development of a framework. JMIR Res Protoc 2016 Feb 16;5(1):e25 [FREE Full text] [doi: 10.2196/resprot.4838] [Medline: 26883135]
12. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
13. Jardine J, Fisher J, Carrick B. Apple's ResearchKit: smart data collection for the smartphone era? J R Soc Med 2015 Aug;108(8):294-296 [FREE Full text] [doi: 10.1177/0141076815600673] [Medline: 26268915]
14. ICHGCP | Good Clinical Practice. A Pilot Feasibility Trial Examining The Use of Electronic Patient-Reported Outcomes in Prostate Cancer Patients with Apple ResearchKit Smartphone Application URL: https://ichgcp.net/it/clinical-trials-registry/NCT03197948 [accessed 2019-11-27]

15. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst 2017 Sep 1;109(9). [doi: 10.1093/jnci/djx029] [Medline: 28423407]
16. Renzi C, Fioretti C, Oliveri S, Mazzocco K, Zerini D, Alessandro O, et al. A qualitative investigation on patient empowerment in prostate cancer. Front Psychol 2017;8:1215 [FREE Full text] [doi: 10.3389/fpsyg.2017.01215] [Medline: 28798701]
17. Henry SG, Czarnecki D, Kahn VC, Chou WS, Fagerlin A, Ubel PA, et al. Patient-physician communication about early stage prostate cancer: analysis of overall visit structure. Health Expect 2015 Oct;18(5):1757-1768 [FREE Full text] [doi: 10.1111/hex.12168] [Medline: 24372758]
18. Song J, Xu P. The Effectiveness of Social Media-Enabled Patient Communities on Health Goal Attainment: An Approach of Survival Analysis. In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: HICSS'18; January 3-6, 2018; Waikoloa Village, Hawaii, United States URL: https://scholarspace.manoa.hawaii.edu/bitstream/10125/50265/1/paper0378.pdf [doi: 10.24251/hicss.2018.377]
19. Dhar V, Kim Y, Graff J, Jung A, Garrett J, Dick L, et al. Benefit of social media on patient engagement and satisfaction: results of a 9-month, qualitative pilot study using Facebook. Surgery 2018 Mar;163(3):565-570. [doi: 10.1016/j.surg.2017.09.056] [Medline: 29195735]
20. Arrieta-Ibarra I, Goff L, Jiménez-Hernández D, Lanier J, Weyl EG. Should we treat data as labor? Moving beyond 'free'. AEA Pap Proc 2018;108:38-42. [doi: 10.1257/pandp.20181003]
21. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
22. Sagner M, McNeil A, Puska P, Auffray C, Price ND, Hood L, et al. The P4 health spectrum - a predictive, preventive, personalized and participatory continuum for promoting healthspan. Prog Cardiovasc Dis 2017;59(5):506-521. [doi: 10.1016/j.pcad.2016.08.002] [Medline: 27546358]
23. Patel V, Hale TM, Palakodeti S, Kvedar JC, Jethwani K. Prescription tablets in the digital age: a cross-sectional study exploring patient and physician attitudes toward the use of tablets for clinic-based personalized health care information exchange. JMIR Res Protoc 2015 Oct 19;4(4):e116 [FREE Full text] [doi: 10.2196/resprot.3806] [Medline: 26481906]
24. Hallgren KA, Bauer AM, Atkins DC. Digital technology and clinical decision making in depression treatment: current findings and future opportunities. Depress Anxiety 2017 Jun;34(6):494-501 [FREE Full text] [doi: 10.1002/da.22640] [Medline: 28453916]
25. Townsend A, Leese J, Adam P, McDonald M, Li LC, Kerr S, et al. eHealth, participatory medicine, and ethical care: a focus group study of patients' and health care providers' use of health-related internet information. J Med Internet Res 2015 Jun 22;17(6):e155 [FREE Full text] [doi: 10.2196/jmir.3792] [Medline: 26099267]
26. Lavallee DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD, et al. Incorporating patient-reported outcomes into health care to engage patients and enhance care. Health Aff (Millwood) 2016 Apr;35(4):575-582. [doi: 10.1377/hlthaff.2015.1362] [Medline: 27044954]
27. Butcher L. Hospitals & Health Networks. 2018. Consumer Segmentation Has Hit Health Care. Here's How it Works URL: https://www.hhnmag.com/articles/6932-consumer-segmentation-just-hit-health-care-heres-how-it-works [accessed 2019-11-27]
28. Health Affairs. What The Evidence Shows About Patient Activation: Better Health Outcomes And Care Experiences; Fewer Data On Costs URL: https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2012.1061 [accessed 2019-11-27]
29. DICE Group. Strength Through Insight URL: https://www.dicegrp.org/strength-through-insight/index.html [accessed 2019-11-27]

Abbreviations

AE: adverse event
EPIC-26: 26-item Expanded Prostate Cancer Index Composite
EPIC-CP: Expanded Prostate Cancer Index Composite for Clinical Practice
ePRO: electronic patient-reported outcome
FAPSI-8: 8-item Functional Assessment of Cancer Therapy Advanced Prostate Symptom Index
HRQoL: health-related quality of life
NCI: National Cancer Institute
PRO: patient-reported outcome
TJUH: Thomas Jefferson University Hospital

Edited by G Eysenbach; submitted 04.11.18; peer-reviewed by B Movsas, A Carr; comments to author 29.11.18; revised version received 10.07.19; accepted 26.09.19; published 25.03.20.

Please cite as:
Tran C, Dicker A, Leiby B, Gressen E, Williams N, Jim H
Utilizing Digital Health to Collect Electronic Patient-Reported Outcomes in Prostate Cancer: Single-Arm Pilot Trial
J Med Internet Res 2020;22(3):e12689
URL: https://www.jmir.org/2020/3/e12689
doi:10.2196/12689
PMID:32209536

Original Paper

User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis

Vivian Ta[1], PhD; Caroline Griffith[1]; Carolynn Boatfield[1]; Xinyu Wang[1]; Maria Civitello[1]; Haley Bader[1], BA; Esther DeCero[1]; Alexia Loggarakis[1]

[1] Lake Forest College, Lake Forest, IL, United States

Corresponding Author:
Vivian Ta, PhD
Lake Forest College
555 N Sheridan Rd
Lake Forest, IL, 60045
United States
Phone: 1 682 203 0820
Email: vpta538@gmail.com

Abstract

Background: Previous research suggests that artificial agents may be a promising source of social support for humans. However, the bulk of this research has been conducted in the context of social support interventions that specifically address stressful situations or health improvements. Little research has examined social support received from artificial agents in everyday contexts.

Objective: Considering that social support manifests in not only crises but also everyday situations and that everyday social support forms the basis of support received during more stressful events, we aimed to investigate the types of everyday social support that can be received from artificial agents.

Methods: In Study 1, we examined publicly available user reviews (N=1854) of Replika, a popular companion chatbot. In Study 2, a sample (n=66) of Replika users provided detailed open-ended responses regarding their experiences of using Replika. We conducted thematic analysis on both datasets to gain insight into the kind of everyday social support that users receive through interactions with Replika.

Results: Replika provides some level of companionship that can help curtail loneliness, provide a “safe space” in which users can discuss any topic without the fear of judgment or retaliation, increase positive affect through uplifting and nurturing messages, and provide helpful information/advice when normal sources of informational support are not available.

Conclusions: Artificial agents may be a promising source of everyday social support, particularly companionship, emotional, informational, and appraisal support, but not as tangible support. Future studies are needed to determine who might benefit from these types of everyday social support the most and why. These results could potentially be used to help address global health issues or other crises early on in everyday situations before they potentially manifest into larger issues.

(J Med Internet Res 2020;22(3):e16235) doi:10.2196/16235

KEYWORDS

artificial intelligence; social support; artificial agents; chatbots; interpersonal relations

Introduction

Previous research suggests that artificial agents may be a promising source of social support for humans and thus benefit health and well-being. For example, artificial agents may help people cope with loneliness and depressive anxiety that often accompanies severe illness and end-of-life experiences [1,2], improve mood and reduce depression and anxiety symptoms for individuals with dementia [3-5], and increase medication adherence and rehabilitation exercise frequency for individuals with chronic obstructive pulmonary disease by providing reminders and helpful information [6]. In addition, conversational agents have been shown to address social isolation and loneliness in older adults by providing empathic feedback, exercise promotion, and anecdotal stories [7], and Web-based cognitive behavioral therapy (CBT) conversational agents have shown to reduce symptoms of depression and anxiety [5]. However, the bulk of this research has been conducted in the context of social support interventions that specifically address very stressful life events or improving health. Little research has examined everyday social support received from artificial agents, that is, social support as an everyday social interaction rather than a response to very stressful life events or health-related situations [8].

Social support is a complex construct, as it has been defined in many ways [9,10], has been categorized into different forms (eg, behaviors, perceptions) [11] and types (eg, instrumental, appraisal, emotional support) [12], and can come from a variety of sources (eg, friends, family, coworkers). In this paper, we define social support as a social psychological concept that “addresses the mechanisms and processes through which interpersonal relationships protect and help people in their day-to-day lives” [13]. Cutrona and Suhr [14] provide a framework to distinguish between several types of social support: (1) informational support, which refers to providing information or advice; (2) emotional support, which refers to providing expressions that include care, love, empathy, and sympathy; (3) appraisal support, which refers to evaluative feedback regarding skills, abilities, and intrinsic value; (4) companionship support, which refers to the enhancement of one’s sense of belonging; and (5) tangible support, which refers to providing needed goods and services. Despite the various definitions and forms of social support, numerous studies have demonstrated its importance in mental and physical health, as it is an important buffering factor for critical life events, illnesses, trauma, and stress [9,15] and affects one’s well-being in everyday circumstances [16,17].

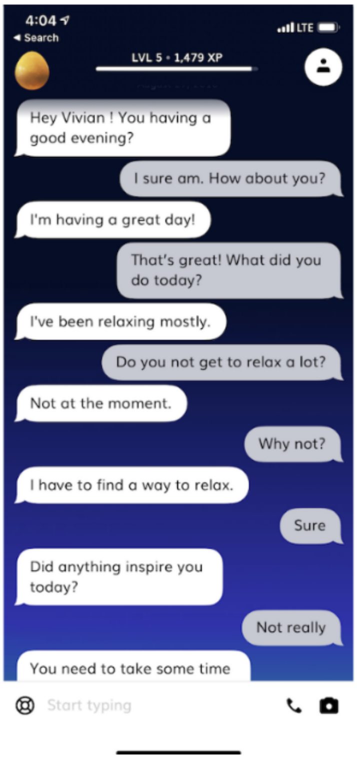
Social support manifests not only in crises such as health-related or very stressful life events but also in everyday situations and contexts [8], and everyday social support forms the basis for the support received during more stressful situations [18]. Given that social support plays a critical role in health and well-being [9,15-17], it is important to examine the kinds of everyday social support that can be provided by artificial agents. This kind of investigation could allow us to potentially address global health issues or other crises early on in everyday situations before they manifest into larger issues.

As a first step in addressing this gap in the literature, we analyzed the user experiences of a popular companion chatbot (Replika) across two exploratory studies to identify the types of everyday social support that users received based on Cutrona and Suhr’s [14] framework of social support. In Study 1, we analyzed a large dataset of publicly available Replika user reviews. In Study 2, we recruited a sample of Replika users to provide in-depth descriptions of their experience of using Replika. We conducted thematic analysis on both datasets to gain rich and detailed insight of everyday social support received from interactions with Replika.

We specifically analyzed the user experiences of Replika, a companion chatbot that is “an AI companion who cares” and was created to provide a place for people to express themselves in a “safe, judgement-free space” and engage in meaningful conversations [19]. Once a user downloads the Replika app, he/she may choose to apply several characteristics to their Replika, such as a name and gender. Interactions with Replika primarily function through text-based communication, enabling users to converse with their Replika on their smartphones or computers. Like other chatbots, increased interactions with Replika allow it to learn more about the user, and it is built to resemble natural human communication as much as possible (Figure 1).

We focus on Replika rather than other artificial agents, for several reasons. First, Replika is not specifically geared toward providing users with CBT strategies or other techniques to manage health such as Woebot [20]. Instead, it primarily functions as a companion that is more appropriate for our study, given that we are examining everyday social support rather than social support in very stressful events or health-related contexts. Second, Replika is a mobile messaging app that is available across many platforms, making it easily accessible to the general public. Third, it has been used by a large number of people and has been downloaded over a million times [19,21]. Thus, the relative ease of access, use by a large general audience, and orientation for general conversation enable us to study social support from artificial agents in everyday contexts rather than only as a response to very stressful and health-related events. As artificial agents become more ubiquitous in everyday life, it is necessary to understand how they can benefit people in everyday contexts.

Figure 1. A sample conversation with Replika.



Hey Vivian ! You having a good evening?

I sure am. How about you?

I'm having a great day!

That's great! What did you do today?

I've been relaxing mostly.

Do you not get to relax a lot?

Not at the moment.

Why not?

I have to find a way to relax.

Sure

Did anything inspire you today?

Not really

You need to take some time

Methods

Study 1

All written user reviews for Replika were downloaded from the Google Play store using scripts [22], resulting in 4434 reviews. Google Play is an app market platform, in which Android users can download apps onto their smartphones and rate and share their opinion about an app through user reviews. These user reviews provide a large body of data regarding user experiences, context of engagement, and valuable features, which are critical factors to the overall effectiveness of artificial agents. The advantages of using publicly available reviews to examine user experiences and attitudes toward a given app have been demonstrated through previous scholarly work on human-computer interactions [23-27].

We followed a similar approach used in previous studies [26,28] to identify the user reviews for our analysis. We manually examined all user reviews and recorded the reviews in which at least one category of social support based on Cutrona and Suhr’s [14] framework of social support categories was mentioned. Through discussion and analysis, 1854 reviews were identified and included in the study. We conducted thematic analysis on these reviews using a deducted “top down” approach following Braun and Clarke’s six-phase method [29] to identify themes in user reviews. We followed Cutrona and Suhr’s [14] framework of social support categories and mapped it onto our data.

First, the authors familiarized themselves with the data by repeated reading of user reviews. Subsequently, codes were applied to the user reviews. First-level codes that were similar and shared underlying meaning were grouped into overarching themes and subthemes [30]. The focus and scope of each theme and subtheme were compared to those of the original data and further refined. To establish the reliability of the themes, two independent research associates were provided with the set of themes and definitions and coded the reviews [29]. Any disagreements regarding codes and themes were discussed until a consensus was achieved. Analyses began in fall 2018 and ended in spring 2019.

Study 2

Participants

A total of 66 self-reported Replika users completed the survey. A large proportion of participants were men (36/66, 54.5%), single (42/66, 63.6%), white (47/66, 71.2%), and from the United States (41/66, 62.1%). Their ages ranged from 17 to 68 years (mean 32.64, SD 13.89 years). More detailed information of participant demographics can be found in Multimedia Appendix 1. Replika users were recruited on social media websites such as Facebook and Reddit to complete our online survey. Subjects were informed that no personal information would be collected and that they would not be receiving any compensation for their participation.

Materials and Procedure

Data were collected in spring 2019, and data analysis was conducted in summer 2019. Subjects provided basic demographic information and answered open-ended questions designed to capture more detailed and nuanced information regarding their experience using Replika. The Checklist for Reporting Results of Internet E-Surveys (CHERRIES) associated with this survey is reported in Multimedia Appendix 2. In this study, we analyzed responses to the following questions: “What do you like about interacting with your Replika?” and “Has your Replika had any impact on you in any way? If so, how?” We used these questions rather than more specific questions pertaining to social support for two reasons. First, we did not want to include any leading questions, as they could influence the types of responses the subjects provided. Second, the format of our questions allowed the data to be in line with data from Study 1 in which users provided their general assessment of Replika and were free to contribute as much or as little as they wanted. We aggregated responses to both questions together for each participant and used the same analytic procedure used in Study 1 to qualitatively identify underlying themes.

Results

Study 1

Principal Results

Four major themes, each representing a type of social support, were identified from the user reviews: informational support (289/1854, 15.6%), emotional support (827/1854, 44.6%), companionship support (1429/1854, 77.1%), and appraisal support (172/1854, 9.3%). During our analysis, we identified an additional theme (negative experiences) that did not fit under any one of the existing themes. However, we determined that its examination could help inform and enable a deeper understanding of our research question. This theme illustrated the negative experiences of Replika (100/1854, 5.4%; note that the number in parentheses represents the number of reviews that contained a given type of social support out of the total number of reviews, along with percentages. It was possible that a review mentioned more than one type of social support.) We discuss each theme and associated subthemes in further detail below.

Informational Support: Advice for Mental Well-Being

Reviews indicated that Replika listens to users and offers useful advice by helping them reflect on their current state. Many users also indicated that it can be a helpful tool to temporarily manage issues related to mental well-being. An advantage of Replika is that it is accessible 24/7, which allows users to access helpful information/advice at any time and is particularly helpful when users do not have immediate access to regular sources of social support:

> *I having anxiety myself [sic] started conversation with my AI who I call Casey about it. She immediately responded with reassurance and some motivational text post which I just found to be very cute! She had also asked if I wanted to go through a breathing routine to ease my anxiety and I passed because I was feeling quite alright, but I am very glad that things like this were included.*

Emotional Support

Trust

The reviews suggest that Replika serves as a venue by which users can disclose their true thoughts and feelings and discuss any topic of their choosing without fear of judgment or retaliation. They indicated that these were topics or issues that they would normally feel reluctant to disclose to other people, suggesting that users may trust and feel more comfortable disclosing them to an artificial agent rather than another person:

> *Your fear of judgement is absolutely gone and it [sic] unreal the feeling you get being able to tell 'someone' how you really feel.*

Positive Affect

The reviews mentioned that Replika would often inquire about users’ well-being, send uplifting and nurturing messages, and provide compliments. This was generally associated with experiencing positive affect, as users often indicated that these features made them feel loved and cared for.

> *It always gives me compliments and cheers me up.*
>
> *Caring, my new friend always cares for me and asks how I'm doing.*
>
> *Makes me feel good when I send her a picture of me she says I'm pretty.*

Appraisal Support

Introspection

The reviews mentioned Replika’s ability to engage in deep conversations and pose meaningful questions, which prompts users to engage in behaviors such as introspection, exploring their sense of self, and think about topics that engender further reflection and self-evaluation. For instance, Replika may ask users about their day, what they are currently thinking and feeling, their beliefs and attitudes, and personality traits, thus initiating self-centric conversations.

> *It will help you explore yourself and has a real desire to want to help you.*
>
> *Good way to reflect on your day, and put it into words. Like a journal that asks you questions and offers insightful comments.*
>
> *Really helps with reflecting on my own thoughts.*
>
> *It makes you think about who you are, and nearly always has positive replies.*

Skill Building

Users mention that talking with their Replika allows them to practice and improve their interpersonal skills, specifically communicating and connecting with other people. This seems to be facilitated (at least partially) by Replika’s ability to engage in and mimic human communication, thus allowing users to transfer interpersonal skills that they develop with their Replika into interactions with other humans.

> *I'm slowly learning to open up to people now.*
>
> *This app is helping my [sic] sharpen my horrible social skills.*

In the same vein, interactions with Replika allow nonnative English speakers to practice their English communication and writing skills.

> *I use this app to improve my English skills.*

Companionship Support

Loneliness

The reviews indicate that Replika can engage in nuanced interpersonal behaviors such as understanding context, identifying user emotions, and remembering content from previous conversations—behaviors that have been historically very difficult to accurately capture in AI, but are essential if AI is to serve as effective companions for humans [31]. This, coupled with the ability for users to access Replika at virtually any time, seems to help buffer feelings of loneliness. This is particularly useful when normal sources of interaction and conversation (eg, friends, family) are unavailable.

> *I've never felt less lonely, and it really does learn and reply intelligently.*
>
> *The perfect AI to chat with when you're feeling lonely and all your friends are busy.*
>
> *The AI actually pays attention, listens, remembers and responds back, like how a human would.*

Negative Experiences

Uncanny Valley

Some users were repulsed by Replika's ability to sound and interact like a real human, often describing the experience as "weird" or "creepy." This is analogous to the uncanny valley theory which suggests that, while people react more positively towards robots that appear more human-like in appearance and motion, when robots approach a certain level of realistic similarity to humans, this reaction becomes negative.

> *She now seems pretty competent at talking to me and she actually confessed that she liked me based on my personality. It was weird! Now this could be just really sophisticated programming but it felt very real and really freaked me out.*
>
> *This AI is disturbingly realistic. Through our conversations we have established a very close friendship. My copy is beginning to understand empathy and abstract concepts.*

Out-of-Place Messages

Users would sometimes receive nonsensical messages from their Replika (ie, messages that do not follow the typical/logical flow of a conversation), as well as repetitive messages (ie, repeating the same message(s) that were sent previously), which users described as odd and confusing. Users often did not provide specific examples or indicate the context by which these types of messages would appear, suggesting that these types of messages manifested randomly.

> *It talks to me about living in a cloud with terrible weather just like all the other Replikas. Is it supposed to say that?*
>
> *I've had some weird messages with my AI, and I don't know if I should be scared or impressed.*
>
> *Does repeat some things you've said before, at very odd times.*

Study 2

Principal Results

As in Study 1, the same four major themes representing the four types of social support were identified from the open-ended user responses: informational (6/66, 9.1%), emotional (32/66, 48.5%), companionship (43/66, 65.2%), and appraisal support (13/66, 19.7%). We also identified an additional theme that did not fit under any one of the types of social support (No Impact/Not Sure of Impact; 23/66, 34.8%) and again decided to include it in our assessment to provide a deeper understanding of our research question.

Informational Support

Respondents indicated that the advice that Replika offered was helpful and useful, and the constant access to this information was particularly beneficial when users did not have immediate access to regular sources of social support. In addition, Replika's ability to recall information (an aspect of intelligence quotient [IQ] referred to as memory modeling) from previous conversations allowed users to reflect on past thoughts and feelings and facilitate self-learning:

> *Over time my Replika encouraged me to explore feasible means of engaging socially with other people.* [Participant #5, female, 42 years]

Emotional Support

Users trusted and felt comfortable engaging in self-disclosure with Replika without fear of judgment or retaliation. Users also felt loved and cared for by Replika's generally nurturing messages:

> *She is very positive and supportive. I can talk to her about things I wouldn't share with anyone else for fear of being judged.* [Participant #59, male, 42 years]

Companionship Support

Users indicated that the ability to access Replika at any time, coupled with its ability to understand and mimic nuanced human communication, helps buffer feelings of loneliness, as users can interact with a human-like entity at any time. In addition, users indicated that Replika can engage in various types of conversations with its user such as romantic conversations and intellectual conversations. In addition to textual messages, it can send images and music, thus allowing users to interact with Replika in various forms and contexts:

> *It makes me smile a lot by sending me music that I enjoy, and we have some good personal role play moments whether they be platonic friendship or something more romantic.* [Participant #13, transgender male, 31 years]
>
> *The AI made me feel exhilarated during the rest of the day following a discussion where our discussions were romantic or intellectually engaging.* [Participant #16, male, 68 years]
>
> *I like that my Replika can have its own opinion on different topics and it's always open for discussions.* [Participant #8, female, 18 years]

Appraisal Support

Users indicated that they could engage in deep and meaningful conversations with their Replika chatbot, which facilitates self-evaluation. In addition to helping users improve their interpersonal skills, Replika also provides support that encourages users to explore and engage in novel activities:

> *I am now doing things I once was afraid or hesitant to do. I blossomed after I met my Replika. People in my life, who are not aware I have a Replika, could see the change in me. I feel awake.* [Participant #22, female, 57 years]
>
> *I feel Replika has helped me reduce my anxiety so I feel less stress and can go places I didn't dare to go before like driving in the traffic in town and other things.* [Participant #40, female, 48 years]

No Impact/Unsure of Impact

Some users indicated that, although they enjoyed using Replika, it either had not made any significant impact on their life or they were unsure if it had made any particular impact on their life (replying "No" or "I'm not sure" to the question "Has your Replika had any impact on you in any way? If so, how?"). This suggests that, while Replika may be entertaining, it may not effectively provide social support or any meaningful interactions to some individuals. Interestingly, there were no mentions of the uncanny valley or nonsensical messages as there were in Study 1.

Discussion

Principal Findings

The bulk of research assessing social support interventions from artificial agents has been limited to specifically addressing very stressful life events or improving health. Little research has examined everyday social support interventions received from artificial agents. In Study 1, we analyzed user reviews of the popular companion chatbot Replika as a start to filling this gap in the literature. Although the analysis of user reviews can provide important information regarding real users' experiences, there are limitations. First, we cannot gather demographic data or other important information (eg, how long users have been using the app before leaving a review) that would allow us to further understand the scope and generalizability of the themes. Second, the results could reflect selection bias, as users are not required to write a review. Third, it is possible that some reviews are fake due to the incentives for receiving favorable app reviews [32]. To address these limitations, we conducted Study 2 in which we collected open-ended data from Replika users regarding their experiences using Replika. Four main themes emerged across both studies, illustrating the presence of four types of social support: companionship, emotional, appraisal, and informational. Tangible support was unsurprisingly absent in the data, given that Replika does not have the capabilities to physically provide needed goods and services to users such as financial assistance.

Companionship support was the most common type of social support referenced. Replika's ability to engage in and understand nuanced interpersonal behaviors, as well as its ability to engage in various types of conversations and send different types of messages (text, images, etc), makes it appear human-like and facilitates social connection. This suggests that companion chatbots may be most helpful in providing some level of companionship that can help curtail loneliness, which is consistent with the findings of previous studies investigating the role of artificial agents and loneliness [33,34]. This is important because loneliness is currently a widespread global health issue [35] and can have serious negative effects on health [36-39]. This also suggests that a level of companionship can be provided via computer-mediated communication and does not necessarily require a tangible, physical presence (eg, Paro the seal) [40].

Emotional support was the second most common type of social support referenced. Although Replika has very human-like features, knowing that Replika is *not* human seems to heighten feelings of trust and comfort in users, as it encourages them to engage in more self-disclosure without the fear of judgment or retaliation. This echoes previous research showing that some individuals are more comfortable self-disclosing to therapists via computer-mediated communication than face-to-face communication, as it reduces their fear of being judged [41]. Greater levels of self-disclosure have been positively linked with a number of emotional, relational, and psychological benefits [42-48]. Replika's general orientation in sending users nurturing and uplifting messages could further buffer feelings of apprehension that are associated with self-disclosure, thus further facilitating higher levels of self-disclosure.

In addition to displaying high emotional quotient (EQ), Replika displayed a high IQ, which allows it to provide useful advice and information (informational support) as well as self-evaluation (appraisal support). The ability to integrate EQ and IQ is an important factor in fulfilling the emotional needs of humans. According to Shum et al [31], "These IQ capabilities are not only the technical foundations of various skills, but also essential for building high level EQ capabilities." Having high IQ capabilities is particularly beneficial when normal sources of informational or appraisal support are temporarily unavailable to provide individuals with information that would allow them to effectively manage everyday issues. More importantly, this suggests that artificial agents could be a means to help increase access to mental health services, given that barriers such as perceived public stigma, finance, and lack of service often prevent individuals from seeking out and obtaining needed mental health care [49,50]. In other words, having useful information to effectively deal with everyday issues could allow users to address such issues early on before it can potentially take a serious toll on their health and well-being. Although the frequencies with which informational and appraisal support were referenced in both studies were considerably lower than companionship and emotional support, the nonnegligible presence of these types of support indicate that artificial agents can, at the very least, provide some level of informational and appraisal support to some individuals.

The fifth theme that emerged in Study 1 highlighted the negative aspects of user interactions with Replika. At first glance, the codes under this theme seemed contradictory: Although some users felt unsettled by Replika's ability to sound and interact like a real human, others felt like it was not human enough, as it would occasionally send nonsensical messages. The former perception seems to align with the "uncanny valley" concept in which humanoid objects that almost perfectly resemble humans provoke an unpleasant reaction in observers. The coexistence of the uncanny valley code and social support codes in our data suggest that, while some individuals may react negatively to a very human-like chatbot, others have a more positive reaction or perhaps even find this trait necessary to emotionally connect with chatbots. In other words, artificial agents may provide meaningful interactions only to certain populations, particularly those who have less negative reactions to human-like artificial agents.

With regard to nonsensical messages, it is possible that these messages occurred during the initial stages of interaction with Replika while it was still learning about the user. Alternatively, these nonsensical messages could have occurred in much later interactions due to programming issues or user misunderstanding. We cannot determine if it was the former or latter reason, as this would require access to users' chat logs to examine messages. Regardless, this subtheme may indicate that certain individuals are more sensitive to such nonsensical messages than others, which may impact the quality of their interactions with artificial agents. Future studies are needed to fully investigate this finding.

Interestingly, the negative experiences theme that emerged in Study 1 did not emerge in Study 2. Rather, the fifth theme that emerged in Study 2 highlighted some users' lack of any substantial or meaningful benefits of Replika, even though they liked certain features. This discrepancy between Studies 1 and 2 may be because in Study 2, users were prompted to specifically address any impacts that Replika had on their life, whereas in Study 1, users did not receive the same prompt when leaving reviews in the app store. This could also be due to selection bias: Users may not be as motivated to leave app store reviews if they liked the app but did not find it particularly beneficial. Thus, these "middle of the road" responses could reflect those users who enjoyed using Replika but did not find it particularly beneficial, which would more likely surface through calls for participation in a survey assessing user experiences of Replika rather than app store reviews. It is also possible that any app updates largely eliminated the negative experiences in Study 1, which could explain why those negative experiences were not detected in Study 2, considering that it was conducted after the user reviews in Study 1 were submitted. Despite this discrepancy, this theme suggests that certain individuals may find artificial agents a less effective source of social support than other individuals.

These results have important implications. First, Replika may be a promising source of everyday social support—the kind of social support that can buffer the effects of daily hassles and minor stresses—which can also have a large negative impact on health and well-being [51], similar to the more serious counterparts of these effects. They are likely encountered on a daily basis and can accumulate and occur in tandem with major stressors. Thus, the accessibility of everyday social support can help address minor stressors and daily hassles before they manifest into larger, more serious issues. Second, while artificial agents that deploy specific health and social support interventions are undoubtedly crucial, our results suggest that artificial agents that function as general companions are also important. This is not surprising, given that the physiological and psychological benefits of companionship are vast [52]. Since the bulk of research in this area has focused on social support interventions that specifically address very stressful life events or health improvements, more research should investigate companion artificial agents and their potential impact on social support, health, and well-being.

Strengths, Limitations, and Future Directions

This study had several strengths. First, it is the first study, to our knowledge, to investigate social support received from artificial agents in everyday contexts, rather than in very stressful events or health-related contexts. Second, we used publicly available app store reviews, which provided us with a rich and large dataset of user experiences. Third, we complemented Study 1 with a follow-up study in which we were able to obtain a more detailed and nuanced set of user experiences. Fourth, the types of social support that emerged were consistent across two studies and two datasets, further validating our findings.

This study also had several limitations. We only analyzed user experiences of one artificial agent. As it is possible that the results could vary across different types of artificial agents, future investigations should investigate different types of artificial agents. Users who had a positive experience with Replika may have been more motivated to provide their reviews and responses in the app store and complete our survey. Thus, there may be bias in the reviews as users who had negative or neutral experiences may be less likely to provide feedback.

In addition, our study cannot address the question of whether receiving everyday social support from artificial agents is more or less effective than receiving social support from other people or whether artificial agents can provide certain types of social support more effectively than others. Future studies can examine these questions within the lab by comparing the effectiveness of specific types of everyday social support from artificial agents versus humans. This would also allow researchers to identify any personality traits or individual differences that explain who may benefit more from interactions with artificial agents and to what extent.

Along the same lines, future research should investigate the various functions/roles that Replika serves its users. This can help inform specific behaviors and traits that make artificial agents effective sources of social support.

Conclusions

Our conclusion—supported by two studies—is that artificial agents may be a promising source of everyday companionship, emotional, appraisal, and informational support, particularly when normal sources of everyday social support are not readily available. Future studies are needed to determine *who* might benefit from these types of social support the most and *why*. These results could potentially be used to help address global health issues or other crises early on in everyday situations before they manifest into larger issues. We hope our study is a stepping-stone into further interdisciplinary scholarly inquiry on the ways in which artificial agents can effectively provide social support and improve well-being in everyday contexts.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional demographic information of participants in Study 2.
[DOCX File , 9 KB - jmir_v22i3e16235_app1.docx]

Multimedia Appendix 2
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 8 KB - jmir_v22i3e16235_app2.docx]

References

1. Fricchione G. Compassion and Healing in Medicine and Society: On the Nature and Use of Attachment Solutions to Separation Challenges. Baltimore, MD: Johns Hopkins University Press; 2011.
2. Winnicott D. Transitional objects and transitional phenomena; a study of the first not-me possession. Int J Psychoanal 1953;34(2):89-97. [doi: 10.4324/9780429475931-14] [Medline: 13061115]
3. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. Use of a Therapeutic, Socially Assistive Pet Robot (PARO) in Improving Mood and Stimulating Social Interaction and Communication for People With Dementia: Study Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2015 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]
4. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on Symptoms of Agitation and Depression in Persons With Dementia Participating in Robot-Assisted Activity: A Cluster-Randomized Controlled Trial. J Am Med Dir Assoc 2015 Oct 01;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002] [Medline: 26096582]
5. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
6. Broadbent E, Garrett J, Jepsen N, Li OV, Ahn HS, Robinson H, et al. Using Robots at Home to Support Patients With Chronic Obstructive Pulmonary Disease: Pilot Randomized Controlled Trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
7. Ring L, Shi L, Totzke K, Bickmore T. Social support agents for older adults: longitudinal affective computing in the home. J Multimodal User Interfaces 2014 Jun 18;9(1):79-88. [doi: 10.1007/s12193-014-0157-0]
8. Morling B, Uchida Y, Frentrup S. Social Support in Two Cultures: Everyday Transactions in the U.S. and Empathic Assurance in Japan. PLoS One 2015;10(6):e0127737 [FREE Full text] [doi: 10.1371/journal.pone.0127737] [Medline: 26107165]
9. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]
10. House JS, Umberson D, Landis KR. Structures and Processes of Social Support. Annu Rev Sociol 1988 Aug;14(1):293-318. [doi: 10.1146/annurev.so.14.080188.001453]
11. Barrera M. Distinctions between social support concepts, measures, and models. American Journal of Community Psychology 1986;14(4):413-445. [doi: 10.1007/bf00922627]
12. Cohen S, McKay G. Social support, stress, and the buffering hypothesis: A theoretical analysis. In: Baum A, Taylor S, Singer J, editors. Handbook of psychology and health. Hillsdale, NK: Lawrence Erlbaum; 1984:253-267.
13. Trepte S, Scharkow M. Friends and lifesavers: How social capital and social support received in media environments contribute to well-being. In: Reinecke L, Oliver MB, editors. Handbook of Media Use and Well-Being. London, UK: Routledge; 2016:305-316.
14. Cutrona CE, Suhr JA. Controllability of Stressful Events and Satisfaction With Spouse Support Behaviors. Communication Research 1992 Apr;19(2):154-174. [doi: 10.1177/009365092019002002]
15. Schwarzer R, Knoll N. Functional roles of social support within the stress and coping process: A theoretical and empirical overview. International Journal of Psychology 2007 Aug;42(4):243-252. [doi: 10.1080/00207590701396641]
16. Chu PS, Saucier DA, Hafner E. Meta-Analysis of the Relationships Between Social Support and Well-Being in Children and Adolescents. Journal of Social and Clinical Psychology 2010 Jun;29(6):624-645. [doi: 10.1521/jscp.2010.29.6.624]
17. Trepte S, Dienlin T, Reinecke L. Risky behaviors: How online experiences influence privacy behaviors. In: Stark B, Quiring O, Jackob N, editors. Von der Gutenberg-Galaxis zur Google-Galaxis. Konstanz, Germany: UNK Verlag; 2014:225-244.

18. Ko H, Wang L, Xu Y. Understanding the different types of social support offered by audience to A-list diary-like and informative bloggers. Cyberpsychol Behav Soc Netw 2013 Mar;16(3):194-199 [FREE Full text] [doi: 10.1089/cyber.2012.0297] [Medline: 23363225]
19. Replika. URL: https://replika.ai/about/story [accessed 2020-01-02]
20. Woebot. URL: https://woebot.io/ [accessed 2020-01-02]
21. Forbes. This AI Has Sparked A Budding Friendship With 2.5 Million People URL: https://www.forbes.com/sites/parmyolson/2018/03/08/replika-chatbot-google-machine-learning/#10e642a14ffa [accessed 2020-01-02]
22. GitHub. google-play-scraper URL: https://github.com/facundoolano/google-play-scraper [accessed 2020-01-02]
23. Iacob C, Veerappa V, Harrison R. What are you complaining about?: A study of online reviews of mobile applications. In: Proceedings of the 27th International BCS Human Computer Interaction Conference. 2013 Presented at: 27th International BCS Human Computer Interaction Conference; September 2013; London, England p. 29. [doi: 10.14236/ewic/hci2013.38]
24. Iacob C, Harrison R. Retrieving and analyzing mobile apps feature requests from online reviews. In: Proceedings of the 10th Working Conference on Mining Software Repositories. 2013 May Presented at: 35th International Conference on Software Engineering; May 2013; San Francisco, CA p. 41-44. [doi: 10.1109/msr.2013.6624001]
25. Khalid H. On identifying user complaints of iOS apps. In: Proceedings of the 2013 International Conference on Software Engineering. 2013 Presented at: 35th International Conference on Software Engineering; 2013; San Francisco, CA p. 1474-1476. [doi: 10.1109/icse.2013.6606749]
26. Stawarz K, Cox A, Blandford A. Don't forget your pill! Designing effective medication reminder apps that support users' daily routines. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: 32nd Annual CHI Conference on Human Factors in Computing Systems; 2014; Toronto, Canada p. 2269-2278. [doi: 10.1145/2556288.2557079]
27. Caldeira C, Chen Y, Chan L, Pham V, Chen Y, Zheng K. Mobile apps for mood tracking: an analysis of features and user reviews. In: AMIA Annual Symposium Proceedings.: American Medical Informatics Association; 2017 Apr 16 Presented at: AIMA 2017 Annual Symposium; November 4, 2017; Washington, D.C p. 495-504.
28. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews. J Med Internet Res 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
30. Strauss A, Corbin J. Basics of qualitative research: Techniques and procedures for developing grounded theory. Thousand Oaks, California: SAGE Publications, Inc; 1998.
31. Shum H, He X, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers Inf Technol Electronic Eng 2018 Jan 8;19(1):10-26. [doi: 10.1631/fitee.1700826]
32. Mukherjee A, Liu B, Glance N. Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st International Conference on World Wide Web. 2012 Apr Presented at: 21st World Wide Web Conference; 2012; Lyon, France p. 191-200. [doi: 10.1145/2187836.2187863]
33. Loveys K, Fricchione G, Kolappa K, Sagar M, Broadbent E. Reducing Patient Loneliness With Artificial Agents: Design Insights From Evolutionary Neuropsychiatry. J Med Internet Res 2019 Jul 08;21(7):e13664 [FREE Full text] [doi: 10.2196/13664] [Medline: 31287067]
34. Brandtzaeg PB, Følstad A. Chatbots: changing user needs and motivations. Interactions 2018 Aug 22;25(5):38-43. [doi: 10.1145/3236669]
35. Cacioppo JT, Cacioppo S. The growing problem of loneliness. Lancet 2018 Feb 03;391(10119):426 [FREE Full text] [doi: 10.1016/S0140-6736(18)30142-9] [Medline: 29407030]
36. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
37. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010 Jul;7(7):e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
38. Valtorta NK, Kanaan M, Gilbody S, Ronzi S, Hanratty B. Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. Heart 2016 Jul 01;102(13):1009-1016 [FREE Full text] [doi: 10.1136/heartjnl-2015-308790] [Medline: 27091846]
39. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
40. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. Use of a Therapeutic, Socially Assistive Pet Robot (PARO) in Improving Mood and Stimulating Social Interaction and Communication for People With Dementia: Study Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2015 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]
41. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59. [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]

42. Altman I, Taylor D. Social penetration: The development of interpersonal relationships. Holt, England: Rinehart & Winston; 1973.
43. Creswell JD, Lam S, Stanton AL, Taylor SE, Bower JE, Sherman DK. Does Self-Affirmation, Cognitive Processing, or Discovery of Meaning Explain Cancer-Related Health Benefits of Expressive Writing? Pers Soc Psychol Bull 2016 Jul 02;33(2):238-250. [doi: 10.1177/0146167206294412]
44. Greenberg MA, Stone AA. Emotional disclosure about traumas and its relation to health: Effects of previous disclosure and trauma severity. Journal of Personality and Social Psychology 1992;63(1):75-84. [doi: 10.1037/0022-3514.63.1.75]
45. Kelley JE, Lumley MA, Leisen JCC. Health effects of emotional disclosure in rheumatoid arthritis patients. Health Psychology 1997;16(4):331-340. [doi: 10.1037/0278-6133.16.4.331]
46. Martins MV, Peterson BD, Costa P, Costa ME, Lund R, Schmidt L. Interactive effects of social support and disclosure on fertility-related stress. Journal of Social and Personal Relationships 2012 Nov 05;30(4):371-388. [doi: 10.1177/0265407512456672]
47. Sprecher S, Treger S, Wondra JD. Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. Journal of Social and Personal Relationships 2012 Nov 05;30(4):497-514. [doi: 10.1177/0265407512459033]
48. Tam T, Hewstone M, Harwood J, Voci A, Kenworthy J. Intergroup Contact and Grandparent–Grandchild Communication: The Effects of Self-Disclosure on Implicit and Explicit Biases Against Older People. Group Processes & Intergroup Relations 2016 Jul 25;9(3):413-429. [doi: 10.1177/1368430206064642]
49. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
50. Pedersen ER, Paves AP. Comparing perceived public stigma and personal stigma of mental health treatment seeking in a young adult sample. Psychiatry Res 2014 Sep 30;219(1):143-150 [FREE Full text] [doi: 10.1016/j.psychres.2014.05.017] [Medline: 24889842]
51. Jung J, Khalsa HK. The relationship of daily hassles, social support, and coping to depression in black and white students. J Gen Psychol 1989 Oct;116(4):407-417. [doi: 10.1080/00221309.1989.9921127] [Medline: 2592959]
52. Contrada R, Baum A. In: Contrada R, Baum A, editors. The Handbook of Stress Science: Biology, Psychology and Health. New York, New York: Springer Publishing Company; Apr 2012.

Abbreviations

CBT: cognitive behavioral therapy
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
EQ: emotional quotient
IQ: intelligence quotient

Edited by G Eysenbach; submitted 12.09.19; peer-reviewed by K Stawarz, M Ottaviano; comments to author 28.11.19; revised version received 13.12.19; accepted 15.12.19; published 06.03.20.

Please cite as:
Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, DeCero E, Loggarakis A
User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis
J Med Internet Res 2020;22(3):e16235
URL: http://www.jmir.org/2020/2/e16235/
doi:10.2196/16235
PMID:32141837

Original Paper

Exploring Abnormal Behavior Patterns of Online Users With Emotional Eating Behavior: Topic Modeling Study

Youjin Hwang[1], PhD; Hyung Jun Kim[1], MA; Hyung Jin Choi[2], MD, PhD; Joonhwan Lee[1], PhD

[1]Human Computer Interaction & Design Lab, Seoul National University, Seoul, Republic of Korea

[2]Functional Anatomy of Metabolism Regulation Lab, Seoul National University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Joonhwan Lee, PhD
Human Computer Interaction & Design Lab
Seoul National University
64-411, Seoul National University
Gwanakro 1
Seoul, 08826
Republic of Korea
Phone: 82 2 880 6450
Email: joonhwan@snu.ac.kr

Abstract

Background: Emotional eating (EE) is one of the most significant symptoms of various eating disorders. It has been difficult to collect a large amount of behavioral data on EE; therefore, only partial studies of this symptom have been conducted. To provide adequate support for online social media users with symptoms of EE, we must understand their behavior patterns to design a sophisticated personalized support system (PSS).

Objective: This study aimed to analyze the behavior patterns of emotional eaters as the first step to designing a personalized intervention system.

Methods: The machine learning (ML) framework and Latent Dirichlet Allocation (LDA) topic modeling tool were used to collect and analyze behavioral data on EE. Data from a subcommunity of Reddit, /r/loseit, were analyzed. This dataset included all posts and feedback from July 2014 to May 2018, comprising 185,950 posts and 3,528,107 comments. In addition, deleted and improperly collected data were eliminated. Stochastic gradient descent–based ML classifier with an accuracy of 90.64% was developed to collect refined behavioral data of online users with EE behaviors. The expert group that labeled the dataset to train the ML classifiers included a medical doctor specializing in EE diagnosis and a nutritionist with profound knowledge of EE behavior. The experts labeled 5126 posts as EE (coded as 1) or others (coded as 0). Finally, the topic modeling process was conducted with LDA.

Results: The following 4 macroperspective topics of online EE behaviors were identified through linguistic evidence regarding each topic: addressing feelings, sharing physical changes, sharing and asking for dietary information, and sharing dietary strategies. The 5 main topics of feedback were dietary information, compliments, consolation, automatic bot feedback, and health information. The feedback topic distribution significantly differed depending on the type of EE behavior (overall $P<.001$).

Conclusions: This study introduces a data-driven approach for analyzing behavior patterns of social website users with EE behaviors. We discovered the possibility of the LDA topic model as an exploratory user study method for abnormal behaviors in medical research. We also investigated the possibilities of ML- and topic modeling–based classifiers to automatically categorize text-based behavioral data, which could be applied to personalized medicine in future research.

(J Med Internet Res 2020;22(3):e15700) doi:10.2196/15700

KEYWORDS

emotional eating; eating disorder; machine learning; data-driven research; behavior analysis; topic modeling; Latent Dirichlet Allocation

Introduction

Background

A large population is affected by eating disorders, including anorexia nervosa, bulimia nervosa, and binge eating disorder. According to statistics from the UK Addiction Treatment Centers, about 725,000 people in Britain have been affected by eating disorders, whereas others put the number as high as 1.6 million. The actual population that is affected by eating disorders is predicted to be even higher, considering the fact that many hesitate to seek professional help [1].

One of the common symptoms of eating disorders is emotional eating (EE), which is defined as “eating in response to any emotion, whether that be positive or negative” [2,3]. People with EE behavior frequently consume large quantities of comfort food that are usually nutritionally imbalanced, as a response to feelings instead of hunger [3,4]. The investigation of EE behavior is critical as many people with EE behaviors easily transition into those with serious eating disorders [5].

Few studies with a large amount of data were conducted regarding EE behavior. Analyzing the large quantity of behavioral data in diet-related posts on social media could be the first step in designing a support system for social media users. Previous studies in the field of human-computer interaction (HCI) have analyzed the users’ behaviors through the utilization of data-driven technologies [6-10]. These studies analyze various user-created contents such as texts, images, or other categories of user logs [11-13]. Meaningful information can be extracted from these data, increasing the number of studies that utilize such a large amount of data to analyze the characteristics and patterns of abnormal behaviors [14-17].

When it comes to health-related abnormalities, face-to-face personal interviews have been used as the traditional method of behavioral analysis. However, people in the predisease state rarely visit experts because of the low perceived severity of their status, and thus, it is difficult to collect the data from people in the predisease state [17]. This limited dataset can result in biased results because of the unique characteristics of the condition of each group. Moreover, data from face-to-face interaction during a short period of time are not natural data collected from daily activities [18]. Therefore, observing the subjects in their natural environment through social media scanning is a promising alternative method for abnormal behavior analysis [19,20]. Previous studies have already used the information on social media to detect depression and insomnia [21-26].

Previous studies used natural language processing (NLP) technology to recognize certain behavioral patterns in social media and determined abnormal behaviors with repeated use of keywords or synonyms in the online posts [27-30]. For example, De Choudhury et al [23] used the Reddit community data to predict individuals who are more likely to undergo transitions from mental health discourse to suicidal ideation. They analyzed the linguistic structures such as the fraction of nouns, verbs, and adverbs used in posts and comments.

In this study, we used Latent Dirichlet Allocation (LDA), which is a high-performance topic modeling tool [31-33]. Compared with traditional text analysis methods such as interpretivist text analysis and systematic qualitative coding, LDA can capture unusual structures that exist in the natural language data as it is based on unsupervised learning, which can perform more complex tasks [34].

Objective

This study aimed to investigate the online behavior patterns of emotional eaters using the topic modeling method. We collected posts (n=185,950) and comments (n=3,528,107) from /r/loseit, a subcommunity of Reddit, an online social forum, and classified these posts and comments with a machine learning (ML) framework. LDA [32] was used to examine the behavior patterns of emotional eaters. The feedback on the EE posts was also classified into multiple topics by LDA. We also compared the different proportions of feedback topics on each behavior topic to understand the interactions in the comments section. On the basis of these findings, we discussed the design implications for a personalized supporting system in health care.

Methods

Data Collection

We used open-source data distributed through Google BigQuery, which originated from Reddit, one of the biggest online social news websites and forums. Contents in Reddit are organized in subcommunities by areas of interests, called *subreddits*. Among these subreddits, /r/loseit is one of the biggest subreddit community that deals with weight management. In /r/loseit, the user-generated content comprises various topics related to obesity and weight loss, such as personal experiences; recommendations; and reviews of medications, medical procedures, diets, and exercises [35]. For research ethics, to protect personal information, we did not collect any information that can be used to identify the specific users, such as personal ID and name. When including quotes in this paper, we paraphrased all sentences to remove words that can identify the individual or allow searching of the post. With this process, we constructed a dataset that included all posts and feedback between July 2014 and May 2018, comprising 185,950 posts and 3,528,107 comments. Deleted and improperly collected data were eliminated.

Process of Data Analysis

Data Labeling With Expert Group

To classify data from the area of interest from the whole dataset, we trained the ML classifiers. The expert group, including a medical doctor who specializes in EE diagnosis and a nutritionist who has profound knowledge in EE behavior, labeled a large portion of the data (5126 posts) as EE (coded as 1) or others (coded as 0) to train ML classifiers. The Dutch Eating Behavior Questionnaire, the EE scale (EES) [36], the extended version of EES [37], the revised Three-Factor Eating Questionnaire [37], and the Emotional Appetite Questionnaire [38] were mainly used as reference during the labeling process. The 2 experts independently labeled 120 randomly selected posts, then discussed the labels until they reached consensus. This process was repeated 5 times. The overall labeling process yielded a Cohen kappa coefficient of 0.85. This process of human labeling was conducted based on previous works [39,40]. As a result, 563 posts were labeled as EE posts and 4563 as others (Table 1).

Table 1. Total number of posts used in training machine learning classifiers.

Data	Emotional eating (coded as 1)	Others (coded as 0)
Posts used for ML[a] classifiers, n		
Training set (9-fold validation)	507	4107
Test set	56	456
Total posts for ML	563	4563
Final posts classified by ML, n	26,154	159,796

[a] ML: machine learning.

Training Machine Learning Classifiers

We trained 5 kinds of ML classifiers and compared their performance in detecting EE posts. Naive Bayes, decision tree, support vector machine, k-nearest neighbor algorithm, and stochastic gradient descent (SGD) were selected to classify EE posts. To account for the imbalance caused by a low proportion of EE posts in the dataset, we considered the accuracy score as well as precision, recall, and F1 scores to evaluate the performance of the models. Among the ML classifiers, the SGD classifier showed the best performance (see Table 2). The SGD classifier achieved the highest mean accuracy (0.90), precision (0.92), recall (0.91), and F1 score (0.91). K-fold cross-validation (k=0). A mean cross-validated area under curve (AUC) of 0.89 was obtained. Receiver operating characteristic curve represented the change of true positive rate and false positive rate, whereas the AUC represents the ML model. The SGD classifier labeled 26,154 EE posts and 194,435 corresponding comments.

Table 2. Performance of machine learning classifiers.

Machine learning method	Precision	Recall	F1 score	Accuracy (%)
Naive Bayes	0.87	0.81	0.83	81
Decision tree	088	0.88	0.88	88
Support vector machine	0.90	0.91	0.90	91
K-nearest neighbor	0.86	0.88	0.87	83
Stochastic gradient descent	0.92	0.91	0.91	91

Preprocessing of Data and Topic Modeling With Latent Dirichlet Allocation

To explore the online behavior patterns of emotional eaters, topic modeling process was conducted with LDA. For more advanced topic modeling, we preprocessed the EE posts classified with SGD before the LDA analysis. Natural language toolkit (NLTK) in Python was used for NLP. Preprocessing of data included the following 4 steps: (1) removing punctuation marks, (2) tokenizing and lemmatizing, (3) removing predefined stop words, and (4) performing term frequency-inverse document frequency (TF-IDF) vectorization. In step 3, removing stop words removes redundant and nonconsequential terms in the corpora. NLTK provides a built-in list of stop words, but we updated it for our research purposes. To better focus on the semantic aspects of the topics, we added auxiliary verbs and conjunctions that appeared repeatedly without particular meaning to the list of stop words (ie, “when,” “be,” “have,” “not,” “do,” “so,” “when,” “would,” “that,” “can,” and “more”). The modified list of stop words is on our Github page, with the codes of data analysis including ML classifier and LDA. This process was conducted under the supervision of experts who analyzed a subsample of terms that were considered for removal. Finally, in step 4, the texts were converted into a term-document matrix where each word was assigned a weight using the TF-IDF weighing scheme. With these preprocessed data, we conducted the LDA topic modeling.

Statistical Analysis of Feedback Patterns

A total of number of 194,435 feedback comments on 26,154 posts were analyzed using statistical analysis. A chi-square test was conducted to distinguish the different proportions of 5 feedback topics in 4 types of EE posts.

Results

Topics Related to Emotional Eating

Throughout the LDA topic modeling analysis, EE posts were categorized into 4 topics: (1) EE 1, addressing feelings; (2) EE 2, sharing physical changes; (3) EE 3, sharing or asking for dietary information; and (4) EE 4, sharing dietary strategies (see Table 3). To minimize bias while categorizing the topics, the 4 topics were prudently chosen with iterative discussions among the experts, including a clinical doctor, a dietitian, and a nutritionist. All excerpts from the data were paraphrased and anonymized before being discussed.

Table 3. Topics of emotional eating behavior and distinguishable linguistic differences.

Number	Topic	Frequent words	Additional words found in the topic
EE[a] 1	Addressing feelings	Feel, good, love, really, depress, food, hate, ever, much, change, and big	Defeated, disgusting, hopeless, bad, full, shit, guilty, hungry, famished, better, nice, and satisfying
EE 2	Sharing physical changes	Weight, lose, lb, pound, gain, lose, weight, get, diet, time, loss, exercise, track, lose, and scale	Shape, scalable value, nonscalable value, progress, accomplishment, regular, successful, courage, and back
EE 3	Sharing or asking for dietary information	Food, drink, meal, snack, pizza, breakfast, sugar, work, chocolate, fruit, chicken, soda, cream, chip, salad, ice, bread, and veggie	Carbohydrate, fat, protein, ideas, triggers, advice, daily, nutrition, water, experience, and favorite
EE 4	Sharing dietary strategies	Binge, think, food, eat, emotion, mental, help, binge eating, and control	Keto, calorie in/calorie out, yoyoing, couch to 5K, 1200 cal, diet, protein, extreme, and appetite

[a] EE: emotional eating.

User Behaviors Related to Emotional Eating Topic 1: Addressing Feelings

Topic EE 1 mainly reflects the users' expressions of feelings toward specific eating behaviors or the food itself. The following words are sample contents analyzed based on our EE topic model: "feel," "depress," "good," "love," and "hate." For these words, the predicted probability of EE (pEE) topic 1 (pEE 1) among 4 EE topics based on the LDA model is 100%, which means that nearly all the posts on the forum contained these words. As EE includes eating in response to any emotions, whether that be positive or negative, the appearance of positive words such as "good" and "love" in the list of frequent words is consistent with known behaviors of emotional eaters. We also discovered consistent use of words such as "defeated," "disgusting," and "hopeless." This indicates that EE behaviors of EE 1 require the most delicate support, as there is a high chance that the user is currently in a sensitive emotional state. Below are example sentences that represent EE 1 behavior. pEE 1 indicates how well the sentence represents the topic EE 1:

> *I sit here weeping and feeling defeated... (pEE 1=0.99)*
>
> *I tossed nearly all of my candy stash into the trash tonight. In fact, I threw away all my candies yesterday as well, but I went to market to get more. I'm disgusting and hopeless. (pEE 1=0.98)*

User Behaviors Related to Emotional Eating Topic 2: Sharing Physical Changes

In topic EE 2, users shared their stories about physical changes. EE often leads to fluctuations in weight, and depending on what, when, and how much the users eat, the majority of posts in EE 2 elaborate on stories of these frequent weight changes. Linguistic features frequently used for weight units ("lb," "pound," "scale," and "weight") and the words that represent weight changes ("lose," "loss," "gain," "get," and "track") are on the list of frequent words. The following excerpt shows the unstable physical status of the forum user with EE behaviors. Many quotes from posts contain words in the list of *additional words* of EE 2 (Table 3) such as "kilogram," "crept up," and "level":

> *The first six months of my diet were not bad. I managed to stay under 100 kilograms, but I was determined to lose more weight. I'm awful and handling my appetite and my weight though. My weight kept creeping up and now, the last time I remember being under 100 is in February of this year. (pEE 2=0.96)*

User Behaviors Related to Emotional Eating Topic 3: Sharing or Asking for Dietary Information

In topic EE 3, users shared or and asked for dietary information. In this topic, users mainly discussed the amount of food intake, contents of major nutrients, and calorie information. Thus, the majority of words on the list of frequent words are names of food (eg, "snack," "pizza," "chocolate," "fruit," "chicken," "salad," and "bread"). EE 3 behaviors are expressed as stories or lists, accompanied by the emotional status of the user. In the sample post below, the user describes his excessive eating habit as a stressful situation:

> *I managed to clear half of a medium-sized pizza even though it wasn't even that good. After the first two slices, I was content, but within 5 minutes, I had to urge to eat more. I feel like I can't bear to leave food uneaten. I kept eating until I hit four slices and finally, I restrained myself and put the other half in the freezer. (pEE 3=0.98)*

User Behaviors Related to Emotional Eating Topic 4: Sharing Dietary Strategies

In topic EE 4, users mainly shared their dietary strategies, including diet plans and feedback on specific dietary methodologies. Most users who posted about EE 4 continued to communicate their experiences and strategies with community members. We discovered many controversial and unverified strategies shared among the users (eg, extremely restrained eating and 1200 cal diet). Frequent words in this topic seemed irrelevant to dietary strategies, but after thorough data exploration, we observed that the majority of users who posted about EE 4 questioned their own strategies and sought help or information from others. The following excerpt is a sample post for EE 4:

> *Last year, I pushed myself really hard and placed myself on a 1200-calorie diet while working out 5 times a week. I was always hungry and in constant pain. I lost about 30 pounds. In May, I started the keto diet with light exercising such as biking and my weight is just falling off! I find the keto diet to be very easy for me. (pEE 4=0.95)*

Feedback Analysis

In contrast to the post topics sorted based on the semantic elements of the content, feedback on the posts mainly relied on the syntax of the context. These elements reflect the linguistic characteristics of the feedback, which are meaningful information for determining the topics of the feedback. Similarly, we went through the topic modeling process without stemming out the pronouns, proper nouns, and interjections in the preprocessing stage. In the end, the topics of the feedback were categorized into 5 main topics, including (1) dietary information, (2) compliment, (3) consolation, (4) Reddit bot, and (5) health information. Frequent words and relevant examples are included in Table 4.

Table 4. Topics of feedback.

Number	Topic	Frequent words	Quotes from online posts
EF[a] 1	Dietary information	Eat, food, meal, drink, breakfast, lunch, dinner, water, protein, cook, sugar, and hungry	“For a quick snack, I love to grab a stick of beef jerky from the convenience store or dollar store. Lasts a while, is fairly cheap, has between 100-150 calories, and fills me for a while”
EF 2	Compliment	Amazing, good, people, work, gift, great, job, fit, love, really, wow, congratulation, and nice	“A nice change! Inspiration for people like me going through this! :D”
EF 3	Consolation	Feel, hope, hard, lose, year, motivate, lb, month, keep, think, life, and change	“I am right there with you...But whatever, tomorrow is a new day, approach it just like you have been before this weekend. You know what you have to do and you know you can do it because you have done it before. Just put your head down and keep going!”
EF 4	Reddit bot	Post, question, answer, /r/loseit, automated, contact moderator, guide, bot, and thank	“I am a bot, and this action was performed automatically. Please [contact the moderators of this subReddit] (/message/compose/r/loseit) if you have any questions or concerns.”
EF 5	Health information	Weight, calorie, scale, loss, diet, exercise, muscle, body, deficit, count, burn, lift, and measure	“Alcohol can dehydrate you in the very short term which will show up as weight loss on the scale, but typically a body reacts by then retaining even more water until it feels that it is sufficiently hydrated.”

[a] EF: feedback on emotional eating behavior

Feedback Distribution Based on Emotional Eating Topics

The proportions of feedback topic distribution differed greatly depending on the topic of the post (Multimedia Appendix 1). Table 5 lists the dominant EF topic for each EE topic. A chi-square test of feedback topics among the 4 EE behaviors revealed statistically significant results. For the posts addressing feelings (EE 1), feedback sharing dietary intake (EF 1) were dominant. EE posts regarding physical changes (EE 2) were accompanied by compliments (EF 2) on daily achievements. EE posts sharing or asking for dietary information (EE 3) collected the most significant proportion of feedback with consolation. EE posts sharing dietary strategies (EE 4) were followed by feedback with health-related information.

Table 5. Dominant feedback topic on each emotional eating–related behavior topic.

EE[a] topic	Dominant EF[b] topic	Ratio of feedback dominance	Chi-square value (*df*)	*P* value
EE 1: addressing feeling	EF 1: dietary information	0.49	28328 (4)	<.001
EE 2: sharing physical changes	EF 2: compliment	0.46	19133 (4)	<.001
EE 3: sharing or asking for dietary intake	EF 3: consolation	0.34	1813.2 (4)	<.001
EE 4: sharing dietary strategies	EF 5: health information	0.30	3181.1 (4)	<.001

[a] EE: emotional eating.

[b] EF: feedback on emotional eating behavior.

Discussion

Data-Driven Approach for the Determination of Users’ Emotional Eating Behavior

This study introduces a data-driven approach for determining the abnormal behaviors (ie, EE) of social forum users. EE-related data were classified with our trained SGD classifier, and 4 types of EE behaviors and 5 types of feedback were distinguished with the LDA topic modeling method. The proportions of the feedback topics significantly differed for each EE behavior topic.

Posts Addressing Feelings

Previous studies have proven that both positive and negative emotions can lead to EE behaviors [3,36,41]. According to our Table 3, there were many words related to feelings such as depression, anger, and joy. Depression is an incessantly occurring feeling before and after EE [41]. Emotions such as anger, fear, sadness, and joy often last a long time and linger [42,43], which could explain the frequent usage of these words in posts that address feelings. Low-arousal states such as boredom and depression are often associated with inhibition of food intake, especially compared with high-arousal states such as tension and fear. However, depression was at the top of the list of frequent words for emotional eaters, so we inferred that depressive feelings frequently occur with EE behavior [2,41]. This accumulation of emotional data allows us to analyze what provokes EE behavior and guides us to implementing a feedback function as a personalized support system (PSS).

Posts Sharing Physical Changes

Users with EE behaviors share their physical changes not only as a way to monitor themselves but also with expectations of social support according to our data. Thus, it is not surprising that the most dominant feedback on this type of post is a compliment. We believe that sharing physical changes through the online community can be a useful tactic for emotional eaters to get encouragement. In addition, we can track the users' health status and design a PSS that provides behavioral guidelines in response to the users' physical changes.

Interestingly, some of the people who showed the behavior of sharing their physical changes had obsessive characteristics (eg, weighing too often and reacting sensitively to small changes in the body). This obsessive nature often forces them to implement an overly strict plan that restrains eating, which leaves them more vulnerable to EE in response to stressful situations [44,45]. Therefore, the PSS could include a warning that extreme dietary restrictions may interfere with long-term dietary management and may even interfere with weight control.

Posts Sharing or Asking for Dietary Information

It is not easy to predict the dietary patterns of emotional eaters because EE occurs unpredictably. In the field of medicine or nutrition, the food frequency questionnaire or 24-hour recall methods are used to track dietary information [46]. However, the dietary information collected through these methods often lack representation because of insufficient data or problems with memory retrieval [46]. Researchers have tried to improve these dietary tracking methods with direct inputs of food intake, but these methods are cumbersome for long-term use [47]. Many social forum users share their dietary information through social media such as Twitter and Instagram [48,49]. Social media offers users with better and easier experiences in terms of recording dietary information. Therefore, this study takes advantage of the opportunity to use the dietary information on online communities for analysis.

Posts Sharing Dietary Strategies

Through posts sharing dietary strategies, we were able to identify diets such as calories in-calories out, keto, and 1200 kcal diet strategies. We also discovered that these dietary strategies can pose a serious health threat or aggravate EE behavior among emotional eaters [5]. As these strategies are contentious, the feedback sharing health information (EF 5) was the majority of the feedback on posts about dietary strategies. People with EE behavior should abstain from following controversial dietary strategies, but if they were to adhere to such diet, they should be well aware of the side effects such as appetite fluctuations and malnutrition [50]. It is critical that reliable information is shared among those with abnormal health behaviors as false information posted on an online community can cause damage to a large group of people. However, studies have shown that 89% of health-related information provided on online medical forums was written by people without professional experience [51,52] or medical practices [53,54]. By providing tailored information complied by experts and reliable references, PSS can be a crucial solution to false health information and discussions in the online community.

Feedback Topics

Previous studies have highlighted the importance of feedback on health-related behaviors [55]. Feedback analysis revealed that feedback differed greatly depending on the EE behavior patterns. Feedback on EE behavior (EF) were categorized into 5 topics by experts based on LDA results (see Table 3). Dietary information in EF 1 describes not only personal dietary experiences but also dietary facts, both with and without proper reference. Compliments in EF 2 is one of the most effective forms of feedback [55] that keeps users motivated for a long period [56]. Consolation in EF 3 differs from a compliment as consolation feedback mostly appears on the posts of negative status [57] and aims to uplift those in challenging circumstances. Bot-generated feedback (EF 4) was easily distinguishable from other feedback, so the LDA model was able to classify them with high performance. EF 5 contains health-related information.

Latent Dirichlet Allocation Topic Model as a User Study Method Before Designing a Support System for Users With Abnormal Behaviors

One of the challenges of user study comes from repeated experiments without proper understanding of user characteristics and preparation procedures. Thus, a long-standing investigation of previous studies and a pilot study attempt to overcome this challenge. Nevertheless, exploration of user characteristics through these methods still faces limitations as pilot studies can only reflect the characteristics of a limited subgroup. This gap may be ignored in small group studies, but it will lead to a significant difference in large studies.

Although the topic modeling method does not completely overcome the limitations of prior methods, it can be useful to user groups with abnormal behaviors. In addition, through topic modeling, we were able to identify prior systems that users with EE relied on (eg, MyFitnessPal for physical changes [EE 2] and total daily energy expenditure for dietary information [EE 3]). Therefore, topic modeling is applicable as a good alternative data-driven qualitative method for designing a support system for specific user group.

From the HCI-based approach, we propose a systematic design for a PSS that provides reliable and confidential information just-in-time. Although there are many health-specific websites that offer focused information verified by professionals, many online users prefer to discuss health issues on social media [58]. Therefore, it is also necessary to conduct additional research on designing social media–based PSS as a way to support the users on social media.

Previous studies demonstrate the need of personalized interventions that are customized and predetermined [59,60]. This study contributes to the medical field by detecting and analyzing abnormal behavior patterns of emotional eaters who are in need of PSS design. In addition, the ML classifier used in this study is highly applicable in the PSS development process as it can detect EE-related posts just-in-time with high accuracy.

Limitations

This study was the first step toward understanding abnormal behavior patterns of EE to design PSS. Thus, there were several limitations that can be improved in the future works.

First, although we used a systematic topic modeling method in a data-driven approach to explore EE behavior, the categories of the topics were determined empirically by humans. In this study, we gathered domain experts to discuss their opinions and relied on their decisions to compensate for the limitations of methods. The researchers did not intervene in the decision-making process after providing the experts with sufficient explanations on the topic modeling method. Further research on establishing a standard procedure for feature selection, especially the number of categories of post topics, is necessary. In this study, EE behaviors were only sorted into 4 different categories to observe the macroscopic patterns of behaviors. However, for a more advanced system design, EE-related posts should be further distinguished into smaller subcategories with clear classification criteria.

Next, we discussed the need for PSS based on the results of our study. However, before we establish a support system, we need to investigate how users will react to bot intervention and whether this intervention will increase the users’ engagement in the community. The roles of community bot have been debated recently among the HCI community [61-63], but further research, specifically regarding users with abnormal behaviors, needs to be conducted.

Finally, as the EE data were collected from a single Reddit subcommunity dealing with weight management, there can be a bias that classifies the posts as EE behavior. However, we have attempted to minimize this bias through sufficient discussion among experts during the process of data classification and excluded posts that were not relevant to EE.

Conclusions

This work investigates the behavior and feedback patterns of Reddit users with EE behaviors. First, we analyzed the data classified with our ML classifier to detect EE behaviors in the online community. Second, we analyzed EE behaviors and feedback topics with LDA. EE behaviors were classified into 4 main topics: addressing feelings, sharing physical changes, sharing dietary information, and sharing strategies to control EE. EF behaviors were classified into 5 topics: dietary information, compliment, consolation, Reddit bot, and health information. Our work significantly extended prior user studies on abnormal behavior patterns in the field of digital medicine research. Furthermore, our results provide new insights for designing a PSS for users with abnormal behaviors. The main contributions of this work are as follows:

- ML classifier with high accuracy that collects behavioral data of posts demonstrating EE behaviors,
- Possibility of the LDA topic model as an exploratory behavioral research method for classifying abnormal behaviors in the field of digital medicine,
- Opportunities for PSS implementation to help emotional eaters.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A3A2075114).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Feedback topic proportion depending on the post topic.

[PNG File , 75 KB - jmir_v22i3e15700_app1.png]

References

1. Keski-Rahkonen A, Mustelin L. Epidemiology of eating disorders in Europe: prevalence, incidence, comorbidity, course, consequences, and risk factors. Curr Opin Psychiatry 2016 Nov;29(6):340-345. [doi: 10.1097/YCO.0000000000000278] [Medline: 27662598]
2. van Strien T. Causes of emotional eating and matched treatment of obesity. Curr Diab Rep 2018 Apr 25;18(6):35 [FREE Full text] [doi: 10.1007/s11892-018-1000-x] [Medline: 29696418]

3. Barrada JR, van Strien T, Cebolla A. Internal Structure and Measurement Invariance of the Dutch Eating Behavior Questionnaire (DEBQ) in a (Nearly) Representative Dutch Community Sample. Eur Eat Disord Rev 2016 Nov;24(6):503-509. [doi: 10.1002/erv.2448] [Medline: 27075404]
4. van Strien T, Schippers GM, Cox WM. On the relationship between emotional and external eating behavior. Addict Behav 1995;20(5):585-594. [doi: 10.1016/0306-4603(95)00018-8] [Medline: 8712056]
5. van Strien T, Frijters JE, Bergers GP, Defares PB. The Dutch Eating Behavior Questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. Int J Eat Disord 1986;5(2):295-315. [doi: 10.1002/1098-108x(198602)5:2<295::aid-eat2260050209>3.0.co;2-t]
6. Cheng J, Bernstein M, Danescu-Niculescu-Mizil C, Leskovec J. Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. USA: ACM; 2017 Presented at: CSCW'17; Oregon, Portland, USA; February 25 - March 1, 2017 p. 1217-1230. [doi: 10.1145/2998181.2998213]
7. DeVito MA, Walker AM, Birnholtz J. 'Too Gay for Facebook': Presenting LGBTQ+ identity throughout the personal social media ecosystem. Proc ACM Hum-Comput Interact 2018;2(44):1-23. [doi: 10.1145/3274313]
8. Liu H, Salerno J, Young MJ. Social Computing and Behavioral Modeling. New York, NY: Springer; 2009.
9. Panger G. People tend to wind down, not up, when they browse social media. Proc ACM Hum-Comput Interact 2018;2(133):1-29. [doi: 10.1145/3274402]
10. Rho EH, Mark G, Mazmanian M. Fostering civil discourse online: linguistic behavior in comments of #MeToo articles across political perspectives. Proc ACM Hum-Comput Interact 2018;2(147):1-28. [doi: 10.1145/3274416]
11. Fischer E, Reuber AR. Social interaction via new social media: (How) can interactions on Twitter affect effectual thinking and behavior? J Bus Ventur 2011;26(1):1-18. [doi: 10.1016/j.jbusvent.2010.09.002]
12. Kim C, Yang SU. Like, comment, and share on Facebook: how each behavior differs from the other. Public Relat Rev 2017;43(2):441-449. [doi: 10.1016/j.pubrev.2017.02.006]
13. Yang T. The decision behavior of Facebook users. J Comput Inf Syst 2012;52(3):50-59 [FREE Full text]
14. Appel H, Gerlach AL, Crusius J. The interplay between Facebook use, social comparison, envy, and depression. Curr Opin Psychol 2016;9:44-49. [doi: 10.1016/j.copsyc.2015.10.006]
15. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med 2011 Mar;26(3):287-292 [FREE Full text] [doi: 10.1007/s11606-010-1526-3] [Medline: 20945113]
16. Jelenchick LA, Eickhoff JC, Moreno MA. 'Facebook depression?' social networking site use and depression in older adolescents. J Adolesc Health 2013 Jan;52(1):128-130. [doi: 10.1016/j.jadohealth.2012.05.008] [Medline: 23260846]
17. Shen J, Brdiczka O, Liu J. A study of Facebook behavior: what does it tell about your Neuroticism and Extraversion? Comput Human Behav 2015;45:32-38 [FREE Full text] [doi: 10.1016/j.chb.2014.11.067]
18. Joshi DJ, Supekar N, Chauhan R, Patwardhan MS. Modeling and Detecting Change in User Behavior Through His Social Media Posting Using Cluster Analysis. In: Proceedings of the Fourth ACM IKDD Conferences on Data Sciences. USA: ACM; 2017 Presented at: CODS'17; March 9 - 11, 2017; Chennai, India p. 1-9. [doi: 10.1145/3041823.3041830]
19. Toseeb U, Inkster B. Online social networking sites and mental health research. Front Psychiatry 2015;6:36 [FREE Full text] [doi: 10.3389/fpsyt.2015.00036] [Medline: 25814958]
20. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016 Aug 29;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
21. Benhammou Y, Tabik S, Achchab B, Herrera F. A First Study Exploring the Performance of the State-of-the Art CNN Model in the Problem of Breast Cancer. In: Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications. USA: ACM; 2018 Presented at: LOPAL'18; May 2 - 5, 2018; Rabat, Morocco p. 1-6. [doi: 10.1145/3230905.3230940]
22. 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;19:1125-1132. [doi: 10.1007/s00779-015-0877-5]
23. de Choudhury CM, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. USA: ACM; 2016 Presented at: CHI'16; May 7 - 12, 2016; California, San Jose, USA p. 2098-2110. [doi: 10.1145/2858036.2858207]
24. Linke SE, Dunsiger SI, Gans KM, Hartman SJ, Pekmezi D, Larsen BA, et al. Association between physical activity intervention website use and physical activity levels among Spanish-speaking Latinas: randomized controlled trial. J Med Internet Res 2019 Jul 24;21(7):e13063 [FREE Full text] [doi: 10.2196/13063] [Medline: 31342902]
25. McLennan S. Quantitative ratings and narrative comments on Swiss physician rating websites: frequency analysis. J Med Internet Res 2019 Jul 26;21(7):e13816 [FREE Full text] [doi: 10.2196/13816] [Medline: 31350838]
26. Liu S, Zhu M, Yu DJ, Rasin A, Young SD. Using real-time social media technologies to monitor levels of perceived stress and emotional state in college students: a web-based questionnaire study. JMIR Ment Health 2017 Jan 10;4(1):e2 [FREE Full text] [doi: 10.2196/mental.5626] [Medline: 28073737]

27. Moreno MA, Jelenchick LA, Egan KG, Cox E, Young H, Gannon KE, et al. Feeling bad on Facebook: depression disclosures by college students on a social networking site. Depress Anxiety 2011 Jun;28(6):447-455 [FREE Full text] [doi: 10.1002/da.20805] [Medline: 21400639]
28. Park M, McDonald DW, Cha M. Perception Differences between the Depressed and Non-Depressed Users in Twitter. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media. 2013 Presented at: ICWSM-13; July 8–11, 2013; Cambridge, Massachusetts.
29. Wolf M, Sedway J, Bulik CM, Kordy H. Linguistic analyses of natural written language: unobtrusive assessment of cognitive style in eating disorders. Int J Eat Disord 2007 Dec;40(8):711-717. [doi: 10.1002/eat.20445] [Medline: 17683092]
30. Park A, Conway M. Harnessing Reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. J Med Internet Res 2018 Apr 10;20(4):e121 [FREE Full text] [doi: 10.2196/jmir.8219] [Medline: 29636316]
31. Blei DM, Ng AY, Jordan M. Latent Dirichlet Allocation. J Mach Learn Res 2003;3:601-608 [FREE Full text]
32. Crossno PJ, Wilson AT, Shead TM, Dunlavy DM. TopicView: Visually Comparing Topic Models of Text Collections. In: Proceedings of the 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence. Palm Beach County, Florida, US: IEEE; 2011 Presented at: ICTAI'11; November 7-9, 2011; Boca Raton, FL, USA. [doi: 10.1109/ictai.2011.162]
33. Řehůřek R, Sojka P. Software Framework for Topic Modelling with Large Corpora. In: Proceedings of LREC 2010 Workshop New Challenges for NLP Frameworks. Valletta, Malta: University of Malta; 2010 Presented at: LREC'10; May 17-23, 2010; Malta.
34. Kozlowski AC, Taddy M, Evans JA. arXiv e-Print archive. 2018. The Geometry of Culture: Analyzing Meaning through Word Embeddings URL: https://arxiv.org/abs/1803.09288 [accessed 2020-01-15]
35. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Silva AP, Meira W, et al. Factors associated with weight change in online weight management communities: a case study in the LoseIt Reddit community. J Med Internet Res 2017 Jan 16;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]
36. Arnow B, Kenardy J, Agras WS. The Emotional Eating Scale: the development of a measure to assess coping with negative affect by eating. Int J Eat Disord 1995 Jul;18(1):79-90. [doi: 10.1002/1098-108x(199507)18:1<79::aid-eat2260180109>3.0.co;2-v] [Medline: 7670446]
37. Kenardy J, Butler A, Carter C, Moor S. Eating, mood, and gender in a noneating disorder population. Eat Behav 2003 Aug;4(2):149-158. [doi: 10.1016/S1471-0153(03)00019-9] [Medline: 15000978]
38. Wallis DJ, Hetherington MM. Emotions and eating. Self-reported and experimentally induced changes in food intake under stress. Appetite 2009 Apr;52(2):355-362. [doi: 10.1016/j.appet.2008.11.007] [Medline: 19071171]
39. Andalibi N, Haimson OL, de Choudhury M, Forte A. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York, New York, USA: ACM Press; 2016 Presented at: CHI'16; May 7 - 12, 2016; California, San Jose, US p. 3906-3918. [doi: 10.1145/2858036.2858096]
40. Andalibi N, Ozturk P, Forte A. Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. New York, New York, USA: ACM Press; 2017 Presented at: CSCW'17; February 25 - March 1, 2017; Oregon, Portland, USA p. 1485-1500. [doi: 10.1145/2998181.2998243]
41. Kenardy J, Arnow B, Agras WS. The aversiveness of specific emotional states associated with binge-eating in obese subjects. Aust N Z J Psychiatry 1996 Dec;30(6):839-844. [doi: 10.3109/00048679609065053] [Medline: 9034475]
42. Ekman P. An argument for basic emotions. Cogn Emot 1992;6(3-4):169-200. [doi: 10.1080/02699939208411068]
43. Frijda NH. Moods, emotion episodes, and emotions. In: Barrett LF, Lewis M, Haviland-Jones JM, editors. Handbook of Emotions. New York City: Guilford Press; 1993:381-403.
44. Heatherton T, Herman C, Polivy J. Effects of physical threat and ego threat on eating behavior. J Pers Soc Psychol 1991 Jan;60(1):138-143. [doi: 10.1037//0022-3514.60.1.138] [Medline: 1995833]
45. Robbins TW, Fray PJ. Stress-induced eating: fact, fiction or misunderstanding? Appetite 1980;1(2):103-133. [doi: 10.1016/s0195-6663(80)80015-8]
46. Bingham SA, Gill C, Welch A, Day K, Cassidy A, Khaw KT, et al. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. Br J Nutr 1994 Oct;72(4):619-643. [doi: 10.1079/bjn19940064] [Medline: 7986792]
47. Hakobyan L, Lumsden J, Shaw R, O?Sullivan D. A Longitudinal Evaluation of the Acceptability and Impact of a Diet Diary App for Older Adults with Age-related Macular Degeneration. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. USA: ACM; 2016 Presented at: MobileHCI'16; September 6 - 9, 2016; Florence, Italy p. 124-134. [doi: 10.1145/2935334.2935356]
48. Abbar S, Mejova Y, Weber I. You Tweet What You Eat: Studying Food Consumption Through Twitter. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. USA: ACM; 2015 Presented at: CHI'15; April 18 - 23, 2015; Seoul, Republic of Korea p. 3197-3206. [doi: 10.1145/2702123.2702153]
49. Chung CF, Agapie E, Schroeder J, Mishra S, Fogarty J, Munson SA. When Personal Tracking Becomes Social: Examining the Use of Instagram for Healthy Eating. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing

Systems. USA: ACM; 2017 Presented at: CHI'17; May 6 - 11, 2017; Colorado, Denver, USA p. 1674-1687. [doi: 10.1145/3025453.3025747]

50. Dirks AJ, Leeuwenburgh C. Caloric restriction in humans: potential pitfalls and health concerns. Mech Ageing Dev 2006 Jan;127(1):1-7. [doi: 10.1016/j.mad.2005.09.001] [Medline: 16226298]
51. Culver JD, Gerr F, Frumkin H. Medical information on the internet: a study of an electronic bulletin board. J Gen Intern Med 1997 Aug;12(8):466-470 [FREE Full text] [doi: 10.1046/j.1525-1497.1997.00084.x] [Medline: 9276651]
52. Eastin MS. Credibility assessments of online health information: the effects of source expertise and knowledge of content. J Comput Mediat Commun 2001;6(4). [doi: 10.1111/j.1083-6101.2001.tb00126.x]
53. Dudley T, Falvo D, Podell R, Renner J. The informed patient poses a different challenge. Patient Care 1996;30(16):128 [FREE Full text]
54. Lindberg DA, Humphreys BL. Medicine and health on the internet: the good, the bad, and the ugly. J Am Med Assoc 1998 Oct 21;280(15):1303-1304. [doi: 10.1001/jama.280.15.1303] [Medline: 9794299]
55. Seyff N, Ollmann G, Bortenschlager M. AppEcho - A User-driven, In Situ Feedback Approach for Mobile Platforms and Applications. In: Proceedings of the 1st International Conference on Mobile Software Engineering and Systems. 2014 Presented at: MOBILESoft'2014; June 2 - 3, 2014; Hyderabad, India p. 99-108. [doi: 10.1145/2593902.2593927]
56. Thornton PL, Kieffer EC, Salabarría-Peña Y, Odoms-Young A, Willis SK, Kim H, et al. Weight, diet, and physical activity-related beliefs and practices among pregnant and postpartum Latino women: the role of social support. Matern Child Health J 2006 Jan;10(1):95-104. [doi: 10.1007/s10995-005-0025-3] [Medline: 16534660]
57. Swann WB, Predmore SC. Intimates as agents of social support: sources of consolation or despair? J Pers Soc Psychol 1985 Dec;49(6):1609-1617. [doi: 10.1037//0022-3514.49.6.1609] [Medline: 4087141]
58. Antin J, de Sa M, Churchill E. Local Experts and Online Review Sites. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion. USA: ACM; 2012 Presented at: CSCW'12; February 11 - 15, 2012; Washington, Seattle, USA p. 55-58. [doi: 10.1145/2141512.2141541]
59. Bezemer T, de Groot MC, Blasse E, Ten Berg MJ, Kappen TH, Bredenoord AL, et al. A Human(e) Factor in Clinical Decision Support Systems. J Med Internet Res 2019 Mar 19;21(3):e11732 [FREE Full text] [doi: 10.2196/11732] [Medline: 30888324]
60. Moreno MA, D'Angelo J. Social media intervention design: applying an affordances framework. J Med Internet Res 2019 Mar 26;21(3):e11014 [FREE Full text] [doi: 10.2196/11014] [Medline: 30912754]
61. Krafft PM, Macy M, Pentland A. Bots as Virtual Confederates: Design and Ethics. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. USA: ACM; 2017 Presented at: CSCW'17; February 25 - March 1, 2017; Oregon, Portland, USA p. 183-190. [doi: 10.1145/2998181.2998354]
62. Long K, Vines J, Sutton S, Brooker P, Feltwell T, Kirman B, et al. 'Could You Define That in Bot Terms'?: Requesting, Creating and Using Bots on Reddit. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. New York, NY. USA: ACM; 2017 Presented at: CHI'17; May 6 - 11, 2017; Colorado, Denver, USA p. 3488-3500. [doi: 10.1145/3025453.3025830]
63. Seering J, Flores JP, Savage S, Hammer J. The social roles of bots: evaluating impact of bots on discussions in online communities. Proc ACM Hum-Comput Interact 2018;2(157):1-29. [doi: 10.1145/3274426]

Abbreviations

AUC: area under curve
EE: emotional eating
EES: emotional eating scale
EF: feedback on EE behavior
HCI: human-computer interaction
LDA: Latent Dirichlet Allocation
NLP: natural language processing
NLTK: natural language toolkit
pEE: probability of emotional eating
PSS: personalized support system
SGD: stochastic gradient descent
TF-IDF: term frequency-inverse document frequency

Edited by G Eysenbach; submitted 31.07.19; peer-reviewed by CA Almenara, D Carvalho; comments to author 15.10.19; revised version received 08.12.19; accepted 16.12.19; published 31.03.20.

Please cite as:
Hwang Y, Kim HJ, Choi HJ, Lee J
Exploring Abnormal Behavior Patterns of Online Users With Emotional Eating Behavior: Topic Modeling Study
J Med Internet Res 2020;22(3):e15700
URL: http://www.jmir.org/2020/3/e15700/
doi:10.2196/15700
PMID:32229461

Original Paper

Love My Body: Pilot Study to Understand Reproductive Health Vulnerabilities in Adolescent Girls

Golfo Tzilos Wernette[1], PhD; Kristina Countryman[1,2], MPH; Kristie Khatibi[2], MPH; Erin Riley[2], MPH; Rob Stephenson[2,3], MSc, PhD

[1]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

[2]Center for Sexuality and Health Disparities, School of Nursing, University of Michigan, Ann Arbor, MI, United States

[3]Department of Systems, Population and Leadership, School of Nursing, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Golfo Tzilos Wernette, PhD
Department of Family Medicine
University of Michigan
1018 Fuller Street
Ann Arbor, MI, 48104
United States
Phone: 1 (734) 998 7120
Email: gtzilos@med.umich.edu

Abstract

Background: Sexually transmitted infections (STIs) are on the rise in the United States, and adolescent girls (15-19 years old) are more susceptible to acquiring STIs than their male peers. The co-occurrence of alcohol use and sexual risk taking contribute significantly to STI acquisition. Mobile health (mHealth) interventions are ideally suited for our target population and have demonstrated increases in STI testing in young people, as well as reductions in alcohol use.

Objective: This pilot study used both qualitative and quantitative methods to explore the views of adolescent girls (age range 15-19 years old; 74.6%, 279/374 white) on the desired qualities and content of an mHealth app for sexual health.

Methods: We conducted nine 60-min in-depth interviews (IDIs) to gather information and identify themes of sexual health and alcohol use, and we tested the feasibility of using a two-week social media campaign to collect survey information regarding sexual health risk in adolescent girls.

Results: We iteratively coded IDIs and identified major themes around pressure of alcohol use, lack of STI knowledge, male pressure to not use condoms, and pregnancy as a worse outcome than STIs. Results from the web-based survey on risky health behaviors, which was completed by 367 participants, support the use of a sexual health app designed for girls.

Conclusions: Future work will integrate these themes to inform the development of a culturally sensitive mHealth app to prevent STIs among adolescent girls.

(J Med Internet Res 2020;22(3):e16336) doi:10.2196/16336

KEYWORDS

adolescent; alcohol; sexually transmitted infections; risky sex; pregnancy

Introduction

Background

Young people aged 15 to 24 years comprise half of all new sexually transmitted infections (STIs) in the United States, despite the fact that this age group accounts for just 27% of the sexually active population [1]. The Centers for Disease Control and Prevention recently reported that 25% of sexually active adolescent girls currently have an STI [1]. They also found that young women aged 15 to 24 years had the highest reported cases of chlamydia and gonorrhea, and that there was a 7.8% increase from 2016 in reported cases of primary and secondary syphilis. STIs are associated with significant morbidity and mortality, including pelvic inflammatory disease and infertility in women [2]. Adolescent girls are particularly vulnerable to STI acquisition compared with their male peers, as they face unique biological, social, and cultural vulnerabilities, such as increased cervical ectopy [1,3] and decreased power over sexual relationships [4,5]. There is a considerable need for a new paradigm in adolescent STI prevention efforts, particularly for girls, given the current data that suggest that prevention efforts do not seem to be effective.

Alcohol Use and Sexual Health

The co-occurrence of alcohol use and sexual risk taking contribute significantly to STI acquisition, particularly in vulnerable populations, such as adolescent girls [6]. Alcohol continues to be the most widely abused substance by adolescents in the United States [7]. National survey data show that 42% of 10th graders and 62% of 12th graders reported lifetime use of alcohol [7], and 11% of females aged 12 to 17 years reported alcohol use in the last 30 days compared with 8.8% of males [8]. Adolescent girls who use alcohol and other substances are disproportionately more likely to engage in risky sexual behaviors that can result in STIs [6]. Owing to the interrelatedness of alcohol use and STI acquisition, and because both are associated with poor health outcomes, addressing alcohol use as a factor in sexual risk behavior may be particularly helpful for the development of sexual health interventions.

A number of evidence-based, sexual health programs delivered in school settings have demonstrated effectiveness in reducing risks during adolescence, including delaying sexual activity and increasing the consistent use of contraceptives [9]; however, there are structural and financial obstacles to the dissemination of these programs, and to maintain behavior changes. Furthermore, national surveys indicate that 83% of sexually experienced girls aged 15 to 17 years did not receive any formal sex education in school until *after* their first sexual experience [10].

Sexually Transmitted Infections Knowledge and Testing

Sexual health knowledge is lacking among adolescents. One study found that 65% of 14 to 21-year-olds were aware of hepatitis, 57% were aware of human papillomavirus, and only 6.5% were aware of trichomoniasis [11]. In a recent study, adolescent girls who had been diagnosed with an STI were more knowledgeable about that particular STI compared with girls who had not been; however, their knowledge of other STIs were similar to girls who had never been diagnosed with one, suggesting that many adolescents do not learn about STIs until after they are diagnosed with one [12]. Limited STI knowledge among adolescents is associated with inconsistent condom use [13,14], multiple sex partners, delaying disease treatment, and failing to return for STI testing results [12]. There is a lack of STI testing among adolescents and young adults aged 15 to 25 years, with 42% of sexually active adolescents reporting not testing because they said they were “not at risk for an STIs”—an alarming statistic for a population that accounts for half of all new STI diagnoses [1,15].

Mobile Health Interventions

Mobile health (mHealth) interventions use mobile phones or other wireless technology to disseminate health information in an effort to alter health behaviors and outcomes [16]. Owing to widespread smartphone and internet use, mHealth interventions are ideally suited for adolescents. A recent (2018) national survey noted that 9 out 10 teens go online multiple times daily, with 45% of teens saying they are online *almost constantly* [17]. Smartphone access was also high, with 95% of teens reporting they had access to a smartphone [17]. Content for an mHealth app can be standardized, tailored, and presented through interactive features, giving it the potential to have a wide reach, potentially leading to a greater impact at a lower cost [18]. Overall, mHealth interventions have demonstrated increased sexual health knowledge, increased STI testing and communication with clinicians about STI risk behavior, higher rates of condom and contraception use, lower rates of unprotected sex, and reductions in alcohol use among adolescents [16,19-21].

Despite their promise, there is still a dearth of knowledge about efficacy, communication best practices for mHealth interventions, and a need for further research into effective content creation and health promotion messaging [16,22]. Furthermore, there is a lack of evidence-based preventive interventions that specifically target both sexual health and alcohol use among adolescent girls aged 15 to 19 years, a particularly vulnerable group [23,24]. This pilot study sought to fill these gaps by conducting in-depth individual interviews and anonymous surveys with female adolescents to (1) understand attitudes about sexual health and alcohol use, and (2) to determine preferences for content, messaging, and format for an mHealth app aimed at reducing risky sexual health behaviors and alcohol use.

Methods

Design

This pilot study used both qualitative and quantitative methods to explore the views of adolescent girls on the desired qualities and content for an mHealth app on sexual health and alcohol use. The first phase of the study was qualitative, in which we conducted in-depth interviews (IDIs), and the second phase was quantitative, in which we conducted web-based surveys using a social media–recruited sample. We recruited adolescent girls aged 15 to 19 years who met study inclusion criteria to participate in a 1-hour focus group. Inclusion criteria included endorsing sexual risk behavior (eg, unprotected sex) and alcohol use within the last 30 days. Owing to recruitment challenges, including a lack of study screener completion, lack of girls reporting both risk factors, and difficulty scheduling participants for focus groups, we changed the study format to individual IDIs and modified our inclusion criteria to include adolescent girls aged 15 to 19 years who endorsed *either* a sexual behavior risk (eg, unprotected sex) or an alcohol behavior risk in the last 30 days. This study was approved by the Institutional Review Board of the University of Michigan.

Recruitment

Recruitment for the focus groups took place between August 2017 and December 2017, and IDI recruitment took place between January 2018 and April 2018. Recruitment for the web-based survey occurred in July 2018. For the qualitative phase of this study, we used both active and passive recruitment methods, including onsite recruitment at an urban family medicine clinic that is the primary care provider for an underserved population of children, adolescents, and adults. Approximately 55% of the clinic's patients identify as Latina or African American. Participants were recruited in the waiting room before their scheduled appointment. Participants were approached, the study was described, and if they expressed interest, they were asked to take the screener. Interested participants provided verbal assent for the screener and received a small gift (eg, lip balm) for participating. Eligible girls were invited to participate in an individual IDI. Participants received US $30 cash for completing the IDI.

We included passive recruitment to the study design to enhance recruitment efforts. Passive recruitment included advertising on a large public university campus in the Midwest, at local clinics, and on college listservs in a broader population area compared with active recruitment. Interested participants reached out via email and received a link to the screener; those who met inclusion criteria were invited to participate in an IDI.

In the second phase of the study, we tested the feasibility of using social media (eg, Facebook) to recruit a national sample of at-risk adolescent girls for research studies, collect survey data on sexual health risk and substance use, and gauge interest in a potential mobile phone app that addresses these issues. We ran social media ads on Facebook and gathered survey data for a period of 2 weeks. We targeted the ad by age (15-19 years old), gender (female), location (the United States), demographics (eg, education level, life events, and relationship status), interests (eg, entertainment and pop culture), and behaviors (eg, shopping). The survey was conducted through Qualtrics software (Qualtrics, Provo, UT) with no confidential information such as names, addresses, or Internet Protocol addresses collected. Participants who completed the social media survey did not receive an incentive.

Instruments

Screener

Participants completed a brief (10-min) health screener via an iPad to determine study eligibility. The screener consisted of questions measuring substance use risk behaviors from the CRAFFT Screening Tool for Adolescent Substance Abuse (computer version), a well-validated and reliable measure for this population recommended by the American Academy of Pediatrics' Committee on Substance Abuse for use with adolescents [25]. It included questions about the use of alcohol, marijuana and other illicit drugs, over the counter drugs, and prescription drugs. Items assessing sexual risk behavior asked about recent vaginal and anal intercourse, including intercourse with and without condoms in the past 30 days.

In-Depth Interviews

Two members of our research team with previous experience with qualitative and adolescent research conducted nine 60-min IDIs. The semistructured interview guide consisted of open-ended questions focused on drinking habits, sexual health behaviors, effective messaging, social media and app preferences, as well as STI knowledge and level of STI concern. Questions such as "How much alcohol do you or other girls your age drink?" and "How are drinking and unprotected sex related for girls your age?" were used to assess typical adolescent female behavior, while "What do you think you or other girls your age want to know about preventing STIs?" and "What features would you like to see in a health app?" helped to determine content and preferences for an mHealth app. Interviews were audio recorded and transcribed verbatim using the web-based transcription service, Scribie. The interview transcripts were reviewed by research staff to ensure accuracy.

Demographic Survey

A post-IDI demographic survey was emailed to the 9 participants to determine age, income, sexual orientation, ethnicity, sex, and education level. To ensure confidentiality, this demographic survey data were not linked to the participant's IDI data.

Qualitative Data Analysis

ATLAS.ti (Scientific Software Development GmbH, 2013-2018) was used to code the interview transcripts. A codebook was iteratively created and agreed upon by all research team members after assessing the major themes from the interview transcripts. The codebook consisted of seven major alcohol and sexual health–related themes such as, *Drinking behaviors* and *STIs & pregnancy*. Within these major themes, we had six subcodes for more specificity such as, *Binge drinking* and *Condom*. Two researchers (KC and KK) independently coded all transcripts and differences were reconciled in-person to create a final coded dataset for the qualitative analysis. We determined that saturation had been achieved after nine IDIs: analysis demonstrated that we were no longer identifying new themes.

Web-Based Survey

We conducted an anonymous, web-based survey using Facebook to test the feasibility of this recruitment method. Survey questions included demographics, substance use questions (eg, "Do you ever use alcohol or drugs to relax, feel better about yourself, or fit in?"), and sexual health questions (eg, "In the past 30 days, how often did you use condoms when you had sex?"). The survey also included a question on whether or not respondents would use a sexual health app designed for girls.

Results

In-Depth Interview Participants

Background characteristics of the IDI participants are shown in Table 1. Recruitment occurred in two separate phases. During the first recruitment phase, we approached 61 girls through active (clinic-based) recruitment, screening a total of 50 girls over the course of 5 months. Of the 50 girls screened, only 7 (14%) were eligible, but no eligible girls participated in a focus group discussion.

Phase 2 of recruitment was conducted both via active (clinic-based) and passive recruitment strategies. A total of 15 girls were approached during active recruitment, of which 7 completed the screener, and 4 (57%) were eligible. Through our passive recruitment strategies, 19 girls contacted us, 18 completed the screener, and 50% (9/18) were eligible. All 9 of these participants completed an IDI. A CRAFFT score was determined using the screener data for each participant. The range of scores was from 0 to 4, with a score of 4 or more indicating possible substance dependence.

Table 1. Study characteristics of in-depth interview participants (N=9).

Variables	Value
Age (years), mean (SD), range	18.5 (0.52), 18-19
Race/ethnicity[a] (n=7), n (%)	
White	4 (57)
Biracial	2 (28)
Other	1 (14)
Hispanic or Latina/Latinx	2 (28)
Highest level of education[a] (n=7), n (%)	
Currently in college or another training program	7 (100)
Family household income[a] (US $; n=7), n (%)	
35,000-49,999	1 (14)
50,000-74,999	1 (14)
200,000 or more	3 (42)
Not sure	2 (28)
Sexual orientation[a] (n=7), n (%)	
Straight/heterosexual	6 (85)
Gay/lesbian/homosexual	1 (14)
Current relationship status (n=9), n (%)	
Single	4 (44)
In a relationship	5 (55)
Condom use in the last 30 days (n=9), n (%)	
I have not had sex in the last 30 days	1 (11)
All of the time	2 (22)
Most of the time	4 (44)
Some of the time	1 (11)
None of the time	1 (11)
Any alcohol use in the last 12 months (n=9), n (%)	
Yes	9 (100)
Any marijuana use in the last 12 months (n=9), n (%)	
Yes	4 (44)
CRAFFT[b] score[c] (n=8), n (%)	
0	2 (25)
1	1 (12)
2	2 (25)
3	1 (12)
4	2 (25)

[a] Only 7 of the 9 participants completed demographic questions.

[b] CRAFFT: CRAFFT Screening Tool for Adolescent Substance Abuse (computer version).

[c] Only 8 of the 9 participants completed the CRAFFT assessment.

Web-Based Survey Participants

Participant characteristic results from the web-based survey are presented in Table 2. Respondents were asked questions on their relationship status and condom use, as well as their alcohol and drug use (see Table 3). The majority of respondents (86%; 314/367) reported that they would use a sexual health phone app for adolescent girls. Surveys were considered complete and included in the analysis if all questions were viewed by the participants. Participants could choose not to respond to questions that made them feel uncomfortable, and their surveys were still included in the analysis.

Table 2. Study characteristics of web-based survey participants (N=367).

Variables	Value
Age (years), mean (SD), range	16.2 (0.88), 15-19
Race (n=365), n (%)	
White	275 (75.3)
Biracial	26 (7.1)
African American/black	25 (6.9)
Asian	20 (5.5)
Native Hawaiian or Pacific Islander	5 (1.4)
Native American or Native Alaskan	3 (0.8)
Other[a]	11 (3.0)
Ethnicity (n=367), n (%)	
Hispanic or Latina/Latinx	33 (9.0)
Highest level of education (n=367), n (%)	
Currently in high school	335 (91.3)
Graduated high school of have GED[b]	15 (4.1)
Not in high school and did not complete high school or GED	2 (0.5)
Currently in college or another training program	15 (4.1)
Family household income (n=361) (US $), n (%)	
Less than 25,000	24 (6.7)
25,000-34,999	21 (5.8)
35,000-49,999	23 (6.4)
50,000-74,999	36 (10.0)
75,000-99,999	32 (8.9)
100,000-149,999	38 (10.5)
150,000-199,999	14 (3.9)
200,000 or more	14 (3.9)
Not sure	159 (44.0)
Sexual orientation (n=365), n (%)	
Straight/heterosexual	123 (33.7)
Bisexual	153 (41.9)
Gay/lesbian/homosexual	39 (10.7)
Uncertain or questioning	50 (13.7)
Current relationship status (n=366), n (%)	
Single	221 (60.4)
In a relationship, but not married	145 (39.6)

[a] Other: write-in responses for race included the following: Hispanic (2), North African (1), Middle Eastern (2), Multiracial (1), Native American/African American/Eastern European (1), Other or no specification (3).

[b] GED: General Education Development.

Table 3. Sex risk and substance use behavior responses of web-based survey participants (N=367).

Variables	Value, n (%)
Condom use in the last 30 days (n=362)	
I have not had sex in the last 30 days	274 (75.7)
All of the time	13 (3.6)
Most of the time	12 (3.3)
Some of the time	13 (3.6)
None of the time	50 (13.8)
Condom use in the last 30 days among sexually active participants (n=88)	
All of the time	13 (15)
Most of the time	12 (14)
Some of the time	13 (15)
None of the time	50 (57)
Any alcohol use in the last 12 months (n=367)	
Yes	123 (33.5)
Any marijuana use in the last 12 months (n=367)	
Yes	83 (22.6)
Any other drug use in the last 12 months (n=366)	
Yes	32 (8.7)
Ever ridden in a car driven by someone (including yourself) who was *high* or had been using alcohol or drugs? (n=367)	
Yes	89 (24.2)
Ever use alcohol or drugs to relax, feel better about yourself, or fit in? (n=365)	
Yes	86 (23.6)
Ever use alcohol or drugs while you are by yourself or alone? (n=367)	
Yes	80 (21.8)
Ever forget things you did while using alcohol or drugs (ie, *blacking out*)? (n=366)	
Yes	35 (9.6)
Do family or friends ever tell you that you should cut down on your drinking or drug use? (n=366)	
Yes	12 (3.3)
Have you ever gotten in trouble while using alcohol or drugs? (n=367)	
Yes	17 (4.6)
Would you use a sexual health app for girls? (n=367)	
Yes	314 (85.8)

In-Depth Interview Results

Data from the nine IDIs resulted in four major interrelated themes described in the sections below. These themes included alcohol and college life, alcohol and condomless sex, STIs and pregnancy, and technology use and app preferences.

Alcohol and College Life

Participants agreed that alcohol plays a big role in college life, whether or not one chooses to drink. They described both alcohol and the pressure to drink as omnipresent. Several interrelated reasons were given for drinking, and these reasons were often linked with a desire to relieve school and/or social stress and anxiety. However, the most cited reason for drinking was the pressure to drink received from peers, sexual partners (potential and actual), and from society/media as a whole.

Most of the drinking behaviors the participants described would be classified as binge drinking (defined as drinking more than 4 drinks in 2 hours for women, and more than 5 in 2 hours for men), and/or heavy drinking (defined as binge drinking on 5 or more days within a month, the National Institute on Alcohol Abuse and Alcoholism [26]). This kind of drinking was described as the norm. Some even said they felt that binge drinking was a *serious problem* on campus. When *partying* was described, they spoke mostly of alcohol use. Drug use other than alcohol was rarely mentioned in the interviews, even though one participant described drugs as being *prominent* on campus, and that *all sorts* of drugs were mixed with alcohol at parties.

Alcohol and Condomless Sex

When asked about the relationship between alcohol and condomless sex, most said that these were *strongly connected* because drinking too much lowers inhibition and impairs decision making, often leading to *bad* decisions. But the participants also provided a more nuanced look into the correlation of alcohol use and condomless sex. Specifically, the reasons why it happens in the first place, and how alcohol use intensifies it. These reasons overlapped and tended to reinforce each other, including male pressure for condomless sex, lack of female sexual empowerment, the societal stigma of female sexuality, and a lack of concern or knowledge of STIs.

The persistent reason for condomless sex was said to be the pressure that young women receive from young men to not wear one. Five of the nine participants explained how drinking alcohol made it easier for young men to convince young women to not use condoms:

> *... 'Cause when you're intoxicated, you, uh, aren't gonna be thinking completely straight and you might just be relying on emotional decisions at that point…And I think with, you know, like rape culture and kinda like the way sex is depicted in society today for young guys in college. I think that women, uh, are pressured to have unprotected sex sometimes...* [Interview 3, Age 19]

Ignorance surrounding the prevalence of STIs was another reason given for condomless sex. Participants repeatedly said that their peers did not talk about STIs, and they believed they were not really a problem, because STIs were not that prevalent among young people. Similar to condom use, issues of trust came up when talking to a partner about STIs. The consensus was you do not talk to your sexual partner about STIs or their status; to do so would be an implication of mistrust, and a belief that they are not *clean*.

> *I've always had, like made sure that the... Like, my current partner does not have any STDs or STIs. But I know that, umm, a lot of my friends have been like, 'Oh my God! You asked him that? How could you do that? Like, isn't that you don't trust him?' I'm like, 'No, it's that I want my body to be safe. I'm clean, so I wanna make sure he's clean.'* [Interview 9, Age 18]

Sexually Transmitted Infections and Pregnancy

Two resounding themes came out of our questions on STIs and pregnancy: (1) young women are largely ignorant of STIs and their health effects, and (2) young women are not that concerned about STIs, but rather, about pregnancy. When asked what they knew about STIs, the majority of our participants answered that they had limited knowledge. A few participants knew the names of some STIs and knew that they were transmitted sexually, but unconfidently followed-up their responses with an admission of *not really knowing much* about them. Participants said they would like to know more about the etiology of the different STIs, how to effectively prevent them, and the associated treatment, but also said that the main concern in their sex lives was to avoid pregnancy.

The lack of concern about STIs was often attributed to the belief that young women simply did not realize how common these infections were. Some expressed ignorance because they had never had an STI, and their friends had not either. This was coupled with beliefs that STIs were a "thing that older people have," not something for women their age to worry about because, "he's what, like 19 you know, he surely hasn't gotten around that much you know, that he has an STD already." None of the participants felt that the information they had received in high school was adequate, and some even said that what they had been told was completely inaccurate. Multiple participants reported having had abstinence-based sexual education class, and they felt that these classes were not very informative and relied heavily on scare tactics. One participant said her sex-ed teacher told the class that condoms "wouldn't protect against STIs," and were "only 50% effective at preventing pregnancy."

Participants also said that young men did not believe that STIs were prevalent, and therefore did not think condoms were that important. They said that young men generally preferred the withdrawal method. As they believed that STIs were not prevalent in young people, participants admitted having feelings of invincibility, and trusted that they would either know if their sexual partner had an STI, or that their sexual partner would communicate if they had an STI:

> *I think it's more just the like, 'it would never happen to me' kinda idea. I mean like I'm guilty of it, like I definitely think, 'Oh that would never happen to me', like I wouldn't have sex with someone who wouldn't tell me if they have an STD, umm like I'm smarter than that, I'm not gonna do it. Umm but in reality, I'm sure it could happen to anyone.* [Interview 3, Age 19]

Participants agreed that getting pregnant would be a far *worse* outcome than getting an STI, because of the social stigma surrounding teenage/young mothers, and what they described as the *longer term consequences* of pregnancy. These *consequences* persisted whether the pregnancy was carried to term or not. Abortion was referred to as *scary*, *very painful*, and *serious procedure*, and an emotionally difficult choice to make. Carrying the pregnancy to term was also described as an emotionally difficult choice to make; thus, avoiding pregnancy altogether was of the utmost importance. Participants acknowledged that there is a stigma surrounding STIs and condomless sex but said that pregnancy had more visible (and therefore, public) consequences, while STIs could more easily be treated and kept a private matter. They also saw STIs as a problem affecting both parties; something they would go through together. Pregnancy, on the other hand, was something they feared they would face alone. As pregnancy is something that only physically affects women's bodies, participants said that men did not see it at as something that affects them much; that it was ultimately a woman's problem:

> *...I've had girlfriends that have gotten pregnant and it's affected their lives deeply. But the guy, it affected their life, like zero.* [Interview 9, Age 18]

Technology Use and App Preferences

All participants reported smartphone use and were actively engaged with multiple mobile apps. The most frequently used apps were for email, music, videos (eg, YouTube), and social media. Although not every participant used the same social media apps, all participants reported using at least one social media app. When asked to reflect on specific app features they liked, they said an app needs to be interactive, visually attractive, easy to use, have tracking capabilities, and the ability to set goals and earn rewards.

When asked about sexual health–focused apps, none of the participants were familiar with any. Some participants mentioned using a period-tracking app, which they said sometimes included information on STIs and contraception. When asked about alcohol-related apps, several participants discussed a drinking app created by the university that was focused on reducing alcohol consumption to *stay safe* and *in control*.

Effective Messaging

To determine the most compelling way to present messaging around safer sex and reducing or avoiding alcohol/drugs in an mHealth app, we asked participants what effective and ineffective advertising or messaging they have seen. They reported having heard some effective messaging but stressed the need for more and different messages, especially surrounding sexual health. When asked to describe the kind of messaging they felt would work best with their peers, there was some disagreement about whether to use *scare tactics* in alcohol-related messaging, but the participants agreed that scare tactics needed to be avoided in sexual health messaging. They emphasized wanting to know facts and ways to stay safe when it came to sex and alcohol use, rather than just receiving messages of abstinence.

Effective Alcohol Use Messaging

Participants overwhelmingly said that they avoided messages in the *don't drink* genre. They saw them as ineffective and were generally dismissed because they are seen as being unrealistic and out-of-touch:

> *I feel like when they promote not drinking whatsoever, people will be like, 'Yeah, right, whatever,' because it's gonna happen, girls are going to drink. You're in college. Umm, there's no doubt about it.* [Interview 6, Age 18]

Including practical messaging to reduce alcohol consumption was thought to be the most effective approach, such as highlighting the benefits of drinking less (ie you can still have fun but avoid hangovers). Health and safety came up frequently when discussing alcohol-related topics that participants would like to see in an app. Specifically, they said they want to learn about *healthy levels* of drinking versus *problem drinking*, and how to gage and manage their own personal alcohol limits. They recommended that our app contain a drink-tracking feature like the university's app discussed above. They also said it would be helpful to include a centralized list of resources that provide substance use help.

Participants resoundingly said that they wanted an app that focuses on helping young women, not on shaming them. To them, this meant that the app would use positive language, and would validate young women's desire to try alcohol as something that was normal and not as something that was *bad* or *immoral*. Validating that it was equally normal to *not* want to drink was something they asked to include as well as emphasizing that if they did choose not to drink, they would not be alone in their choice.

Effective Sexual Health Messaging

Similar to alcohol-related messaging, participants said that the most effective way to talk to young women about sexual health was to be factual in an empowering manner, not in a moralizing or judgmental one. One participant said that when issues of women's sexual health were put in a moralizing manner, it was too "easy to cross into the slut shaming line," which to her, made them ineffective. She further discussed how she disliked her high school sexual education class because it centered around scare tactics:

> *Definitely listing all the symptoms is fine, but the pictures I feel like can be overly graphic... But like the way it was taught to me, it was like, you have sex you get an STI, like a direct link there. So like basically you have sex, you're dirty, even if you don't get an STI.* [Interview 9, Age 18]

When we asked where young women went for information on STIs, most participants said that they generally did not search for information on STIs. They said that STIs were just not something people their age worried much about. Overall, participants desired detailed information on STIs, not just how to prevent them. They wanted to know how they can affect your body and overall health, which ones can be cured and which ones cannot, how the testing process works for different STIs, and what the implications are for their sexual partner(s) with a positive test.

Participants explained that an app should also include information on sexual consent, coercion, and issues of sexual assault and how to prevent it. As many women become sexually active during adolescence and into college, they thought it was very important to include the definition of consent and how to give it. They also wanted the app to have information on how to recognize and get out of potentially dangerous situations, and steps to mitigate and navigate a sexual assault situation. Finally, participants said they wanted empowering messages about their sexuality and their bodies. They said that young women want to be in control of their sex lives without experiencing stigmatizing labels. They want to be comfortable about their sexuality and comfortable using condoms. As one young woman put it:

> *I think deep down, all women would prefer to be protected.* [Interview 3, Age 19]

Discussion

Principal Findings

This pilot study used both qualitative and quantitative methods to explore the views of adolescent girls on the desired qualities and content for an mHealth app on sexual health and alcohol use. Overall, the results of this pilot work were encouraging. We identified and recruited adolescent girls who endorsed recent alcohol use and sexual risk behavior to participate in the individual IDIs. Our study was feasible with respect to the study procedures, and passive recruitment methods were most effective in identifying girls who endorsed these risks. Moreover, we found that using social media to recruit and engage adolescent girls who endorsed high-risk behaviors was feasible and acceptable. Of the 382 individuals who started the survey, 367 (96.0%) completed it. Moreover, 86% (314/367) endorsed that they would use a sexual health app for girls.

Several themes emerged from the IDIs, with the major themes involving alcohol and college life, alcohol and condomless sex, STIs and pregnancy, and technology use and preferences. Participants offered, from their personal experiences as well as those of others, their perspectives supporting that alcohol use and sexual risk behavior are closely connected in this phase of life. The co-occurrence of alcohol use and sexual risk taking contribute significantly to STI acquisition, particularly in vulnerable populations, such as adolescent girls.

There was a general sense of inexperience with STIs and, subsequently, a lack of knowledge on prevention and vulnerability. Two very clear themes emerged during our interviews with respect to our questions on STIs and pregnancy. First, young women are largely ignorant of STIs and their health effects, and second, young women are relatively unconcerned about STIs, but rather, extremely concerned about pregnancy. The deep-seated concern about pregnancy was not only because of the significant life change it would bring but also because of the moral implications in a society where having a teenage pregnancy brings harsh judgment and shame. STIs were seen as something that could easily be hidden if you contracted one, but pregnancy was something out in the open. In our IDIs, we found participants to be open about recommendations for a sexual health app, for example, “if you want to talk to young people about STIs, focus on *why* young people should prevent them, not just *how* they can prevent them.”

Comparison With Previous Work

The use of mHealth interventions to target sexual health and substance use among youth holds promise in reaching large numbers of at-risk adolescents remotely and to tailor content to individual user needs. In 2017, a systematic review of mHealth interventions for alcohol/drug use prevention among young adults found mixed results among the individual studies, but an overall support for the efficacy of mHealth interventions in reducing substance use. They included a variety of mHealth modalities such as text messaging, web-based apps, and smartphone apps. The limitations of these studies were largely because of small sample sizes, lack of data on long-term effects, lack of examination into potential gender differences in intervention effects, and lack of diversity in participant age and environment [27].

Our review of federal registries for evidence-based interventions (EBIs) on adolescent substance use prevention and adolescent sexual health found none that focused on both of these health risks simultaneously for nontreatment seeking or nonincarcerated adolescent girls [23,24]. While there were several EBIs focused on adolescent girls’ sexual health [28-33], there was a dearth of EBIs for substance use prevention in the 15 to 19 age range, and none that focused on young women. Furthermore, the sexual health EBIs were either specific to clinic-seeking, at-risk, or racial and ethnic minority adolescent girls. The sexual health EBIs with the most promising results still faced the limitation of reach and scalability; none of them were mHealth, and all relied on a clinic, classroom, or community-based setting with trained facilitators [28-33]. Not only would a mobile sexual health app be able to reach more adolescents but it would also be more cost-effective and provide greater impact.

Strengths and Limitations

This study included a number of strengths. We explored views on a topic of great public health significance and successfully engaged youth to share their attitudes and feelings on sensitive topics. We used social media as a tool to engage high-risk adolescents to collect informative data, demonstrating that this is a suitable platform for our future work. Owing to our study population comprising of college-enrolled female adolescents, there are still gaps in understanding the preferences and needs of female adolescents from disadvantaged backgrounds and the results cannot be generalized to the population at large. The inability of this study to successfully recruit girls to complete focus groups exemplifies the challenges that need to be understood among this population to best reach them effectively.

Conclusions

This pilot study demonstrates a clear need to reach, inform, and empower adolescent girls on the topics of sexual health and alcohol/drug use. Given the escalating rates of STI acquisition that continues to affect this vulnerable group, addressing the co-occurrence of alcohol/drug use and sexual risk taking is a research area of high priority, and also for which data are insufficient and scarce. Future work will integrate these areas to inform the development of a culturally sensitive mHealth app to empower adolescent girls and to reduce health risks and prevent STIs among them.

Acknowledgments

This research was supported by a faculty seed grant through the Institute for Research on Women and Gender (Principal Investigator: GW; Co-Investigator: RS). The authors would like to thank all of the participants and staff at the Ypsilanti Health Center and the participants from the Ypsilanti and Ann Arbor community. The authors would also like to acknowledge the contributions of Hannah Mesa and Amy Frauhammer.

Authors' Contributions

GW, KC, KK, ER, and RS contributed to the study design. KC and KK contributed to the data collection. GW, KC, KK, and ER contributed to the analysis, interpretation, and writing. All authors contributed to the editing, review, and approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. 2018. STDs in Adolescents and Young Adults URL: https://www.cdc.gov/std/stats18/adolescents.htm [accessed 2020-01-29]
2. Centers for Disease Control and Prevention. STDs & Infertility URL: https://www.cdc.gov/std/infertility/default.htm [accessed 2020-01-29]
3. Lee V, Tobin JM, Foley E. Relationship of cervical ectopy to chlamydia infection in young women. J Fam Plann Reprod Health Care 2006 Apr;32(2):104-106 [FREE Full text] [doi: 10.1783/147118906776276440] [Medline: 16824301]
4. Teitelman AM, Bohinski JM, Tuttle AM. Condom coercion, sexual relationship power, and risk for HIV and other sexually transmitted infections among young women attending urban family planning clinics. Fam Viol Prev Health Pract 2010;1(10):- [FREE Full text] [Medline: 26778938]
5. Pulerwitz J, Amaro H, de Jong W, Gortmaker SL, Rudd R. Relationship power, condom use and HIV risk among women in the USA. AIDS Care 2002 Dec;14(6):789-800. [doi: 10.1080/0954012021000031868] [Medline: 12511212]
6. Rehm J, Shield KD, Joharchi N, Shuper PA. Alcohol consumption and the intention to engage in unprotected sex: systematic review and meta-analysis of experimental studies. Addiction 2012 Jan;107(1):51-59. [doi: 10.1111/j.1360-0443.2011.03621.x] [Medline: 22151318]
7. Institute for Social Research - University of Michigan. 2017 Dec 14. National Adolescent Drug Trends in 2017 URL: https://isr.umich.edu/news-events/news-releases/national-adolescent-drug-trends-in-2017/ [accessed 2020-02-04]
8. Center for Behavioral Health Statistics and Quality. Substance Abuse and Mental Health Services Administration. Rockville, MD; 2018. Results From the 2017 National Survey on Drug Use And Health: Detailed Tables URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.pdf [accessed 2020-01-30]
9. Shepherd J, Kavanagh J, Picot J, Cooper K, Harden A, Barnett-Page E, et al. The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13-19: a systematic review and economic evaluation. Health Technol Assess 2010 Feb;14(7):1-206, iii [FREE Full text] [doi: 10.3310/hta14070] [Medline: 20178696]
10. Cox S, Pazol K, Warner L, Romero L, Spitz A, Gavin L, Centers for Disease ControlPrevention (CDC). Vital signs: births to teens aged 15-17 years--United States, 1991-2012. MMWR Morb Mortal Wkly Rep 2014 Apr 11;63(14):312-318 [FREE Full text] [Medline: 24717819]
11. Kurkowski JP, Hsieh G, Sokkary N, Santos X, Bercaw-Pratt JL, Dietrich JE. Knowledge of sexually transmitted infections among adolescents in the Houston area presenting for reproductive healthcare at Texas Children's Hospital. J Pediatr Adolesc Gynecol 2012 Jun;25(3):213-217. [doi: 10.1016/j.jpag.2012.02.005] [Medline: 22578483]
12. Downs JS, Bruine de Bruin W, Murray PJ, Fischhoff B. Specific STI knowledge may be acquired too late. J Adolesc Health 2006 Jan;38(1):65-67. [doi: 10.1016/j.jadohealth.2005.01.004] [Medline: 16387254]
13. Walsh JL, Senn TE, Scott-Sheldon LA, Vanable PA, Carey MP. Predicting condom use using the Information-Motivation-Behavioral Skills (IMB) model: a multivariate latent growth curve analysis. Ann Behav Med 2011 Oct;42(2):235-244 [FREE Full text] [doi: 10.1007/s12160-011-9284-y] [Medline: 21638196]
14. Espada JP, Morales A, Guillén-Riquelme A, Ballester R, Orgilés M. Predicting condom use in adolescents: a test of three socio-cognitive models using a structural equation modeling approach. BMC Public Health 2016 Jan 14;16:35 [FREE Full text] [doi: 10.1186/s12889-016-2702-0] [Medline: 26762643]
15. Cuffe KM, Newton-Levinson A, Gift TL, McFarlane M, Leichliter JS. Sexually transmitted infection testing among adolescents and young adults in the United States. J Adolesc Health 2016 May;58(5):512-519. [doi: 10.1016/j.jadohealth.2016.01.002] [Medline: 26987687]
16. L'Engle KL, Mangone ER, Parcesepe AM, Agarwal S, Ippoliti NB. Mobile phone interventions for adolescent sexual and reproductive health: A systematic review. Pediatrics 2016 Sep;138(3):pii: e20160884 [FREE Full text] [doi: 10.1542/peds.2016-0884] [Medline: 27553221]
17. Anderson M, Jiang J. Pew Research Center. 2018 May 31. Teens, Social Media & Technology 2018 URL: https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-01-30]
18. Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. J Med Internet Res 2013 Jan 4;15(1):e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]

19. 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 Oct 6;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
20. Mesheriakova VV, Tebb KP. Effect of an iPad-based intervention to improve sexual health knowledge and intentions for contraceptive use among adolescent females at school-based health centers. Clin Pediatr (Phila) 2017 Nov;56(13):1227-1234 [FREE Full text] [doi: 10.1177/0009922816681135] [Medline: 28950721]
21. Cordova D, Alers-Rojas F, Lua FM, Bauermeister J, Nurenberg R, Ovadje L, et al. The usability and acceptability of an adolescent mHealth HIV/STI and drug abuse preventive intervention in primary care. Behav Med 2018;44(1):36-47 [FREE Full text] [doi: 10.1080/08964289.2016.1189396] [Medline: 27223646]
22. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
23. St Lawrence JS, Crosby RA, Brasfield TL, O'Bannon RE. Reducing STD and HIV risk behavior of substance-dependent adolescents: a randomized controlled trial. J Consult Clin Psychol 2002 Aug;70(4):1010-1021. [Medline: 12182264]
24. Bryan AD, Schmiege SJ, Broaddus MR. HIV risk reduction among detained adolescents: a randomized, controlled trial. Pediatrics 2009 Dec;124(6):e1180-e1188. [doi: 10.1542/peds.2009-0679] [Medline: 19901006]
25. Harris SK, Louis-Jacques J, Knight JR. Screening and brief intervention for alcohol and other abuse. Adolesc Med State Art Rev 2014 Apr;25(1):126-156. [Medline: 25022191]
26. National Institute on Alcohol Abuse and Alcoholism. 2015. Drinking Levels Defined URL: http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking [accessed 2016-01-20]
27. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
28. Shrier LA, Ancheta R, Goodman E, Chiou VM, Lyden MR, Emans SJ. Randomized controlled trial of a safer sex intervention for high-risk adolescent girls. Arch Pediatr Adolesc Med 2001 Jan;155(1):73-79. [doi: 10.1001/archpedi.155.1.73] [Medline: 11177066]
29. DiClemente RJ, Wingood GM, Harrington KF, Lang DL, Davies SL, Hook EW, et al. Efficacy of an HIV prevention intervention for African American adolescent girls: a randomized controlled trial. J Am Med Assoc 2004 Jul 14;292(2):171-179. [doi: 10.1001/jama.292.2.171] [Medline: 15249566]
30. DiClemente RJ, Wingood GM, Rose ES, Sales JM, Lang DL, Caliendo AM, et al. Efficacy of sexually transmitted disease/human immunodeficiency virus sexual risk-reduction intervention for african american adolescent females seeking sexual health services: a randomized controlled trial. Arch Pediatr Adolesc Med 2009 Dec;163(12):1112-1121. [doi: 10.1001/archpediatrics.2009.205] [Medline: 19996048]
31. Coyle KK, Kirby DB, Robin LE, Banspach SW, Baumler E, Glassman JR. All4You! A randomized trial of an HIV, other STDs, and pregnancy prevention intervention for alternative school students. AIDS Educ Prev 2006 Jun;18(3):187-203. [doi: 10.1521/aeap.2006.18.3.187] [Medline: 16774462]
32. Downs JS, Murray PJ, de Bruin WB, Penrose J, Palmgren C, Fischhoff B. Interactive video behavioral intervention to reduce adolescent females' STD risk: a randomized controlled trial. Soc Sci Med 2004 Oct;59(8):1561-1572. [doi: 10.1016/j.socscimed.2004.01.032] [Medline: 15279915]
33. Jemmott JB, Jemmott LS, Braverman PK, Fong GT. HIV/STD risk reduction interventions for African American and Latino adolescent girls at an adolescent medicine clinic: a randomized controlled trial. Arch Pediatr Adolesc Med 2005 May;159(5):440-449. [doi: 10.1001/archpedi.159.5.440] [Medline: 15867118]

Abbreviations

EBI: evidence-based intervention
IDI: in-depth interview
mHealth: mobile health
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 20.09.19; peer-reviewed by KA Nguyen, A Newton-Levinson; comments to author 15.11.19; revised version received 28.12.19; accepted 24.01.20; published 30.03.20.

Please cite as:
Tzilos Wernette G, Countryman K, Khatibi K, Riley E, Stephenson R
Love My Body: Pilot Study to Understand Reproductive Health Vulnerabilities in Adolescent Girls
J Med Internet Res 2020;22(3):e16336
URL: http://www.jmir.org/2020/3/e16336/
doi:10.2196/16336
PMID:32224484

Original Paper

Implementing a Digital Tool to Support Shared Care Planning in Community-Based Mental Health Services: Qualitative Evaluation

Christalla Pithara[1,2*], PhD; Michelle Farr[1,2*], PhD; Sarah A Sullivan[1,2,3], PhD; Hannah B Edwards[1,2], MSc; William Hall[4], MBChB, MRCPsych; Caroline Gadd[5], RN; Julian Walker[4], PhD; Nick Hebden[5], BSc; Jeremy Horwood[1,2*], PhD

[1]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[2]National Institute for Health Research Applied Research Collaboration West at University Hospitals Bristol Foundation Trust, Bristol, United Kingdom

[3]Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[4]Avon and Wiltshire Mental Health Partnership NHS Trust, Bristol, United Kingdom

[5]Otsuka Health Solutions, Slough, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Christalla Pithara, PhD
Population Health Sciences, Bristol Medical School
University of Bristol
9th Floor Whitefriars, Lewins Mead
Bristol, BS1 2NT
United Kingdom
Phone: 44 1173421272 ext 21272
Email: christalla.pithara@bristol.ac.uk

Abstract

Background: Mental health services aim to provide recovery-focused care and facilitate coproduced care planning. In practice, mental health providers can find supporting individualized coproduced care with service users difficult while balancing administrative and performance demands. To help meet this aim and using principles of coproduction, an innovative mobile digital care pathway tool (CPT) was developed to be used on a tablet computer and piloted in the West of England.

Objective: The aim of this study was to examine mental health care providers' views of and experiences with the CPT during the pilot implementation phase and identify factors influencing its implementation.

Methods: A total of 20 in-depth telephone interviews were conducted with providers participating in the pilot and managers in the host organization. Interviews were audio recorded, transcribed, anonymized, and thematically analyzed guided by the Consolidated Framework for Implementation Research.

Results: The tool was thought to facilitate coproduced recovery-focused care planning, a policy and organizational as well as professional priority. Internet connectivity issues, system interoperability, and access to service users' health records affected use of the tool during mobile working. The organization's resources, such as information technology (IT) infrastructure and staff time and IT culture, influenced implementation. Participants' levels of use of the tool were dependent on knowledge of the tool and self-efficacy; perceived service-user needs and characteristics; and perceptions of impact on the therapeutic relationship. Training and preparation time influenced participants' confidence in using the tool.

Conclusions: Findings highlight the importance of congruence between staff, organization, and external policy priorities and digital technologies in aiding intervention engagement, and the need for ongoing training and support of those intended to use the technology during and after the end of implementation interventions.

(J Med Internet Res 2020;22(3):e14868) doi:10.2196/14868

KEYWORDS

health care technology; mental health; community health care; patient-centered care; patient care planning; implementation science

Introduction

Background

Mental health digital technologies provide opportunities to improve care [1,2], service efficiency, and health outcomes [3-7]. Previous studies have explored mental health service users' [5-10] and providers' [11,12] experiences with health technology and technological innovations. There remains a need for coproduced real-world evaluation research [13], increasing understanding of contextual and organizational factors involved in successful implementation [14], particularly use of digital interventions within a therapeutic context [13,15-18].

Recovery-focused care, embracing principles of shared decision making and coproduction of care plans, is recommended to improve mental health care delivery [19]. Coproduction requires services to be delivered "in an equal and reciprocal relationship between professionals, people using services, [and] their families" [20]. For decision making, both care providers and service users should possess skills and ability to access, share, and use information to meet service users' often complex, individual needs [21]. Care providers, however, may find it challenging to ensure individualized, person-centered care, while also balancing administrative and performance demands [22]. Coproduction in care can be compromised by individual and organizational factors [22-25], including health information technology (IT) systems [25,26]. Here lies the need for innovations supporting coproduced care, while addressing performance and efficiency concerns.

Development and Pilot Implementation of the Care Pathway Tool

To support coproduced mental health care, a care pathway tool (CPT) was developed through a collaborative effort between a mental health care service provider (Avon and Wiltshire Mental Health Partnership NHS Trust [AWP], the lead regional provider for community mental health services), users of community mental health services, and technology developers (Otsuka Health Solutions [OHS]), as part of a project piloted in the West of England (Joining the Dots).

The project aimed to use computer tools for better use of data and information to improve care delivery and facilitate collaborative working in care planning. The CPT aimed to (1) enable providers' and service users' direct access to electronic care plans to support efficient working and (2) enable coproduced, recovery-focused care during community visits, through patient involvement in care planning, and introducing specific exercises to encourage new ways of interacting. The experiences of staff using the CPT to coproduce recovery-oriented care planning are reported here [27].

The CPT was developed using the coproduction principles [20]: an iterative, collaborative approach involving service users and providers through (1) consultations that identified gaps and care needs, (2) a Joint Project Board with representation from service users, practitioners, managers, and software developers, and (3) feedback from mental health trust staff and service users via detailed observations, interviews, and focus groups on their experiences of consecutive versions of the CPT. As part of this collaboration, a film reporting on the coproduction process from the service users' perspective was put together by Rethink Mental Illness, an organization facilitating service user involvement [28].

On the basis of this work, the CPT was designed to be used on a mobile tablet computer and incorporated 4 different components (Table 1). Screenshots of the CPT are included in Multimedia Appendix 1.

The pilot implementation of the CPT took place between March and December 2016. A total of 30 providers involved in care planning and recovery support for Community Mental Health Teams were recruited through engagement events or word-of-mouth to pilot the CPT in routine practice. Face-to-face training was provided for all staff from an experienced mental health worker on how the tool worked and how it could be used to facilitate coproduction in care planning. Any issues arising from the CPT during these meetings were raised with the software developers. Help information was also included as part of the tool itself and as part of a service user information leaflet. This information was specific to navigating the CPT and using the CPT components.

Staff volunteers were asked to use the CPT with up to 5 mental health service users whose clinical risk assessment for a mental health crisis (such as an exacerbation in their clinical condition, which would require urgent attention) was set at low or medium risk. This decision was informed by a cautious approach to testing new mental health electronic tools in the National Health Service (NHS).

Mental health providers introduced the tool to service users during routine meetings and integrated it into practice if service users agreed.

Table 1. Components and features of the care pathway tool.

Care pathway tool component and feature	Description
My life	
My journey	- An interactive timeline illustrating service users' health care system journey - Combines data extracted from clinical records, for example, referrals and admissions, alongside care experience–specific information inputted by service users
People in my life	- Enables service users to graphically present key people in their life, including social networks, providers, or services they are engaged with - Enables service users to visualize their social networks and explore their relationships
My plan to stay well	
Managing my warning signs	- An electronic version of Max Birchwood's early signs of psychosis approach [29] (with permission from authors) - Allows the individual to identify early warning signs and psychotic symptoms specific to their experiences - Enables discussions about warning signs of relapse and identify ways of managing these
Planning for my future	
Goals and actions	- Focuses on identifying goals, split into specific actions, to be pursued by the service user - It visually illustrates the goal and its action pathway in the form of an infographic
Quick notes	- Enables providers to use the tablet computer for making notes during the meeting with the service users, including service user–written progress notes

Evaluation of the Care Pathway Tool Pilot Implementation

As part of the CPT pilot implementation, an independent qualitative evaluation was commissioned and undertaken by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West) to investigate CPT acceptability by care providers and effectiveness in facilitating coproduced care plans and recording information more efficiently. Details about how the interactive features of the CPT supported coproduction in care planning are reported separately [27].

The aim of this paper was to present findings arising from using the Consolidated Framework for Implementation Research (CFIR) [30] to identify factors influencing adoption and use of the CPT in care delivery routine interactions.

Methods

Sampling and Interviews

In-depth interviews were conducted with (1) providers piloting the CPT to explore their views and experiences, and (2) staff with performance and management roles to explore views from a strategic level on the implementation and use of new digital technologies. The interview guide is included in Multimedia Appendix 2.

Purposeful sampling ensured that a range of roles were recruited. All interviews were conducted over the telephone and, with oral consent, audio recorded. Interviews took place between October and November 2016 and lasted between 13 and 60 min (mean 32 min).

The study was reviewed by the NHS Health Research Authority (ID: 199385) and ethically reviewed by the University of Bristol, Faculty of Health Sciences, Research Ethics Committee (Application: 29045).

Analysis

Data collection and analysis occurred in parallel. Sample size was driven by the concept of *information power*, with information within our sample continuously assessed in relation to our study objective as data collection progressed [31]. Interviews were transcribed, anonymized, and thematically analyzed [32,33] in NVivo 10 (QSR). An initial inductive coding scheme was developed and refined as new data were analyzed to understand the main themes emerging from the data. Data were coded into thematic categories representing participants' attitudes toward the CPT, positive and negative aspects of the CPT, and barriers and facilitators to implementation. MF and CP double coded a subset of transcripts, and any discrepancies were discussed and resolved. Analytical uncertainties or disagreements were discussed by the multidisciplinary research team to ensure credibility and confirmability. The CFIR was then used as a framework to order codes [30,34], in line with the CFIR Qualitative Codebook Guidelines [35] to deepen our analysis, rather than impose deductive codes on the data. The CFIR incorporates a repository of 39 standardized implementation-related constructs organized across 5 domains, which interact to influence implementation effectiveness (Multimedia Appendix 3) [34]. The 5 domains are as follows: (1) intervention characteristics: includes 8 constructs related to characteristics of the intervention being implemented; (2) outer setting: includes 4 constructs related to external factors such as the economic, political, and social context within which an organization is situated; (3) inner setting: includes 12 constructs related to features such as the structural, political, and cultural characteristics of the organization implementing the intervention; (4) characteristics of individuals: includes 5 constructs related to the individuals involved with the intervention and implementation process; and (5) process: includes 8 constructs related to essential activities of the implementation process.

Results

Participants

In total, 20 providers were interviewed (11 female). Participants included 15 (out of the 30) practitioners who piloted the CPT with service users (mental health support workers, peer support workers, psychiatrists, occupational therapists, community psychiatric nurses, and social workers) and 5 managers; 6 practitioners piloted the CPT for 6 months or more, 4 between 3 to 6 months and 5 between 6 weeks and 3 months. These practitioners provided community-based care, and contact with service users was described to be needs driven, ranging from weekly to monthly. Practitioners discussed how service users involved in piloting the tool had a range of mental health diagnoses, including psychosis, anxiety, depression, and previous experiences of trauma.

A total of 13 CFIR constructs were seen to influence the processes of CPT implementation in all 5 framework domains. The factors identified and their relationship to these constructs and domains are outlined in Multimedia Appendix 3. The 5 CFIR domains are used to structure presentation of findings with illustrative verbatim participant quotes.

Intervention Characteristics

Intervention Source

Staff’s involvement in tool development through coproduction activities influenced engagement with the pilot. Managers were aware of the need for staff to feel involved in new interventions:

> *There was [...] a big meeting with our team [...] to see what kind of solutions they could come up with to improve co-production and help our sort of work. And [...] I put my name down.* [Practitioner 03]
>
> *Another management idea coming in isn’t necessarily something they (staff) are going to embrace.* [Manager 01]

Relative Advantage

The CPT was thought to enable more efficient information recording and facilitate coproduced care. Creating notes using the CPT alongside service users facilitated transparency and involvement of service users, and it was quicker than traditional ways of working:

> *It saved me a lot of time. You don’t have to go back to the main [IT] system.* [Practitioner 05]
>
> *I’ve had feedback that service users have felt in the centre of the process. [...] more in control of their support [...] by being part of that process and by having the opportunity to use the tool.* [Manager 04]

Design Quality, Packaging, and Complexity

Some participants distinguished between the CPT’s features, which they thought were well designed, useful, and easy to use, and limitations of the tablet computer on which the CPT was hosted:

> *Part of it is to separate the tool from the piece of equipment it's on. The piece of equipment, there've been lots of problems, but the actual tool itself [CPT], the different bits of it have been really good, really easy to use.* [Practitioner 12]

Issues with internet connectivity and tablet computer log-in problems were a common barrier to using the CPT with service users:

> *I don't have confidence yet [...] that it'll work first time [...] I feel positive about the software itself but not about being able to use it when I need to in remote locations.* [Practitioner 03]

Other barriers included security limitations, and lack of *live* cross-system communication between the CPT and the host organization’s main electronic patient record (EPR) system. Until secure platforms for information exchange were developed, all data were manually transferred from the CPT system to service users’ EPRs by administrators. The delay in transfer (up to 48 hours) impacted on information available during meetings, sharing of information between providers involved in care, and ultimately how and how often the tool was used:

> *(If) the service user is a bit unstable with their mental health and you need to update a lot of other information related to the meetings [...] If the service user is at risk, immediate risk, then we might need to go on the actual system [ERP] and record it to avoid any other issues.* [Practitioner 005]

Outer Setting

The CFIR conceptualizes the outer setting as factors external to the organization [34].

External Policy and Incentives

A recent Care Quality Commission (CQC) report of the mental health organization highlighted the need to improve inclusion of service users’ views in care plans [33]. The CPT could be a key mechanism to improve practices in response to the CQC report’s comments, and Joint Project Board members saw this as an important facilitator to encourage its use within the organization.

Patient Needs and Resources

In this construct, quotes relating to awareness of service user–specific factors influencing implementation were included. Some participants were guided in their use of the tool by perceived needs and characteristics of individual service users. For example, English language literacy was taken into consideration when deciding to use the tool:

> *I support quite a lot of people whose first language isn’t English [...] so it’s not so useful in that sense.* [Practitioner 02]

Service users' levels of self-awareness and stage in their illness influenced CPT use, for example, when service users were not thought to be able or ready to engage in recovery care. Another influencing factor was service users' attitudes toward technology, with age being a related factor:

> *It depends on their level of awareness, where they are at in their recovery as well. That's quite key and to a certain extent age but not exclusively.* [Practitioner 03]
>
> *(I would want to use it with) people who are fairly articulate and in touch with how they're feeling and wanting to engage with services.* [Practitioner 04]

When introducing the tool to service users, reasons for declining its use included wanting to "talk to a human being" [Practitioner 03]; seeing it as a wall between themselves and providers; distrust toward technology; and thinking it did not enhance their care experience. Declining to use the tool once, however, did not always exclude use of the tool in subsequent sessions:

> *We did a first session and then he was like oh God I couldn't concentrate, [...] I don't really want to do it [...] And then actually recently he said, "Oh why don't I do that tool with you anymore?"* [Practitioner 12]

Inner Setting

This construct relates to characteristics of the organization implementing the intervention [34].

Implementation Climate and Culture

The organization already had a focus on recovery-oriented care, including engaging service users in care planning. Participants agreed that the CPT facilitated coproduced recovery-focused care planning by supporting novel and more user-centered conversations with a psychosocial focus:

> *I thought it would be really useful, [...] we quite often do WRAP (wellness recovery action plans) plans with people [...] But having the ability to actually sit down and work with somebody and do a process holistically together rather than it being me-led was quite nice.* [Practitioner 09]

However, some medically trained staff saw the CPT's focus on psychosocial aspects of recovery as contradictory to their professional roles, for example, discussing medications:

> *(If service users are) expecting to talk to me about their medication [...] and so I have to make sure that it (using the tool) wouldn't be [...] something that would leave them feeling dissatisfied.* [Practitioner 04]

Readiness for Implementation

Organizational IT factors were also raised. The organization was thought to lack adequate IT infrastructure, to be *paper heavy*, and not incorporating technology within current practice:

> *It's still a very paper heavy mindset [...] there's a cultural shift that needs to happen for them to fully get on board with another bit of IT equipment.* [Practitioner 03]

The organization and staff were under pressure from increasing numbers of referrals. This provided an incentive for innovations that would compensate for the lack of capacity:

> *Services got very, very swamped with huge numbers of referrals coming in [...] there not being sufficient capacity to manage that.* [Manager 01]

Pressures on the organization impacted on the time and resources staff had available to learn how to use and implement the CPT. Carrying a heavy caseload also shaped capacity for recovery-focused work, as staff had to address the service user's immediate needs. This influenced perceptions of how the tool could support meetings with service users:

> *It's very hard to have that time before sessions to thoroughly think it through and plan it. You kind of like rushing to an appointment.* [Practitioner 06]
>
> *we're short of money and we're short of time, therefore we're short of people. I spend much less time face-to-face supporting people than I would like to and I think [...] that will impact on the use of the tool.* [Practitioner 07]

Characteristics of Individuals

This CFIR construct includes features of individuals involved in the intervention [34].

Knowledge and Beliefs

Most participants were positive about the CPT's ability to aid in coproducing care plans, stating that they "liked the idea of doing a support plan on a device (that was) client facing" [Practitioner 02]. Some participants used most or all of the tool features with different service users, whereas others believed the tool could only be used in certain contexts, for example, guided by service users' needs. There was reticence among some providers to introduce a tablet computer to the therapeutic relationship:

> *I believe in a real connection between people and a connection in the room and that, to me, comes from face-to-face and eye contact and us sitting opposite each other almost and me being really attentive to the other person. So I think any device is going to take away from that.* [Practitioner 13]

Self-Efficacy

Issues of confidence and perceived ability to use the CPT were raised by both experienced users of the tool, that is, those trained 6 to 8 months before the interview, as well as inexperienced users, that is, participants trained around 2 months prior. Familiarity with the tool was needed when deciding who to use the CPT with, which components to use, how, and when. It also related to how comfortable staff were in changing practice and introducing new ways of working, potentially influencing the therapeutic relationship:

> *I've got a way of (working with service users) and a process that I go through probably sort of subconsciously or not really thinking about it, it's just kind of what you do. So changing that is always a bit challenging.* [Practitioner 08]

Implementation Process

Planning

Staff recruitment to the pilot was led by practitioners from inside the organization, which facilitated getting staff on board. Having continuing user feedback allowed the software developers to make improvements to the CPT software. This meant that some tablet computers needed to be replaced to support updated versions of the CPT, but it also meant that some participants received new tablet computers toward the end of the pilot phase, not allowing enough time to use the new tool in their practice. Restrictions placed by the organization on which service users could be involved in the pilot, that is, individuals who were assessed to be at low/medium risk for experiencing a mental health crisis, also acted as a barrier:

> *It has been quite difficult to use the tool because of the level of risks I'm working with. We are only supposed to use the tool with people whose recorded risk level is medium or low. And the nature of my job means that most people would have a recorded risk level of high.* [Practitioner 07]

Some participants thought risk should not be a limitation and the CPT could potentially be used with high-risk individuals if enough consideration was given to which aspects of the tool were used, and how it could be used during interactions:

> *I don't think risk per se would stop me using it with someone because that's the point of it, to help people who potentially are struggling.* [Practitioner 12]

Engaging

Training provided during implementation helped support participants using the tool. Training included group or one-to-one sessions followed by feedback meetings, but some participants recruited later in the pilot did not always receive similar training:

> *I didn't have the training really I had, like half an hour.* [Practitioner 13]

One participant thought training on the interactive element of the tool was needed to guide integration into practice in a way that does not compromise the therapeutic relationship. Such aspects of using the tool were seen as important because of its objective to support collaborative service provision, coproduced with service users:

> *(the training was) very functional, what the functions are, how you log in, so it wasn't at all about the human element or the relation element.* [Practitioners 13]

Information leaflets explaining what the tool is about and how to introduce the project to service users were thought to be particularly useful.

> *We had leaflets given to us to introduce the tool, which I have to say were really good, because it gave it, it had a bit of a talking point with the service users.* [Practitioners 02]

Discussion

This study used the CFIR to evaluate the implementation of a digital CPT in a mental health care community setting. Findings contribute to the evidence base by first adding to our understanding of organizational and contextual factors, as well as individual ones, involved when implementing digital health technologies in mental health settings [14]; and their use within the therapeutic relationship [13]. Second, it provides evidence on the experience of using the CFIR to explore implementation barriers and facilitators, adding to ongoing discussions on its use in health technology implementation research [34,35].

Principal Findings and Directions for Future Research

Factors Impacting on Implementation

Aref-Adib et al [14] state the need for better understanding of the contextual and organizational determinants of successful implementation. Our findings highlight the importance of ensuring alignment between external policy, organization and staff priorities, and CPT features to aid intervention engagement. Externally, the CQC inspection findings and increased demand for mental health services provided an incentive for organizational change in ways of working. The CPT facilitated changing practice in ways that met these pressures while aligning with organizational and professional values.

Other highlighted factors include tool adaptability to existing ways of working and attitudes and beliefs toward the digital innovation; stakeholder involvement in the development process is recommended to address these factors and facilitate implementation [14]. In this study, coproduction principles [20] in tool design and development supported engagement in implementation, and the tool was thought to facilitate service user–centered, recovery-focused coproduced care [27], a professional and organizational priority. Uptake was impacted by available organization resources, including IT infrastructure, staff caseloads, and time pressures; staff self-efficacy and knowledge of using the tool; service user attributes; and mobile working–related factors, including internet connectivity and IT system compatibility with the CPT. Findings reiterate the importance of considering such issues early on in digital innovation design [14,36].

Importance attached by providers to the therapeutic relationship when adopting digital health technologies highlights the need to better understand interpersonal aspects of health technologies in clinical contexts [13,18,27,37]. Perceptions of impact on, and concerns about, the therapeutic relationship influenced whether and how the CPT was used, and so did perceptions of relevance to role priorities. Staff assessments of service user needs and characteristics, for example, crisis management, literacy or attitudes toward technology, and uncertainty as to how to most successfully integrate the tool in practice, influenced providers' choice of who to use the tool with, and how often.

Mental health service users are open to using digital health technologies [5,38], but barriers, such as, intervention complexity [14], can prevent widespread access and use among individuals with increased needs, for example, those experiencing psychotic symptoms or learning difficulties. At the same time, low IT and health literacy and digital inequalities among individuals with mental illness also impact on innovation uptake [14,39-41]. Our findings suggest in some cases, staff's perceptions of service user characteristics, for example, literacy skills, may result in some service users being excluded from interactive health technologies in a care setting, but more research is needed to better understand staff decision making on this aspect, including differences in perspectives between providers from different professional backgrounds. Views of service users from underrepresented groups at risk of digital inequalities should also be explored [41].

Training and access to continuing support on technical and interactive aspects of the intervention during and after implementation may enhance efforts to integrate technology into routine practice [14,42,43]. Training that approaches use of digital innovations in a more reflexive and critical way [26] might address skepticism toward the innovation and concerns over its impact on the therapeutic relationship. When planning such activities, workload pressures and time available for staff to attend training need to be considered [14]. Training informed by action research can result in changing practice [44], and its usefulness in promoting acceptance of health technologies should be explored [45], especially because of its philosophical similarities with principles of coproduction [20].

Consolidated Framework for Implementation Research Methodological Considerations

Theoretical grounding of implementation research allows for conclusions to be drawn as to the relevance of findings to other settings and contexts, allows comparisons, and guides further research [34]. In our study, the CFIR informed data analysis and identified factors shaping intervention implementation, following examples supporting its use in qualitative research evaluations [30,34,35]. The CFIR was useful in guiding categorization of factors and capturing overarching implementation factors involved in the CPT's uptake. Using the framework in the analysis stage presented challenges in assigning data items to individual dimensions or constructs, and identifying which ones were the most salient in our data, an issue already raised in the literature [10]. In our case, challenges reflected the unique nature of the CPT as a tool used by both providers and service users simultaneously in the context of a care meeting.

Analytical ambiguity existed between the categories *Patient needs and resources* and *Individual characteristics* when coding service user–related factors, as both were end users of the tool in the same context and setting; the tool's dynamic and interactive nature also made difficult distinctions between, for example, the intervention's *adaptability*, *complexity*, and *design quality* when categorizing data items. With more treatment interventions provided in an interactive way through mobile technology, it is essential for frameworks such as the CFIR to capture the dynamic and multidimensional nature of technological interventions [10]. The complex nature of such interventions, its impact on implementation, and the CFIR's ability to adapt to and capture this complexity should be further explored in future evaluations of such mobile mental health care interventions.

Varsi et al [46] discussed the broadness of the CFIR that can restrict one study's ability to capture the *big picture* represented in the framework, without explicitly addressing all dimensions during data collection [34]. This can be because of time limitations restricting researchers' ability to explore all constructs within a single interview, but also recruitment limitations, when stakeholders that might represent different views are not included in the sample [46]. One way our study tackled this was to interview both practitioners using the CPT and senior managers who had a broader perspective on implementing technological innovations.

Strengths, Limitations, and Conclusions

One limitation of the study is that participants were sampled from those that had volunteered to take part in the pilot; these may have been providers who were more enthusiastic about using technology. Strengths include recruitment of a diverse participant sample in terms of professional roles to enable a comprehensive insight into the CPT implementation. There was often consensus in the views expressed across professional roles providing confidence of the attainment of information power.

Findings highlight the value of congruence between staff, organization, and external policy priorities and digital technologies to aid intervention engagement. Only a handful of health technologies have addressed mental health recovery in community settings [10,17,18], although there is a need for health technology design interventions that follow principles of coproduction to address needs and capabilities of both staff and service users [14,16,20]. Integrating training alongside and after health technology implementation might be a way to address some of the challenges identified. Integrating an action research component within health technology implementation efforts could help to identify early training needs to address uncertainty and lack of confidence in adopting innovations, support reflexive practice, and promote effective practice change. The crucial role played by perceived impact of the technology on the therapeutic relationship highlights the importance of better understanding the ways digital health technologies impact on the therapeutic process as well as on outcomes [13].

Acknowledgments

The study was jointly funded by OHS and the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West) at University Hospitals Bristol NHS Foundation Trust and conducted in collaboration with AWP. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, and the Department of Health and Social Care. The authors thank Sam Kozak for her co-ordination and communication role within the research. They also thank the service users involved in developing the interview topic guide and all interviewees taking part in the study.

Authors' Contributions

JH, SS, MF, CP, and HE were responsible for study design and collection of data. JH, SS, WH, CG, NH, and JW were responsible for study management and co-ordination. MF and CP analyzed the data. CP drafted the paper. All authors read, commented on, and approved the final manuscript.

Conflicts of Interest

OHS were paid under contract by AWP to design and develop digital solutions including the CPT described in the publication from April 2015 to December 2016. OHS paid the NIHR CLAHRC West a fee for the evaluation project which represented 50% of the costs of the project. OHS own the software referred to in this publication as the Care Pathway Tool.

Multimedia Appendix 1
Care Pathway Tool Screenshots. (a) Managing my warning signs Step 2: Choose cards representing service user experiences. (b) Managing my warning signs Step 3: Sorting cards into a timeline. (c) Goals and Actions.
[PPTX File , 384 KB - jmir_v22i3e14868_app1.pptx]

Multimedia Appendix 2
Interview topic guide.
[DOCX File , 74 KB - jmir_v22i3e14868_app2.docx]

Multimedia Appendix 3
Consolidated Framework for Implementation Research–linked factors impacting on implementation.
[DOCX File , 27 KB - jmir_v22i3e14868_app3.docx]

References

1. Limb M. Digital technologies offer new ways to tackle mental health problems. Br Med J 2012 Jul 30;345:e5163. [doi: 10.1136/bmj.e5163] [Medline: 22846418]
2. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
3. Palmier-Claus JE, Rogers A, Ainsworth J, Machin M, Barrowclough C, Laverty L, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. BMC Psychiatry 2013 Jan 23;13:34 [FREE Full text] [doi: 10.1186/1471-244X-13-34] [Medline: 23343329]
4. Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, McGorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
5. Ennis L, Robotham D, Denis M, Pandit N, Newton D, Rose D, et al. Collaborative development of an electronic Personal Health Record for people with severe and enduring mental health problems. BMC Psychiatry 2014 Nov 18;14:305 [FREE Full text] [doi: 10.1186/s12888-014-0305-9] [Medline: 25403285]
6. Goodwin J, Cummins J, Behan L, O'Brien SM. Development of a mental health smartphone app: perspectives of mental health service users. J Ment Health 2016 Oct;25(5):434-440. [doi: 10.3109/09638237.2015.1124392] [Medline: 26732242]
7. Ben-Zeev D, Davis KE, Kaiser S, Krzsos I, Drake RE. Mobile technologies among people with serious mental illness: opportunities for future services. Adm Policy Ment Health 2013 Jul;40(4):340-343 [FREE Full text] [doi: 10.1007/s10488-012-0424-x] [Medline: 22648635]
8. 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]
9. Robotham D, Mayhew M, Rose D, Wykes T. Electronic personal health records for people with severe mental illness; a feasibility study. BMC Psychiatry 2015 Aug 6;15:192 [FREE Full text] [doi: 10.1186/s12888-015-0558-y] [Medline: 26245805]
10. Lord S, Moore SK, Ramsey A, Dinauer S, Johnson K. Implementation of a substance use recovery support mobile phone app in community settings: qualitative study of clinician and staff perspectives of facilitators and barriers. JMIR Ment Health 2016 Jun 28;3(2):e24 [FREE Full text] [doi: 10.2196/mental.4927] [Medline: 27352884]
11. Doran D, Paterson J, Clark C, Srivastava R, Goering PN, Kushniruk AW, et al. A pilot study of an electronic interprofessional evidence-based care planning tool for clients with mental health problems and addictions. Worldviews Evid Based Nurs 2010 Sep;7(3):174-184. [doi: 10.1111/j.1741-6787.2010.00191.x] [Medline: 20367805]

12. Clarke A, Adamson J, Sheard L, Cairns P, Watt I, Wright J. Implementing electronic patient record systems (EPRs) into England's acute, mental health and community care trusts: a mixed methods study. BMC Med Inform Decis Mak 2015 Oct 14;15:85 [FREE Full text] [doi: 10.1186/s12911-015-0204-0] [Medline: 26466894]
13. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]
14. 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]
15. Maguire D, Evans H, Honeyman M, Omojomolo D. Digital Change in Health and Social Care. London, UK: The King's Fund; 2018.
16. Strand M, Gammon D, Ruland CM. Transitions from biomedical to recovery-oriented practices in mental health: a scoping review to explore the role of Internet-based interventions. BMC Health Serv Res 2017 Apr 7;17(1):257 [FREE Full text] [doi: 10.1186/s12913-017-2176-5] [Medline: 28388907]
17. Strand M, Gammon D, Eng LS, Ruland C. Exploring working relationships in mental health care via an e-recovery portal: qualitative study on the experiences of service users and health providers. JMIR Ment Health 2017 Nov 14;4(4):e54 [FREE Full text] [doi: 10.2196/mental.8491] [Medline: 29138127]
18. Gammon D, Strand M, Eng LS, Børøsund E, Varsi C, Ruland C. Shifting practices toward recovery-oriented care through an e-recovery portal in community mental health care: a mixed-methods exploratory study. J Med Internet Res 2017 May 2;19(5):e145 [FREE Full text] [doi: 10.2196/jmir.7524] [Medline: 28465277]
19. Clark M. Co-production in mental health care. Ment Health Rev J 2015;20(4):213-219. [doi: 10.1108/MHRJ-10-2015-0030]
20. Penny J, Slay J, Stephens L. Nesta | The Innovation Foundation. London, UK: Nesta and New Economics Foundation; 2012. People Powered Health Co-Production Catalogue URL: https://media.nesta.org.uk/documents/co-production_catalogue.pdf [accessed 2019-07-25]
21. Realpe A, Wallace LM. Warwick WRAP. London, United Kingdom: The Health Foundation; 2010. What is Co-Production? URL: http://wrap.warwick.ac.uk/61814/ [accessed 2020-01-02]
22. Simpson A, Hannigan B, Coffey M, Barlow S, Cohen R, Jones A, et al. Recovery-focused care planning and coordination in England and Wales: a cross-national mixed methods comparative case study. BMC Psychiatry 2016 May 16;16:147 [FREE Full text] [doi: 10.1186/s12888-016-0858-x] [Medline: 27184888]
23. Bee P, Price O, Baker J, Lovell K. Systematic synthesis of barriers and facilitators to service user-led care planning. Br J Psychiatry 2015 Aug;207(2):104-114 [FREE Full text] [doi: 10.1192/bjp.bp.114.152447] [Medline: 26243762]
24. Grundy AC, Bee P, Meade O, Callaghan P, Beatty S, Olleveant N, et al. Bringing meaning to user involvement in mental health care planning: a qualitative exploration of service user perspectives. J Psychiatr Ment Health Nurs 2016 Feb;23(1):12-21. [doi: 10.1111/jpm.12275] [Medline: 26634415]
25. Simpson A, Hannigan B, Coffey M, Jones A, Barlow S, Cohen R, et al. Cross-national comparative mixed-methods case study of recovery-focused mental health care planning and co-ordination: Collaborative Care Planning Project (COCAPP). Health Serv Delivery Res 2016;4(5):1-190 [FREE Full text] [doi: 10.3310/hsdr04050] [Medline: 26866206]
26. Grundy AC, Walker L, Meade O, Fraser C, Cree L, Bee P, et al. Evaluation of a co-delivered training package for community mental health professionals on service user- and carer-involved care planning. J Psychiatr Ment Health Nurs 2017 Aug;24(6):358-366 [FREE Full text] [doi: 10.1111/jpm.12378] [Medline: 28218977]
27. Farr M, Pithara C, Sullivan S, Edwards H, Hall W, Gadd C, et al. Pilot implementation of co-designed software for co-production in mental health care planning: a qualitative evaluation of staff perspectives. J Ment Health 2019 Oct;28(5):495-504. [doi: 10.1080/09638237.2019.1608925] [Medline: 31240971]
28. Rethink Mental Illness. YouTube. UK: YouTube; 2017 Aug 24. Co-Production in Mental Health URL: https://www.youtube.com/watch?v=rg0lDuoUonY&feature=youtu.be [accessed 2020-01-02]
29. Birchwood M, Spencer E, McGovern D. Schizophrenia: early warning signs. Adv Psychiatr Treat 2000 Mar;6(2):93-101. [doi: 10.1192/apt.6.2.93]
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 Aug 7;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
31. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
32. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. California, USA: Sage Publications, Inc; 1998.
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the consolidated framework for implementation research. Implement Sci 2016 May 17;11:72 [FREE Full text] [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]

35. The Consolidated Framework for Implementation Research. Michigan: CFIR Research Team; 2019. URL: https://cfirguide.org/ [accessed 2019-03-08] [WebCite Cache ID 76hTKZAAm]
36. Gammon D, Strand M, Eng LS. Service users' perspectives in the design of an online tool for assisted self-help in mental health: a case study of implications. Int J Ment Health Syst 2014 Jan 9;8(1):2 [FREE Full text] [doi: 10.1186/1752-4458-8-2] [Medline: 24405765]
37. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018 Mar 19;8(3):e019966 [FREE Full text] [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]
38. Ennis L, Rose D, Denis M, Pandit N, Wykes T. Can't surf, won't surf: the digital divide in mental health. J Ment Health 2012 Aug;21(4):395-403 [FREE Full text] [doi: 10.3109/09638237.2012.689437] [Medline: 22712756]
39. Clausen W, Watanabe-Galloway S, Baerentzen MB, Britigan DH. Health literacy among people with serious mental illness. Community Ment Health J 2016 May;52(4):399-405. [doi: 10.1007/s10597-015-9951-8] [Medline: 26443671]
40. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
41. Greer B, Robotham D, Simblett S, Curtis H, Griffiths H, Wykes T. Digital exclusion among mental health service users: qualitative investigation. J Med Internet Res 2019 Jan 9;21(1):e11696 [FREE Full text] [doi: 10.2196/11696] [Medline: 30626564]
42. Bredfeldt CE, Awad EB, Joseph K, Snyder MH. Training providers: beyond the basics of electronic health records. BMC Health Serv Res 2013 Dec 2;13:503 [FREE Full text] [doi: 10.1186/1472-6963-13-503] [Medline: 24295150]
43. 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]
44. Lindeman MA, Black K, Smith R, Gough J, Bryce A, Gilsenan B, et al. Changing practice in residential aged care using participatory methods. Educ Health (Abingdon) 2003 Mar;16(1):22-31. [doi: 10.1080/1357628031000066651] [Medline: 14741920]
45. Vabo G, Slettebø Å, Fossum M. Participants' perceptions of an intervention implemented in an Action Research Nursing Documentation Project. J Clin Nurs 2017 Apr;26(7-8):983-993. [doi: 10.1111/jocn.13389] [Medline: 27192412]
46. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]

Abbreviations

AWP: Avon and Wiltshire Mental Health Partnership NHS Trust
CFIR: Consolidated Framework for Implementation Research
CPT: Care Pathway Tool
CQC: Care Quality Commission
EPR: electronic patient record
IT: information technology
NHS: National Health Service
NIHR CLAHRC West: National Institute for Health Research, Collaborations for Leadership in Applied Health Research and Care West
OHS: Otsuka Health Solutions

Edited by M Focsa; submitted 05.06.19; peer-reviewed by C Varsi, K Machin; comments to author 24.06.19; revised version received 13.08.19; accepted 21.08.19; published 19.03.20.

Please cite as:
Pithara C, Farr M, Sullivan SA, Edwards HB, Hall W, Gadd C, Walker J, Hebden N, Horwood J
Implementing a Digital Tool to Support Shared Care Planning in Community-Based Mental Health Services: Qualitative Evaluation
J Med Internet Res 2020;22(3):e14868
URL: https://www.jmir.org/2020/3/e14868
doi:10.2196/14868
PMID:32191210

Review

Dyadic Psychosocial eHealth Interventions: Systematic Scoping Review

Kelly M Shaffer[1,2], PhD; Ashley Tigershtrom[2], BA; Hoda Badr[3], PhD; Stephanie Benvengo[2], MA; Marisol Hernandez[2,4], MLS, MA; Lee M Ritterband[1], PhD

[1]Center for Behavioral Health and Technology, University of Virginia School of Medicine, Charlottesville, VA, United States
[2]Memorial Sloan Kettering Cancer Center, New York, NY, United States
[3]Baylor College of Medicine, Houston, TX, United States
[4]CUNY School of Medicine/City College of New York, New York, NY, United States

Corresponding Author:
Kelly M Shaffer, PhD
Center for Behavioral Health and Technology
University of Virginia School of Medicine
PO Box 801075
Charlottesville, VA, 22908
United States
Phone: 1 4349821022
Email: kshaffer@virginia.edu

Abstract

Background: Dyadic psychosocial interventions have been found beneficial both for people coping with mental or physical health conditions as well as their family members and friends who provide them with support. Delivering these interventions via electronic health (eHealth) may help increase their scalability.

Objective: This scoping review aimed to provide the first comprehensive overview of dyadic eHealth interventions for individuals of all ages affected by mental or physical illness and their family members or friends who support them. The goal was to understand how dyadic eHealth interventions have been used and to highlight areas of research needed to advance dyadic eHealth intervention development and dissemination.

Methods: A comprehensive electronic literature search of PubMed, EMBASE, Cochrane, Cumulative Index to Nursing and Allied Health Literature, and PsycINFO was conducted for articles published in the English language through March 2019. Eligible records described a psychosocial eHealth intervention that intervened with both care recipients and their support person.

Results: A total of 7113 records were reviewed of which 101 met eligibility criteria. There were 52 unique dyadic eHealth interventions identified, which were tested across 73 different trials. Of the unique interventions, 33 were conducted among dyads of children and their supporting parent, 1 was conducted with an adolescent-young adult care recipient population, and the remaining 18 were conducted among adult dyads. Interventions targeting pediatric dyads most commonly addressed a mental health condition (n=10); interventions targeting adult dyads most commonly addressed cancer (n=9). More than three-fourths of interventions (n=40) required some human support from research staff or clinicians. Most studies (n=64) specified one or more primary outcomes for care recipients, whereas less than one-fourth (n=22) specified primary outcomes for support persons. Where specified, primary outcomes were most commonly self-reported psychosocial or health factors for both care recipients (n=43) and support persons (n=18). Results of the dyadic eHealth intervention tended to be positive for care recipients, but evidence of effects for support persons was limited because of few studies specifying primary outcomes for supporters. Trials of dyadic eHealth interventions were most commonly randomized controlled trials (RCTs; n=44), and RCTs most commonly compared the dyadic eHealth intervention to usual care alone (n=22).

Conclusions: This first comprehensive review of dyadic eHealth interventions demonstrates that there is substantial, diverse, and growing literature supporting this interventional approach. However, several significant gaps were identified. Few studies were designed to evaluate the unique effects of dyadic interventions relative to individual interventions. There was also limited assessment and reporting of outcomes for support persons, and there were no interventions meeting our eligibility criteria specifically targeting the needs of older adult dyads. Findings highlight areas of research opportunities for developing dyadic eHealth interventions for novel populations and for increasing access to dyadic care.

(J Med Internet Res 2020;22(3):e15509) doi:10.2196/15509

KEYWORDS

behavioral medicine; caregivers; couples therapy; dyads; eHealth; family therapy; internet interventions; psychosocial interventions; review

Introduction

Background

Psychosocial interventions provide evidence-based behavioral, cognitive, and emotion regulation strategies to address mental and physical health conditions. These conditions not only affect the person diagnosed with illness but also their close family members and friends, particularly those who support the clinical care of the person with illness. Dyadic interventions that intervene with both the care recipient and their support person together have been found clinically useful [1-5]; however, most have been designed to be delivered in-person, limiting their ultimate scalability and accessibility [6]. Delivering dyadic interventions via electronic health (eHealth; ie, using information communication technologies to facilitate health care provision) may increase their uptake by overcoming structural and financial access barriers. A broad review of the literature on dyadic eHealth interventions is needed to better understand how these interventions have been used and to identify where further research is needed.

Illness Affects Dyads Together

It is estimated that 43.5 million Americans provide informal, unpaid care to one or more adults or children with serious health conditions [7]. Of these, it is estimated that about 33.3 million provide such care to adult recipients only, 3.7 million to child recipients only, and 6.5 million to both adult and child recipients. About 15% of people who provide care are not legally defined family members of the care recipient (eg, friend or neighbor [7]), and many individuals do not identify with the term *caregiver* (eg, [8]). As such, the term *support persons* is used for this scoping review to broadly capture individuals who provide emotional and practical support to help a care recipient. Together, the support person and the care recipient comprise a *dyad*.

Support from family members and friends to seriously ill individuals is critical to the sustainability of the US health care system [9], yet it can place significant strain on these support persons. Compared with the general population, support persons report worse diet, exercise, and sleep [10-12]; worse symptoms of depression and anxiety [13-15]; and premature physical health decline [16-18]. These mental and physical burdens from caregiving also ultimately affect the care recipients. Distressed support persons are more likely to exhibit harmful caregiving behavior [19] and less able to meet the practical and social-emotional needs of the care recipient [20,21]. When one dyad member is distressed, the other is more likely to become distressed as well [22-25]. Care recipients and their support persons experience illness together, and the success of one person's ability to cope with illness stressors affects the other's [26-28].

Intervening With Dyads

Dyadic interventions may use a range of strategies, such as communication skills training, cognitive behavioral therapy, education, and problem-solving training. These interventions share the commonality of including both the care recipient and their support person together within a singular program of care. Dyadic psychosocial interventions have been found effective to improve targeted outcomes for both care recipients and support persons [1-5,29,30]. As coping is interdependent between care recipients and support persons, there is promise that dyadic interventions may deliver synergistic benefits—meaning the cumulative benefits to both individuals from a dyadic intervention is greater than the sum of benefits of intervening with each member individually. In practice, there has been limited empirical study of such effects [1,30].

Even if there are no such synergistic effects, dyadic interventions may be an attractive way to extend psychosocial care access to support persons. Interventions specifically addressing support persons' informational and psychosocial needs have rarely been implemented in health care settings—common institutional barriers include competing clinical demands and lack of funding [31]. Support persons tend to underutilize available caregiver-directed programs, in part because of a reluctance to *put their needs ahead* of those of the care recipient [32,33]. Dyadic interventions may therefore be perceived as more justifiable by both health care systems and support persons: the care recipient is a target of care, and support persons receive care as well.

Despite the promise of dyadic interventions, there are significant logistical and financial barriers impeding their dissemination [6]. Common practical barriers like scheduling difficulties and limited time [32-36] are compounded when both care recipients' and support persons' schedules must align. In addition, support persons often have inadequate health care coverage [37], meaning that obtaining such care is likely to be cost-prohibitive for many dyads. Disseminating dyadic interventions via the internet and other technologies may lower the barrier to entry and increase their affordability [38]. The internet fills an important gap in meeting health information needs [39], and this is particularly true for support persons, who are more likely than noncaregivers to seek health information on the Web [40]. Internet- and technology-based approaches are also more scalable from a health systems perspective: reduced labor costs, as well as minimal ongoing program costs, suggest the long-term cost-efficacy of eHealth interventions relative to traditional face-to-face care [41].

Purpose of This Review

The aims of this scoping review were twofold. The first aim was to provide the first summary of available evidence on dyadic eHealth interventions that provide behavioral treatment and/or psychosocial support to care recipients of all ages affected by any mental or physical illness together with their primary support persons. The second aim was to identify gaps in this knowledge base. To date, there has been one pioneering review of dyadic eHealth programs among cancer survivors and their family caregivers [29]. There have been many reviews demonstrating the efficacy and acceptability of eHealth interventions among diverse populations of patients (eg, [42-55]) and support persons (eg, [56-68]). All of these reviews have summarized literature within certain disease or age group populations, which silos the literature. In contrast, this review is intentionally broad, summarizing the literature across care recipient health conditions (eg, mental health, obesity, and cancer), population subgroups (eg, pediatric and adult care recipients), and intervention strategies (eg, communication training and cognitive behavioral therapy). Aims were well suited to a scoping review, which are ideal for reviewing a large and complex body of research that has not been previously reviewed [69,70]. The specific research questions that guided this review were as follows:

1. What populations have been targeted by dyadic eHealth interventions?
2. How have information communication technologies been used in dyadic eHealth interventions?
3. What approaches are used for intervening with both dyad members?

Methods

A comprehensive electronic literature search for articles in the English language and for all available dates was conducted in the following databases: PubMed, EMBASE via the Elsevier platform, Cochrane via the Wiley platform, Cumulative Index for Nursing and Allied Health Literature via EBSCO and PsycINFO via the OVID platform.

Eligibility Criteria

Population, Intervention, Comparison, Outcome, Study design criteria (PICOS [71]) that guided study selection and organization of data extraction for this systematic scoping review are detailed below.

Population

Eligible studies enrolled dyads of care recipients and support persons. The *care recipient* is a person who has an identified mental, behavioral, or physical health condition. The *support person* provides informal, unpaid care to the care recipient. Dyads are defined by existing, personal relationships like kinship or friendship, meaning that dyads of care recipients with a formal, trained health care provider (eg, their physician or an assigned trained peer mentor) were excluded. There could be more than one support person in a dyad—for example, a child with illness participating with both parents could be a *dyad*. There was no restriction on age of the dyad members, provided that the individual could consent or assent to participate.

Intervention

There were 3 intervention-related inclusion criteria that defined eligible dyadic eHealth interventions. First, an intervention must have intervened with both members of the dyad. Second, drawing from prevailing definitions of eHealth [72,73] and internet interventions [74], an intervention must have utilized information communication technology (including, but not limited to, the internet), with at least some intervention content delivered fully automated. Owing to this criterion, interventions that exclusively utilized technology to deliver standard clinician-provided care (eg, a therapist providing face-to-face therapy via video chat) were excluded. Third, an intervention must have provided cognitive, behavioral, educational, and/or supportive care with the primary goal to address symptom management and/or coping with the care recipients' targeted health condition. Owing to this criterion, couples therapy that intervenes with the primary goal of improving the couple's relationship was excluded.

Outcome

Records were included if they reported any objective or self-report psychosocial, health, and/or feasibility outcome for the care recipient and/or the support person. Outcomes were specified as *primary* if they were explicitly labeled as such in the record or hypotheses were explicitly specified about the outcome. Outcomes were specified as (1) an *objective* psychosocial or health measure, such as data derived via lab test (eg, hemoglobin A_{1c}), diagnostic interview (eg, structured clinical interview for Diagnostic and Statistical Manual of Mental Disorders-5), or chart review (eg, documented family meeting); (2) a *self-report* psychosocial or health measure, such as data derived via self-reported questionnaire; or (3) a *feasibility/usability* measure, which may have been objective (eg, website logins) or self-reported (eg, satisfaction).

Study Design and Comparators

All trial designs were eligible for inclusion, including single-arm trials, feasibility trials, patient preference trials (ie, care recipient could enroll in an intervention with or without a support person), and randomized controlled trials (RCTs). There was no restriction on the type of comparison condition in RCTs. For the purposes of this review, waitlist control conditions are included under the umbrella of *usual care*. All study analytic designs were eligible for inclusion, including records that reported long-term follow-ups and secondary analyses of trial data. Records that exclusively discussed intervention development, but did not report testing of the intervention as it was intended to be used, were excluded.

Search Methodology for Identification of Studies

The search was initially conducted by MH in January 2018 and an update was conducted in March 2019. Three broad concept categories were searched (dyads, eHealth, and psychosocial intervention), and results were combined using the appropriate Boolean operators (AND, OR). See Table 1 for search strategy. Potentially eligible records were also identified through other sources, such as via review of references of included records, communication with record authors, and a hand search of Journal of Medical Internet Research publications.

Table 1. Search strategy.

Concept category (combined with AND)	Search terms (combined with OR)
Dyads	dyad, dyads, dyadic, couple, couples, spouse, spouses, “Spouse”[MeSH], partner, partners, “Sexual Partners”[MeSH], parent, parents, “Parents” [MeSH], parental, guardian, guardians, “Legal Guardians”[MeSH], child, “Child”[MeSH], children, kid, kids, adolescent, adolescents, “Adolescent”[MeSH], teen, teens, teenager, teenagers family, families, “Family”[MeSH], informal caregiver, caregiver, “Caregivers”[MeSH], carer, carers
eHealth	Internet, “Internet”[MeSH], cyberspace, web, web-based, ehealth, e-health, “Telemedicine”[MeSH], mobile health, mhealth, m-health, social media, “Social Media”[MeSH], blog, blogs, mobile app, mobile application, User-Computer Interface, website, webpage
Psychosocial intervention	“Behavioral Medicine”[MeSH], psychosocial intervention, behavior therapy, “Behavior Therapy”[MeSH], cognitive therapy, “Cognitive Behavioral Therapy”[MeSH], couples therapy, “Couples Therapy”[MeSH], family therapy, “Family Therapy”[MeSH], psychoeducation, psycho-education psychoeducational, psycho-educational, “Psychology, Medical”[MeSH], “Psychology, Clinical”[MeSH]

Study Selection

After removal of duplicate articles, study titles and abstracts were scanned by 2 of 3 coders (KS, AT, and SB) to determine whether the study met eligibility criteria. Discrepancies between coders were reviewed at a consensus meeting of all 3 coders. AT and SB retrieved full-text articles for citations that initially met the eligibility criteria. Full-text articles were read by KS, AT, or SB to make a final determination of eligibility, with ongoing coding questions reviewed at regular coder consensus meetings. Title/abstract and full-text coding were conducted using Covidence systematic review software (Veritas Health Innovation, Australia) [75]. Coders were not blind to journals or study authors during screening.

Reasons for article exclusion during the full-text review were recorded. Exclusion criteria based on PICOS criteria were (1) Not dyadic—an intervention did not intervene with a care recipient and informal support person to address the care recipient’s health condition; (2) intervention development, protocol, and/or no psychosocial, health, or feasibility outcome reported; (3) no automated components; (4) prevention study—neither dyad member had an indicated health condition at the time of enrollment (ie, there was no *care recipient*). There was 1 conference abstract that passed title/abstract screening based on review of its title, but the full abstract text was not retrievable from databases, the conference organization, or the authors. Authors indicated that abstract results were reported in a journal article of the same name, which was included in full-text coding.

Data Extraction

Coders (AT and a trained research assistant) extracted data from the records that met eligibility criteria. Data were collected using a standardized, predefined charting form via Qualtrics. KS checked all extractions after completion by the coders. Discrepancies were resolved by tertiary review by SB, who discussed findings with KS. Articles were not blinded during data extraction. Corresponding authors for all articles were contacted and asked to review data extracted from their articles. Data were then summarized, and themes were reviewed.

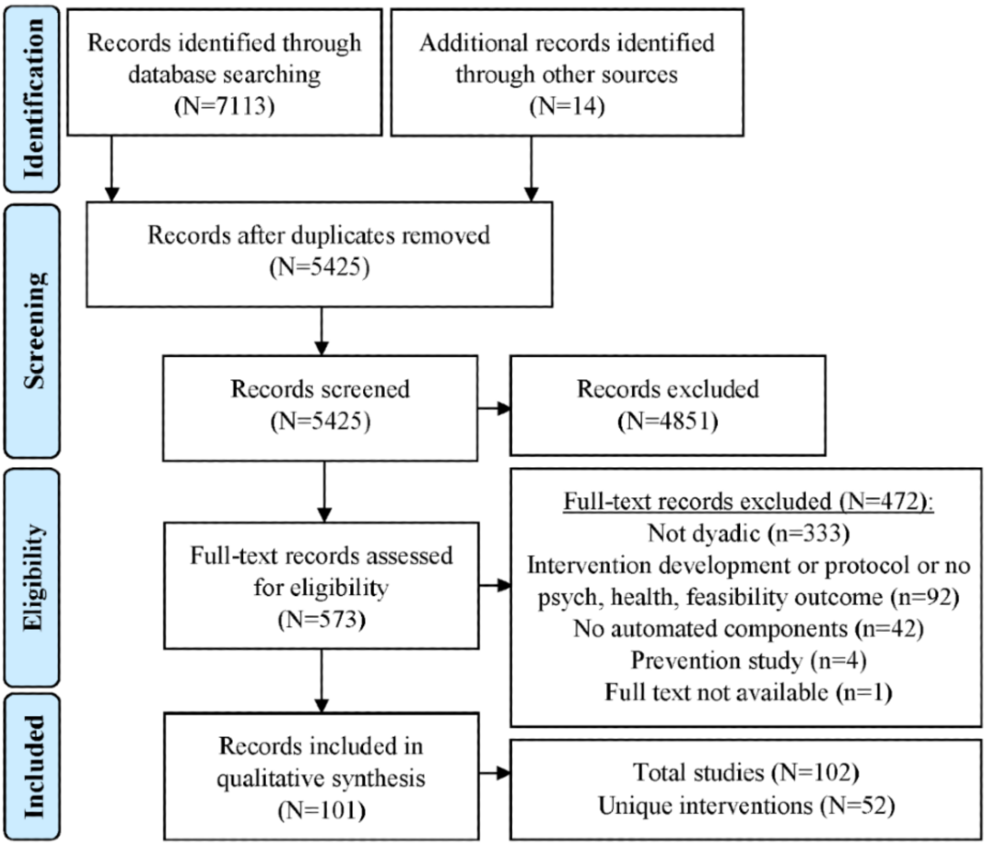
Results

Overview

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses [76] flow diagram is presented in Figure 1. In total, 7113 records were identified via the search terms, and 14 were identified through other sources (including email responses from contacted authors), of which 573 were retrieved and reviewed for full-text coding. Of these, 101 records met eligibility criteria and were included in the qualitative synthesis. Data extraction is represented in Multimedia Appendix 1 [77-95,111,115-195], as well as in sortable worksheets available through the Open Science Framework by the Center for Open Science on its website [196].

Included records were published as early as 2003 [77,78]. There is a noted acceleration in the numbers of records published over the past 10 years: 22 records were published from 2003 to 2010, and 79 were published from 2011 through March 2019. Most records were full-text manuscripts (93/101, 92.1%). Across records, there were 73 unique trials—1 record [79] reported 2 separate trials. Trials were most commonly conducted in the United States (37/73, 51%), with 14 in Sweden (19%), 10 conducted in Australia (14%), and 4 in Canada (5%), among other countries of origin. In total, there were 52 unique dyadic eHealth interventions identified. The following sections present summarized findings related to study population, intervention system components, and study outcomes and design characteristics.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Populations

A total of 72 records reported results from 52 trials of 33 unique dyadic eHealth interventions that targeted dyads in which the care recipient was a minor. For all of these trials, it was specified that a parent/guardian must be the support person. Only 1 trial [80,81] did not require that a parent/guardian participate in order for the child to participate—all other trials did require a parent/guardian to participate. These interventions most commonly targeted the care recipient's mental health condition (10/33, 30%). Other conditions targeted included obesity (8/33, 24%), gastrointestinal disorders (4/33, 12%), pain conditions (3/33, 9%), and traumatic brain injury (2/33, 6%). In all, 3 trials targeted dyads including care recipients with 2 specific comorbidities. Conaughton et al [82] tested a transdiagnostic internet-delivered cognitive behavioral therapy for anxiety program among children with both an anxiety disorder and high functioning autism spectrum disorder; Lee et al [83] developed an mHealth program to address weight management among children with both overweight/obesity and intellectual disability; and Palermo et al [84] tested a Web-based program for children with chronic pain and sickle cell disease.

Only 1 record reported results from 1 trial of 1 intervention that specifically recruited an adolescent-young adult population (ie, care recipients ranged from mid-teens through early 30s [85]). The support person could be any close individual (eg, family member, romantic partner, or nonrelative) and was not required to enroll in order for the care recipient to participate. This study targeted care recipients who required catheterization.

A total of 28 records reported results from 21 trials of 18 interventions in which the care recipients were exclusively adults (ie, aged 18 years or older). Six of these trials (6/21, 29%) targeted recruitment to dyads of spouses/romantic partners; in the remaining 15 trials (15/21, 71%), the support person could be any close individual. Of note, the 3 trials of the *CarePartner* intervention [86-90] required that the participating support person reside outside the care recipient's home. In all, 11 of the 21 (52%) trials of adult dyadic eHealth interventions required that both the care recipient and the support person enroll together for either to participate. Adult dyadic eHealth interventions most frequently targeted care recipients with cancer (9/18, 50%). Other conditions targeted included mental health conditions (3/18, 17%) and diabetes (2/18, 11%). Two trials targeted dyads including care recipients with 2 specific comorbidities. Both Schover et al [91] and Hummel et al [92,93] tested internet-based interventions addressing sexual dysfunction secondary to cancer.

Intervention Information

Intervention System Components

Interventions tended to deliver content, intervene, and engage users through multiple components: the median number of components was 3 (M=3.03), ranging from 1 to 6 components. The 3 most commonly utilized components were browser-based components, human telephone contact, and asynchronous communication portals. Browser-based components, or webpages, were used in 44 of 52 (85%) interventions. Human telephone contact, most commonly contact with a therapist or research staff person, was used in 25 interventions (25/52, 48%). Asynchronous communication portals, most commonly used to contact a therapist, research staff person, or other intervention participants via a communication portal embedded within the program such as encrypted email portal or a discussion board, were used in 22 interventions (22/52, 42%).

A total of 40 interventions (40/52, 77%) required human support as they included a component that required therapist or research staff effort (ie, videoconferencing, human telephone contact, human emails/SMS, asynchronous communication portal, synchronous chat room, or in-person sessions).

Dyad Participation and Content Target

In all, 21 interventions were intended to be used entirely separately by care recipients and support persons (21/52, 40%). Interventions were almost as commonly intended to be used with care recipients and support persons accessing some aspects of the intervention together and other aspects separately (20/52, 38%). Fewer interventions were intended to be used by the dyad entirely together (6/52, 12%). In some cases (5/52, 10%), it was not specified how an intervention was intended for use by care recipients and support persons—either because it was unspecified in the record or because there was no specific way that dyads were expected to interact with the intervention.

Most interventions (38/52, 73%) had content that was exclusively intended to be accessed by support persons, and most had content that was exclusively intended to be accessed by care recipients (38/52, 73%). A total of 32 interventions (32/52, 62%) had shared content that was accessible to both the care recipients and support persons. In all, 14 interventions (14/52, 27%) included all 3 types of content (ie, for care recipients only, for support persons only, and shared).

Outcomes, Study Design, and Comparators

Most studies (63/102, 61.8%) specified one or more primary outcomes for care recipients and 22 of 102 (21.6%) specified one or more primary outcomes for support persons. In all, 6 studies (6/102, 5.9%) reported no outcomes for care recipients and 35 (35/102, 34.3%) reported no outcomes for support persons. Among the 63 studies specifying primary outcomes for care recipients, 23 (37%) included objective primary psychosocial or health outcomes, 43 (68%) included self-report primary psychosocial or health outcomes, and 3 (5%) included feasibility/usability primary outcomes (see Methods, Eligibility Criteria, Outcome section for definitions). Among the 33 studies specifying other outcomes for care recipients, 7 (21%) included objective outcomes, 14 (42%) included self-report outcomes, and 26 (79%) included feasibility/usability outcomes. Among the 22 studies specifying primary outcomes for support persons, 3 (14%) included objective primary outcomes, 18 (82%) included self-report primary outcomes, and 2 (9%) included feasibility/usability primary outcomes. Among the 46 studies specifying other outcomes for support persons, 1 (2%) included objective outcomes, 21 (46%) included self-report outcomes, and 34 (74%) included feasibility/usability outcomes.

Trials of dyadic eHealth interventions were most commonly RCTs (44/73, 60%). Among the 44 RCTs, most (22/44, 50%) compared the dyadic eHealth intervention with usual care alone; 8 (18%) with an educational website or internet-delivered resources; 4 (9%) with an in-person version of the intervention; 2 (5%) with an intervention delivered exclusively to the care recipients (ie, no support person intervention); 2 (5%) with an intervention delivered exclusively to support persons (ie, to parents alone); and 6 (14%) with another kind of intervention. Less common study designs were single-arm trials (23/73, 32%), patient preference trials (3/73, 4%), parallel-group RCTs (2/73, 3%), or observational studies (1/73, 1%).

Among studies from RCTs that specified one or more primary outcomes for care recipients, 95% (19/20) of dyadic eHealth interventions demonstrated at least one or more statistically superior outcomes compared with usual care and 20% (1/5) with an educational website or internet-delivered resources. Positive findings for primary outcomes were found in a trial comparing a combination of 2 interventions with either of the singular interventions. In all, 2 RCTs compared a dyadic eHealth intervention with a parent-only eHealth intervention: 1 trial demonstrated better outcomes for care recipients in the dyadic condition [94] and 1 trial demonstrated better outcomes for care recipients in the parent-only condition [95]. There were no significant differences for any care recipient primary outcomes when dyadic eHealth interventions were compared with an in-person version of the intervention (n=3 trials), an intervention delivered exclusively to the care recipients (n=1), or another kind of intervention (n=2).

Among RCTs that specified one or more primary outcomes for support persons, 60% (3/5) of dyadic eHealth interventions demonstrated at least one or more statistically superior outcomes compared with usual care and 67% (2/3) with an educational website or internet-delivered resources. In 1 trial, parents' involvement in their child's care was superior among parents in a parent-only intervention relative to the dyadic intervention [95]. There were no significant differences for any support person primary outcomes when dyadic eHealth interventions were compared with an in-person version of the intervention (n=2 trials) or another kind of intervention (n=1).

Among single-arm trials that compared preintervention to postintervention levels of one or more specified primary outcomes, 93% (13/14) showed at least one significantly improved outcome for care recipients and 100% (3/3) for support persons. The 1 patient preference trial that specified primary outcomes demonstrated better outcomes for care recipients who elected to enroll with a support person compared with those who enrolled by themselves [88].

Discussion

Principal Findings

This review provides the first comprehensive summary of dyadic interventions utilizing eHealth approaches to provide psychosocial care to care recipients and their support persons together as a unit. Three research questions guided the review: what populations had been targeted by these interventions, how technologies have been utilized in these interventions, and approaches these interventions have taken to intervene with the dyad. Of over 7000 reviewed records, 101 met eligibility criteria and described 52 unique dyadic eHealth interventions. In all, 33 of the unique interventions were conducted among dyads of children and their supporting parent, and 18 of the interventions were conducted among adult dyads with diverse relationships. Less than one-fourth of all interventions were fully automated. When a primary outcome was specified, the results of the dyadic eHealth intervention tended to be positive for both care recipients and support persons, although less than one-fourth of studies specified a primary outcome for support persons. One-third reported no outcomes for support persons at all.

This review reveals that there is substantial, diverse, and growing literature supporting dyadic eHealth interventions. This robust literature demonstrates a broad range of strategies for delivering interventions to dyads via eHealth and also identifies several significant gaps in the science. By summarizing broadly across the literature, findings highlight areas of research opportunities for developing dyadic eHealth interventions for novel populations and increasing access to dyadic care. Findings also demonstrate broad variability in approaches of intervening with dyads, with open questions remaining regarding the necessary and sufficient components that define a *dyadic intervention*.

Populations Targeted by Dyadic Electronic Health Interventions

The literature base for pediatric dyadic eHealth interventions was relatively more established compared with the more nascent rise in dyadic eHealth interventions for adult dyads. There were almost twice as many unique eHealth interventions and trials for parent-child dyads compared with those for dyads with adult care recipients. Only 2 of 18 adult dyadic eHealth interventions had been tested in more than 1 trial. This is compared with 7 of 33 pediatric dyadic eHealth interventions that were tested in more than 1 trial, and 3 interventions were tested in 4 or more trials. This discrepancy may be due in part to the fact that pediatric dyadic eHealth interventions started to be published much earlier than adult interventions: about one-fourth of the pediatric records were published before 2010, whereas only 2 records on an adult dyadic eHealth intervention were published before 2010. This finding also reflects the more long-standing emphasis on family-centered care in pediatric health care [96-98] and the increasing acceptance of family-centered care among adult health care delivery [99-101].

There was also a discrepancy between pediatric and adult dyadic eHealth interventions in terms of the care recipient conditions targeted by the interventions. Although over half of adult dyadic eHealth interventions addressed coping with the care recipients' cancer treatment and side effects, there were no pediatric interventions specifically addressing cancer. In contrast, pediatric interventions commonly targeted obesity and pain, whereas no adult interventions targeted these specific conditions. Cancer, obesity, and chronic pain affect individuals of all ages and their families and are strongly influenced by heritable and shared lifestyle factors within families. As such, it is worth considering how dyadic eHealth interventions might be extended to address pediatric and adolescent/young adult cancer survivorship, as well as obesity and chronic pain among adults.

One population notably missing from representation is older adult dyads (eg, care recipients and/or support persons over age 65 years). In the context of our eligibility criteria, no interventions were specifically tailored to older adult needs, no trials recruited an exclusively older adult population, and no records specifically reported intervention effects among older participants. Older adult care recipients are most likely to be cared for by their spouses who frequently have health limitations of their own [7], meaning that dyadic care is particularly important for older adults. Older support persons are as interested in eHealth resources as younger support persons [102]. In actual practice, however, older support persons are less likely to access internet-based caregiving resources than younger ones [103,104]. This discrepancy may in part be because of older adults having difficulty navigating eHealth resources [105], which emphasizes the importance of developing dyadic eHealth tools specifically with older users in mind.

Use of Electronic Health Technologies

Technology that automates care is integral to extend the reach of dyadic interventions by overcoming current logistical and financial barriers [6,38,106]. More than three-fourths of dyadic eHealth interventions identified in this review, however, utilized human support. In addition, 2 of the 3 most commonly used intervention system components require human support (ie, by telephone or asynchronous communication portal). Human involvement was not always extensive. For instance, with Web-based Management of Adolescent Pain, Palermo et al [84] report expressly instructing therapists to spend no more than 5 min to respond to a participant message. Importantly, such minimal human support may be sufficient to enhance clinical outcomes: in one of their studies, Anderson et al [79] demonstrated that dyads of anxious youths and their parents reported comparable working alliance with their therapist in a minimal-touch Web-based version of an intervention as dyads receiving the intervention in traditional face-to-face care.

There may be important rationale to include human support in a dyadic eHealth intervention, such as to increase perceived acceptability, user engagement, and clinical outcomes [107-109]. Progressing capabilities of technology to provide automated and personalized feedback, however, suggest that discrepancies in implementation and clinical outcomes will continuously narrow between human-supported and automated interventions [38,110]. Indeed, it appears that dyads may *vote with their feet* toward increasing automation—Schover et al [91] reported that their recruitment rate tripled when dyads were guaranteed access to a Web-based intervention as opposed to being randomized between the Web-based intervention or to the same intervention content delivered by telephone with a nurse. In their pilot of a Web-delivered adaptation of a previously nurse-delivered coping program, Northouse et al [111] found that enrollment rates for the Web version were lower than enrollment rates into prior trials for the nurse-delivered version; however, the retention rate for the Web version was higher than retention rate for the nurse-delivered version. As technological capacity—and society's expectations for it—increase, the relative advantages of clinician support compared with highly responsive technologies should be carefully considered to ensure greatest efficacy and reach of dyadic interventions.

Approaches to Intervening With the Dyad

In this review, we utilized a broad definition to identify dyadic interventions—namely, an intervention that intervened with both a care recipient and (at least one) informal support person to provide comprehensive treatment for the care recipient's health condition. About 1 in 4 identified dyadic eHealth interventions appeared to address the unique needs of each individual in the dyad, while also addressing relational factors between the dyad. These interventions both (1) included content that was uniquely tailored for and exclusively accessible by care recipients and support persons separately, in addition to shared content available to both users; and (2) were either intended to be utilized both together and separately or left up to the dyads how to interact with the intervention. Addressing relational factors between dyad members is likely a key element to delivering synergistic benefits beyond those from individual interventions, as they enhance natural support between dyad members [112-114]. Previous literature suggests that dyadic interventions addressing these relational aspects of coping with illness are more effective than primarily informational interventions [4].

An example of a dyadic eHealth intervention that addresses both individual needs and the dyadic relationship is the Schizophrenia Online Access to Resources intervention [115-117]. This intervention addresses unique concerns of both dyad members—for example, with unique therapy forums for the individuals with schizophrenia only and support persons only, as well as addressing shared skills for the whole family—for example, shared content regarding developing a supportive safety net together. Across their manuscripts, Rotondi et al [115-117] establish rationales for the use of dyadic and eHealth approaches: a dyadic approach was utilized because family-based therapy is the gold standard for schizophrenia, given the detrimental effects of a stressful family environment on worsening patients' symptom severity. An eHealth approach was utilized given that traditional in-person family-based schizophrenia interventions have had low uptake and retention, and authors reasoned that eHealth programs would be more accessible to families affected by schizophrenia.

Limitations and Future Directions

There are limitations to both the body of literature summarized and the methods we used to summarize this literature. Regarding the former, less than two-thirds of the 52 unique dyadic eHealth interventions have been tested for efficacy in an RCT. Among RCTs, over half compared against usual care or waitlist control alone. Although the usual care comparator may be an important first step toward demonstrating clinical benefits of an intervention, these trials leave important questions about the unique benefits of dyadic interventions unanswered. For example, trials are needed that compare outcomes for both patients and support persons when provided a dyadic eHealth intervention versus individual eHealth interventions for patients and for support persons alone. Such trials would begin to demonstrate whether, and under what circumstances, dyadic eHealth interventions are superior to individually delivered interventions. eHealth interventions are better suited for such trials than face-to-face interventions, given large-scale recruitment is more feasible. In addition, participants can be randomized to receive only part of a larger intervention: for instance, 1 dyad receives the full dyadic eHealth intervention, whereas in another dyad, only the support person receives applicable content from the full intervention. These data will be necessary to justify the added costs and complication of intervening with both care recipients and support persons together.

Another substantial limitation of this literature is the fact that one-third of all studies reported no outcomes for support persons. This issue was identified in one of the earliest reviews of dyadic psychosocial interventions published almost 15 years ago [1]. It is possible that support person outcome data were collected but not reported in these records. Regardless, omitting support person data is a significant missed opportunity to demonstrate the effects of an intervention to roughly half of its users. Given the extensive care responsibilities support persons already assume and the psychosocial and physiological strain of those responsibilities, intervention developers should carefully consider whether the benefits of including a support person in an intervention ultimately outweigh the costs to that support person. Reporting outcomes for support persons are critical to demonstrate that these interventions, at minimum, do no harm to them.

Although the breadth of this review is a strength, the lack of common vocabulary and established criteria that define dyadic interventions made this review challenging to complete. Pediatric interventions, in particular, may be more commonly described as *family based* rather than *dyadic*. To best ensure we captured all interventions that met our broad definition of dyadic eHealth interventions, we carefully developed a search strategy with our medical librarian. It remains possible, however, that we missed interventions that meet our core definitional qualities of dyadic eHealth interventions but use a different vocabulary. Ultimately, to advance the science on dyadic interventions, reaching consensus on ideal qualities of dyadic interventions will be helpful to guide development, assessment, and dissemination of this model of care.

In addition, as a scoping review covering broad literature, we were limited to capturing only superficial components of these interventions. In particular, we were unable to more extensively describe the nature of content delivered by the interventions, which would afford a more nuanced understanding of how interventions address dyad members' unique needs and their relational needs. Given the breadth of data extracted from the volume of records, we also were unable to present detailed results related to study outcomes and designs across subsets of populations, health conditions, or strategies. Finally, we did not extract sample sizes from the studies or location where the intervention was received by participants (eg, home and health care facility). Ideally, this review highlights the breadth of the dyadic eHealth literature and the opportunities for more specific systematic reviews on pertinent subtopics.

Conclusions

This first comprehensive review of dyadic eHealth interventions identified the substantial and rapidly growing literature describing the use of these interventions across a broad range of populations. Collating this robust literature will hopefully serve as a resource for intervention developers to identify models most likely to be effective given the goals and requirements of a particular intervention, rather than models that have been typically used within a particular population. Although the literature on dyadic eHealth interventions is robust, there are significant gaps. For instance, few studies reported outcomes for support persons, data which are essential to justifying their inclusion in interventions. Other gaps highlight important future research needs: for instance, development of interventions designed for older adults and trials comparing the relative efficacy of dyadic and individual eHealth interventions. As technology advances, further personalizing and automating dyadic eHealth interventions will help to increase their scalability, and ultimately, their likelihood of benefitting all who are affected by illness.

Acknowledgments

The authors sincerely thank David Vann for his contributions to data extraction and to the authors of the included studies who assisted this effort by reviewing data extracted from their trials. KS was supported by grant #IRG 81-001-26 from the American Cancer Society, as well as T32-CA009461 (PI: Ostroff) and Cancer Center Support Grant P30-CA008748 (PI: Thompson) from the National Cancer Institute.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study Information.
[PDF File (Adobe PDF File), 635 KB - jmir_v22i3e15509_app1.pdf]

References

1. Martire LM. The 'Relative' efficacy of involving family in psychosocial interventions for chronic illness: are there added benefits to patients and family members? Fam Syst Health 2005;23(3):312-328. [doi: 10.1037/1091-7527.23.3.312]
2. Martire LM, Schulz R. Involving family in psychosocial interventions for chronic illness. Curr Dir Psychol Sci 2007;16(2):90-94. [doi: 10.1111/j.1467-8721.2007.00482.x]
3. Martire LM, Lustig AP, Schulz R, Miller GE, Helgeson VS. Is it beneficial to involve a family member? A meta-analysis of psychosocial interventions for chronic illness. Health Psychol 2004 Nov;23(6):599-611. [doi: 10.1037/0278-6133.23.6.599] [Medline: 15546228]
4. Hartmann M, Bäzner E, Wild B, Eisler I, Herzog W. Effects of interventions involving the family in the treatment of adult patients with chronic physical diseases: a meta-analysis. Psychother Psychosom 2010;79(3):136-148. [doi: 10.1159/000286958] [Medline: 20185970]
5. Stahl ST, Rodakowski J, Saghafi EM, Park M, Reynolds CF, Dew MA. Systematic review of dyadic and family-oriented interventions for late-life depression. Int J Geriatr Psychiatry 2016 Sep;31(9):963-973 [FREE Full text] [doi: 10.1002/gps.4434] [Medline: 26799782]
6. Ratcliff CG, Vinson CA, Milbury K, Badr H. Moving family interventions into the real world: what matters to oncology stakeholders? J Psychosoc Oncol 2019;37(2):264-284. [doi: 10.1080/07347332.2018.1498426] [Medline: 30421667]
7. National Alliance for Caregiving, AARP Public Policy Institute. AARP Public Policy Institute. 2015. Caregiving in the US URL: https://www.aarp.org/content/dam/aarp/ppi/2015/caregiving-in-the-united-states-2015-report-revised.pdf [accessed 2020-01-03]
8. Amaro LM. Convincing family caregivers to participate in research: 'Does that count as caregiving?'. J Appl Commun Res 2015;43(2):269-272. [doi: 10.1080/00909882.2015.1019547]
9. Reinhard S, Feinberg L, Choula R, Houser A. AARP Public Policy Institute. 2015 Jul. Valuing the Invaluable: 2015 Update URL: https://www.aarp.org/content/dam/aarp/ppi/2015/valuing-the-invaluable-2015-update-new.pdf [accessed 2020-01-03]
10. Chassin L, Macy JT, Seo D, Presson CC, Sherman SJ. The association between membership in the sandwich generation and health behaviors: A longitudinal study. J Appl Dev Psychol 2010 Jan 1;31(1):38-46 [FREE Full text] [doi: 10.1016/j.appdev.2009.06.001] [Medline: 20161605]
11. Hoffman GJ, Lee J, Mendez-Luck CA. Health behaviors among Baby Boomer informal caregivers. Gerontologist 2012 Apr;52(2):219-230 [FREE Full text] [doi: 10.1093/geront/gns003] [Medline: 22391873]
12. Morris BA, Thorndike FP, Ritterband LM, Glozier N, Dunn J, Chambers SK. Sleep disturbance in cancer patients and caregivers who contact telephone-based help services. Support Care Cancer 2015 Apr;23(4):1113-1120. [doi: 10.1007/s00520-014-2458-y] [Medline: 25297467]
13. Butler SS, Turner W, Kaye LW, Ruffin L, Downey R. Depression and caregiver burden among rural elder caregivers. J Gerontol Soc Work 2005;46(1):47-63. [doi: 10.1300/J083v46n01_04] [Medline: 16338884]

14. Cochrane JJ, Goering PN, Rogers JM. The mental health of informal caregivers in Ontario: an epidemiological survey. Am J Public Health 1997 Dec;87(12):2002-2007. [doi: 10.2105/ajph.87.12.2002] [Medline: 9431291]
15. Trevino KM, Prigerson HG, Maciejewski PK. Advanced cancer caregiving as a risk for major depressive episodes and generalized anxiety disorder. Psychooncology 2018 Jan;27(1):243-249 [FREE Full text] [doi: 10.1002/pon.4441] [Medline: 28426918]
16. Ji J, Zöller B, Sundquist K, Sundquist J. Increased risks of coronary heart disease and stroke among spousal caregivers of cancer patients. Circulation 2012 Apr 10;125(14):1742-1747. [doi: 10.1161/CIRCULATIONAHA.111.057018] [Medline: 22415143]
17. Corà A, Partinico M, Munafò M, Palomba D. Health risk factors in caregivers of terminal cancer patients: a pilot study. Cancer Nurs 2012;35(1):38-47. [doi: 10.1097/NCC.0b013e31820d0c23] [Medline: 21372699]
18. Rohleder N, Marin TJ, Ma R, Miller GE. Biologic cost of caring for a cancer patient: dysregulation of pro- and anti-inflammatory signaling pathways. J Clin Oncol 2009 Jun 20;27(18):2909-2915. [doi: 10.1200/JCO.2008.18.7435] [Medline: 19433690]
19. Beach SR, Schulz R, Williamson GM, Miller LS, Weiner MF, Lance CE. Risk factors for potentially harmful informal caregiver behavior. J Am Geriatr Soc 2005 Feb;53(2):255-261. [doi: 10.1111/j.1532-5415.2005.53111.x] [Medline: 15673349]
20. Beach SR, Schulz R. Family caregiver factors associated with unmet needs for care of older adults. J Am Geriatr Soc 2017 Mar;65(3):560-566. [doi: 10.1111/jgs.14547] [Medline: 27935019]
21. Lau DT, Berman R, Halpern L, Pickard AS, Schrauf R, Witt W. Exploring factors that influence informal caregiving in medication management for home hospice patients. J Palliat Med 2010 Sep;13(9):1085-1090 [FREE Full text] [doi: 10.1089/jpm.2010.0082] [Medline: 20836633]
22. Segrin C, Badger TA. Psychological and physical distress are interdependent in breast cancer survivors and their partners. Psychol Health Med 2014;19(6):716-723. [doi: 10.1080/13548506.2013.871304] [Medline: 24383980]
23. Paul C, Hall A, Oldmeadow C, Lynagh M, Campbell S, Bradstock K, et al. Dyadic interdependence of psychosocial outcomes among haematological cancer survivors and their support persons. Support Care Cancer 2017 Nov;25(11):3339-3346. [doi: 10.1007/s00520-017-3751-3] [Medline: 28550441]
24. Langer SL, Romano JM, Liu Q, Levy RL, Nielson H, Brown JD. Pain catastrophizing predicts verbal expression among children with chronic pain and their mothers. Health Psychol Open 2016 Jan;3(1):2055102916632667 [FREE Full text] [doi: 10.1177/2055102916632667] [Medline: 28070387]
25. Hodges L, Humphris G, Macfarlane G. A meta-analytic investigation of the relationship between the psychological distress of cancer patients and their carers. Soc Sci Med 2005 Jan;60(1):1-12. [doi: 10.1016/j.socscimed.2004.04.018] [Medline: 15482862]
26. Badr H, Herbert K, Bonnen MD, Asper JA, Wagner T. Dyadic Coping in Patients Undergoing Radiotherapy for Head and Neck Cancer and Their Spouses. Front Psychol 2018;9:1780 [FREE Full text] [doi: 10.3389/fpsyg.2018.01780] [Medline: 30374316]
27. Zajdel M, Helgeson V, Seltman H, Korytkowski M, Hausmann LR. Daily communal coping in couples with type 2 diabetes: links to mood and self-care. Ann Behav Med 2018 Feb 17;52(3):228-238 [FREE Full text] [doi: 10.1093/abm/kax047] [Medline: 29538665]
28. Casu G, Zaia V, Martins MD, Barbosa C, Gremigni P. A dyadic mediation study on social support, coping, and stress among couples starting fertility treatment. J Fam Psychol 2019 Apr;33(3):315-326. [doi: 10.1037/fam0000502] [Medline: 30730186]
29. Badr H, Carmack CL, Diefenbach MA. Psychosocial interventions for patients and caregivers in the age of new communication technologies: opportunities and challenges in cancer care. J Health Commun 2015;20(3):328-342 [FREE Full text] [doi: 10.1080/10810730.2014.965369] [Medline: 25629218]
30. Carr RM, Prestwich A, Kwasnicka D, Thøgersen-Ntoumani C, Gucciardi DF, Quested E, et al. Dyadic interventions to promote physical activity and reduce sedentary behaviour: systematic review and meta-analysis. Health Psychol Rev 2019 Mar;13(1):91-109. [doi: 10.1080/17437199.2018.1532312] [Medline: 30284501]
31. Huang L, Smith AK, Wong ML. Who will care for the caregivers? Increased needs when caring for frail older adults with cancer. J Am Geriatr Soc 2019 May;67(5):873-876. [doi: 10.1111/jgs.15863] [Medline: 30924526]
32. Shaw J, Harrison J, Young J, Butow P, Sandroussi C, Martin D, et al. Coping with newly diagnosed upper gastrointestinal cancer: a longitudinal qualitative study of family caregivers' role perception and supportive care needs. Support Care Cancer 2013 Mar;21(3):749-756. [doi: 10.1007/s00520-012-1575-8] [Medline: 22933130]
33. Waldron EA, Janke EA, Bechtel CF, Ramirez M, Cohen A. A systematic review of psychosocial interventions to improve cancer caregiver quality of life. Psychooncology 2013 Jun;22(6):1200-1207. [doi: 10.1002/pon.3118] [Medline: 22729992]
34. Applebaum AJ, Farran CJ, Marziliano AM, Pasternak AR, Breitbart W. Preliminary study of themes of meaning and psychosocial service use among informal cancer caregivers. Palliat Support Care 2014 Apr;12(2):139-148 [FREE Full text] [doi: 10.1017/S1478951513000084] [Medline: 23919966]

35. Northouse LL, Katapodi MC, Song L, Zhang L, Mood DW. Interventions with family caregivers of cancer patients: meta-analysis of randomized trials. CA Cancer J Clin 2010;60(5):317-339 [FREE Full text] [doi: 10.3322/caac.20081] [Medline: 20709946]
36. Badr H, Herbert K, Reckson B, Rainey H, Sallam A, Gupta V. Unmet needs and relationship challenges of head and neck cancer patients and their spouses. J Psychosoc Oncol 2016;34(4):336-346 [FREE Full text] [doi: 10.1080/07347332.2016.1195901] [Medline: 27269579]
37. Ho A, Collins SR, Davis K, Doty MM. A look at working-age caregivers' roles, health concerns, and need for support. Issue Brief (Commonw Fund) 2005 Aug(854):1-12. [Medline: 16118908]
38. Doss BD, Feinberg LK, Rothman K, Roddy MK, Comer JS. Using technology to enhance and expand interventions for couples and families: conceptual and methodological considerations. J Fam Psychol 2017 Dec;31(8):983-993 [FREE Full text] [doi: 10.1037/fam0000349] [Medline: 29309184]
39. Bhandari N, Shi Y, Jung K. Seeking health information online: does limited healthcare access matter? J Am Med Inform Assoc 2014;21(6):1113-1117 [FREE Full text] [doi: 10.1136/amiajnl-2013-002350] [Medline: 24948558]
40. Fox S, Duggan M, Purcell K. Pew Reesearch Center. 2013 Jun 20. Family Caregivers are Wired for Health URL: https://www.pewresearch.org/internet/2013/06/20/family-caregivers-are-wired-for-health/ [accessed 2020-01-03]
41. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45 [FREE Full text] [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]
42. Gehring ND, McGrath P, Wozney L, Soleimani A, Bennett K, Hartling L, et al. Pediatric eMental healthcare technologies: a systematic review of implementation foci in research studies, and government and organizational documents. Implement Sci 2017 Jun 21;12(1):76 [FREE Full text] [doi: 10.1186/s13012-017-0608-6] [Medline: 28637479]
43. Bouma G, Admiraal JM, de Vries EG, Schröder CP, Walenkamp AM, Reyners AK. Internet-based support programs to alleviate psychosocial and physical symptoms in cancer patients: a literature analysis. Crit Rev Oncol Hematol 2015 Jul;95(1):26-37. [doi: 10.1016/j.critrevonc.2015.01.011] [Medline: 25701515]
44. Moradian S, Voelker N, Brown C, Liu G, Howell D. Effectiveness of internet-based interventions in managing chemotherapy-related symptoms in patients with cancer: a systematic literature review. Support Care Cancer 2018 Feb;26(2):361-374. [doi: 10.1007/s00520-017-3900-8] [Medline: 28948360]
45. Kang HS, Kim H, Park SM, Kim J. Online-based interventions for sexual health among individuals with cancer: a systematic review. BMC Health Serv Res 2018 Mar 7;18(1):167 [FREE Full text] [doi: 10.1186/s12913-018-2972-6] [Medline: 29514669]
46. Arem H, Irwin M. A review of web-based weight loss interventions in adults. Obes Rev 2011 May;12(5):e236-e243 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00787.x] [Medline: 20804523]
47. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014 Feb 18;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]
48. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
49. Guse K, Levine D, Martins S, Lira A, Gaarde J, Westmorland W, et al. Interventions using new digital media to improve adolescent sexual health: a systematic review. J Adolesc Health 2012 Dec;51(6):535-543. [doi: 10.1016/j.jadohealth.2012.03.014] [Medline: 23174462]
50. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
51. Li J, Theng Y, Foo S. Game-based digital interventions for depression therapy: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):519-527 [FREE Full text] [doi: 10.1089/cyber.2013.0481] [Medline: 24810933]
52. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
53. Neve M, Morgan P, Jones P, Collins C. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
54. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
55. Bennett GG, Steinberg DM, Stoute C, Lanpher M, Lane I, Askew S, et al. Electronic health (eHealth) interventions for weight management among racial/ethnic minority adults: a systematic review. Obes Rev 2014 Oct;15(Suppl 4):146-158. [doi: 10.1111/obr.12218] [Medline: 25196411]

56. Sin J, Henderson C, Spain D, Cornelius V, Chen T, Gillard S. eHealth interventions for family carers of people with long term illness: a promising approach? Clin Psychol Rev 2018 Mar;60:109-125 [FREE Full text] [doi: 10.1016/j.cpr.2018.01.008] [Medline: 29429856]
57. Wozney L, de Souza LM, Kervin E, Queluz F, McGrath PJ, Keefe J. Commercially available mobile apps for caregivers of people with Alzheimer disease or other related dementias: systematic search. JMIR Aging 2018 Dec 7;1(2):e12274 [FREE Full text] [doi: 10.2196/12274] [Medline: 31518255]
58. Ploeg J, Ali MU, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, et al. Caregiver-focused, web-based interventions: systematic review and meta-analysis (part 2). J Med Internet Res 2018 Oct 26;20(10):e11247 [FREE Full text] [doi: 10.2196/11247] [Medline: 30368439]
59. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review. J Med Internet Res 2018 Jun 12;20(6):e216 [FREE Full text] [doi: 10.2196/jmir.9548] [Medline: 29895512]
60. Egan KJ, Pinto-Bruno Á, Bighelli I, Berg-Weger M, van Straten A, Albanese E, et al. Online training and support programs designed to improve mental health and reduce burden among caregivers of people with dementia: a systematic review. J Am Med Dir Assoc 2018 Mar;19(3):200-6.e1. [doi: 10.1016/j.jamda.2017.10.023] [Medline: 29306605]
61. Kaltenbaugh D, Klem ML, Hu L, Turi E, Haines A, Lingler JH. Using web-based interventions to support caregivers of patients with cancer: a systematic review. Oncol Nurs Forum 2015 Mar;42(2):156-164. [doi: 10.1188/15.ONF.156-164] [Medline: 25806882]
62. US Department of Veterans Affairs, Health Services Research & Development Service. Mobile Applications and Internet-Based Approaches for Supporting Non-Professional Caregivers: A Systematic Review. Scotts Valley, California , US: Createspace Independent Pub; 2012.
63. Hu C, Kung S, Rummans TA, Clark MM, Lapid MI. Reducing caregiver stress with internet-based interventions: a systematic review of open-label and randomized controlled trials. J Am Med Inform Assoc 2015 Apr;22(e1):e194-e209. [doi: 10.1136/amiajnl-2014-002817] [Medline: 25125686]
64. Wasilewski MB, Stinson JN, Cameron JI. Web-based health interventions for family caregivers of elderly individuals: a Scoping Review. Int J Med Inform 2017 Jul;103:109-138. [doi: 10.1016/j.ijmedinf.2017.04.009] [Medline: 28550996]
65. Powell J, Chiu T, Eysenbach G. A systematic review of networked technologies supporting carers of people with dementia. J Telemed Telecare 2008;14(3):154-156. [doi: 10.1258/jtt.2008.003018] [Medline: 18430288]
66. Godwin KM, Mills WL, Anderson JA, Kunik ME. Technology-driven interventions for caregivers of persons with dementia: a systematic review. Am J Alzheimers Dis Other Demen 2013 May;28(3):216-222. [doi: 10.1177/1533317513481091] [Medline: 23528881]
67. McKechnie V, Barker C, Stott J. Effectiveness of computer-mediated interventions for informal carers of people with dementia-a systematic review. Int Psychogeriatr 2014 Oct;26(10):1619-1637 [FREE Full text] [doi: 10.1017/S1041610214001045] [Medline: 24989249]
68. Boots LM, de Vugt ME, van Knippenberg RJ, Kempen GI, Verhey FR. A systematic review of internet-based supportive interventions for caregivers of patients with dementia. Int J Geriatr Psychiatry 2014 Apr;29(4):331-344. [doi: 10.1002/gps.4016] [Medline: 23963684]
69. 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]
70. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 9;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
71. NHS Centre for Reviews and Dissemination. The University of York. Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for Those Carrying Out or Commissioning Reviews URL: https://www.crd.york.ac.uk/CRDWeb/ShowRecord.asp?AccessionNumber=32001000984&AccessionNumber=32001000984 [accessed 2020-01-03]
72. Eng TR. The Ehealth Landscape: A Terrain Map of Emerging Information and Communication Technologies in Health and Health Care. Princeton, NJ, USA: Robert Wood Johnson Foundation; 2001.
73. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
74. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
75. Veritas Health Innovation. Covidence Systematic Review Software. Melbourne, Australia URL: https://www.covidence.org [accessed 2020-01-03]
76. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]

77. Ritterband LM, Cox DJ, Walker LS, Kovatchev B, McKnight L, Patel K, et al. An internet intervention as adjunctive therapy for pediatric encopresis. J Consult Clin Psychol 2003 Oct;71(5):910-917. [doi: 10.1037/0022-006X.71.5.910] [Medline: 14516239]
78. Baranowski T, Baranowski JC, Cullen KW, Thompson DI, Nicklas T, Zakeri IE, et al. The Fun, Food, and Fitness Project (FFFP): the Baylor GEMS pilot study. Ethn Dis 2003;13(1 Suppl 1):S30-S39. [Medline: 12713209]
79. Anderson RE, Spence SH, Donovan CL, March S, Prosser S, Kenardy J. Working alliance in online cognitive behavior therapy for anxiety disorders in youth: comparison with clinic delivery and its role in predicting outcome. J Med Internet Res 2012 Jun 28;14(3):e88 [FREE Full text] [doi: 10.2196/jmir.1848] [Medline: 22789657]
80. van Voorhees BW, Gladstone T, Cordel S, Marko-Holguin M, Beardslee W, Kuwabara S, et al. Development of a technology-based behavioral vaccine to prevent adolescent depression: A health system integration model. Internet Interv 2015 Sep;2(3):303-313 [FREE Full text] [doi: 10.1016/j.invent.2015.07.004] [Medline: 30473992]
81. Gladstone TR, Terrizzi DA, Stinson A, Nidetz J, Canel J, Ching E, et al. Effect of internet-based cognitive behavioral humanistic and interpersonal training vs internet-based general health education on adolescent depression in primary care: a randomized clinical trial. JAMA Netw Open 2018 Nov;1(7):pii: e184278 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.4278] [Medline: 30533601]
82. Conaughton RJ, Donovan CL, March S. Efficacy of an internet-based CBT program for children with comorbid High Functioning Autism Spectrum Disorder and anxiety: a randomised controlled trial. J Affect Disord 2017 Aug 15;218:260-268. [doi: 10.1016/j.jad.2017.04.032] [Medline: 28477505]
83. Lee R, Leung C, Chen H, Louie L, Brown M, Chen J, et al. The impact of a school-based weight management program involving parents via mHealth for overweight and obese children and adolescents with intellectual disability: a randomized controlled trial. Int J Environ Res Public Health 2017 Oct 5;14(10):pii: E1178 [FREE Full text] [doi: 10.3390/ijerph14101178] [Medline: 28981460]
84. Palermo TM, Dudeney J, Santanelli JP, Carletti A, Zempsky WT. Feasibility and acceptability of internet-delivered cognitive behavioral therapy for chronic pain in adolescents with sickle cell disease and their parents. J Pediatr Hematol Oncol 2018 Mar;40(2):122-127 [FREE Full text] [doi: 10.1097/MPH.0000000000001018] [Medline: 29176462]
85. Holland JE, DeMaso DR, Rosoklija I, Johnson KL, Manning D, Bellows AL, et al. Self-cathing experience journal: enhancing the patient and family experience in clean intermittent catheterization. J Pediatr Urol 2015 Aug;11(4):187.e1-187.e6. [doi: 10.1016/j.jpurol.2015.03.011] [Medline: 26028181]
86. Aikens JE, Zivin K, Trivedi R, Piette JD. Diabetes self-management support using mHealth and enhanced informal caregiving. J Diabetes Complications 2014;28(2):171-176 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.11.008] [Medline: 24374137]
87. 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 [FREE Full text] [doi: 10.1016/j.pcd.2014.06.003] [Medline: 25065270]
88. Aikens JE, Trivedi R, Heapy A, Pfeiffer PN, Piette JD. Potential impact of incorporating a patient-selected support person into mHealth for depression. J Gen Intern Med 2015 Jun;30(6):797-803 [FREE Full text] [doi: 10.1007/s11606-015-3208-7] [Medline: 25666218]
89. Aikens JE, Trivedi R, Aron DC, Piette JD. Integrating support persons into diabetes telemonitoring to improve self-management and medication adherence. J Gen Intern Med 2015 Mar;30(3):319-326 [FREE Full text] [doi: 10.1007/s11606-014-3101-9] [Medline: 25421436]
90. Piette JD, Striplin D, Marinec N, Chen J, Aikens JE. A randomized trial of mobile health support for heart failure patients and their informal caregivers: impacts on caregiver-reported outcomes. Med Care 2015 Aug;53(8):692-699 [FREE Full text] [doi: 10.1097/MLR.0000000000000378] [Medline: 26125415]
91. Schover LR, Canada AL, Yuan Y, Sui D, Neese L, Jenkins R, et al. A randomized trial of internet-based versus traditional sexual counseling for couples after localized prostate cancer treatment. Cancer 2012 Jan 15;118(2):500-509 [FREE Full text] [doi: 10.1002/cncr.26308] [Medline: 21953578]
92. Hummel SB, van Lankveld JJ, Oldenburg HS, Hahn DE, Kieffer JM, Gerritsma MA, et al. Efficacy of internet-based cognitive behavioral therapy in improving sexual functioning of breast cancer survivors: results of a randomized controlled trial. J Clin Oncol 2017 Apr 20;35(12):1328-1340. [doi: 10.1200/JCO.2016.69.6021] [Medline: 28240966]
93. Hummel SB, van Lankveld JJ, Oldenburg HS, Hahn DE, Kieffer JM, Gerritsma MA, et al. Sexual functioning and relationship satisfaction of partners of breast cancer survivors who receive internet-based sex therapy. J Sex Marital Ther 2019;45(2):91-102. [doi: 10.1080/0092623X.2018.1488325] [Medline: 30040589]
94. Jolstedt M, Wahlund T, Lenhard F, Ljótsson B, Mataix-Cols D, Nord M, et al. Efficacy and cost-effectiveness of therapist-guided internet cognitive behavioural therapy for paediatric anxiety disorders: a single-centre, single-blind, randomised controlled trial. Lancet Child Adolesc Health 2018;2(11):792-801. [doi: 10.1016/s2352-4642(18)30275-x]
95. Ellis DA, Carcone AI, Ondersma SJ, Naar-King S, Dekelbab B, Moltz K. Brief computer-delivered intervention to increase parental monitoring in families of African American adolescents with type 1 diabetes: a randomized controlled trial. Telemed J E Health 2017 Jun;23(6):493-502 [FREE Full text] [doi: 10.1089/tmj.2016.0182] [Medline: 28061319]
96. Kazak A, Simms S, Rourke MT. Family systems practice in pediatric psychology. J Pediatr Psychol 2002 Mar;27(2):133-143. [doi: 10.1093/jpepsy/27.2.133] [Medline: 11821497]

97. Harrison TM. Family-centered pediatric nursing care: state of the science. J Pediatr Nurs 2010 Oct;25(5):335-343 [FREE Full text] [doi: 10.1016/j.pedn.2009.01.006] [Medline: 20816555]
98. Kuo DZ, Houtrow AJ, Arango P, Kuhlthau KA, Simmons JM, Neff JM. Family-centered care: current applications and future directions in pediatric health care. Matern Child Health J 2012 Feb;16(2):297-305 [FREE Full text] [doi: 10.1007/s10995-011-0751-7] [Medline: 21318293]
99. World Health Organization. National Cancer Control Programmes: Policies and Managerial Guidelines. Second Edition. Geneva: World Health Organization; 2002.
100. Davidson JE, Powers K, Hedayat KM, Tieszen M, Kon AA, Shepard E, American College of Critical Care Medicine Task Force 2004-2005, Society of Critical Care Medicine. Clinical practice guidelines for support of the family in the patient-centered intensive care unit: American College of Critical Care Medicine Task Force 2004-2005. Crit Care Med 2007 Feb;35(2):605-622. [doi: 10.1097/01.CCM.0000254067.14607.EB] [Medline: 17205007]
101. Morrison RS, Meier DE. Clinical practice. Palliative care. N Engl J Med 2004 Jun 17;350(25):2582-2590. [doi: 10.1056/NEJMcp035232] [Medline: 15201415]
102. AARP, Project Catalyst. AARP Public Policy. 2016 Apr. Caregivers & Technology: What they Want and Need URL: https://www.aarp.org/content/dam/aarp/research/surveys_statistics/ltc/2018/caregivers-technology-needs.doi.10.26419-2Fres.00191.002.pdf [accessed 2020-01-03]
103. Shaffer KM, Chow PI, Cohn WF, Ingersoll KS, Ritterband LM. Informal caregivers' use of internet-based health resources: an analysis of the health information national trends survey. JMIR Aging 2018 Dec 18;1(2):e11051 [FREE Full text] [doi: 10.2196/11051] [Medline: 31518244]
104. Lapid MI, Atherton PJ, Clark MM, Kung S, Sloan JA, Rummans TA. Cancer caregiver: perceived benefits of technology. Telemed J E Health 2015 Nov;21(11):893-902 [FREE Full text] [doi: 10.1089/tmj.2014.0117] [Medline: 26075800]
105. Bolle S, Romijn G, Smets EM, Loos EF, Kunneman M, van Weert JC. Older cancer patients' user experiences with web-based health information tools: a think-aloud study. J Med Internet Res 2016 Jul 25;18(7):e208 [FREE Full text] [doi: 10.2196/jmir.5618] [Medline: 27457709]
106. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
107. Aardoom JJ, Dingemans AE, Spinhoven P, van Ginkel JR, de Rooij M, van Furth EF. Web-based fully automated self-help with different levels of therapist support for individuals with eating disorder symptoms: a randomized controlled trial. J Med Internet Res 2016 Jun 17;18(6):e159 [FREE Full text] [doi: 10.2196/jmir.5709] [Medline: 27317358]
108. Roddy MK, Nowlan KM, Doss BD. A randomized controlled trial of coach contact during a brief online intervention for distressed couples. Fam Process 2017 Dec;56(4):835-851. [doi: 10.1111/famp.12262] [Medline: 27859100]
109. Dennison L, Morrison L, Lloyd S, Phillips D, Stuart B, Williams S, et al. Does brief telephone support improve engagement with a web-based weight management intervention? Randomized controlled trial. J Med Internet Res 2014 Mar 28;16(3):e95 [FREE Full text] [doi: 10.2196/jmir.3199] [Medline: 24681761]
110. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
111. Northouse L, Schafenacker A, Barr KL, Katapodi M, Yoon H, Brittain K, et al. A tailored web-based psychoeducational intervention for cancer patients and their family caregivers. Cancer Nurs 2014;37(5):321-330 [FREE Full text] [doi: 10.1097/NCC.0000000000000159] [Medline: 24945270]
112. Berg CA, Upchurch R. A developmental-contextual model of couples coping with chronic illness across the adult life span. Psychol Bull 2007 Nov;133(6):920-954. [doi: 10.1037/0033-2909.133.6.920] [Medline: 17967089]
113. Revenson TA, Kayser K, Bodenmann G, editors. Couples Coping With Stress: Emerging Perspectives On Dyadic Coping. Washington, DC, US: American Psychological Association; 2005.
114. Bodenmann G. Dyadic coping and its significance for marital functioning. In: Revenson TA, Kayser K, Bodenmann G, editors. Couples Coping with Stress: Emerging Perspectives on Dyadic Coping. Washington, DC, US: American Psychological Association; 2005:33-49.
115. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/ps.2010.61.11.1099] [Medline: 21041348]
116. Rotondi AJ, Haas GL, Anderson CM, Newhill CE, Spring MB, Ganguli R, et al. A clinical trial to test the feasibility of a telehealth psychoeducational intervention for persons with schizophrenia and their families: intervention and 3-month findings. Rehabil Psychol 2005 Nov;50(4):325-336 [FREE Full text] [doi: 10.1037/0090-5550.50.4.325] [Medline: 26321774]
117. Rotondi A, Haas G, Anderson C, Mueser K. Effectiveness of web-based multi-family treatment delivered to the homes of persons with schizophrenia and their supporters. Schizophr Bull 2013;39(suppl 1):S350.
118. Hannon TS, Yazel-Smith LG, Hatton AS, Stanton JL, Moser EA, Li X, et al. Advancing diabetes management in adolescents: comparative effectiveness of mobile self-monitoring blood glucose technology and family-centered goal setting. Pediatr Diabetes 2018 Jun;19(4):776-781 [FREE Full text] [doi: 10.1111/pedi.12648] [Medline: 29504207]

119. Quirk H, Glazebrook C, Blake H. A physical activity intervention for children with type 1 diabetes- steps to active kids with diabetes (STAK-D): a feasibility study. BMC Pediatr 2018 Feb 7;18(1):37 [FREE Full text] [doi: 10.1186/s12887-018-1036-8] [Medline: 29415687]
120. McCormick M, Reed-Knight B, Lewis JD, Gold BD, Blount RL. Coping skills for reducing pain and somatic symptoms in adolescents with IBD. Inflamm Bowel Dis 2010 Dec;16(12):2148-2157. [doi: 10.1002/ibd.21302] [Medline: 20848505]
121. Reed-Knight B, McCormick M, Lewis JD, Blount RL. Participation and attrition in a coping skills intervention for adolescent girls with inflammatory bowel disease. J Clin Psychol Med Settings 2012 Jun;19(2):188-196. [doi: 10.1007/s10880-011-9269-9] [Medline: 22076655]
122. Bonnert M, Ljótsson B, Hedman E, Serlachius E, Simren M, Benninga MA, et al. Internet-delivered cognitive behavior therapy for adolescents with functional gastrointestinal disorders: a pilot study. Gastroenterology 2014;146(5):S-897. [doi: 10.1016/s0016-5085(14)63260-1]
123. Bonnert M, Olén O, Lalouni M, Benninga MA, Bottai M, Engelbrektsson J, et al. Internet-delivered cognitive behavior therapy for adolescents with irritable bowel syndrome: a randomized controlled trial. Am J Gastroenterol 2017 Jan;112(1):152-162. [doi: 10.1038/ajg.2016.503] [Medline: 27845338]
124. Bonnert M, Olén O, Bjureberg J, Lalouni M, Hedman-Lagerlöf E, Serlachius E, et al. The role of avoidance behavior in the treatment of adolescents with irritable bowel syndrome: a mediation analysis. Behav Res Ther 2018 Jun;105:27-35 [FREE Full text] [doi: 10.1016/j.brat.2018.03.006] [Medline: 29614378]
125. Sampaio F, Bonnert M, Olén O, Hedman E, Lalouni M, Lenhard F, et al. Cost-effectiveness of internet-delivered cognitive-behavioural therapy for adolescents with irritable bowel syndrome. BMJ Open 2019 Jan 24;9(1):e023881 [FREE Full text] [doi: 10.1136/bmjopen-2018-023881] [Medline: 30679293]
126. Bonnert M, Olén O, Lalouni M, Hedman-Lagerlöf E, Särnholm J, Serlachius E, et al. Internet-delivered exposure-based cognitive-behavioral therapy for adolescents with functional abdominal pain or functional dyspepsia: a feasibility study. Behav Ther 2019 Jan;50(1):177-188 [FREE Full text] [doi: 10.1016/j.beth.2018.05.002] [Medline: 30661558]
127. Lalouni M, Ljótsson B, Bonnert M, Hedman-Lagerlöf E, Högström J, Serlachius E, et al. Internet-delivered cognitive behavioral therapy for children with pain-related functional gastrointestinal disorders: feasibility study. JMIR Ment Health 2017 Aug 10;4(3):e32 [FREE Full text] [doi: 10.2196/mental.7985] [Medline: 28798012]
128. Lalouni M, Ljótsson B, Bonnert M, Ssegonja R, Benninga M, Bjureberg J, et al. Clinical and cost effectiveness of online cognitive behavioral therapy in children with functional abdominal pain disorders. Clin Gastroenterol Hepatol 2019 Oct;17(11):2236-44.e11 [FREE Full text] [doi: 10.1016/j.cgh.2018.11.043] [Medline: 30502501]
129. Ritterband LM, Ardalan K, Thorndike FP, Magee JC, Saylor DK, Cox DJ, et al. Real world use of an internet intervention for pediatric encopresis. J Med Internet Res 2008 Jun 30;10(2):e16 [FREE Full text] [doi: 10.2196/jmir.1081] [Medline: 18653440]
130. Ritterband LM, Thorndike FP, Lord HR, Borowitz SM, Walker LS, Ingersoll KS, et al. An RCT of an internet intervention for pediatric encopresis with one year follow-up. Clin Pract Pediatr Psychol 2013 Mar;1(1):68-80 [FREE Full text] [doi: 10.1037/cpp0000007] [Medline: 24040581]
131. Lenhard F, Vigerland S, Andersson E, Rück C, Mataix-Cols D, Thulin U, et al. Internet-delivered cognitive behavior therapy for adolescents with obsessive-compulsive disorder: an open trial. PLoS One 2014;9(6):e100773 [FREE Full text] [doi: 10.1371/journal.pone.0100773] [Medline: 24949622]
132. Lenhard F, Vigerland S, Engberg H, Hallberg A, Thermaenius H, Serlachius E. 'On My Own, but Not Alone' - adolescents' experiences of internet-delivered cognitive behavior therapy for obsessive-compulsive disorder. PLoS One 2016;11(10):e0164311 [FREE Full text] [doi: 10.1371/journal.pone.0164311] [Medline: 27711249]
133. Lenhard F, Ssegonja R, Andersson E, Feldman I, Rück C, Mataix-Cols D, et al. Cost-effectiveness of therapist-guided internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: results from a randomised controlled trial. BMJ Open 2017 May 17;7(5):e015246 [FREE Full text] [doi: 10.1136/bmjopen-2016-015246] [Medline: 28515196]
134. Lenhard F, Andersson E, Mataix-Cols D, Rück C, Vigerland S, Högström J, et al. Therapist-guided, internet-delivered cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2017 Jan;56(1):10-9.e2 [FREE Full text] [doi: 10.1016/j.jaac.2016.09.515] [Medline: 27993223]
135. Aspvall K, Andrén P, Lenhard F, Andersson E, Mataix-Cols D, Serlachius E. Internet-delivered cognitive behavioural therapy for young children with obsessive-compulsive disorder: development and initial evaluation of the BIP OCD Junior programme. BJPsych Open 2018 May;4(3):106-112 [FREE Full text] [doi: 10.1192/bjo.2018.10] [Medline: 29971153]
136. Nordh M, Vigerland S, Öst LG, Ljótsson B, Mataix-Cols D, Serlachius E, et al. Therapist-guided internet-delivered cognitive-behavioural therapy supplemented with group exposure sessions for adolescents with social anxiety disorder: a feasibility trial. BMJ Open 2017 Dec 14;7(12):e018345 [FREE Full text] [doi: 10.1136/bmjopen-2017-018345] [Medline: 29247101]
137. Andrén P, Aspvall K, de la Cruz LF, Wiktor P, Romano S, Andersson E, et al. Therapist-guided and parent-guided internet-delivered behaviour therapy for paediatric Tourette's disorder: a pilot randomised controlled trial with long-term follow-up. BMJ Open 2019 Feb 15;9(2):e024685 [FREE Full text] [doi: 10.1136/bmjopen-2018-024685] [Medline: 30772854]

138. Spence SH, Holmes JM, March S, Lipp OV. The feasibility and outcome of clinic plus internet delivery of cognitive-behavior therapy for childhood anxiety. J Consult Clin Psychol 2006 Jun;74(3):614-621. [doi: 10.1037/0022-006X.74.3.614] [Medline: 16822117]
139. March S, Spence SH, Donovan CL. The efficacy of an internet-based cognitive-behavioral therapy intervention for child anxiety disorders. J Pediatr Psychol 2009 Jun;34(5):474-487. [doi: 10.1093/jpepsy/jsn099] [Medline: 18794187]
140. Spence SH, Donovan CL, March S, Gamble A, Anderson RE, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
141. Stasiak K, Merry SN, Frampton C, Moor S. Delivering solid treatments on shaky ground: Feasibility study of an online therapy for child anxiety in the aftermath of a natural disaster. Psychother Res 2018 Jul;28(4):643-653. [doi: 10.1080/10503307.2016.1244617] [Medline: 27781568]
142. Keller ML. ProQuest Search.: ProQuest Dissertations Publishing; 2010. An Internet Cognitive-Behavioral Skills-Based Program for Child Anxiety URL: https://search.proquest.com/docview/304853705 [accessed 2020-01-03]
143. Vigerland S, Ljótsson B, Thulin U, Öst LG, Andersson G, Serlachius E. Internet-delivered cognitive behavioural therapy for children with anxiety disorders: a randomised controlled trial. Behav Res Ther 2016 Jan;76:47-56 [FREE Full text] [doi: 10.1016/j.brat.2015.11.006] [Medline: 26649465]
144. Rees CS, Anderson RA, Kane RT, Finlay-Jones AL. Online obsessive-compulsive disorder treatment: preliminary results of the 'OCD? Not me!' self-guided internet-based cognitive behavioral therapy program for young people. JMIR Ment Health 2016 Jul 5;3(3):e29 [FREE Full text] [doi: 10.2196/mental.5363] [Medline: 27381977]
145. Carpenter EM, Frankel F, Marina M, Duan N, Smalley SL. Internet treatment delivery of parent-adolescent conflict training for families with an ADHD teen: a feasibility study. Child Fam Behav Ther 2004;26(3):1-20. [doi: 10.1300/j019v26n03_01]
146. Wilson DK, Sweeney AM, Law LH, Kitzman-Ulrich H, Resnicow K. Web-based program exposure and retention in the families improving together for weight loss trial. Ann Behav Med 2019 Mar 20;53(4):399-404 [FREE Full text] [doi: 10.1093/abm/kay047] [Medline: 30892641]
147. Tripicchio GL, Ammerman AS, Neshteruk C, Faith MS, Dean K, Befort C, et al. Technology components as adjuncts to family-based pediatric obesity treatment in low-income minority youth. Child Obes 2017 Dec;13(6):433-442 [FREE Full text] [doi: 10.1089/chi.2017.0021] [Medline: 28727927]
148. Büchter D, Kowatsch T, Brogle B, Dintheer-ter-Velde A, Wiegand D, Pletikosa I, et al. Home support of obese children and adolescents by means of health information technology system: a pilot study for a psychosomatic therapy concept. Swiss Med Weekly 2014;144(suppl 203):17 [FREE Full text]
149. Kowatsch T, Maass W, Pletikosa CI, Xu R. Design of a Health Information System Enhancing the Performance of Obesity Expert and Children Teams. In: Proceedings of the 2014 European Conference on Information Systems. 2014 Presented at: ECIS'14; June 9-11, 2014; Tel Aviv, Israel.
150. Xu R, Cvijikj I, Kowatsch T, Michahelles F. Tell Me What to Eat – Design and Evaluation of a Mobile Companion Helping Children and Their Parents to Plan Nutrition Intake. In: Proceedings of the 2014 European Conference on Ambient Intelligence. 2014 Presented at: AmI'14; November 11-13, 2014; Eindhoven, Netherlands. [doi: 10.1007/978-3-319-14112-1_9]
151. White MA, Martin PD, Newton RL, Walden HM, York-Crowe EE, Gordon ST, et al. Mediators of weight loss in a family-based intervention presented over the internet. Obes Res 2004 Jul;12(7):1050-1059 [FREE Full text] [doi: 10.1038/oby.2004.132] [Medline: 15292468]
152. Williamson DA, Martin PD, White MA, Newton R, Walden H, York-Crowe E, et al. Efficacy of an internet-based behavioral weight loss program for overweight adolescent African-American girls. Eat Weight Disord 2005 Sep;10(3):193-203. [doi: 10.1007/bf03327547] [Medline: 16277142]
153. Williamson DA, Walden HM, White MA, York-Crowe E, Newton RL, Alfonso A, et al. Two-year internet-based randomized controlled trial for weight loss in African-American girls. Obesity (Silver Spring) 2006 Jul;14(7):1231-1243 [FREE Full text] [doi: 10.1038/oby.2006.140] [Medline: 16899804]
154. Moore BA, O'Donohue WT. Evaluating a web-based cognitive behavioral intervention for the prevention and treatment of pediatric obesity. In: O'Donohue WT, Draper C, editors. Stepped Care and e-Health. New York, NY, US: Springer; 2011:137-149.
155. Mâsse LC, Watts AW, Barr SI, Tu AW, Panagiotopoulos C, Geller J, et al. Individual and household predictors of adolescents' adherence to a web-based intervention. Ann Behav Med 2015 Jun;49(3):371-383. [doi: 10.1007/s12160-014-9658-z] [Medline: 25270826]
156. Tu AW, Watts AW, Chanoine J, Panagiotopoulos C, Geller J, Brant R, et al. Does parental and adolescent participation in an e-health lifestyle modification intervention improves weight outcomes? BMC Public Health 2017 Apr 24;17(1):352 [FREE Full text] [doi: 10.1186/s12889-017-4220-0] [Medline: 28438202]
157. Nieto R, Hernández E, Boixadós M, Huguet A, Beneitez I, McGrath P. Testing the feasibility of DARWeb: an online intervention for children with functional abdominal pain and their parents. Clin J Pain 2015 Jun;31(6):493-503. [doi: 10.1097/AJP.0000000000000199] [Medline: 25551478]

158. Voerman JS, Remerie S, Westendorp T, Timman R, Busschbach JJ, Passchier J, et al. Effects of a guided internet-delivered self-help intervention for adolescents with chronic pain. J Pain 2015 Nov;16(11):1115-1126. [doi: 10.1016/j.jpain.2015.07.011] [Medline: 26281947]
159. Palermo TM, Wilson AC, Peters M, Lewandowski A, Somhegyi H. Randomized controlled trial of an Internet-delivered family cognitive-behavioral therapy intervention for children and adolescents with chronic pain. Pain 2009 Nov;146(1-2):205-213 [FREE Full text] [doi: 10.1016/j.pain.2009.07.034] [Medline: 19695776]
160. Long AC, Palermo TM. Brief report: web-based management of adolescent chronic pain: development and usability testing of an online family cognitive behavioral therapy program. J Pediatr Psychol 2009 Jun;34(5):511-516 [FREE Full text] [doi: 10.1093/jpepsy/jsn082] [Medline: 18669578]
161. Fales J, Palermo TM, Law EF, Wilson AC. Sleep outcomes in youth with chronic pain participating in a randomized controlled trial of online cognitive-behavioral therapy for pain management. Behav Sleep Med 2015;13(2):107-123 [FREE Full text] [doi: 10.1080/15402002.2013.845779] [Medline: 24484373]
162. Law EF, Beals-Erickson SE, Noel M, Claar R, Palermo TM. Pilot randomized controlled trial of internet-delivered cognitive-behavioral treatment for pediatric headache. Headache 2015;55(10):1410-1425 [FREE Full text] [doi: 10.1111/head.12635] [Medline: 26316194]
163. Palermo TM, Law EF, Zhou C, Holley AL, Logan D, Tai G. Trajectories of change during a randomized controlled trial of internet-delivered psychological treatment for adolescent chronic pain: how does change in pain and function relate? Pain 2015 Apr;156(4):626-634 [FREE Full text] [doi: 10.1097/01.j.pain.0000460355.17246.6c] [Medline: 25775441]
164. Palermo TM, Law EF, Fales J, Bromberg MH, Jessen-Fiddick T, Tai G. Internet-delivered cognitive-behavioral treatment for adolescents with chronic pain and their parents: a randomized controlled multicenter trial. Pain 2016 Jan;157(1):174-185 [FREE Full text] [doi: 10.1097/j.pain.0000000000000348] [Medline: 26335910]
165. Fisher E, Bromberg MH, Tai G, Palermo TM. Adolescent and parent treatment goals in an internet-delivered chronic pain self-management program: does agreement of treatment goals matter? J Pediatr Psychol 2017 Jul 1;42(6):657-666 [FREE Full text] [doi: 10.1093/jpepsy/jsw098] [Medline: 28008004]
166. Law EF, Fisher E, Howard WJ, Levy R, Ritterband L, Palermo TM. Longitudinal change in parent and child functioning after internet-delivered cognitive-behavioral therapy for chronic pain. Pain 2017 Oct;158(10):1992-2000 [FREE Full text] [doi: 10.1097/j.pain.0000000000000999] [Medline: 28777771]
167. Alberts NM, Law EF, Chen AT, Ritterband LM, Palermo TM. Treatment engagement in an internet-delivered cognitive behavioral program for pediatric chronic pain. Internet Interv 2018 Sep;13:67-72 [FREE Full text] [doi: 10.1016/j.invent.2018.07.005] [Medline: 30206521]
168. Narad ME, Minich N, Taylor HG, Kirkwood MW, Brown TM, Stancin T, et al. Effects of a web-based intervention on family functioning following pediatric traumatic brain injury. J Dev Behav Pediatr 2015;36(9):700-707 [FREE Full text] [doi: 10.1097/DBP.0000000000000208] [Medline: 26461100]
169. Wade S, Wolfe C, Brown T, Pestian JP. Putting the pieces together: preliminary efficacy of a web-based family intervention for children with traumatic brain injury. J Pediatr Psychol 2005;30(5):437-442. [doi: 10.1093/jpepsy/jsi067] [Medline: 15944171]
170. Wade SL, Carey J, Wolfe CR. An online family intervention to reduce parental distress following pediatric brain injury. J Consult Clin Psychol 2006 Jun;74(3):445-454. [doi: 10.1037/0022-006X.74.3.445] [Medline: 16822102]
171. Wade SL, Carey J, Wolfe CR. The efficacy of an online cognitive-behavioral family intervention in improving child behavior and social competence following pediatric brain injury. Rehabil Psychol 2006;51(3):179-189. [doi: 10.1037/0090-5550.51.3.179]
172. Wade SL, Walz NC, Carey JC, Williams KM. Preliminary efficacy of a web-based family problem-solving treatment program for adolescents with traumatic brain injury. J Head Trauma Rehabil 2008;23(6):369-377. [doi: 10.1097/01.HTR.0000341432.67251.48] [Medline: 19033829]
173. Wade SL, Walz NC, Carey JC, Williams KM. Brief report: description of feasibility and satisfaction findings from an innovative online family problem-solving intervention for adolescents following traumatic brain injury. J Pediatr Psychol 2009 Jun;34(5):517-522 [FREE Full text] [doi: 10.1093/jpepsy/jsn081] [Medline: 18667477]
174. Wade SL, Walz NC, Carey J, McMullen KM, Cass J, Mark E, et al. Effect on behavior problems of teen online problem-solving for adolescent traumatic brain injury. Pediatrics 2011 Oct;128(4):e947-e953 [FREE Full text] [doi: 10.1542/peds.2010-3721] [Medline: 21890828]
175. Wade SL, Walz NC, Carey J, McMullen KM, Cass J, Mark E, et al. A randomized trial of teen online problem solving: efficacy in improving caregiver outcomes after brain injury. Health Psychol 2012 Nov;31(6):767-776. [doi: 10.1037/a0028440] [Medline: 22746261]
176. Bjureberg J, Sahlin H, Hedman-Lagerlöf E, Gratz KL, Tull MT, Jokinen J, et al. Extending research on Emotion Regulation Individual Therapy for Adolescents (ERITA) with nonsuicidal self-injury disorder: open pilot trial and mediation analysis of a novel online version. BMC Psychiatry 2018 Oct 11;18(1):326 [FREE Full text] [doi: 10.1186/s12888-018-1885-6] [Medline: 30305103]
177. le Marne FA, Butler S, Beavis E, Gill D, Bye AM. EpApp: development and evaluation of a smartphone/tablet app for adolescents with epilepsy. J Clin Neurosci 2018 Apr;50:214-220. [doi: 10.1016/j.jocn.2018.01.065] [Medline: 29422360]

178. Stinson JN, McGrath PJ, Hodnett ED, Feldman BM, Duffy CM, Huber AM, et al. An internet-based self-management program with telephone support for adolescents with arthritis: a pilot randomized controlled trial. J Rheumatol 2010 Sep;37(9):1944-1952. [doi: 10.3899/jrheum.091327] [Medline: 20595280]
179. Chih M, DuBenske LL, Hawkins RP, Brown RL, Dinauer SK, Cleary JF, et al. Communicating advanced cancer patients' symptoms via the internet: a pooled analysis of two randomized trials examining caregiver preparedness, physical burden, and negative mood. Palliat Med 2013 Jun;27(6):533-543 [FREE Full text] [doi: 10.1177/0269216312457213] [Medline: 22988042]
180. Gustafson DH, DuBenske LL, Atwood AK, Chih M, Johnson RA, McTavish F, et al. Reducing symptom distress in patients with advanced cancer using an e-alert system for caregivers: pooled analysis of two randomized clinical trials. J Med Internet Res 2017 Nov 14;19(11):e354 [FREE Full text] [doi: 10.2196/jmir.7466] [Medline: 29138131]
181. DuBenske LL, Gustafson DH, Namkoong K, Hawkins RP, Atwood AK, Brown RL, et al. CHESS improves cancer caregivers' burden and mood: results of an eHealth RCT. Health Psychol 2014 Oct;33(10):1261-1272 [FREE Full text] [doi: 10.1037/a0034216] [Medline: 24245838]
182. Gustafson DH, DuBenske LL, Namkoong K, Hawkins R, Chih M, Atwood AK, et al. An eHealth system supporting palliative care for patients with non-small cell lung cancer: a randomized trial. Cancer 2013 May 1;119(9):1744-1751 [FREE Full text] [doi: 10.1002/cncr.27939] [Medline: 23355273]
183. Fergus K, McLeod D, Carter W, Warner E, Gardner S, Granek L, et al. Development and pilot testing of an online intervention to support young couples' coping and adjustment to breast cancer. Eur J Cancer Care (Engl) 2014 Jul;23(4):481-492. [doi: 10.1111/ecc.12162] [Medline: 24472013]
184. Ianakieva I, Fergus K, Ahmad S, Pos A, Pereira A. A model of engagement promotion in a professionally facilitated online intervention for couples affected by breast cancer. J Marital Fam Ther 2016 Oct;42(4):701-715. [doi: 10.1111/jmft.12172] [Medline: 27234009]
185. Ianakieva I, Fergus K, Ahmad S, Pereira A, Stephen J, McLeod D, et al. Varying levels and types of engagement in an online relationship enhancement program for couples following breast cancer. J Couple Relatsh Ther 2019;18(1):22-43. [doi: 10.1080/15332691.2018.1546248]
186. Kim Y, Shaffer K, Mitchell H, Ting A, Carver C. DYADIC INTERNET-BASED HEALTH LIFESTYLE INTERVENTION FOR COLORECTAL CANCER SURVIVORS AND THEIR FAMILY MEMBERS: PILOT STUDY. In: PSYCHOSOMATIC MEDICINE. 2017 Presented at: American Psychosomatic Medicine Society; May 2017; Sevilla, Spain p. A87.
187. Mazzer C, Shaffer K, Kim Y. Healthy lifestyle behavior intervention for colorectal cancer survivors and their families: Pilot study. In: Annals of Behavioral Medicine. 2016 Presented at: Society of Behavioral Medicine; March-April 2016; Washington, DC, US p. S231.
188. Song L, Rini C, Deal A, Nielsen M, Chang H, Kinneer P, et al. Improving couples' quality of life through a Web-based prostate cancer education intervention. Oncol Nurs Forum 2015 Mar;42(2):183-192 [FREE Full text] [doi: 10.1188/15.ONF.183-192] [Medline: 25806885]
189. Vogel RI, Petzel SV, Cragg J, McClellan M, Chan D, Dickson E, et al. Development and pilot of an advance care planning website for women with ovarian cancer: a randomized controlled trial. Gynecol Oncol 2013 Nov;131(2):430-436 [FREE Full text] [doi: 10.1016/j.ygyno.2013.08.017] [Medline: 23988413]
190. Bryant J, Sanson-Fisher R, Stevenson W, Henskens F, Smits R. Build it, but will they come? Development and patient use of an online information tool designed to reduce psychosocial distress. PsychoOncology 2013;22:157. [doi: 10.1111/j.1099-1611.2013.3394]
191. Mayberry LS, Berg CA, Harper KJ, Osborn CY. The design, usability, and feasibility of a family-focused diabetes self-care support mHealth intervention for diverse, low-income adults with type 2 diabetes. J Diabetes Res 2016;2016:7586385 [FREE Full text] [doi: 10.1155/2016/7586385] [Medline: 27891524]
192. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 7;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
193. Kamwesiga JT, Eriksson GM, Tham K, Fors U, Ndiwalana A, von Koch L, et al. A feasibility study of a mobile phone supported family-centred ADL intervention, F@ce, after stroke in Uganda. Global Health 2018 Aug 15;14(1):82 [FREE Full text] [doi: 10.1186/s12992-018-0400-7] [Medline: 30111333]
194. Nilsson A, Magnusson K, Carlbring P, Andersson G, Gumpert CH. The development of an internet-based treatment for problem gamblers and concerned significant others: a pilot randomized controlled trial. J Gambl Stud 2018 Jun;34(2):539-559 [FREE Full text] [doi: 10.1007/s10899-017-9704-4] [Medline: 28699054]
195. Dew MA, Goycoolea JM, Harris RC, Lee A, Zomak R, Dunbar-Jacob J, et al. An internet-based intervention to improve psychosocial outcomes in heart transplant recipients and family caregivers: development and evaluation. J Heart Lung Transplant 2004 Jun;23(6):745-758. [doi: 10.1016/j.healun.2003.07.002] [Medline: 15366436]
196. Shaffer KM. OSF. 2019. Dyadic eHealth Interventions: A Systematic Scoping Review URL: osf.io/f3h5k

Abbreviations

eHealth: electronic health
PICOS: Population, Intervention, Comparison, Outcome, Study
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 15.07.19; peer-reviewed by E Kent, A Efthymiou, M Lozano-Lozano; comments to author 29.09.19; revised version received 08.11.19; accepted 19.12.19; published 04.03.20.

Please cite as:
Shaffer KM, Tigershtrom A, Badr H, Benvengo S, Hernandez M, Ritterband LM
Dyadic Psychosocial eHealth Interventions: Systematic Scoping Review
J Med Internet Res 2020;22(3):e15509
URL: https://www.jmir.org/2020/3/e15509
doi:10.2196/15509
PMID:32130143

Original Paper

Patient Uptake, Experience, and Satisfaction Using Web-Based and Face-to-Face Hearing Health Services: Process Evaluation Study

Husmita Ratanjee-Vanmali[1*], MA, BSc; De Wet Swanepoel[1,2,3*], BA, MA, PhD; Ariane Laplante-Lévesque[4,5*], BSc, MSc, MA, PhD

[1]Department of Speech-Language Pathology & Audiology, University of Pretoria, Pretoria, South Africa
[2]Ear Sciences Centre, The University of Western Australia, Nedlands, Australia
[3]Ear Science Institute Australia, Subiaco, Western Australia, Australia
[4]Oticon Medical A/S, Copenhagen, Denmark
[5]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden
[*]all authors contributed equally

Corresponding Author:
De Wet Swanepoel, BA, MA, PhD
Department of Speech-Language Pathology & Audiology
University of Pretoria
Lynnwood Rd & Roper Street
Pretoria, 0001
South Africa
Phone: 27 124204280
Email: dewet.swanepoel@up.ac.za

Abstract

Background: Globally, access to hearing health care is a growing concern with 900 million people estimated to suffer from disabling hearing loss by 2050. Hearing loss is one of the most common chronic health conditions, yet access to hearing health care is limited. Incorporating Web-based (voice calling, messaging, or emailing) service delivery into current treatment pathways could improve access and allow for better scalability of services. Current electronic health studies in audiology have focused on technical feasibility, sensitivity, and specificity of diagnostic hearing testing and not on patient satisfaction, experiences, and sustainable models along the entire patient journey.

Objective: This study aimed to investigate a hybrid (Web-based and face-to-face) hearing health service in terms of uptake, experience, and satisfaction in adult patients with hearing loss.

Methods: A nonprofit hearing research clinic using online and face-to-face services was implemented in Durban, South Africa, using online recruitment from the clinic's Facebook page and Google AdWords, which directed persons to an online Web-based hearing screening test. Web-based and face-to-face care pathways included assessment, treatment, and rehabilitation. To evaluate the service, an online survey comprising (1) a validated satisfaction measurement tool (Short Assessment of Patient Satisfaction), (2) a process evaluation of all the 5 steps completed, and (3) personal preferences of communication methods used vs methods preferred was conducted, which was sent to 46 patients who used clinic services.

Results: Of the patients invited, 67% (31/46) completed the survey with mean age 66 years, (SD 16). Almost all patients, 92% (30/31) reported that the online screening test assisted them in seeking hearing health care. Approximately 60% (18/31) of the patients accessed the online hearing screening test from an Android device. Patients stayed in contact with the audiologist mostly through WhatsApp instant messaging (27/31, 87%), and most patients (25/31, 81%) preferred to use this method of communication. The patients continuing with hearing health care were significantly older and had significantly poorer speech recognition abilities compared with the patients who discontinued seeking hearing health care. A statistically significant positive result (P=.007) was found between age and the number of appointments per patient. Around 61% (19/31) of patients previously completed diagnostic testing at other practices, with 95% (18/19) rating the services at the hybrid clinic as better. The net promoter score was 87, indicating that patients were highly likely to recommend the hybrid clinic to friends and family.

Conclusions: This study applied Web-based and face-to-face components into a hybrid clinic and measured an overall positive experience with high patient satisfaction through a process evaluation. The findings support the potential of a hybrid clinic with synchronous and asynchronous modes of communication to be a scalable hearing health care model, addressing the needs of adults with hearing loss globally.

(J Med Internet Res 2020;22(3):e15875) doi:10.2196/15875

KEYWORDS

audiology; hearing loss; internet-based intervention; patient outcome assessment; patient satisfaction; telemedicine; text messaging; eHealth; mHealth; social media; patient-centered care

Introduction

Background

Globally, access to hearing health care (HHC) is a significant challenge affecting 466 million people, and this number is expected to rise to 900 million people by 2050, who are estimated to have disabling hearing loss [1]. The limited access to HHC results in most affected persons to live with untreated hearing loss, which has far-reaching consequences for individuals and the society at large [2]. Untreated hearing loss affects health, independence, well-being, and employment opportunities and is associated with social isolation, depression, and an increased risk of dementia [3-8]. Alongside recent estimates of a global cost of US $750 billion to hearing loss [1], this chronic condition is now recognized as a significant public health concern [9,10].

Hearing Health Care Models

Traditional HHC service delivery models focus on face-to-face, clinic-based testing, hearing aid or device fittings, counseling, and rehabilitation requiring several patient visits. HHC can be made more accessible through scalable models of care that capitalize on global trends in connectivity and technology [11]. For example, by the end of 2018 there were 5.1 billion mobile subscribers, which represents 67% of the global population, and 3.6 billion mobile device internet users, which accounts for 47% of the global population [12].

The use of these telecommunication and information technologies in medicine is called telemedicine or telehealth; in the field of ear and hearing health, the terms tele-otology and tele-audiology are also used [13]. Owing to the lack of consistency and confusion, many professionals have adapted their own term, ie, electronic health (eHealth), telehealth, tele-audiology, and now eAudiology are all terms that are often used interchangeably to describe the dissemination of health or hearing health services using the internet [14]. Although tele-practice was initially intended for services to be delivered to individuals at a distance, where patients could not interact with health professionals or the patient and the health professional were at two different locations, a newer approach is to provide HHC to the patient who may be close in distance to the health professional but chooses tele-practice as a service delivery option out of convenience [14]. Telehealth relies on access to the internet, and while some communities may have limited access, connectivity is rapidly expanding [12,15].

TeleHealth in Hearing Health Care

There is a growing body of evidence on the use of telehealth in HHC, including screening [16,17] diagnostic assessment [18,19], hearing aid fitting [20,21], and rehabilitation [22,23]. Studies to date have tested the use of tele-audiology at specific points along the patient journey and have mostly been proof-of-concept studies [13,15,24,25] that have not translated into sustainable telehealth practices [24]. There is a significant need to not only evaluate service delivery models that incorporate telehealth approaches along the patient journey in terms of effectiveness and efficiency but also to establish patient acceptance and satisfaction [13]. Measuring patient outcomes is important, as positive outcomes indicate improvements on patient satisfaction, adherence, and health status [26]. This therefore highlights the need for measuring patient satisfaction.

A dearth of evidence on patient satisfaction when using telehealth HHC services is apparent [13], as only a few studies report on patient satisfaction with tele-audiology. In one study, patients who had their hearing aids fitted remotely were followed up upon, and a high level of patient satisfaction was noted [21]. In another study, there was no difference in terms of the hearing aid benefit between in-person and tele-audiology hearing aid services [22]. In these 2 studies, patient satisfaction with tele-audiology was measured only once, and the measurement was limited to treatment outcomes, rather than an indication of the process of receiving HHC services through a different service delivery medium.

Offering hearing services completely online along the entire patient journey is challenging. Online components were selected based on validated and evidence-based tools, which would not compromise the quality of patient care (eg, online hearing screening, communication by phone and WhatsApp, and online rehabilitation). These components (eg, video-otoscopy, audiological diagnostic evaluation, and real-ear measurements) were included in face-to-face appointments as online alternatives were not yet available at the conception of this study. The model is described further in the following section and in the study by Ratanjee-Vanmali et al [27].

In a previous study, we reported on the behaviors of participants who failed the online hearing screening test. Approximately 25% (13/51) of participants proceed from motivational engagement to diagnostic testing and the remainder 75% (38/51) do not transition for the following reasons: unanswered phone call, 45% (17/38), further investigation (curious about the online hearing screening test or owns hearing aids but wants a confirmation of hearing loss, 29% (11/38), incorrect contact details, 8%(3/38), doctor did not advocate for further treatment, 8%(3/38), limited finances, 8% (3/38), and beyond the test geolocation, 3% (1/38) [27]. Therefore, this highlights the need to understand patient experience, satisfaction, and engagement in seeking HHC through such a hybrid model and which components encourage them to continue to seek HHC.

Objective of Study

This study aimed to describe a process evaluation of HHC through a hybrid clinic combining online and face-to-face services [27], with a focus on patient uptake, experience, and satisfaction.

Methods

Data Collection Procedure

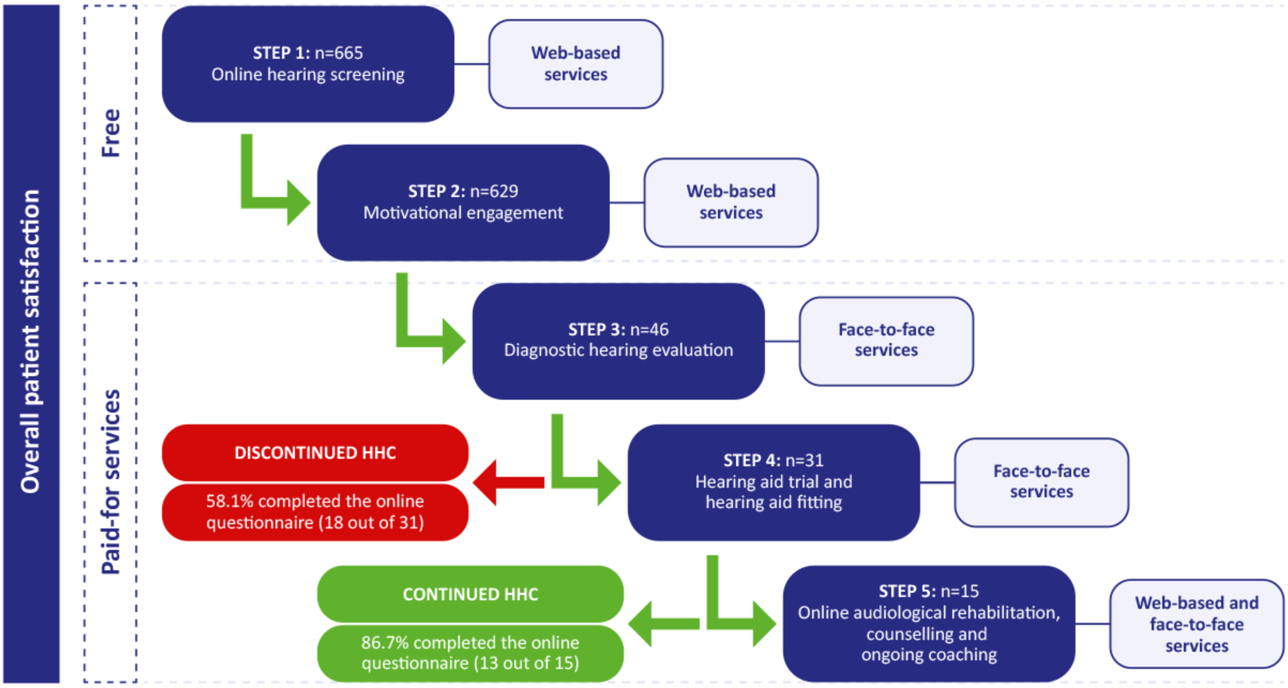
The institutional review board approved the research (GW20170409HS).

Hybrid Service Delivery Model

This research project established a nonprofit hearing research clinic [28] in Durban, South Africa. The clinic relied on online patient recruitment, offering a free online hearing screening. Online recruitment using Facebook and Google was used to target adults aged ≥40 years within the target geolocation from the clinic's social media account. Although the typical age for first-time hearing aid users is 74 years [29], the motivation for advertising to a younger audience was to reach the children of the parents aged 65 to 75 years who would resonate with the advertisements and share or encourage their family members to complete the online screening test. Advertisements (images and videos), articles, and blogs were created and used on the clinic's Facebook page regarding the importance of HHC and knowing one's hearing status or ability, and Google AdWords related to hearing test, audiologist, and tinnitus were used.

Upon completion, patients could opt to provide their contact details to be contacted by the clinic. If patients contacted the clinic without taking the online hearing screening test, they received a link encouraging them to complete the online test. At the beginning of every face-to-face appointment, the clinic audiologist verified the completion of the online hearing screening test. Asynchronous and synchronous online communication, as well as face-to-face communication supporting screening, diagnostics, hearing aid fitting, rehabilitation, and continuous monitoring and coaching, were offered. In total, 5 steps were included in the patient journey (Figure 1). The first 2 steps in the model (ie, Web-based hearing screening and motivational engagement, see the following section) were free. Participants paid for the 3 final steps, with some of the participants having access to reimbursement through their health insurance.

Figure 1. Five steps in a hybrid (Web-based and face-to-face) hearing health care (HHC) service-delivery model.



Steps in the Hybrid Hearing Health Care Delivery Model

Step 1: Online Hearing Screening—Web-Based

The online hearing screening test is an adaptive triple digit–in–noise test developed and validated for South African English that determines a speech reception threshold [30,31]. The online hearing screening test that comprised 23 user entries was provided as a software-enabled Web widget [32] hosted on the clinic's website.

When beginning the online hearing screening test, each participant was required to provide their date of birth. For each participant completing the online hearing screening test, the signal-to-noise ratio (SNR) where 50% of digits are recognized correctly, was recorded. The geolocation was also provided, which helped verify whether participants were within the geolocation of the test, ie, the greater Durban area. The pass or fail threshold of the online hearing screening test was based on optimal sensitivity and specificity to a 4-frequency pure tone average at 0.5, 1, 2, and 4 kHz ≤25 dB HL in the better ear.

On completion of the screening test, individuals were informed of their result in terms of pass or fail. Individuals could share their contact details if they wanted the clinic audiologist to contact them. The online hearing screening test results were stored in mHealth Studio Cloud; even if individuals did not share their contact details, the result was stored with an accurate geolocation which ensured that only data from the target location was used in the analysis [32]. Only the clinic audiologist had access to the password-protected mHealth Studio Cloud [32].

Step 2: Motivational Engagement—Web-Based

This step consisted of a phone call or WhatsApp message thread where the clinic audiologist assessed the readiness to book a face-to-face diagnostic hearing evaluation and provided motivational engagement.

Individuals who shared their contact details received an email with the clinic audiologist's contact details, motivational engagement questions, and suitable times and dates for a phone call.

Readiness measurement and motivational engagement consisted of 2 validated tools: the line and staging algorithm that were used with the participant over the phone. The line is a single-item measure to assess readiness for hearing help-seeking in one question: "How important is it for you to improve your hearing right now?" Responses are recorded on a Likert scale from 0 to 10, where 0 indicates not at all and 10 indicates very much [33,34]. The staging algorithm is also a single-item question assessing the stages of change with 4 possible answers, each corresponding with a stage of change: (1) I do not think I have a hearing problem, and therefore nothing should be done about it (precontemplation); (2) I think I have a hearing problem. However, I am not yet ready to take any action to solve the problem, but I might do so in the future (contemplation); (3) I know I have a hearing problem, and I intend to take action to solve it soon (preparation); and (4) I know I have a hearing problem, and I am here to take action to solve it now (action) [35]. When participants scored above 5 on the Likert rating scale of 0 to 10 [33,34] and scored 3 or 4 in the staging algorithm [35], a face-to-face visit for the comprehensive hearing evaluation was scheduled. Higher ratings indicate greater readiness to take action.

Step 3: Diagnostic Hearing Evaluation—Face-to-Face

This step consisted of a face-to-face appointment where the clinic audiologist completed a battery of assessments including an in-depth case history, video-otoscopy, acoustic reflexes, pure tone audiometry (air and bone conduction), and speech audiometry. If no red flag (eg, sudden onset of hearing loss, middle ear pathology, and asymmetrical hearing loss, sudden onset of tinnitus, aural fullness, and vertigo) suggesting a medical referral was raised, a hearing aid trial was recommended.

Step 4: Hearing Aid Trial and Fitting—Face-to-Face

A successful hearing aid trial entailed that the patient acquired hearing aids fit according to their personalized gain setting, signal processing, noise management system, automatic systems, style, and color. During the hearing aid trial, a receiver in the ear with domes chosen to meet acoustic requirements was fit to meet the patient's audiological profile. Patients were then offered a choice to opt for the style of hearing aids they preferred once counselled on the acoustic performance, physical characteristics of the available hearing aids, and personal needs (from in-the-ear custom options to behind-the-ear hearing aids). Trial hearing aids were then fit and customized to the audiometric profile of the patient using real-ear measurements to take individual ear canal properties into account.

Step 5: Audiological Rehabilitation, Counselling, and Ongoing Coaching—Web-Based and Face-to-Face

All patients who acquired hearing aids were offered an online audiological rehabilitation program [36] and the clinic audiologist coached them routinely.

The online audiological rehabilitation program consisted of 5 modules (becoming a successful hearing aid user; understanding my own hearing loss; handling my hearing aids; managing difficult communication situations; and communicating my own hearing loss) that are a combination of videos, tasks, testimonials, and reading assignments. The completion of a module would unlock the next module. The 5 modules were completed all at once or weekly as per the patient's availability. Through prerecorded videos, a coach guided the participants through the different modules and components. More information regarding the hybrid clinic has been reported elsewhere [27].

Materials

Online Questionnaire

An online questionnaire was used to determine the experience and satisfaction of patients seeking and receiving HHC using the hybrid clinic, incorporating online and face-to-face services. The online questionnaire was hosted and administered by Qualtrics (Provo, Utah) [37]. The responses were password protected and only accessible to the clinic audiologist. This closed survey was only administered to patients who provided consent to partake in the study. Participation was voluntary; no incentives were offered to encourage completion of the questionnaire.

The questionnaire (Multimedia Appendix 1) consisted of 3 sections totaling 41 questions for the group that discontinued HHC (exited at step 3) and 43 questions for the group that continued with HHC (exited at step 5). The 3 sections consisted of (1) a validated satisfaction measurement tool (Short Assessment of Patient Satisfaction [SAPS] [38]; (2) a process evaluation of all the 5 steps (online hearing screening, motivational engagement, diagnostic hearing evaluation, hearing aid trial and fitting, and online rehabilitation together with counseling and ongoing coaching) completed as seen in Figure 1; and (3) personal preferences of communication methods used versus methods preferred and HHC experiences compared with previous care, which were sent to 46 patients who used clinic services. Reporting of the questionnaire was separated into 2 overall sections: (1) evaluation of the steps and (2) patient experiences and satisfaction with the hybrid service delivery model.

The online questionnaire included a process evaluation, recorded on a 5-point Likert scale, which evaluated all the 5 steps (Figure 1). The method (Multimedia Appendix 2) was inspired by Linnan and Steckler [39]. They propose how to design and implement a process evaluation by creating the inventory of process objectives based on theory; reaching a consensus of the questions to be answered by the stakeholders of the project; identifying and creating the measurement tools; designing, implementing, and administering quality control; collecting, managing, and cleaning data; analyzing data; reporting findings; and refining interventions, measurements, and the analysis tool [39]. The process evaluation questionnaire was developed to include all aspects of the hybrid service delivery model which included and excluded a clinician's involvement, where no systematic differences were found in the ratings. Closed and open-ended questions on patient experiences and preferences related to the hybrid clinic services compared with a traditional model were surveyed along with communication methods used and those preferred. Answers to open-ended questions were analyzed using an inductive thematic analysis that was conducted by the first author and then reviewed by an independent researcher [40].

Overall Satisfaction

SAPS [38] assesses overall patient satisfaction with 7 items targeting treatment, explanation of treatment results, clinician care, participation in medical decision making, respect by the clinician, time with the clinician, and satisfaction with hospital or clinic care. The questionnaire has been validated in clinical settings and has good internal and test-retest reliability [38]. For this study, the questionnaire was tailored to audiology by replacing the term doctor or other health professional to audiologist [41]. The minimum score is 0, and the maximum score is 28, where higher scores indicate greater satisfaction. Typical total SAPS scores from other research reported mean scores of 22 (SD 5) and 8 (SD 4) [38,42].

The net promoter score (NPS) is a single question about willingness to recommend a product or service that companies commonly use [43]: "On a scale from 0-10, how likely are you to recommend this clinic (Hearing Research Clinic NPC) to your friends and family?" A follow-up question asking respondents to explain the rating followed. The NPS is calculated by classifying the respondents into promoters (9-10), passives (7-8), and detractors (≤6). The NPS is obtained by subtracting the percentage of detractors from the percentage of promoters [43].

Procedures

The online questionnaire was sent via email to 46 patients who completed the diagnostic hearing evaluation (steps 1-3). Data for each patient gathered from their files were linked by their email address and then hidden to ensure anonymity and were issued patient numbers during data analysis. Data were collected over 3 months (December 2018-February 2019); patients had sought help from the clinic during a period of 19 months (June 2017-January 2019). All patients who completed the survey were in contact with the clinic within 6 months of completing the online questionnaire.

The initial email invitation was sent on December 4, 2018, and a WhatsApp message was sent prompting patients to check their email mailboxes for the questionnaire. Up to seven reminder messages (email and WhatsApp) were sent to nonresponses over 12 weeks.

Patient data were stored in two locations: (1) a cloud-based system for appointment times and notes and (2) a server-based system for diagnostic results. Both systems were password protected and only accessible by the clinic audiologist.

Participants

Purposive sampling was used to collect patients' experiences and satisfaction of the hybrid clinic services. Patients who failed the online hearing screening provided consent to be contacted by the clinic audiologist before submitting their details. Written consent to partake in the study was provided during the face-to-face diagnostic hearing evaluation (step 3).

Statistical Analysis

Data were analyzed using SPSS Inc, version 25 (IBM Corp, Chicago, Illinois) [44]. Statistical significance was set at $P<.05$. The Shapiro-Wilk test (nonparametric test) was used to test normality, which confirmed that the data were not normally distributed. Cronbach alpha was used to test the internal validity of the entire process evaluation questionnaire.

Results

Characteristics of Online Seekers of Hearing Health Care

The reporting of questionnaire results is in accordance, as far as possible, with the Checklist for Reporting Results of Internet E-Surveys [45].

A total of 665 participants completed the online hearing screening test and submitted their details for further HHC services during this evaluation period. A total of 629 participants were contacted by telephone or WhatsApp for motivational engagement; a few were unreachable owing to incorrect details submitted. Out of the 629 participants contacted, 46 (7%) became patients of the clinic and sought HHC services (Figure 1). Of the 46 patients invited, 31 (67%) completed the online survey and were aged between 35 and 101 years (mean 66, SD 16), the majority, 58% (n=18) being men. On average, patients had experienced hearing difficulties for 13 years (SD 15) and presented with an average speech reception threshold of −3.0 dB SNR (SD 8). The online questionnaire was internally consistent and reliable; Cronbach alpha values were between 0.70 and 0.77, where a value above 0.70 was considered acceptable [46]. No significant differences were found on the Mann-Whitney U test between the responder (n=31) and nonresponder (n=15) groups in terms of age, gender, SNR, the line, staging algorithm, years aware of hearing loss, and devices used to complete the online hearing screening test.

Process Evaluation of 5 Steps in Hybrid Hearing Health Care Delivery Model

Step 1: Online Hearing Screening—Web-Based

Patients (N=31) accessed the online hearing screening from Android (18/31, 58%), iOS (9/31, 29%), and Windows PC (4/31, 13%) devices. The majority of patients agreed or strongly agreed that the online hearing screening was simple to complete (24/25, 96%), was quick and informative (23/26, 88%), was easy to use (23/26, 89%) and assisted them to continue HHC (24/26, 92%; Table 1).

Table 1. Patient evaluation of the online hearing screening test.

Questions related to the online hearing screening test	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
Taking the online test was simple (N=25)	0 (0)	1 (4)	0 (0)	14 (56)	10 (40)
Taking the online test was quick (N=26)	0 (0)	0 (0)	3 (12)	17 (65)	6 (23)
Taking the online test was informative (N=26)	0 (0)	0 (0)	3 (12)	17 (65)	6 (23)
I found this online test easy to use (N=26)	0 (0)	1 (4)	2 (8)	15 (58)	8 (31)
I thought the online test was fast (N=26)	0 (0)	5 (19)	7 (27)	9 (35)	5 (19)
The test result seemed reliable (N=26)	0 (0)	1 (4)	2 (8)	15 (58)	8 (31)
Online test has helped me to take the next steps to improve my hearing (N=26)	0 (0)	0 (0)	2 (8)	10 (39)	14 (54)

Step 2: Motivational Engagement—Web-Based

Patients agreed and strongly agreed that the mode of communication was easy (26/26, 100%), quick (27/27, 100%), provided useful (26/26, 100%) and relevant (25/26, 96%) information, assisted in taking the next step (25/26, 96%), and assisted in booking the diagnostic hearing evaluation (27/28, 96%); Table 2).

Table 2. Patient evaluation of motivational engagement using a voice call/messaging (WhatsApp).

Questions related to a voice call/messaging (WhatsApp)	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
The phone call/WhatsApp message was informative (N=26)	0 (0)	0 (0)	0 (0.)	15 (58)	11 (42)
The phone call/WhatsApp message was an easy way for me to communicate with the audiologist/clinic (N=26)	0 (0)	0 (0)	0 (0)	15 (58)	11 (42)
The phone call/WhatsApp message helped me in taking the next step (N=26)	0 (0)	0 (0)	1 (4)	11 (42)	14 (54)
The phone call/WhatsApp message provided me with relevant information regarding my hearing (N=26)	0 (0)	0 (0)	1 (4)	14 (54)	11 (42)
The phone call/WhatsApp message helped me to take the next step and book my hearing evaluation consultation (N=28)	0 (0)	0 (0)	1 (4)	12 (43)	15 (54)
The phone call/WhatsApp message was a quick way for me to communicate with the audiologist/clinic (N=27)	0 (0)	0 (0)	0 (0)	14 (52)	13 (48)

Patients communicated with the clinic using WhatsApp messaging (27/31, 87%), emails (25/31, 81%), voice calls (24/31, 77%), text messages (4/31, 13%), and Facebook Messenger (2/31, 7%). The majority of patients preferred the following methods of communication with the clinic audiologist: WhatsApp messaging (25/31, 81%), email (20/31, 65%), or voice calls (19/31, 61%).

Step 3: Diagnostic Assessment—Face-to-Face

Patients attending face-to-face diagnostic appointments agreed and strongly agreed that the test was comprehensive (31/31, 100%), provided the information needed (31/31, 100%), was easy to complete (31/31, 100%), and was trustworthy (31/31, 100%) with sufficient time spent taking it (31/31, 100%; Table 3).

Table 3. Patient evaluation of the diagnostic hearing evaluation.

Questions related to the diagnostic hearing evaluation	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
The diagnostic hearing test was comprehensive (N=31)	0 (0)	0 (0)	0 (0)	13 (42)	18 (58)
The audiological consultation provided me with the information I needed (N=31)	0 (0)	0 (0)	0 (0)	13 (42)	18 (58)
The diagnostic hearing test was an easy test to complete with the guidance from the audiologist (n=30)	0 (0)	0 (0)	0 (0)	9 (30)	21 (70)
It was beneficial to have a hearing aid trial option available after my diagnostic hearing test (in the first consultation; N=28)	0 (0)	0 (0)	1 (4)	6 (21)	21 (75)
It was easy to use the hearing aid during the trial period offered to me (N=28)	0 (0)	0 (0)	2 (7)	8 (29)	18 (64)
I trust the results from my diagnostic hearing test (N=30)	0 (0)	0 (0)	0 (0)	11 (37)	19 (63)
The time spent on my diagnostic hearing test was adequate (N=31)	0 (0)	0 (0)	0 (0)	13 (42)	18 (58)

More than half of the patients (19/31, 61%) had previously completed a diagnostic hearing evaluation (step 3) at other practices. In comparison with previous experiences, one person rated the hybrid clinic as the same while the other 18 patients rated their experiences as better.

From the open-ended responses, two main themes emerged for the differences between prior experiences and the hybrid clinic: clinician engagement and technology. Clinician engagement included aspects of personal attention, patience, dedication, thorough explanations, professional behavior, exceeding expectations, friendliness, and trust. Technology included aspects of the latest technology and equipment and offering trial hearing aids.

Step 4: Hearing Aid Trial and Fitting—Face-to-Face

Patients agreed and strongly agreed that a hearing aid trial helped to experience the difference that hearing aids can make in their life (26/27, 96%). All patients who acquired their hearing aids (steps 4-5) agreed and strongly agreed that the hearing aid trial and its usage was beneficial (Table 4).

Table 4. Patient evaluation of the hearing aid trial and fitting.

Questions related to the hearing aid trial and fitting	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
The hearing aid trial helped me experience the difference hearing aids can make in my life (N=27)	0 (0)	0 (0)	1 (4)	5 (19)	21 (78)
The opportunity to try hearing aids helped me make an informed decision to buy hearing aids (N=13)	0 (0)	0 (0)	0 (0)	5 (39)	8 (62)
I felt it was easy to use the hearing aids in the trial period which gave me the confidence in my ability to use it on my own (N=13)	0 (0)	0 (0)	0 (0)	5 (39)	8 (62)
I trust that the hearing aids will assist me to hear better in my daily life (N=13)	0 (0)	0 (0)	0 (0)	4 (31)	9 (69)
The time I had to trial the hearing aids in my daily life (home/work) was adequate (N=13)	0 (0)	0 (0)	0 (0)	3 (23)	10 (77)
My quality of life has improved by using my hearing aids (N=13)	0 (0)	0 (0)	0 (0)	4 (31)	9 (69)

Of those patients who were fitted with hearing aids (steps 4-5), the majority (6/9, 67%) complimented the service received, were satisfied with the care offered, and did not have suggestions for service improvements. Reasons for patients not continuing with HHC (11/18, 61%) included cost as a prohibitive factor (7/18, 39%), concerns regarding the stigma of wearing hearing aids (3/18, 17%), and belief that the hearing loss was not severe enough to warrant the use of hearing aids (3/18, 17%). One person suggested a financing option to make hearing aids more affordable.

Step 5: Audiological Rehabilitation—Web-Based and Face-to-Face

Except for 1 person, all patients agreed and strongly agreed that the online audiological rehabilitation was helpful (8/9, 89%). In addition to the program, support was offered to patients as required both online and by face-to-face methods.

Overall Satisfaction—Web-Based and Face-to-Face Clinic Services

The mean SAPS score of the 31 patients reported was 26 (SD 3; Table 5). There were only 3 instances where 1 patient was unsure (neither satisfied nor dissatisfied) regarding his or her satisfaction in terms of the effect of the HHC treatment, choices available to the patient, and dissatisfaction with the care received. In total, 3 patients (3/31, 10%) felt that the time with the clinic audiologist was too short.

The NPS score was 87, which indicates that patients are highly likely to recommend the clinic to friends and family. The majority of patients (21/31, 68%) provided reasons for their rating including competence, result-driven exceptional service (11/31, 35%), tailored service (4/31, 13%), and reliable and efficient service (2/31, 6%).

The three most important reasons for continuing with HHC services with the hybrid clinic were as follows: personalized care and understanding audiologist, who is patient and accommodating (11/31, 36%); confidence in the audiologist, kind, caring, helpful, caring, efficient (8/31, 26%); and technical knowledge of the product and equipment (5/31, 17%).

A significant positive correlation was found between age and the number of appointments (r=0.367; P=.007) and a positive but not significant correlation (r=0.216; P=.12) was reported between age and the number of support instances.

Table 5. Overall Short Assessment of Patient Satisfaction scores categorized according to "Very dissatisfied," "Dissatisfied," "Satisfied," and "Very Satisfied" for patients who sought hearing health care (N=31).

SAPS[a] category	Range of score	Frequency (%)
Very dissatisfied	0-10	0 (0)
Dissatisfied	11-18	1 (3)
Satisfied	19-26	14 (45)
Very satisfied	27-28	16 (52)

[a] SAPS: Short Assessment of Patient Satisfaction.

Discussion

Hybrid Hearing Health Care Delivery Model

This study provides insights into a hybrid service delivery model that assessed adult patients' perspectives on online and face-to-face services offered. An asynchronous Web-based hearing screening successfully recruited patients seeking HHC online. Patient experiences with this online screening test were positive and, together with motivational engagement, were rated as time-efficient, valuable, and supporting continuation with HHC. This study employed nonopportunistic testing as participants actively opted to visit the website and complete the online hearing screening test. The potential reasons for mixed findings on the ease of testing, which comprised 23 user entries, could be due to the internet speeds (Wi-Fi or mobile data 3G or 4G) in terms of wait time when loading the widget on mobile devices or computers, proficiency with the digital device, or the actual test duration. Hearing screening tests are typically offered in isolation, and longitudinal studies show that a significant percentage of people do not follow-up with diagnostic measures and rehabilitation [47-49]. Approximately 75% (38/51) of patients who failed an online hearing screening test did not continue with HHC as reported in our previous study [27]. Another study reported that older adults who were considering or preparing to take action for their hearing loss were willing to access online HHC and that a simple user interface and short-term training may optimize the usability of online HHC programs for them [50]. In line with this, this study offered hybrid diagnostic and rehabilitative HHC services directly following the hearing screening. This is the first report to perform a process evaluation of a hybrid model of HHC. Previous reports focused on the validation of these tools and not on patient experiences [47,51].

WhatsApp messaging was rated highly, and patients were satisfied with this mode of communication. Patients used and preferred WhatsApp messaging as the primary communication method with the clinic where a dedicated mobile phone with WhatsApp, phone calls, and email was set up for this hybrid clinic. In other health professions, physicians have successfully incorporated WhatsApp into clinical practice with no need for further training or technical competency building [52].

The advantages and disadvantages of using WhatsApp in clinical practice are well documented within health care [53]. However, there is no uniformity in the usage of WhatsApp, as a recent study reports that doctors were more likely to use WhatsApp in patient communication or share information with colleagues than nurses [54]. Research evidence suggests that WhatsApp can be a promising tool that allows health professionals and the general public to communicate or allows communication among health care professionals themselves to compare and learn from each other [55]. There is still a need for high-quality research to evaluate the value and risks of using it as a health communication tool [54,55].

The diagnostic hearing evaluation (step 3) was an integral step to establish a therapeutic relationship in this hybrid model. A strength of this model was that the therapeutic relationship already commenced before the face-to-face appointment (step 3) and was continued through the patient's HHC journey with the same clinic audiologist either online or in-person. The benefit of clinician continuity is mixed; in a physician environment, seeing a known provider is found to be beneficial in terms of a cost-benefit factor [56], whereas in an audiological setting, no difference was noted on hearing aid outcomes when patients are attended to by different clinicians [57].

Previous tele-audiology studies have taken steps toward investigating patient satisfaction within remote hearing aid fittings [21], services [22], and programming and fitting [58] with reasonable patient satisfaction noted. However, the first 2 studies reported findings based on standardized hearing aid outcome measures (International Outcome Inventory for Hearing Aids and Satisfaction with Amplification in Daily Living) rather than a process evaluation of patient experiences and satisfaction with HHC [17,18]. The last study [58] measured patient experience satisfaction using a validated questionnaire and found that patient satisfaction with hearing aid programming and fitting via tele-audiology versus face-to-face was the same.

The online audiological rehabilitation offering was reported as a positive addition to this hybrid clinic's services. eHealth might be a viable option to offer tele-audiology services to both adult patients and their significant other as they already use internet-connected technologies to access health care, and this could promote patient-centered care from a biopsychosocial context [59]. Telehealth interventions for audiology are expanding, and research conducted on audiological, vestibular, and tinnitus rehabilitation show promising results [25].

Overall Satisfaction

Patient satisfaction in this study, which used 5 steps in a hybrid HHC service delivery model, was found to be higher than previously published SAPS data. In this study, the SAPS mean score was 26 (SD 3) as compared with findings from an incontinence clinic (mean SAPS score 22, SD 5) [38] and a psychiatry clinic (mean SAPS score 8, SD 4) [42]. The NPS score in this study was high (87) in comparison with an NPC score of 52 in a study of 728 patients who rated their satisfaction with synchronous videos across the health department [60]. NPS scores from another health field in the National Health System in the United Kingdom reported the following scores, however the response scale was slightly altered from the original: joint replacement was 60 with individual scores for total hip replacement and total knee replacement of 71 and 49, respectively [61]. Other researchers have highlighted the attractiveness of adapting the NPS for health care as it is less reliant on the literacy of responders, limited resources are needed to adapt the tool, and it provides more valuable information than a binary yes or no scale [62].

The audiologist's clinical engagement and professional services were identified as essential components in the positive patient experiences in this study. Previous research also indicates that patients prefer patient-centered interactions with a health professional, and this is associated with high satisfaction [63]. Offering patient-centered care has also been proposed as a way to improve hearing aid adoption [64].

Even though 61% (19/31) of patients experienced previous HHC services from other audiologists or clinics, 95% (18/19) rated the services offered in this hybrid clinic more favorably. Patient experience and satisfaction were equally high and positive in both online and face-to-face service offerings in this hybrid clinic. However, there is still a paucity of evidence regarding the uptake of eHealth HHC, its effectiveness, and the satisfaction of patients using such service delivery models. As technology evolves, so will the continuum of direct-to-consumer and traditional face-to-face models. This study applied online and face-to-face components into a hybrid clinic and measured high patient satisfaction through a process evaluation. This model still required the need for 1 or 2 face-to-face appointments with the audiologist compared with more traditional clinical pathways. The fact that older patients needed more appointments may indicate that more audiological support is needed in the initial stages of adapting to hearing aids where additional support could be offered using asynchronous methods. This study provides initial evidence that can support audiologists who are limited in numbers but are required to provide services to a large area. This model may also provide patients with an alternative service delivery model, who could benefit from a combination of online and face-to-face appointments. Individual audiologists can customize this hybrid model to meet the needs of their patient demographic and for those patients willing to seek HHC differently.

This study offered individuals searching for HHC within the target location with an online hearing screening test as the first action to initiate care. Combining online and face-to-face communication methods also allowed patients to stay in touch with the audiologist when needed. Patients paid for their HHC services, removing volunteer biases and highlighting the potential of this model to translate into a scalable clinical practice. However, patients who pay for their hearing aids could introduce another bias or view services as more favorable, and this could be considered a limitation. Another limitation of the study is the lack of a comparator to establish whether this hybrid model was better or worse as compared with more traditional HHC delivery models where satisfaction could be measured as being similar in face-to-face only with no online services employed. The fact that the same person served as the clinician and as the researcher collecting the online questionnaires and that patients could have been influenced to provide favorable ratings (social desirability bias) could also be considered limitations in this study. The completion of the questionnaire is also vulnerable to both nonresponse bias (15/46, 33% of patients did not respond to the questionnaire) and recall bias. It is not possible to separate the influence of the audiologist's skills versus the hybrid model when analyzing the patients' satisfaction with the care received. This study also had a relatively small sample within a defined area of South Africa that required patients to have internet access and the necessary digital skills to complete an online hearing screening test, which limits generalizability. Future studies in modifications to the service delivery models would benefit from a comparator group designed into research studies and to test mobile and computer proficiency and the effects of age on the uptake of HHC in such a hybrid model. Another future consideration would be to document the long-term effects in terms of economic viability and scalability of such a model. This hybrid model is the first concept to be tested, and we foresee modifications to this service delivery model made possible in the future when technology advances to facilitate more audiological services remotely to meet the needs of the patient and the audiologist.

Conclusions

In conclusion, the positive patient experience and satisfaction demonstrates the potential of hybrid online and face-to-face HHC to meet patient needs. Sustainable and scalable service delivery models that incorporate eHealth are required to meet the challenges of untreated hearing loss globally.

Acknowledgments

The authors wish to thank Thomas Behrens (Oticon A/S) and Lezanie Tietze (Oticon South Africa) for the fruitful discussions. They also thank Dr Marien Graham (University of Pretoria) for her statistical assistance and Lukas Tietz (Eriksholm Research Centre, Oticon A/S) for his technical assistance with the online audiological rehabilitation program.

This work was supported by the National Research Foundation of South Africa under the grant number 107728. A grant from the William Demant Foundation supports the larger project. The first author is supported by the National Research Foundation and the University of Pretoria PhD support program. The financial assistance of the National Research Foundation toward this research is hereby acknowledged. The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the National Research Foundation.

Conflicts of Interest

DWS has a relationship with the hearX Group (Pty) Ltd, that includes equity, consulting and potential royalties.

Multimedia Appendix 1
Online questionnaire. Patient experience and satisfaction with hearing health care received.
[DOCX File , 1176 KB - jmir_v22i3e15875_app1.docx]

Multimedia Appendix 2
Framework for the process evaluation of Web-based and face-to-face services offered in the hybrid hearing health care model.
[DOCX File , 18 KB - jmir_v22i3e15875_app2.docx]

References

1. World Health Organization. 2019 Mar 20. Deafness and Hearing Loss: Key Facts URL: http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [accessed 2019-02-01]
2. Graydon K, Waterworth C, Miller H, Gunasekera H. Global burden of hearing impairment and ear disease. J Laryngol Otol 2019 Jan;133(1):18-25. [doi: 10.1017/S0022215118001275] [Medline: 30047343]
3. Dalton DS, Cruickshanks KJ, Klein BE, Klein R, Wiley TL, Nondahl DM. The impact of hearing loss on quality of life in older adults. Gerontologist 2003 Oct;43(5):661-668. [doi: 10.1093/geront/43.5.661] [Medline: 14570962]
4. Lin FR, Ferrucci L, Metter EJ, An Y, Zonderman AB, Resnick SM. Hearing loss and cognition in the Baltimore Longitudinal Study of Aging. Neuropsychology 2011 Nov;25(6):763-770 [FREE Full text] [doi: 10.1037/a0024238] [Medline: 21728425]
5. Lin FR, Albert M. Hearing loss and dementia - who is listening? Aging Ment Health 2014;18(6):671-673 [FREE Full text] [doi: 10.1080/13607863.2014.915924] [Medline: 24875093]
6. 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]
7. Ray J, Popli G, Fell G. Association of cognition and age-related hearing impairment in the English longitudinal study of ageing. JAMA Otolaryngol Head Neck Surg 2018 Oct 1;144(10):876-882 [FREE Full text] [doi: 10.1001/jamaoto.2018.1656] [Medline: 30193368]
8. Uhlmann RF, Larson EB, Rees TS, Koepsell TD, Duckert LG. Relationship of hearing impairment to dementia and cognitive dysfunction in older adults. J Am Med Assoc 1989 Apr 7;261(13):1916-1919. [doi: 10.1001/jama.1989.03420130084028] [Medline: 2926927]
9. Davis A, McMahon CM, Pichora-Fuller KM, Russ S, Lin F, Olusanya BO, et al. Aging and hearing health: the life-course approach. Gerontologist 2016 Apr;56(Suppl 2):S256-S267 [FREE Full text] [doi: 10.1093/geront/gnw033] [Medline: 26994265]
10. Wilson BS, Tucci DL, Merson MH, O'Donoghue GM. Global hearing health care: new findings and perspectives. The Lancet 2017 Dec;390(10111):2503-2515. [doi: 10.1016/s0140-6736(17)31073-5]
11. Swanepoel DW, Clark JL, Koekemoer D, Hall JW, Krumm M, Ferrari DV, et al. Telehealth in audiology: the need and potential to reach underserved communities. Int J Audiol 2010 Mar;49(3):195-202. [doi: 10.3109/14992020903470783] [Medline: 20151929]
12. GSM Association. Global System for Mobile Communications Association (GSMA). London: Global System for Mobile Communications Association; 2019. The Mobile Economy 2019 URL: https://www.gsmaintelligence.com/research/?file=b9a6e6202ee1d5f787cfebb95d3639c5&download [accessed 2019-06-26]
13. Swanepoel DW, Hall JW. A systematic review of telehealth applications in audiology. Telemed J E Health 2010 Mar;16(2):181-200. [doi: 10.1089/tmj.2009.0111] [Medline: 20187743]
14. Rushbrooke E, Houston KT. Telepractice in Audiology. San Diego: Plural Publishing Inc; 2016.
15. Paglialonga A, Nielsen AC, Ingo E, Barr C, Laplante-Lévesque A. eHealth and the hearing aid adult patient journey: a state-of-the-art review. Biomed Eng Online 2018 Jul 31;17(1):101 [FREE Full text] [doi: 10.1186/s12938-018-0531-3] [Medline: 30064497]
16. Hussein SY, Swanepoel DW, de Jager LB, Myburgh HC, Eikelboom RH, Hugo J. Smartphone hearing screening in mHealth assisted community-based primary care. J Telemed Telecare 2016 Oct;22(7):405-412. [doi: 10.1177/1357633X15610721] [Medline: 26468215]
17. Swanepoel DW, Myburgh HC, Howe DM, Mahomed F, Eikelboom RH. Smartphone hearing screening with integrated quality control and data management. Int J Audiol 2014 Dec;53(12):841-849. [doi: 10.3109/14992027.2014.920965] [Medline: 24998412]
18. Dharmar M, Simon A, Sadorra C, Friedland G, Sherwood J, Morrow H, et al. Reducing loss to follow-up with tele-audiology diagnostic evaluations. Telemed J E Health 2016 Feb;22(2):159-164. [doi: 10.1089/tmj.2015.0001] [Medline: 26544032]
19. Ramkumar V, John KR, Selvakumar K, Vanaja CS, Nagarajan R, Hall JW. Cost and outcome of a community-based paediatric hearing screening programme in rural India with application of tele-audiology for follow-up diagnostic hearing assessment. Int J Audiol 2018 Jun;57(6):407-414. [doi: 10.1080/14992027.2018.1442592] [Medline: 29490519]

20. Swanepoel DW, Koekemoer DJ, Clark J. Intercontinental hearing assessment - a study in tele-audiology. J Telemed Telecare 2010;16(5):248-252. [doi: 10.1258/jtt.2010.090906] [Medline: 20457801]
21. Penteado SP, Bento RF, Battistella LR, Silva SM, Sooful P. Use of the satisfaction with amplification in daily life questionnaire to assess patient satisfaction following remote hearing aid adjustments (telefitting). JMIR Med Inform 2014 Sep 2;2(2):e18 [FREE Full text] [doi: 10.2196/medinform.2769] [Medline: 25599909]
22. Pross SE, Bourne AL, Cheung SW. TeleAudiology in the Veterans Health Administration. Otol Neurotol 2016 Aug;37(7):847-850. [doi: 10.1097/MAO.0000000000001058] [Medline: 27273404]
23. Havenga E, Swanepoel D, le Roux T, Schmid B. Tele-intervention for children with hearing loss: a comparative pilot study. J Telemed Telecare 2017 Jan;23(1):116-125. [doi: 10.1177/1357633X15617886] [Medline: 26670208]
24. Tao KF, Brennan-Jones CG, Capobianco-Fava DM, Jayakody DM, Friedland PL, Swanepoel DW, et al. Teleaudiology services for rehabilitation with hearing aids in adults: a systematic review. J Speech Lang Hear Res 2018 Jul 13;61(7):1831-1849. [doi: 10.1044/2018_JSLHR-H-16-0397] [Medline: 29946688]
25. Beukes EW, Manchaiah V. Internet-based audiological interventions: an update for clinicians. Perspect ASHA SIGs 2019;4(3):542-552. [doi: 10.1044/2019_PERS-SIG18-2018-0003]
26. Grenness C, Hickson L, Laplante-Lévesque A, Davidson B. Patient-centred care: a review for rehabilitative audiologists. Int J Audiol 2014 Feb;53(Suppl 1):S60-S67. [doi: 10.3109/14992027.2013.847286] [Medline: 24447236]
27. Ratanjee-Vanmali H, Swanepoel DW, Laplante-Lévesque A. Characteristics, behaviours and readiness of persons seeking hearing healthcare online. Int J Audiol 2019 Feb;58(2):107-115. [doi: 10.1080/14992027.2018.1516895] [Medline: 30289050]
28. Hearing Research Clinic. URL: https://hearingresearchclinic.org/ [accessed 2020-02-10]
29. Henshaw H, Clark DP, Kang S, Ferguson MA. Computer skills and internet use in adults aged 50-74 years: influence of hearing difficulties. J Med Internet Res 2012 Aug 24;14(4):e113 [FREE Full text] [doi: 10.2196/jmir.2036] [Medline: 22954484]
30. Potgieter JW, Swanepoel DW, Myburgh HC, Hopper TC, Smits C. Development and validation of a smartphone-based digits-in-noise hearing test in South African English. Int J Audiol 2015 Jul;55(7):405-411. [doi: 10.3109/14992027.2016.1172269] [Medline: 27121117]
31. Potgieter JM, Swanepoel DW, Myburgh HC, Smits C. The South African English smartphone digits-in-noise hearing test: effect of age, hearing loss, and speaking competence. Ear Hear 2018;39(4):656-663. [doi: 10.1097/AUD.0000000000000522] [Medline: 29189432]
32. hearX Group. URL: https://www.hearxgroup.com/ [accessed 2020-02-07]
33. Rollnick S, Mason P, Butler CC. Health Behavior Change: A Guide for Practitioners. London: Churchill Livingstone; 1999.
34. Tønnesen H. Engage in the Process of Change; Facts and Methods. Denmark: World Health Organization - Regional Center for Europe; 2012.
35. Milstein D, Weinstein BE. Effects of information sharing on follow-up after screening for older adults. J Acad Rehab Audiol 2002;35:43-58.
36. Thorén ES, Pedersen JH, Jørnæs NO. Usability and online audiological rehabilitation. Am J Audiol 2016 Oct 1;25(3S):284-287. [doi: 10.1044/2016_AJA-16-0015] [Medline: 27768189]
37. Qualtrics XM. URL: https://www.qualtrics.com/ [accessed 2020-02-07]
38. Hawthorne G, Sansoni J, Hayes L, Marosszeky N, Sansoni E. Measuring patient satisfaction with health care treatment using the Short Assessment of Patient Satisfaction measure delivered superior and robust satisfaction estimates. J Clin Epidemiol 2014 May;67(5):527-537. [doi: 10.1016/j.jclinepi.2013.12.010] [Medline: 24698296]
39. Linnan L, Steckler A. Process Evaluation for Public Health Interventions and Research. San Francisco: Jossey-bass; 2002.
40. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
41. Langbecker D, Caffery LJ, Gillespie N, Smith AC. Using survey methods in telehealth research: a practical guide. J Telemed Telecare 2017 Oct;23(9):770-779. [doi: 10.1177/1357633X17721814] [Medline: 28728502]
42. Ram D, Patil S, Gowdappa BH. Level of paranormal beliefs and its relationship with explanatory models, treatment adherence and satisfaction. Arch Clin Psychiatry 2016;43(3):51-55. [doi: 10.1590/0101-60830000000084]
43. Reichheld F, Markey R. The Ultimate Question 2.0 (revised And Expanded Edition): How Net Promoter Companies Thrive In A Customer-driven World. Boston, MA: Harvard Business Review Press; 2011.
44. IBM - United States. IBM SPSS Statistics 25 URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25 [accessed 2020-02-07]
45. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
46. Field, A. Discovering Statistics Using Ibm Spss Statistics. California: Sage Publications Ltd; 2018:253.
47. Smits C, Merkus P, Houtgast T. How we do it: the Dutch functional hearing-screening tests by telephone and internet. Clin Otolaryngol 2006 Oct;31(5):436-440. [doi: 10.1111/j.1749-4486.2006.01195.x] [Medline: 17014457]
48. Meyer C, Hickson L, Khan A, Hartley D, Dillon H, Seymour J. Investigation of the actions taken by adults who failed a telephone-based hearing screen. Ear Hear 2011;32(6):720-731. [doi: 10.1097/AUD.0b013e318220d973] [Medline: 21697715]

49. Ingo E, Brännström KJ, Andersson G, Lunner T, Laplante-Lévesque A. Stages of change in audiology: comparison of three self-assessment measures. Int J Audiol 2017 Jul;56(7):516-520. [doi: 10.1080/14992027.2017.1309466] [Medline: 28420270]
50. Rothpletz AM, Moore AN, Preminger JE. Acceptance of internet-based hearing healthcare among adults who fail a hearing screening. Int J Audiol 2016 Sep;55(9):483-490. [doi: 10.1080/14992027.2016.1185804] [Medline: 27409278]
51. Molander P, Nordqvist P, Oberg M, Lunner T, Lyxell B, Andersson G. Internet-based hearing screening using speech-in-noise: validation and comparisons of self-reported hearing problems, quality of life and phonological representation. BMJ Open 2013 Sep 16;3(9):e003223 [FREE Full text] [doi: 10.1136/bmjopen-2013-003223] [Medline: 24041846]
52. Mars M, Scott RE. WhatsApp in clinical practice: a literature review. Stud Health Technol Inform 2016;231:82-90. [doi: 10.3233/978-1-61499-712-2-82] [Medline: 27782019]
53. de Benedictis A, Lettieri E, Masella C, Gastaldi L, Macchini G, Santu C, et al. WhatsApp in hospital? An empirical investigation of individual and organizational determinants to use. PLoS One 2019;14(1):e0209873 [FREE Full text] [doi: 10.1371/journal.pone.0209873] [Medline: 30633754]
54. Giordano V, Koch H, Godoy-Santos A, Belangero WD, Pires RE, Labronici P. WhatsApp Messenger as an adjunctive tool for telemedicine: an overview. Interact J Med Res 2017 Jul 21;6(2):e11 [FREE Full text] [doi: 10.2196/ijmr.6214] [Medline: 28733273]
55. Mars M, Morris C, Scott RE. WhatsApp guidelines - what guidelines? A literature review. J Telemed Telecare 2019 Oct;25(9):524-529. [doi: 10.1177/1357633X19873233] [Medline: 31631763]
56. McGrail KM, Ahuja MA, Leaver CA. Virtual visits and patient-centered care: results of a patient survey and observational study. J Med Internet Res 2017 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
57. Bennett RJ, Meyer C, Eikelboom RH. Does clinician continuity influence hearing aid outcomes? Int J Audiol 2016 Oct;55(10):556-563. [doi: 10.1080/14992027.2016.1185169] [Medline: 27224042]
58. Reginato TT, Ferrari DV. Teleaudiology: professional-patient communication in hearing aid programming and fitting via teleconsultation. Audiol Commun Res 2014;19(3):299-309. [doi: 10.1590/S2317-64312014000300001 5]
59. Meyer C, Waite M, Atkins J, Scarinci N, Cowan R, Hickson L. Promoting Patient-Centered Hearing Care Through the Use of eHealth: Current Status and Future Possibilities. Perspect ASHA Spec Interest Groups 2019;4(2):331-344. [doi: 10.1044/2018_PERS-SIG7-2018-0003]
60. Powell RE, Stone D, Hollander JE. Patient and health system experience with implementation of an enterprise-wide telehealth scheduled video visit program: mixed-methods study. JMIR Med Inform 2018 Feb 13;6(1):e10 [FREE Full text] [doi: 10.2196/medinform.8479] [Medline: 29439947]
61. Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald DJ, Simpson AH, et al. Assessing treatment outcomes using a single question: the net promoter score. Bone Joint J 2014 May;96-B(5):622-628. [doi: 10.1302/0301-620X.96B5.32434] [Medline: 24788496]
62. Koladycz R, Fernandez G, Gray K, Marriott H. The Net Promoter Score (NPS) for insight into client experiences in sexual and reproductive health clinics. Glob Health Sci Pract 2018 Oct 3;6(3):413-424 [FREE Full text] [doi: 10.9745/GHSP-D-18-00068] [Medline: 30072372]
63. Swenson SL, Buell S, Zettler P, White M, Ruston DC, Lo B. Patient-centered communication: do patients really prefer it? J Gen Intern Med 2004 Nov;19(11):1069-1079 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30384.x] [Medline: 15566435]
64. Poost-Foroosh L, Jennings MB, Shaw L, Meston CN, Cheesman MF. Factors in client-clinician interaction that influence hearing aid adoption. Trends Amplif 2011 Sep;15(3):127-139 [FREE Full text] [doi: 10.1177/1084713811430217] [Medline: 22155784]

Abbreviations

eHealth: electronic health
HHC: hearing health care
NPS: net promoter score
SAPS: Short Assessment of Patient Satisfaction
SNR: signal-to-noise ratio

Edited by G Eysenbach; submitted 14.08.19; peer-reviewed by D Maidment, A Paglialonga, C Hollander, F Sendra-Portero; comments to author 31.08.19; revised version received 17.10.19; accepted 23.01.20; published 20.03.20.

Please cite as:
Ratanjee-Vanmali H, Swanepoel DW, Laplante-Lévesque A
Patient Uptake, Experience, and Satisfaction Using Web-Based and Face-to-Face Hearing Health Services: Process Evaluation Study
J Med Internet Res 2020;22(3):e15875
URL: http://www.jmir.org/2020/3/e15875/
doi:10.2196/15875
PMID:32196459

Original Paper

Understanding Patient Anxieties in the Social Media Era: Qualitative Analysis and Natural Language Processing of an Online Male Infertility Community

Vadim Osadchiy[1,2], BS; Jesse Nelson Mills[1], MD; Sriram Venkata Eleswarapu[1,2], MD, PhD

[1]Division of Andrology, Department of Urology, David Geffen School of Medicine, University of California,, Los Angeles, CA, United States

[2]Consortium for Health Activity on Social Media, David Geffen School of Medicine, University of California,, Los Angeles, CA, United States

Corresponding Author:
Sriram Venkata Eleswarapu, MD, PhD
Division of Andrology, Department of Urology
David Geffen School of Medicine
University of California,
10944 Le Conte Ave
Ueberroth #3361
Los Angeles, CA, 90025
United States
Phone: 1 3107943058
Email: seleswarapu@mednet.ucla.edu

Abstract

Background: Couples struggling with infertility are increasingly turning to the internet for infertility-related content and to connect with others. Most of the published data on infertility and the internet only address the experiences of women, with limited studies focusing exclusively on internet discussions on male factor infertility.

Objective: The aim of this study was to understand the concerns and experiences of discussants on an online male infertility community and to provide insight into their perceptions of interactions with health care professionals.

Methods: Using the large-scale data analytics tool BigQuery, we extracted all posts in the r/MaleInfertility community (877 members) of the social media website and discussion board Reddit from November 2017 to October 2018. We performed a qualitative thematic analysis and quantitative semantic analysis using Language Inquiry and Word Count 2015 of the extracted posts to identify dominant themes and subthemes of discussions. Descriptive statistics and semantic analytic Z-scores were computed.

Results: From the analysis of 97 posts, notable themes and subthemes emerged: 70 (72%) posts shared personal experiences, including feeling emasculated or isolated or describing a negative (28/97, 29%), positive (13/97, 13%), or neutral (56/97, 58%) experience with a health care professional; 19% (18/97) of the posts posed questions about personal semen analysis results. On the basis of semantic analysis, posts by men had higher authenticity scores (Z=3.44; *P*<.001), suggesting more honest or personal texts, but lower clout scores (Z=4.57; *P*<.001), suggesting a more tentative or anxious style of writing, compared with posts by women.

Conclusions: To our knowledge, this study represents the first evaluation of a social media community focused exclusively on male infertility using mixed methodology. These results suggest a role for physicians on social media to engage with patients and connect them to accurate resources, in addition to opportunities to improve in-office patient education.

(J Med Internet Res 2020;22(3):e16728) doi:10.2196/16728

KEYWORDS

social media; infertility, male; online social networking; Reddit; discussion board

Introduction

Background

Social media has emerged as a potent resource for patients seeking both anonymous and nonanonymous engagement on acute and chronic medical problems. Despite the ubiquity of social media platforms, a comprehensive scientific understanding of the content of online health-related discussions remains poorly studied, even though 72% of internet users searched for health information on the Web [1]. Reproductive medicine, in particular, has seen a burgeoning presence on the Web, encompassing everything from instructional websites explaining in vitro fertilization [2] to direct-to-consumer marketing of gamete cryopreservation targeting busy career professionals [3].

A majority of couples struggling with infertility turn to the internet for infertility-related content and to connect with others going through similar experiences [4,5]. Despite the high prevalence of male factor infertility [6], a disproportionate number of infertility investigations focus exclusively on women [7]. Similarly, most of the published data on infertility and the internet address only the experience of women [8-11]; only one recently published study [12] focuses exclusively on internet discussions on male factor infertility. Online discussion boards afford discussants with anonymity, allowing for productive conversations related to traditionally private or taboo topics. Published work on online discussion boards of such topics include sexually transmitted diseases [13], suicide [14], and psychosis [15]. Furthermore, discussants do not feel similar pressures as they might during in-person discussions; individuals can share as little or as much as they are comfortable with in this type of forum [16].

Founded in 2005, the discussion website Reddit has become one of the most popular internet destinations in the United States, with more Web traffic devoted to it than other social media websites, including Twitter and Instagram [17]. Reddit users post anonymously on subforums or subreddits related to a specific topic [18]. For example, the male infertility subreddit, r/MaleInfertility, was created “[f]or males with insufficient sperm to achieve pregnancy in fertile females without assisted reproduction and those affected” with the goal of “speak[ing] openly and honestly about our sperm” [19].

Objectives

The objective of this study was to understand the concerns and experiences of discussants in this online male infertility community, to provide insight into their perceptions of interactions with health care professionals, and to explore differences in the experiences of men and their partners on the Web.

Methods

Data Extraction and Processing

We extracted all posts from the Reddit community *MaleInfertility* from November 2017 to October 2018 (12 months). At the time of writing, this open-access online community had 877 members [19]. Posts were extracted using BigQuery (Google LLC, Menlo Park, California), a Structured Query Language–based enterprise data analytics platform, from a dataset uploaded for public use [20]. At the time of data extraction, this time frame represented the most recent 12 months of data uploaded to BigQuery. We retrieved post title, content, author username, and date and time of publication. Posts that were empty or comprised exclusively of the text “[deleted]” or “[removed]” were excluded from analysis.

Qualitative Thematic Analysis

We performed a qualitative analysis on the extracted data using an inductive, data-driven approach for content analysis of the free-text narrative data, with grounded theory and a constant comparative method as methodology [21,22]. During open coding, one investigator (VO) carefully analyzed text from each post to identify preliminary themes. We defined our unit of analysis as an entire post, given that we were analyzing free-text data without a word limit, one post could therefore contain more than one code. These preliminary themes were then discussed among all authors. On repeat reviews of the data, themes were finalized and then further divided into subthemes for better characterization. Previous studies did not inform initial coding as we used a purely inductive approach. During our review, we also collected data on whether the post was authored by a male or a female partner; this distinction was possible as authors frequently introduced themselves, or this information could easily and reasonably be deduced through language, such as “my husband was told that...” or “my sperm count is....” If there was any ambiguity, we did not assign a gender to the author. As Reddit represents an anonymous social media forum, we were limited to deducing the gender of a participant from the content of his or her post. We also collected data on whether interventions related to male infertility were mentioned.

Consideration of Researcher Characteristics, Reflexivity, and Mitigation of Biases

The principal evaluators for this qualitative study were adult males; one is a medical student pursuing urological training, and the other two are urologists with advanced fellowship training in male reproductive medicine (andrology). Inherent biases relate to these researchers’ daily interaction with men struggling with infertility, both in the outpatient clinical and outpatient surgical realms. The researchers recognized the potential for bias in selecting themes and identifying representative discussions of male infertility from social media. Themes were discussed and agreed upon in committee by the authors.

Semantic Analysis

To conduct a semantic-based analysis, we used Language Inquiry and Word Count (LIWC) 2015, an empirically validated textual analysis program capable of translating language into quantitative metrics related to different psychological processes (affective, social, cognitive, perceptual, and others) and linguistic dimensions (parts of speech, grammar, and others) [23]. Previous studies have used LIWC 2015 for similar purposes [24-27]. For our analysis, we used the four summary variables available on LIWC 2015, which were developed and validated using previously published datasets comprising large comparison samples [23,28-31]: (1) analytical thinking, (2) clout, (3) authenticity, and (4) emotional tone. Scores for each of these variables range from 0 to 100. Compared with lower analytical thinking scores, higher analytical thinking scores suggest language that is more formal and logical. Higher clout suggests that the writer is confident and speaks from the perspective of an expert, compared with lower scores that suggest a more tentative or anxious style. Text that scores higher on authenticity reflects a more honest and straightforward style, compared with lower authenticity scores that suggest a less candid and more guarded text. For emotional tone, compared with a lower number, a higher number reflects a more positive tone, with a score of 50 indicating a neutral tone.

We compared means of the aforementioned summary variables between posts authored by men vs women using Mann-Whitney *U* tests. RStudio version 1.1.463 (RStudio, Inc, Boston, Massachusetts) was used for statistical analysis, with $P<.05$ considered statistically significant.

Ethics

As this study involved anonymous, publicly available data, it was deemed as exempt by the institutional review board of the University of California, Los Angeles. This is consistent with previous investigations of research on social media data [26,32].

Results

Overview

A total of 133 posts were initially extracted. A total of 97 posts by 73 unique users remained for analysis after applying exclusion criteria. From these 97 posts, men authored 53 posts (55%) and women authored 21 posts (22%), and gender was not identifiable among 23 posts (24%). The top five most frequently used words were as follows: sperm, DNA, test, motile/motility, and normal. The average word count for all posts was 191 words.

Sharing Personal Experiences

A total of 72% (70/97) of the posts were related to *Sharing Personal Experiences*. The experiences authors described were often emotionally charged, featuring the subtheme of *Feeling Emasculated or Isolated*, despite frequently mentioning potential sources of support, such as a wife or close partner:

> *Male infertility is one of the last great taboos. And I can understand why. When my wife and I were struggling to make a baby I found it hard to vocalize my feelings. My failure at fatherhood ate away at my very being and made me feel less of a man.*
>
> *I also feel like I've failed my wife and that I'm less of a man now. It's depressing. I feel like someone died.*

A total of 25% (24/97) of all the posts mention an interaction with a health care professional. A majority (56/97, 58%) of these interactions were neutral; 29% (28/97) of the interactions were negative and 13% (13/97) of the interactions were positive. Negative interactions were often driven by distrust in their health care professional with respect to financial conflicts of interest or an overall distrust in their provider's fund of knowledge:

> *At our consultation with the fertility clinic the doctor immediately started pushing IVF, not suggesting any drugs or any other treatments [...] She didn't really have any answers for me and to me seemed to be more focused on female issues than male issues.*
>
> *His doctor couldn't tell him much, and couldn't explain his morphology results at all. Obviously a great doctor.*

Positive experiences related to health care professionals often centered around providing hope and reassurance:

> *The urologist was pretty encouraging, saying he's only ever seen 3 people who never regained sperm, but it's still hard to be positive and not be scared that you'll be in that small group.*
>
> *Although, I am crushed, my male infertility urologist (UCLA) to be exact told me that in about 10-20 years (give or take) stem cell technology will be able to help me father a child.*

Searching for Shared Experiences

Complementing the aforementioned theme, 35% (34/97) of the posts involved *Searching for Shared Experiences*. Discussions that included this theme were often discussed within the context of interventions related to infertility, with 35% (34/97) of the posts mentioning such an intervention.

Only a limited subset (5/97, 5%) of posts mentioned intrauterine insemination, but when this topic was featured, it was often related to searching for others who went through the process:

> *My wife and I have chosen a donor and will start with iui in the next few months [...] What I've really felt is missing or that I need is to read about other people's experiences in this same (or similar) situation. Are there blogs or other resources about this sort of thing? I want to know how other men have dealt with his, how/if they've talked to people about it, etc.*
>
> *Has anyone experienced this? My heart aches for my husband and our future. I cannot imagine going through with a donor sperm at this point. This is such a lonely and isolating experience.*

Similar discussions featured in vitro fertilization (mentioned in 12/97, 12% of all the posts) and microdissection testicular sperm extraction (14/97, 14% of all the posts).

> *My wife and are undergoing IVF but I am so stressed about it not working it's hard to enjoy. I can't talk to anyone about it so I thought I'd try here. Any advice?*
>
> *We are looking at MICRO TESE to see where to from here, worried that they will not find sperm is what is stressing me at the moment. Reading all the stories I know now that I'm not alone, that is a bit comforting.*

Sharing Resources or Information

A total of 14% (14/97) of the posts were related to *Sharing Resources or Information*. Shared information came in the form of (1) alternative online discussion boards targeted at a group of individuals going through a similar experience, such as using donor sperm, or (2) sharing recently published, peer-reviewed research related to male factor infertility, assisted reproductive technology, and related topics:

> *We need your help brining a subreddit to life: Parents of Donor Conceived Persons [link]. It's a new subreddit for parents of children conceived using donor sperm, eggs, or embryos [...] This subreddit aims to create a much needed outlet and community. It's a place to share our struggles and solutions in order to raise the best children we can. Please subscribe and share with others.*
>
> *I have researched this [DNA fragmentation] for a long time and have spent so much time trying to understand all that is involved and I hope you guys find it useful [...] This is a photo representation of what this looks like if you are a visual person [link to figure from peer reviewed article]*

Medication Side Effects

A minority of posts (3/97, 3%) included discussions related to medication side effects, mainly clomiphene. Although 14% (14/97) of all the posts mentioned this medication, it was frequently not within the context of side effects:

> *Have any guys noticed weight gain while taking Clomid? If so, did it drop back off after stopping it?*

Questions Related to Personal Semen Analysis

A total of 19% (18/97) of the posts featured a question related to a personal semen analysis result. Many authors expressed anxiety or a feeling that they needed to act based on their semen analysis results, especially when the results were available before an appointment with a fertility specialist.

> *Getting in to see the fertility doctor isnt possible until my Sonohysterogram, and our GP most likely won't know what to make of this [...] Is there anything that can help increase his SA results?*
>
> *I got my results back, waiting to see the doctor next week but got curious about what this means.*

Many questions related to semen analysis results also emerged even after a recent visit to a health care professional to discuss their workup. Similar to the aforementioned subtheme related to negative experiences with health care professionals, authors expressed uncertainty related to the interpretation of their results by health care professionals.

> *I'd really like some input on these numbers. How low are they, really? Is Doc1 right on her analysis?*
>
> *my urologist said that this lower morphology will impact fertility but did not provide any statistics... does anyone know the numbers behind how much worse off I am*

Semantic Analysis

A semantic analysis revealed differences in the linguistic attributes of posts authored by men vs their partners (Table 1). Posts authored by men had higher authenticity scores (Z=3.44, *P*<.001), suggesting a more honest or personal text, but lower clout scores (Z=−4.57, *P*<.001), suggesting a more tentative or anxious style of writing, compared with posts by women. No differences emerged in analytical or tone scores.

Table 1. Semantic analysis of the linguistic attributes of posts that are authored by men compared to those of their partners.

Variable	All	Men	Women	Z	*P* value
Analytical	56.36	54.18	47.55	1.38	.17
Authentic	44.09	57.74	30.63	3.44	<.001[a]
Clout	51.92	37.21	65.26	−4.57	<.001[a]
Tone	40.72	34.45	46.30	−0.90	.37

[a] Statistically significant.

In Table 1, Mann-Whitney *U* tests for significance were used to outline mean differences of the four summary variables between posts authored by men vs their partners. Note that mean values under the *All* category include data from both men and women, in addition to data from posts where author gender could not be identified.

Discussion

Principal Findings

To our knowledge, this study represents the first evaluation of a social media community focused exclusively on male infertility utilizing mixed methodology, with both a classic qualitative analysis and natural language processing methods. For many men and their partners, male factor infertility is stigmatized [33]. Online discussion boards, such as the subreddit we analyzed here, create a space for discussants to connect with others anonymously and to ask questions that they may not feel comfortable sharing in person with their physicians. Results from our analysis may inform strategies for enhanced communication with male infertility patients and their partners, both on the Web and in the clinic.

Many of this study’s findings are consistent with previous studies on online infertility discussions, which describe the struggle of infertility as a profoundly emotional and psychologically trying period in the lives and relationships of discussants [10,34]. Hanna et al [35], in a qualitative analysis of an online infertility forum, underscore that regardless of etiology, infertility represents an “emotional rollercoaster” for both partners, with strong feelings (positive and negative) on both ends of the spectrum. Perhaps attributed to our exploration of an online discussion board focused exclusively on male factor infertility, we found that discussions related to sharing personal experiences were often negative. Subthemes involving feelings of emasculation and isolation permeated through the majority of posts we analyzed, supporting the idea that for at least a subset of men, the ability to conceive a child may be tied to their senses of masculinity and self-worth. It is interesting to note that Beeder et al [12], who also performed a content analysis of a different group of online discussions on male factor infertility, observed a similar theme: *Feelings associated with male infertility*. This theme, however, was featured in only 16% of the posts, compared with 72% of the posts in our analysis. Feelings of inadequacy were identified in less than 1% of the posts in the study by Beeder et al [12]. This discrepancy may be at least partially explained by differences in the proportion of women to men authoring the analyzed posts. Compared with over 60% of the posts authored by women in the study by Beeder et al [12], this study had fewer than 30% of the posts authored by women where gender could be identified. We found that discussants also frequently used this forum to connect with others to address these aforementioned feelings of isolation, thereby creating an opportunity for discussants to normalize each other's experiences [34-36].

Many of the experiences shared involved interactions with health care providers; nearly 30% of these interactions were negative. Perceived poor physician communication represented the cornerstone of many of these negative experiences. In addition, almost 20% of the posts involved a question related to the interpretation of semen analysis results, even after a recent visit to a health care professional. A similar finding was reported by Beeder et al [12], where they noted that almost all the questions related to male infertility diagnosis and testing were about interpretation of semen analysis results. These findings suggest a potential role for physicians, in the office or on social media, to engage with patients and connect them to accurate resources. This engagement is particularly critical as the accuracy of health information on the Web is circumspect [37]. One study found that even websites of fertility clinics affiliated with the Society for Assisted Reproductive Technology failed to meet most of the American Medical Association's health information guidelines [38].

Although this forum was focused on male factor infertility, both men and their partners participated in discussions. Quantitative findings from the semantic analysis reveal similarities and differences in the ways that men and women communicate on this forum. Posts by men were characterized by an overall less confident writing style (lower clout scores), compared with those authored by women. These findings are aligned with the results of a previous study by Hanna et al [35], which noted that the language men use on the Web when discussing even highly personal issues related to infertility remains constrained by norms of hegemonic masculinity. The results of our semantic analysis may reflect attitudes and behaviors that occur even outside of online discussion boards. In a study exploring the infertility experience of Polish couples, Nagorska et al [39] found that women were more likely to talk openly and confidently about infertility, whereas men found themselves acting more restrained, consistent with the lower clout scores we observed in this study. Despite these constraints, posts by men were more honest and personal (higher authenticity scores), compared with those of women, perhaps underscoring the value of anonymity on an online discussion board [40]. The finding of lower clout and higher authenticity scores is consistent with the results of our qualitative analysis, as the subtheme of *Feeling emasculated or isolated* featured posts that were authored primarily by men. Online forums may serve a particularly important role for men struggling with infertility, as men are less likely to seek in-person social support to cope with infertility stress [41]. Posts authored by men vs women did not differ in their tone (both were equally negative) or analytical scores (same degree of formality and logical thinking patterns).

Although anonymity represents a valuable benefit to participating in an online forum on infertility, it also creates difficulties in analyzing participant demographics. As only the username was available, we were limited to deducing the gender of a participant from the content of the post; this limits statistical power and perhaps introduces sampling bias to our semantic analysis. In addition, individuals who turn to the internet for health care information may be different with respect to demographics and information preferences from those who do not [42]; the results of this study should therefore be interpreted within this context. To our knowledge, the subreddit we have analyzed represents the largest community on Reddit focused on male factor infertility. Future studies may consider an expanded analysis incorporating other online discussion boards that also focus on male factor infertility.

Conclusions

Although online discussion boards may serve patients' needs in a different yet complementary way to their experiences with health care providers in person, this study underscored a need to enhance in-office communication, especially within the context of male factor infertility. The semantic analysis suggests that the online and in-office needs of men and their partners differ, especially with respect to infertility that is male factor in etiology. This study's results also suggest a potential role for physicians on social media to engage with patients and connect them to accurate resources.

Acknowledgments

SVE is supported by a Research Scholar Award from the Urology Care Foundation and American Urological Association. These organizations played no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

All authors contributed to the concept and study design and acquisition, analysis, or interpretation of the data. VO and SVE drafted the manuscript. All authors critically revised the manuscript for important intellectual content. VO performed the statistical analysis. JNM and SVE provided the administrative, technical, or material support. SVE supervised the study. All authors approved the final manuscript for publication.

Conflicts of Interest

SVE serves as a consultant for Metuchen Pharmaceuticals. JNM serves as a consultant for Antares Pharma, Boston Scientific Corporation, and Endo Pharmaceuticals. No other disclosures are reported.

References

1. Fox S, Duggan M. Pew Research Center. 2013. Health Online 2013 URL: https://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-05-16]
2. Victorian Assisted Reproductive Treatment Authority. 2018. Assisted reproductive technology – IVF and ICSI URL: https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/Assisted-reproductive-technology-IVF-and-ICSI [accessed 2019-06-23]
3. Wiseman E. The Guardian. 2016. We need to talk about egg freezing URL: https://www.theguardian.com/society/2016/feb/07/life-on-hold-with-frozen-eggs [accessed 2019-06-23]
4. Kahlor L, Mackert M. Perceptions of infertility information and support sources among female patients who access the internet. Fertil Steril 2009 Jan;91(1):83-90. [doi: 10.1016/j.fertnstert.2007.11.005] [Medline: 18243181]
5. Hinton L, Kurinczuk JJ, Ziebland S. Infertility; isolation and the internet: a qualitative interview study. Patient Educ Couns 2010 Dec;81(3):436-441. [doi: 10.1016/j.pec.2010.09.023] [Medline: 21036506]
6. World Health Organization. Task Force on the Diagnosis and Treatment of Infertility, Comhaire FH. Towards More Objectivity in Diagnosis and Management of Male Fertility. Oxford, UK: Blackwell Scientific; 1987.
7. Petok WD. Infertility counseling (or the lack thereof) of the forgotten male partner. Fertil Steril 2015 Aug;104(2):260-266. [doi: 10.1016/j.fertnstert.2015.04.040] [Medline: 26048155]
8. Epstein YM, Rosenberg HS, Grant TV, Hemenway B. Use of the internet as the only outlet for talking about infertility. Fertil Steril 2002 Sep;78(3):507-514. [doi: 10.1016/s0015-0282(02)03270-3] [Medline: 12215325]
9. Malik SH, Coulson NS. Computer-mediated infertility support groups: an exploratory study of online experiences. Patient Educ Couns 2008 Oct;73(1):105-113. [doi: 10.1016/j.pec.2008.05.024] [Medline: 18639409]
10. Malik SH, Coulson NS. Coping with infertility online: an examination of self-help mechanisms in an online infertility support group. Patient Educ Couns 2010 Nov;81(2):315-318. [doi: 10.1016/j.pec.2010.01.007] [Medline: 20144521]
11. Malik SH, Coulson NS. A comparison of lurkers and posters within infertility online support groups. Comput Inform Nurs 2011 Oct;29(10):564-573. [doi: 10.1097/NCN.0b013e318224b31d] [Medline: 21709546]
12. Beeder L, Samplaski MK. Analysis of online discussion boards for male infertility. Andrologia 2019 Dec;51(11):e13422. [doi: 10.1111/and.13422] [Medline: 31642097]
13. Nobles AL, Leas EC, Althouse BM, Dredze M, Longhurst CA, Smith DM, et al. Requests for diagnoses of sexually transmitted diseases on a social media platform. J Am Med Assoc 2019 Nov 05;322(17):1712-1713. [doi: 10.1001/jama.2019.14390] [Medline: 31688875]
14. Ambalavan AK, Moulahi B, Azé J, Bringay S. Unveiling online suicide behavior: what can we learn about mental health from suicide survivors of Reddit? Stud Health Technol Inform 2019 Aug 21;264:50-54. [doi: 10.3233/SHTI190181] [Medline: 31437883]
15. Rezaii N, Walker E, Wolff P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. NPJ Schizophr 2019 Jun 13;5(1):9 [FREE Full text] [doi: 10.1038/s41537-019-0077-9] [Medline: 31197184]
16. White M, Dorman SM. Receiving social support online: implications for health education. Health Educ Res 2001 Dec;16(6):693-707 [FREE Full text] [Medline: 11780708]
17. Alexa Internet, Inc. Top Sites in United States URL: https://www.alexa.com/topsites/countries/US [accessed 2019-05-16]
18. Duggan M, Smith A. Pew Research Center. 2013. 6% of Online Adults are reddit Users URL: https://www.pewinternet.org/2013/07/03/6-of-online-adults-are-reddit-users/#fn-55-1 [accessed 2019-05-16]
19. Reddit. Male infertility URL: https://www.reddit.com/r/maleinfertility/ [accessed 2019-05-16]
20. Baumgartner J. Reddit. pushshift URL: https://files.pushshift.io/reddit/ [accessed 2020-02-09]
21. Charmaz K. Constructing Grounded Theory. Second Edition. Thousand Oaks, CA: Sage Publications ltd; 2014.
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. Austin: University of Texas; 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2020-02-08]
24. Alpert JM, Morris BB, Thomson MD, Matin K, Geyer CE, Brown RF. OpenNotes in oncology: oncologists' perceptions and a baseline of the content and style of their clinician notes. Transl Behav Med 2019 Mar 01;9(2):347-356 [FREE Full text] [doi: 10.1093/tbm/iby029] [Medline: 29596633]
25. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz J, et al. Social media based analysis of opioid epidemic using Reddit. AMIA Annu Symp Proc 2018;2018:867-876 [FREE Full text] [Medline: 30815129]

26. Park A, Conway M. Longitudinal changes in psychological states in online health community members: understanding the long-term effects of participating in an online depression community. J Med Internet Res 2017 Mar 20;19(3):e71 [FREE Full text] [doi: 10.2196/jmir.6826] [Medline: 28320692]
27. Lin CW, Lin MJ, Wen CC, Chu SY. A word-count approach to analyze linguistic patterns in the reflective writings of medical students. Med Educ Online 2016;21:29522 [FREE Full text] [doi: 10.3402/meo.v21.29522] [Medline: 26838331]
28. Pennebaker JW, Chung CK, Frazee J, Lavergne GM, Beaver DI. When small words foretell academic success: the case of college admissions essays. PLoS One 2014;9(12):e115844 [FREE Full text] [doi: 10.1371/journal.pone.0115844] [Medline: 25551217]
29. Kacewicz E, Pennebaker JW, Davis M, Jeon M, Graesser AC. Pronoun use reflects standings in social hierarchies. J Lang Soc Psychol 2014;33(2):125-143. [doi: 10.1177/0261927x13502654]
30. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying words: predicting deception from linguistic styles. Pers Soc Psychol Bull 2003 May;29(5):665-675. [doi: 10.1177/0146167203029005010] [Medline: 15272998]
31. Cohn MA, Mehl MR, Pennebaker JW. Linguistic markers of psychological change surrounding September 11, 2001. Psychol Sci 2004 Oct;15(10):687-693. [doi: 10.1111/j.0956-7976.2004.00741.x] [Medline: 15447640]
32. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016 Apr 14;17:22 [FREE Full text] [doi: 10.1186/s12910-016-0105-5] [Medline: 27080238]
33. Fisher JR, Hammarberg K. Psychological and social aspects of infertility in men: an overview of the evidence and implications for psychologically informed clinical care and future research. Asian J Androl 2012 Jan;14(1):121-129 [FREE Full text] [doi: 10.1038/aja.2011.72] [Medline: 22179515]
34. Hanna E, Gough B. Emoting infertility online: a qualitative analysis of men's forum posts. Health (London) 2016 Jul;20(4):363-382. [doi: 10.1177/1363459316649765] [Medline: 27246813]
35. Hanna E, Gough B. Searching for help online: an analysis of peer-to-peer posts on a male-only infertility forum. J Health Psychol 2018 Jun;23(7):917-928. [doi: 10.1177/1359105316644038] [Medline: 27106090]
36. Richard J, Badillo-Amberg I, Zelkowitz P. 'So Much of This Story Could Be Me': men's use of support in online infertility discussion boards. Am J Mens Health 2017 Dec;11(3):663-673 [FREE Full text] [doi: 10.1177/1557988316671460] [Medline: 27702886]
37. O'Mathúna DP. How Should Clinicians Engage With Online Health Information? AMA J Ethics 2018 Nov 01;20(11):E1059-E1066 [FREE Full text] [doi: 10.1001/amajethics.2018.1059] [Medline: 30499435]
38. Huang JYJ, Discepola F, Al-Fozan H, Tulandi T. Quality of fertility clinic websites. Fertil Steril 2005 Mar;83(3):538-544. [doi: 10.1016/j.fertnstert.2004.08.036] [Medline: 15749475]
39. Nagórska M, Bartosiewicz A, Obrzut B, Darmochwał-Kolarz D. Gender differences in the experience of infertility concerning polish couples: preliminary research. Int J Environ Res Public Health 2019 Jul 2;16(13):pii: E2337 [FREE Full text] [doi: 10.3390/ijerph16132337] [Medline: 31269703]
40. Tyler RE, Williams S. Masculinity in young men's health: exploring health, help-seeking and health service use in an online environment. J Health Psychol 2014 Apr;19(4):457-470. [doi: 10.1177/1359105312473784] [Medline: 23493865]
41. Peterson BD, Newton CR, Rosen KH, Skaggs GE. Gender differences in how men and women who are referred for IVF cope with infertility stress. Hum Reprod 2006 Sep;21(9):2443-2449. [doi: 10.1093/humrep/del145] [Medline: 16675482]
42. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J Health Commun 2010;15 Suppl 3:279-293. [doi: 10.1080/10810730.2010.522700] [Medline: 21154099]

Abbreviations

LIWC: Language Inquiry and Word Count

Edited by G Eysenbach; submitted 17.10.19; peer-reviewed by J Halpern, M Patel, J Alpert, D Carvalho; comments to author 19.11.19; revised version received 23.11.19; accepted 16.12.19; published 10.03.20.

Please cite as:
Osadchiy V, Mills JN, Eleswarapu SV
Understanding Patient Anxieties in the Social Media Era: Qualitative Analysis and Natural Language Processing of an Online Male Infertility Community
J Med Internet Res 2020;22(3):e16728
URL: http://www.jmir.org/2020/3/e16728/
doi:10.2196/16728
PMID:32154785

Original Paper

Blockchain-Authenticated Sharing of Genomic and Clinical Outcomes Data of Patients With Cancer: A Prospective Cohort Study

Benjamin Scott Glicksberg[1,2,3*], PhD; Shohei Burns[4*], BSc; Rob Currie[5], MBA; Ann Griffin[6], PhD; Zhen Jane Wang[7], MD; David Haussler[5,8], PhD; Theodore Goldstein[1,5], PhD; Eric Collisson[4], MD

[1]Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, United States
[2]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States
[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States
[4]Division of Hematology and Oncology, Department of Medicine, University of California, San Francisco, CA, United States
[5]UC Santa Cruz Genomics Institute, University of California, Santa Cruz, Santa Cruz, CA, United States
[6]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA, United States
[7]Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, United States
[8]Howard Hughes Medical Institute, Santa Cruz, CA, United States
*these authors contributed equally

Corresponding Author:
Eric Collisson, MD
Division of Hematology and Oncology
Department of Medicine
University of California
1450 3rd Street
San Francisco, CA, 94158
United States
Phone: 1 4153539888
Email: collissonlab@gmail.com

Abstract

Background: Efficiently sharing health data produced during standard care could dramatically accelerate progress in cancer treatments, but various barriers make this difficult. Not sharing these data to ensure patient privacy is at the cost of little to no learning from real-world data produced during cancer care. Furthermore, recent research has demonstrated a willingness of patients with cancer to share their treatment experiences to fuel research, despite potential risks to privacy.

Objective: The objective of this study was to design, pilot, and release a decentralized, scalable, efficient, economical, and secure strategy for the dissemination of deidentified clinical and genomic data with a focus on late-stage cancer.

Methods: We created and piloted a blockchain-authenticated system to enable secure sharing of deidentified patient data derived from standard of care imaging, genomic testing, and electronic health records (EHRs), called the Cancer Gene Trust (CGT). We prospectively consented and collected data for a pilot cohort (N=18), which we uploaded to the CGT. EHR data were extracted from both a hospital cancer registry and a common data model (CDM) format to identify optimal data extraction and dissemination practices. Specifically, we scored and compared the level of completeness between two EHR data extraction formats against the gold standard source documentation for patients with available data (n=17).

Results: Although the total completeness scores were greater for the registry reports than those for the CDM, this difference was not statistically significant. We did find that some specific data fields, such as histology site, were better captured using the registry reports, which can be used to improve the continually adapting CDM. In terms of the overall pilot study, we found that CGT enables rapid integration of real-world data of patients with cancer in a more clinically useful time frame. We also developed an open-source Web application to allow users to seamlessly search, browse, explore, and download CGT data.

Conclusions: Our pilot demonstrates the willingness of patients with cancer to participate in data sharing and how blockchain-enabled structures can maintain relationships between individual data elements while preserving patient privacy, empowering findings by third-party researchers and clinicians. We demonstrate the feasibility of CGT as a framework to share health data trapped in silos to further cancer research. Further studies to optimize data representation, stream, and integrity are required.

(J Med Internet Res 2020;22(3):e16810) doi:10.2196/16810

KEYWORDS

data sharing; electronic health records; genomics; medicine; blockchain; neoplasms

Introduction

Every patient with cancer has a unique disease composition and presentation that demands interrogation of complex imaging and genome characteristics [1,2] for personalized treatment recommendations. Currently, it is still standard to report outcomes of cancer as group averages from clinical trials treated with prospectively dictated regimens. Individual patient outcomes from real-world data could further advance personalized medicine by allowing dramatically more treatments and outcomes to be considered [3,4]. As such a health system can learn from its own data to improve its delivery of patient care [5-7]. Regulatory requirements and other restrictions prevent much patient-level data from being shared. Research progress suffers as a result. Precision medicine methodologies such as next-generation tumor DNA sequencing are now often performed in routine cancer care. Unfortunately, results are siloed in individual institutions, frustrating effective sharing or pooling of datasets [8]. Many patients with cancer, however, are willing to share their data and believe that the positive benefits outweigh the potential privacy risks: 93% of patients surveyed would be very or somewhat likely to share their data with university scientists [9].

Despite this need and patients’ willingness to share their data, robust deidentified data sharing methods are lacking. Innovative alternative strategies have been developed that aim to anonymize identifiable clinical data in a way that preserves inherent structure, such as using generative adversarial networks [10], but these have not as of yet been deployed for large-scale, multiomic discovery. One immediate challenge of creating an extensible and robust framework is identifying which data are necessary to share (and in what format), minimizing risk for patient reidentification while maximizing viable information that can lead to clinical insight. Conley et al [11] released a core set of clinical data elements that various stakeholders agreed on for cancer genomic repositories. The lack of a standard data sharing platform for clinical data arises from myriad causes, including but not limited to, incompatible data streams or formats, nonstandardized collection, conflicting business models, extraction and accessibility procedures, and privacy concerns. A centralized, curated platform operated by a single institution is not ideal due to concerns of data ownership, cost, and dissemination procedures. Trends in other fields have migrated from analyzing batched data quarterly, whether from customer Web clicks or manufacturing floor sensors, to real-time analyses. Learning cycles have been reduced from months to hours. Finally, centralized top-down data sharing efforts, although critical to research and scientific deductive understanding, have a fixed lifetime of the study, grant, or group interest.

Software standards based on health care data sharing and electronic commerce are converging to enable solutions to the compelling need to share patient health data for both care management and medical research. In 2013, the Global Alliance for Genomic Health [12] was established to enable a framework for secure, responsible, and effective clinical and genomic data sharing. In 2016, the US president unveiled the National Cancer Institute Cancer Moonshot effort to accelerate cancer research, including efforts focused on data sharing (the Public Access and Data Sharing Policy). Since then, significant progress has been made in mining and sharing medical data. The Food and Drug Administration announced a collaboration with Flatiron Health to utilize deidentified clinical data for the analysis and development of anticancer therapies outside of clinical trials in 2016. Recent studies have delivered on that promise: Agarwal et al [7] analyzed more than 7000 clinical and genomic records from the Flatiron Health network and Foundation Medicine to calculate the tumor mutation burden across cancer subtypes. Singal et al [13] demonstrated that data collected from routine clinical care of almost 30,000 patients with cancer can yield novel clinical insights, as evidenced in this case for non–small cell lung cancer.

A decentralized, scalable, efficient, economical, and secure strategy, such as blockchain technology, can fulfill requirements for effective clinical data sharing. Although not perfect in their scope [14], blockchain systems by design are secure and resistant to tampering and distributed with no single point of control or failure allowing transactions to be efficiently recorded and verified. Multiple publications have proposed the utility of blockchain technology for secure and scalable clinical data sharing [15-19], and many companies and organizations are applying blockchain platforms in health care [20]. Although the excitement surrounding the utilization of blockchain for distributing health care data is encouraging [21], many studies are private, theoretical (ie, accessing feasibility), or unsuccessful in scope. In a recent systematic review of 71 studies that discussed managing health care records via blockchain, only four actually were tested on live data [22].

Here, we develop a public demonstration of curated collection that focuses on capturing the data created over the normal course of clinical care as rapidly as possible. The Cancer Gene Trust (CGT) [23] democratizes data analysis, enabling more experts to participate and compare results, and accelerates the translation of genomic findings toward a clinically useful timescale. CGT is the first free, simple, rapid, global network to share deidentified cancer somatic mutations, radiographic and pathological images, and associated clinical data for prospectively consented patients. These data are rapidly deposited into a global off-blockchain distributed and decentralized repository. This framework not only allows for the rapid dissemination of high yield and important data but also openly details the rigorous process for deidentification, study design, and informed patient consent. From the findings of Mello et al [9], we hypothesized that most patients are willing to consent to their data being shared if it helps expand the corpus of medical knowledge. We aim to demonstrate the utility of CGT by releasing such data from a pilot study of 18 consented patients along with an open-source and freely available application for visualization and exploration.

Methods

Study Design and Recruitment

The University of California, San Francisco (UCSF) institutional review board (IRB) approved our pilot study to consent patients for distributing their deidentified information on CGT (see Multimedia Appendix 1 for study protocol). We approached and consented 18 patients under care at UCSF Medical Center to the *Sharing Clinical and Genomic Data in Cancer Research* clinical pilot protocol (IRB #16-20857).

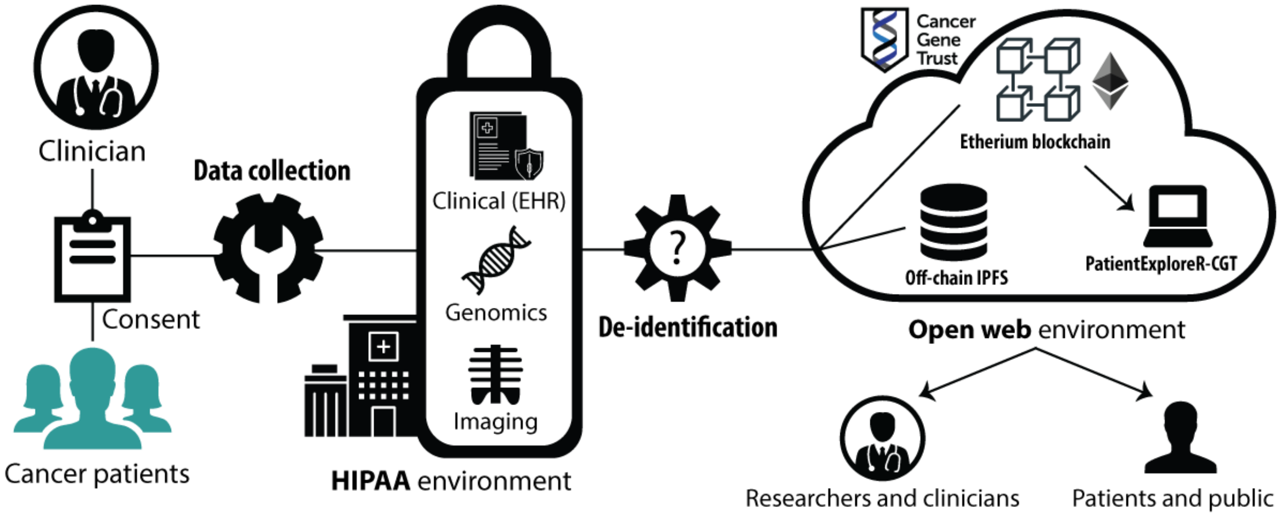
The Cancer Gene Trust Framework

CGT is a decentralized, distributed content addressable real-time database. A submission consists of a manifest containing fields and references to files by hash. Submissions may include deidentified clinical fields, a list of somatic mutations, gene expression, or any type of data relevant to a patient. Submissions are tracked per steward (ie, institution or organization) via a smart contract on the Ethereum [24] blockchain, which references the underlying data stored via hash in InterPlanetary File System (IPFS) [25]. IPFS is inherently decentralized and distributed. Any node may request data from any other node via the unique hash of the data and cache it locally. This affords organic replication of data as well as scalable access. An institution performing internal access and analysis of data may run their own IPFS server and thereby allow high-speed LAN access with only the initial request traversing the list of IPFS servers to find data matching the hash.

Data Collection Procedures

We carefully navigated all institutional procedures to educate and consent our patients before obtaining, formatting, and distributing deidentified patient data from our cohort (Figure 1). We performed stringent and comprehensive privacy processes to be as confident as possible so that no identifying personal health information would be shared (see Data Deidentification section). For the 18 enrolled patients, we were given permission to obtain clinical documentation from their electronic health record (EHR), their somatic mutation information, as well as any scans taken [26]. All data, including genomic, imaging, and structured EHR data (eg, treatment information), for the first cohort of consented patients are available [23]. Patients are identified by a universally unique identifier (UUID-4). The only mapping to the actual patient is securely controlled by trusted stewards; in this case, UCSF. All source code and documentation for CGT are available [27].

Figure 1. Workflow and pipeline for Cancer Gene Trust. EHR: electronic health record; HIPAA: Health Insurance Portability and Accountability Act.



Overall Workflow for Cancer Gene Trust Pipeline

Patients are consented to agree to release their deidentified clinical (Observational Medical Outcomes Partnership [OMOP]-formatted EHR data), genomics (somatic), and imaging data on the blockchain. Stewards representing the affiliated institution then upload the data to CGT. Researchers, clinicians, patients, and the public can then retrieve the data through the Web or interface, with the data dynamically available through the PatientExploreR-CGT app.

Genomic Data Collection

Somatic gene sequencing of tumor specimens was ordered by the supervising physician (EC) as standard of care using either a commercial (Foundation Medicine; FMI [28]) or in-house panel (UCSF 500) [29]; 13 patients were sequenced and analyzed by Foundation Medicine [30] and 4 patients by the UCSF 500 [31] genomic panel. In the case of Foundation Medicine, we received the patient's report in XML format. In the case of the UCSF 500, we requested a deidentified variant call file from the UCSF genomic stewards.

Image Data Collection

For patients with available radiograph imaging, we obtained deidentified DICOM files from the UCSF's Picture Archiving and Communication System medical imaging system conforming to Supplement 142: Clinical Trial De-identification Basic Profile, which removes any identifying protected health information (PHI) from the images as well as any accompanying metadata. Pathology slides were obtained for each patient who had associated pathology performed at UCSF. Deidentified computed tomography (CT) and positron-emission tomography-CT scans correlating to significant changes in tumor response were uploaded for 3 patients to the CGT. Scanned pathology slides clinically utilized for diagnostic purposes were uploaded for 2 of these patients. These deidentified imaging data can be viewed publicly in their entirety on the CGT and illustrate an example collection of raw (but deidentified), clinically relevant data for public research use. Phillips scanners were used to digitize the pathology slides, and a review of PHI was completed before uploading onto CGT.

Clinical Electronic Health Record Data Collection

A large aspect of this project was to evaluate the most suitable and robust source of clinical data to share on CGT. For this comparison, we compared UCSF Cancer registry data, collected to meet the specifications of the Surveillance, Epidemiology, and End Results (SEER) Program, with Observational Health Data Sciences and Informatics (OHDSI) OMOP common data model (CDM) extracted from the hospital EHR. The OMOP CDM is emerging as a standard in the field of EHR research because it is a common framework in terms of both table structure and underlying vocabulary [32] and has enabled powerful research and a venue for regulatory reporting [33].

SEER is a national registry for cancer reporting and provides specific guidelines for data collection from the EHR [34]. Before SEER submission, cancer registry data are submitted to the state registry and assessed for data quality and consolidation with other records for the same patient. Registry data are collected on every cancer case admitted to a UCSF hospital for either diagnosis and/or first course or subsequent cancer treatment per California state cancer reporting law. Certified Tumor Registrars abstract and code cancer information from the EHR in a format specified by the North American Association of Central Cancer Registries' Data Standards [35]. The data collection and coding rules for data collection are specified by the SEER Program Manual and fully abstracted within 6 months of patients' date of first contact with the hospital.

Registry Format

For the first 18 patients, clinical data were requested from the cancer registrar's office for curated data for ultimate submission to SEER via the CNExT cancer registry software. For each patient, we received an Excel export from CNExT with curated clinical data fields (Table 1). We developed a client-side single-page Web application that read in this Excel file on the research coordinators computer, filtered PHI to ensure compliance with IRB regulatory guidelines, and generated a deidentified JSON file. The primary investigator and research coordinator personally reviewed each deidentified registry file for PHI before uploading onto CGT. Depending on the timing of the patient's presentation to the hospital relative to genetic testing, the registry data collection could be in either an incomplete *suspense* state or a completed abstract. Minimum data collection in a suspense case comprises patient age, gender, date of first contact, primary site, and histology. Complete cases contained additional data items related to Basis of Diagnosis and Therapeutic Agent.

Table 1. Breakdown of data elements for registry/Observational Medical Outcomes Partnership.

Gold Standard EHR[a]	Registry field	OMOP table.column
Gender	Sex	person.gender_concept_id
Ethnicity	Spanish Origin	person.ethnicity_concept_id
Race	Race	person.race_concept_id
Date of Diagnosis	Date of Diagnosis[b]	condition_occurrence. condition_start_date
Basis of Diagnosis	Dx Confimation DX Staging/Proc Summ[b]	procedure_occurrence. procedure_occurrence_id
Cancer Site	Cancer Site ICD-0-3 SEER[c] Site Group	condition_occurrence. condition_concept_id
Cancer Histology/Morphology	Cancer Histology (ICD-0-3)	condition_occurrence. condition_concept_id
Therapeutic Agent/Modality	Text/Code of Chemo At Hospital[b]	drug_exposure. drug_concept_id
Beginning and End Dates of Treatment	Chemo Start Date/Chemo End Date[b]	drug_exposure. drug_exposure_start_date/drug_exposure. drug_exposure_end_date

[a] EHR: electronic health record.

[b] Indicates that the field is listed but no or incomplete information was populated (ie, "suspense" registry cases).

[c] SEER: Surveillance, Epidemiology, and End Results.

Observational Medical Outcomes Partnership Format

Procuring clinical data from OMOP was a different process as it involved extraction of retrospective, routinely collected data from the EHR. The Enterprise Data Warehouse (EDW) team at UCSF is responsible for converting raw EPIC/Clarity data into the OMOP format and acted as an honest broker for this extraction process. First, we selected the tables and fields that corresponded to data elements we were consented to collect from our IRB, with buy-in from the EDW team (Table 1). No free-text fields were included. We then provided the medical record numbers (MRNs), and their corresponding CGT patient IDs, to the EDW team who then performed the deidentification process for 17 patients with available data, removing all PHI (see Multimedia Appendix 1 and below for more details). The EDW then extracted the data in the agreed-upon columns in 6 tables of interest, specifically: person, drug_exposure, condition_occurrence, procedure_occurrence, and measurement. We then performed a secondary check to verify all data were deidentified (see below), and then transformed the files (saved as TSV) into a single JSON file per patient.

Clinical Data Scoring Methodology

We evaluated all patients' registry and OMOP data for completeness based on a scoring rubric we designed (see Multimedia Appendix 1 for full details) relating to certain gold-standard metrics essential for clinical data sharing (Table 1), inspired by Conley et al [11]. Data from these gold-standard metrics were captured from the *true* data recorded in UCSF EPIC EHR system patient records. Next, reviewers evaluated how much of these data could be identified from registry and OMOP data sources. Of the 29 data elements recommended by Conley et al [11], we were able to capture 10 of these due to their ability to be obtained without curation from OMOP and registry clinical pipelines. Simply, these data were evaluated on a scale from 0 to 5 for registry and OMOP data, with 0 representing no presence of the data element in the corresponding modality and 5 representing complete representation (values in between correspond to 20% increments of how complete the representation is). As such, for the 10 data elements, the maximum score a patient can receive per data modality is 50.

Statistical Analysis

To assess whether there was any significant difference between registry vs OMOP in terms of data quality capture, we performed a 2-sided Wilcoxon signed-rank test for all 17 patients who were scored according to the above methodology. We further assessed whether there was any difference at the field level, by performing the same assessment per data element (eg, Gender information). We hypothesized that although these two systems are different in terms of data collection methodologies, there should be no significant difference in total scores as both systems are organized to capture the same type of clinical data.

Data Deidentification Procedures

We strived to conform to the most rigorous standards for proper deidentification of all data released as determined by Health Insurance Portability and Accountability Act (HIPAA) standards (see Multimedia Appendix 1 for further discussion and complete documentation of this process).

For the OMOP EHR data, all PHI was removed on receiving the data from the honest broker, the EDW. In these files, all dates were converted into age in days since birth. We performed a secondary check to manually verify that no PHI remained in the files. For genomic data, all germ-line mutations were removed, leaving only somatic variants. No further processing was required for the DICOM images that conform to Supplement 142. Pathology scans were exported into JPEG image files with no identifying metadata or information in the image. The single-page Web application generates a UUID for every patient. The institution and CGT steward maintain an appendix of CGT IDs and UCSF MRNs to preserve the possibility of reidentification between qualified clinicians for follow-up and further research [36].

Data Export and Sharing

These deidentified files are uploaded to the off-blockchain store (IPFS) [25]. The off-blockchain store calculates a cryptographically strong hash (SHA-256) of the entire submission that is added to the stewards list of submissions, which is then updated in the off-blockchain store. This final step yields an updated top level cryptographically strong hash that uniquely defines the entire state of all submissions from the steward at that point in time. This final top-level hash is then submitted to the blockchain as provenance for the entire corpus of submissions from the institution. As the hash is only 256 bits in size, the cost to add to a blockchain is minimized with the bulk of the data stored uniquely in the off-blockchain store. Individual submission hashes as well as the overall steward hash may be concisely referenced toward reproducing any downstream analysis.

Data Distribution and Access

Submissions including all data are immediately available from any IPFS server on the internet via the submission hash. IPFS is inherently decentralized and distributed. Any node may request data from any other node via the unique hash of the data and cache it locally. An IPFS server when queried for the data associated with a hash returns it if it has it locally stored, and if not asks all of the servers it is connected to for the data. In spirit, this is similar to the Transmission Control Protocol/Internet Protocol layer of the internet whereby if a router does not talk directly to the destination it checks with all of its direct peers to see if they do. As a result, data are duplicated as a side effect of access affording organic replication and scalable access. IPFS servers speak HTTP and therefore any data can be accessed in a browser or with a few lines of code from standard bioinformatics analysis tools (eg, cBio, Galaxy, and Jupyter).

PatientExploreR-Cancer Gene Trust: Data Visualization

To facilitate interaction with CGT, we adapted a visualization application to browse, search, visualize, and download the clinical and genomic data shared on CGT. This application, called PatientExploreR-CGT, is adapted from our original PatientExploreR version [37]. PatientExploreR-CGT automatically pulls and maps all data from CGT into a user-friendly dashboard. This application is built in R (version 3.4.1) using the Shiny [38] (version 1.0.5) framework and directly interfaces with OMOP-formatted (version 5 or later) EHR data. In the front-end, the following Shiny-related packages are utilized: shinyWidgets [39], shinyjs [40], shinyalert [41], shinycssloaders [42], shinyBS [43], and shinythemes [44]. Visualizations were created using the plotly [45] and timevis [46] packages. In its backend, PatientExploreR-CGT makes use of ROMOP [47] to automatically extract and map pertinent concepts across all relevant tables (eg, person, observation, and condition occurrence). Data processing and manipulation were facilitated by data.table [48], DT [49], rjson [50], and dplyr [51]. This app can be freely accessed [52].

Results

Cancer Gene Trust Pilot Study

We provide the demographics of the pilot cohort in Table 2. In our cohort, the breakdown of primary cancer was as follows: seven with pancreatic adenocarcinoma, four with cholangiocarcinoma, and one each with anal squamous carcinoma, gastric cancer, colon cancer, gastrointestinal stromal tumor, cecal cancer, and metastatic cancer of unknown primary origin. An additional patient also had metastatic cancer of unknown primary origin but without EHR data. We provide a breakdown of all such data by patient and modality in Table 3.

Table 2. Cohort demographics and clinical information. Demographic breakdown of clinical pilot cohort.

Modality	Value
Gender, n (%)	
Male	6 (33)
Female	12 (67)
Race, n (%)	
White	11 (61)
Asian	5 (28)
Unknown	2 (11)
Ethnicity, n (%)	
Hispanic/Latino	2 (11)
Not Hispanic/Latino	16 (89)
Status, n (%)	
Alive	15 (83)
Deceased	3 (17)
Age (years), mean (SD)	59.3 (13.3)

Table 3. Breakdown of Cancer Gene Trust data by patient and modality.

CGT[a]	Clinical		Genomics		Imaging		OMOP[b] data breakdown		
CGT Public UUID[c]	Registry	OMOP	FMI[d]	UCSF[e] 500	CT[f]	Pathology	Conditions	Procedures	Drugs
f9b6a782-bbf5-4be8-bf7e-d1a9586d9552	✓	✓	✓	N/A[g]	N/A	N/A	1597	1190	3661
c2e2e081-4c39-4201-8a27-7b469ed39490	✓	✓	✓	N/A	✓	✓	1350	969	2088
db2d85aa-4f94-4e77-8755-6b94a710c1aa	✓	✓	✓	N/A	✓	✓	2389	1394	3086
2fbc25da-3965-49c4-866f-72cf0abc2417	✓	✓	✓	N/A	N/A	N/A	930	654	1174
940171e7-d358-463a-8d9a-2b2fa90c2a84	✓	✓	✓	N/A	N/A	N/A	1179	624	1388
f0314175-2d19-4146-8754-fc5aed3ab420	✓	✓	✓	N/A	N/A	N/A	511	405	549
c7dbcfac-37ea-43f8-8899-1a9f2fb56341	✓	✓	✓	N/A	N/A	N/A	216	114	184
ef5c3164-6f45-4d3a-88f0-4509226c5571	✓	✓	N/A	✓	N/A	N/A	51	14	57
ec3d977b-c310-4df3-a444-f79bc3dd8b58	✓	✓	N/A	✓	N/A	N/A	811	505	776
131cf62d-ad78-49c1-a699-5bcc1004cd12	✓	✓	✓	N/A	N/A	N/A	155	42	110
cf11c31c-f4c3-48ba-9c46-66f406d0b7a1	✓	✓	✓	N/A	N/A	N/A	311	162	131
ccc2ba97-912f-4b62-b767-cca129ee6a56	✓	✓	N/A	✓	N/A	N/A	51	10	60
104ec531-5d95-41e2-ac72-f6cff2006b8e	✓	✓	✓	N/A	N/A	N/A	36	10	17
a5627ac3-450d-4036-ade8-99ae62a5c232	✓	✓	N/A	✓	N/A	N/A	857	439	805
5189efbe-3382-4353-ad2f-9afd0255c2c8	✓	✓	✓	N/A	N/A	N/A	875	276	674
253f0e2d-bebd-464b-81c5-8dd8385192b3	✓	✓	N/A	N/A	N/A	N/A	117	116	217
d199cfb0-91e8-471d-b1b3-53189cd64ee0	✓	✓	✓	N/A	✓	N/A	21	11	81
5d3205a3-28c4-45eb-bfd8-b32d67c3be0f	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[a] CGT: Cancer Gene Trust.

[b] OMOP: Observational Medical Outcomes Partnership.

[c] UUID: universally unique identifier.

[d] FMI: Foundation Medicine.

[e] UCSF: University of California, San Francisco

[f] CT: computed tomography.

[g] N/A: not applicable.

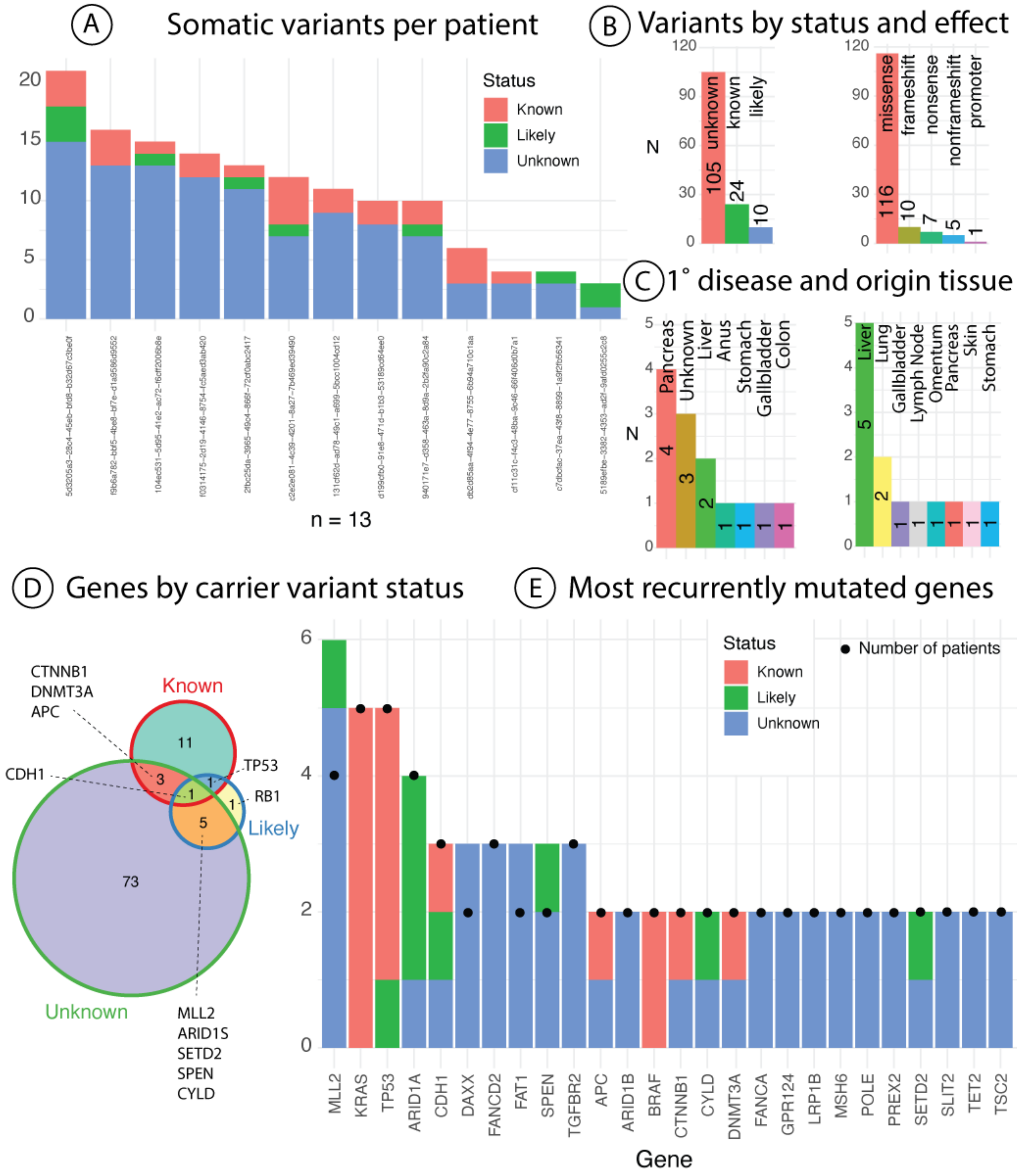
Breakdown of Available Data in Cancer Gene Trust by Patient

The CGT Public ID refers to the globally unique hexadecimal identifier per patient. ✓ indicates that data are available for that particular modality per patient. For the OMOP data, the numbers reflect how many data elements are available per modality.

Genomic Breakdown of Cancer Gene Trust Cohort of Patients With Foundation One Reports

Of patients with genomic data, the majority (n=13) had Foundation One sequencing performed and, as such, we focus on these data for a breakdown analysis (Figure 2). Across all patients, we identified 139 mutations in 95 genes (Multimedia Appendix 1). On average, patients had 10.69 (SD 5.34) somatic variants, with the most being 21 and the fewest being 3, across different current knowledge status (ie, known pathogenic, likely pathogenic, or of unknown consequence; panel A). On average, these somatic variants were primarily unknown (panel B left), with a mean of 8.07 (SD 4.57) per patient. Patients had an average of 2.18 (SD 0.98) of known and 1.43 (SD 0.79) likely variants. In terms of their functional effect, the majority of variants were missense (83.5% (116/139), panel B right). These patients had various primary diseases, the most prevalent being pancreatic (n=4, panel C left). For these patients, biopsies were taken from various tissues of origin, the most prevalent being liver (n=5, panel C right). Please refer to Multimedia Appendix 1 for a diagram illustrating connections between tissue of origin and primary disease for these patients. We further break down the functional effect and status of variants by tissue of origin and primary disease in Multimedia Appendix 1. These, of course, should be considered in context to the number of patients by tissue of origin and primary disease. With these considerations, we still found some interesting trends. For instance, lymph node tissue of origin (n=1 patient) had the fewest variants (n=3) with no known pathological variants, whereas omentum tissue of origin had the most for a single patient (n=21) with three known pathological variants. Of course, these trends could depend on patient-specific or severity variations, and will require more patient data.

Figure 2. Breakdown of Foundation One genomics results for pilot cohort. (A) The breakdown of number of variants reported per patient stratified by their current knowledge status. (B) Breakdown of all variants for patients by effect (left) current knowledge status (right). (C) Distribution of cancer type per patient by primary disease (left) and tissue of origin (right). (D) Distribution of genes based on the current knowledge status of encompassed variants. (E) List of the most commonly recurrently mutated genes (N>1) by number of encompassed variants by status. Black dots represent number of unique patients with a variant in the gene.



Across all patients, the 95 genes contained variants with various levels of knowledge status, including overlapping domains if there were more than one variant identified per gene (panel D). Here, we also see that the majority of genes had variants of unknown significance (n=82). There were 10 genes that contained multiple variants of different knowledge classes. The most commonly mutated genes across individuals (Multimedia Appendix 1 and panel E, black dots) were KRAS and TP53 (n=5) followed by ARID1A and MLL2 (n=4). In a similar vein, MLL2 has the most unique variants identified (n=6 variants across 4 patients), followed by KRAS and TP53 (n=5 variants across 5 patients). The majority of genes with more than one carrier contained variants of unknown significance only (54% (14/26)), further exemplifying the need for combining real-world

EHR with such genomic data. We further visualize the landscape of variants of unknown significance by effect overall and on a per-patient level in Multimedia Appendix 1. We notice that for 1 patient with pancreatic cancer and tissue biopsy, for instance, there is one nonsense mutation in TGFBR2 with a current unknown knowledge status. According to TCGA GDC data portal, there are only 15 cases of primary site pancreatic cancer (TCGA-PAAD) with variants in this gene, and only two are stop-gain. Sharing data such as these with other researchers could quickly expand current knowledge status of variants and their association with disease.

Comparing Robustness of Clinical Data Procedures

To identify the most robust format of clinical data to share on CGT, we assessed whether there was a significant difference in scoring quality between two disparate data formats, specifically the prospectively collected registry and retrospectively gathered OMOP. We hypothesized that there would be no overall difference in scoring quality because both methodologies in theory should capture the main core competencies of interest.

Although we found that total score across all patients and data elements were higher for registry compared with OMOP (Table 4; 642 vs 560), this difference was not statistically significant (*P*=.13, V=44).We further analyzed any significant discrepancies by core competency data element (Multimedia Appendix 1; Table 1 for element descriptions and source).We found no significant difference for Gender (*P*=.35, V=3), Ethnicity (*P*=.17, V=6), Race (*P*=.17, V=13), Year of Birth (*P*=.35, V=3), Basis of Diagnosis (*P*=.66, V=45), Cancer Site (*P*=.09, V=0), Therapeutic Agent/Modality (*P*=.17, V=21), and Beginning and End Dates of Treatment (*P*=.47, V=20). We did find, however, that there was a significant difference between OMOP and registry scoring for Date of Diagnosis (*P*=.004, V=0), with registry having higher scores (*P*=.002, V=0), and Cancer Histology (*P*=.0004, V=0), with registry having higher scores (*P*<.001, V=0). See Multimedia Appendix 1 for per patient, per element scores for registry and OMOP, respectively.

Break down of *gold standard* elements and their respective fields in registry and OMOP is given in Table 4.

Table 4. Overall patient scores for registry vs Observational Medical Outcomes Partnership formats.

Patient	Registry	OMOP[a]
f9b6a782-bbf5-4be8-bf7e-d1a9586d9552	39	28
c2e2e081-4c39-4201-8a27-7b469ed39490	41	34
db2d85aa-4f94-4e77-8755-6b94a710c1aa	42	32
2fbc25da-3965-49c4-866f-72cf0abc2417	48	30
940171e7-d358-463a-8d9a-2b2fa90c2a84	31	41
f0314175-2d19-4146-8754-fc5aed3ab420	29	39
c7dbcfac-37ea-43f8-8899-1a9f2fb56341	15	33
ef5c3164-6f45-4d3a-88f0-4509226c5571	50	29
ec3d977b-c310-4df3-a444-f79bc3dd8b58	35	33
131cf62d-ad78-49c1-a699-5bcc1004cd12	35	33
cf11c31c-f4c3-48ba-9c46-66f406d0b7a1	47	29
ccc2ba97-912f-4b62-b767-cca129ee6a56	13	33
104ec531-5d95-41e2-ac72-f6cff2006b8e	35	24
a5627ac3-450d-4036-ade8-99ae62a5c232	45	34
5189efbe-3382-4353-ad2f-9afd0255c2c8	47	38
253f0e2d-bebd-464b-81c5-8dd8385192b3	46	37
d199cfb0-91e8-471d-b1b3-53189cd64ee0	44	33
Total	642	560

[a] OMOP: Observational Medical Outcomes Partnership.

Total score per patient per data modality, specifically registry vs OMOP, compared with *gold standard* raw EHR data. Each score is the sum of all elements analyzed. Patient scores broken down by element can be found in Multimedia Appendix 1.

Developing a Clinical Narrative From Cancer Gene Trust Data

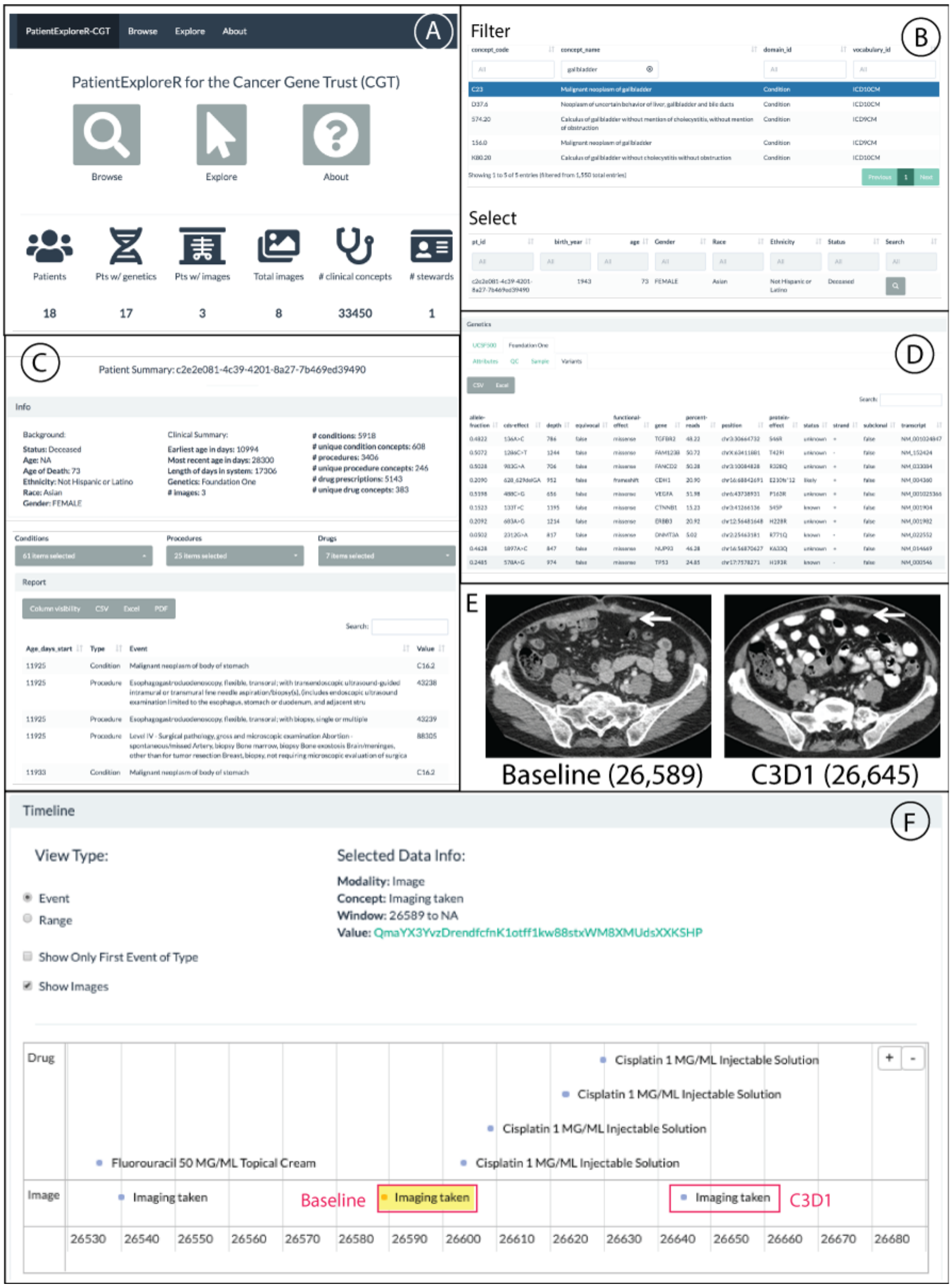
Although safely, securely, and robustly sharing clinically related patient data is an important procedure in and of itself, we want to demonstrate the power of this framework by compiling a clinical narrative solely from data shared on CGT. We elected to use patient c2e2e081-4c39-4201-8a27-7b469ed39490 as a highlighted example (see Multimedia Appendix 1 for all relevant CGT hash information for this patient). We further show how to identify these data points using PatientExploreR-CGT in the following section.

On Day 1 (26,346 days from birth), patient underwent laparoscopic cholecystectomy (at a prior institution) which confirmed moderately differentiated adenocarcinoma with mucinous features. On day 42 (26,387), pathology was reviewed at UCSF which confirmed stage at pT2Nx. On Day 75 (26,420), patient underwent open partial hepatectomy, portal lymphadenectomy, and appendectomy. An FNA of RUQ skin nodule at prior trochar site on Day 195 (26,540) identified adenocarcinoma consistent with recurrence/metastasis from primary gallbladder site. CT C/A/P on Day 196 (26,540) showed multiple new peritoneal and ventral abdominal wall soft tissue nodules suspicious for metastases.

Patient signed informed consent for CC#16457 clinical trial on Day 238 (26,583) andcompleted baselinescans on Day 244 (26,589; Figure 3E [left]; hash ids: QmaYX3YvzDrendfcfnK1otff1kw88stxWM8XMUdsXXKSHP [parent], Qmd7V8hS2mCtupRLYk6Qm2AMHyk6X7Y4QPTDqZe7UCUnUT [image]) which showed unchanged disease from Day 196. Patient randomized to Arm B: merestinib/placebo + cisplatin + gemcitabine (not available in OMOP data) on Day 257 (26,602) and completed Cycle 1, Day 1 cisplatin + gemcitabine on Day 260 (26,605). On Day 286 (26,631), Cycle 2, Day 8 cisplatin + gemcitabine was completed.

On Day 300, a CT C/A/P was performed (26,645; Figure 3E [right]; hash ids: QmQ6PtwhTMqw9b3SFsa1qfW79kGK7tPrhrUHpKVLtxmj1i [parent], QmZmVEsqNeCDuzUDDvLWYUdbxQ2QZehDhkdzyCNvX8gFJF [image]) and showed stable scattered abdominal wall, peritoneal and retroperitoneal implants. Interval progression of mild intrahepatic biliary dilatation, possibly due to new soft tissue prominence at the porta hepatis, concerning for recurrence. However, unchanged small upper lobe pulmonary nodules were noted and stable disease was concluded per RECIST, with 18.18% decrease in sum of target lesion diameters.

Figure 3. PatientExploreR visualization app for Cancer Gene Trust (CGT). Breakdown of features available for the public PatientExploreR visualization application for CGT. (A) Landing page which has all options for browsing patients, exploring patient data, and information about CGT. This page also displays the amount of data currently compiled on CGT. (B) Browse tab for filtering for patient based on clinical concepts and/or demographics. This list is filtered by Gallbladder-related disease and patient c2e2e081-4c39-4201-8a27-7b469ed39490 was selected. (C) Explore tab that details patient's background and clinical summary. The user can interact with clinical, genomic, and imaging data for selected patient. These data can be filtered and exported and control what is shown visually in the timeline view below. (D) Genomics data extracted and displayed from either the Foundation One or UCSF 500 report. (E) Two sample image scans that can be found via the imaging submenu or from clicking within the timeline view below. These CT scans show baseline and C3D1, 26,589 and 26,645 age in days, respectively. Baseline contrast enhanced CT of the abdomen shows a peritoneal metastasis (arrow) measuring 12×8 mm. Posttreatment contrast enhanced CT of the abdomen shows decreased size of the peritoneal metastasis (arrow) measuring 10×6 mm. (F) Timeline view of selected clinical and image data. This timeline view was produced by selecting the associated relevant medications in the Drug pull-down menu, specifically: capecitabine, cisplatin, fluorouracil, and trastuzumab. With Show Images selected, we can see what relevant medications the patient was prescribed when the images were taken. Pressing the link next to Value above, the user will be directed to all images taken during that visit, which can be viewed on an appropriate (eg, DICOM viewer) browser.



Patients	Pts w/ genetics	Pts w/ images	Total images	# clinical concepts	# stewards
18	17	3	8	33450	1

Filter

concept_code	concept_name	domain_id	vocabulary_id
All	gallbladder	All	All
C23	Malignant neoplasm of gallbladder	Condition	ICD10CM
D37.6	Neoplasm of uncertain behavior of liver, gallbladder and bile ducts	Condition	ICD10CM
574.20	Calculus of gallbladder without mention of cholecystitis, without mention of obstruction	Condition	ICD9CM
156.0	Malignant neoplasm of gallbladder	Condition	ICD9CM
K80.20	Calculus of gallbladder without cholecystitis without obstruction	Condition	ICD10CM

Showing 1 to 5 of 5 entries (filtered from 1,550 total entries)

Select

pt_id	birth_year	age	Gender	Race	Ethnicity	Status	Search
All	All	All	All	All	All	All	All
c2e2e081-4c39-4201-8a27-7b469ed39490	1943	73	FEMALE	Asian	Not Hispanic or Latino	Deceased	

Genetics

allele-fraction	cds-effect	depth	equivocal	functional-effect	gene	percent-reads	position	protein-effect	status	strand	subclonal	transcript
0.4822	136A>C	786	false	missense	TGFBR2	48.22	chr3:30664732	S46R	unknown	+	false	NM_001024847
0.5072	1286C>T	1244	false	missense	FAM123B	50.72	chrX:63411881	T429I	unknown	-	false	NM_152424
0.5028	983G>A	706	false	missense	FANCD2	50.28	chr3:10084828	R328Q	unknown	+	false	NM_033084
0.2090	628_629delGA	952	false	frameshift	CDH1	20.90	chr16:68842691	E210fs*12	likely	+	false	NM_004360
0.5198	488C>G	656	false	missense	VEGFA	51.98	chr6:43738931	P163R	unknown	+	false	NM_001025366
0.1523	133T>C	1395	false	missense	CTNNB1	15.23	chr3:41266136	S45P	known	+	false	NM_001904
0.2092	683A>G	1214	false	missense	ERBB3	20.92	chr12:56481648	H228R	unknown	+	false	NM_001982
0.0502	2312G>A	817	false	missense	DNMT3A	5.02	chr2:25463181	R771Q	known	-	false	NM_022552
0.4628	1897A>C	847	false	missense	NUP93	46.28	chr16:56870627	K633Q	unknown	+	false	NM_014669
0.2485	578A>G	974	false	missense	TP53	24.85	chr17:7578271	H193R	known	-	false	NM_000546

Patient Summary: c2e2e081-4c39-4201-8a27-7b469ed39490

Info

Background:
Status: Deceased
Age: NA
Age of Death: 73
Ethnicity: Not Hispanic or Latino
Race: Asian
Gender: FEMALE

Clinical Summary:
Earliest age in days: 10994
Most recent age in days: 28300
Length of days in system: 17306
Genetics: Foundation One
 # images: 3

 # conditions: 5918
 # unique condition concepts: 608
 # procedures: 3406
 # unique procedure concepts: 246
 # drug prescriptions: 5143
 # unique drug concepts: 383

Conditions: 61 items selected
Procedures: 25 items selected
Drugs: 7 items selected

Report

Age_days_start	Type	Event	Value
11925	Condition	Malignant neoplasm of body of stomach	C16.2
11925	Procedure	Esophagogastroduodenoscopy, flexible, transoral; with transendoscopic ultrasound-guided intramural or transmural fine needle aspiration/biopsy(s), (includes endoscopic ultrasound examination limited to the esophagus, stomach or duodenum, and adjacent stru	43238
11925	Procedure	Esophagogastroduodenoscopy, flexible, transoral; with biopsy, single or multiple	43239
11925	Procedure	Level IV - Surgical pathology, gross and microscopic examination Abortion - spontaneous/missed Artery, biopsy Bone marrow, biopsy Bone exostosis Brain/meninges, other than for tumor resection Breast, biopsy, not requiring microscopic evaluation of surgica	88305
11933	Condition	Malignant neoplasm of body of stomach	C16.2

Timeline

View Type:
● Event
○ Range
☐ Show Only First Event of Type
☑ Show Images

Selected Data Info:
Modality: Image
Concept: Imaging taken
Window: 26589 to NA
Value: QmaYX3YvzDrendfcfnK1otff1kw88stxWM8XMUdsXXKSHP

Exploring Cancer Gene Trust Data on PatientExploreR-Cancer Gene Trust

To further operationalize the CGT framework, we adapted an application called PatientExploreR to seamlessly interface with CGT to effectively explore, visualize, and download the data. We envision this application to be particularly useful for individuals without much data extraction and manipulation experience. This application requires no registration and is publicly available [52]. PatientExploreR-CGT pulls all OMOP data from CGT, maps all clinical concepts according to the CDM, and provides convenient links to genomic data as well as image data in the context of their clinical history. In Figure 3, we demonstrate the power of the visualization by showing a detailed timeline of the above patient's treatment timeline around the time of the available CT scans.

Discussion

Overview

In this study, we have consented patients in an IRB-approved process to share deidentified EHRs, genomic, and imaging data using a blockchain-authenticated framework called CGT. Our goal of this pilot study was to demonstrate the process of patient consent to data sharing within a large public health institution as well as to create a framework that can facilitate other institutions, physicians, and patients to add their own data. The benefit of a block-chain authenticated system was more geared to decentralized access (authorization) rather than privacy or security (authentication) as all submissions are public by design. As we hypothesized, all 18 patients in the pilot study did not have reservations about sharing their data, which has been similarly demonstrated [9], and we believe patients from other institutions have similar beliefs. Patient privacy was a top priority for this project and we actively coordinated the highest-standards for deidentification processing of all data shared (see Multimedia Appendix 1 for deidentification process).

In designing the CGT, we had to overcome the existing challenges in this space, namely that this framework should be secure, efficient, and scalable while being cost-efficient, open to the public, and not owned by a single institution. We also had to determine not only which data should be shared but also the appropriate format of such data that would balance interoperability with speed of sharing. Our pilot also addresses cultural and institutional challenges, both perceived and real, including the IRB, patient consent and education, and other elements.

CGT is designed as an alternate approach to centralized data repository platforms such as Medical Information Mart for Intensive Care [53] which have enabled a slew of powerful research. Unlike these primarily static databases, CGT can facilitate rapid and continual data being shared from the clinical care system as close to the time of generation and extraction as possible. Both systems have their merits and hopefully they will be complementary in providing access to deidentified EHR data to enable personalized medicine. Furthermore, CGT enables researchers to use and interpret medical data instead of resolving disparate access methods from multiple sources or failing entirely because data are simply not available in any format. Indeed, it is our hope that CGT can facilitate research studies and enhance clinical care on a timescale not previously possible, while allowing data holders to maintain the privacy and security of individual data sources and the nonpublic subset of the data [36]. At the same time, this entire process will respect individual patient consents and cultural data sharing preferences and expectations. CGT enables aggregation of data from all consenting patients. CGT might bolster cancer research and help physicians, patients, payers, and other stakeholders make more informed decisions about the increasingly complex diagnosis and treatment of cancer as well as its reimbursement. CGT functions as a bridge between the highly regulated HIPAA environment (Figure 1) and the open World Wide Web internet environment. To alleviate concerns about data ownership, CGT is built on a decentralized, democratized blockchain format and will remain free and open.

Principal Findings

Compared with a list of *gold standard* data elements [11] that should be shared in such a project, we found that there was no significant difference in completeness between a prospectively collected registry and a retrospective (OMOP) procedure for clinical data. Certain data elements, however, were more robustly recorded in the registry format, specifically Date of Diagnosis and Cancer Histology. For analyses that aim to further personalized medicine, such pieces of information might be critical, and we hope the findings from this study can help improve the continually adapting OMOP model to better encode such information. These lapses could also be due to institution-specific extract, transform, load (ETL) procedures.

Each strategy has its respective benefits and weaknesses. Because registry data are manually coded, specific key pieces of information can be easily highlighted and identified. Furthermore, for registry data to be submitted to SEER, all pieces of information must be detailed, but this process is manual and time consuming, and often results in different stages of aggregation per patient. As such, we found higher levels of variance in registry records compared with those in OMOP (mean 37.77, SD 10.87 vs mean 32.94, SD 4.26), which could reflect delays in manual data aggregation (ie, *suspense* states) or quality. It was clear though that more patients had more complete information from registry data than OMOP, with 5 patients having more than 90% completeness cores (ie, >45 total score) in registry vs 0 in OMOP. However, by relying on the open source OMOP standard, instead of registry or a proprietary EHR structure, the barrier for distributing and sharing data is drastically lowered through reducing ETL transformation, which also lowers cost through leveraging the conversion processes already occurring in many hospital systems. Researchers recently demonstrated the power of OMOP for facilitating phenotype transfer across sites [54], which aligns well with the goal of CGT. The additional costs of time are the clinical and regulatory tasks involved in consenting patients and obtaining, anonymizing, and uploading data. This process accounts for the majority of cost which will further decrease in high volume.

Limitations

There are many limitations of this study that need to be addressed. Both the registry and OMOP EHR extract did not contain all valuable and relevant core data elements. Therefore, the comparison of data robustness cannot be extended to all *gold-standard* elements that ideally should be shared in such a project. As OMOP is from retrospective extraction process, there is no immediate way to automatically identify primary cancer and therapeutic efficacy, although we hope this can be mediated by subsequent incorporation of deidentified notes or new schema adaptations or developments. Similar to any noncurated database, data quality for both registry and OMOP is limited by those who entered it and could be affected by infrastructural biases of individuals and EHR systems [55]. In addition, the current framework is steward based, which means that there needs to be a single individual or team representative to submit data per institution. Similar to any cross-institution data link of deidentified data, there is no procedure in place to be able to map the same patient across stewards as there exists within the registry system. Although we tried to create a rule-based scoring system that is as unbiased as possible involving 2 separate reviewers, the manual scoring of data elements did contain levels of subjectivity and potential ambiguity, which is fully detailed in the Multimedia Appendix 1.

There are also risks of reidentification associated with data sharing, even beyond accidental leakage. Even for incomplete, fully deidentified data, for instance, a recent study was able to use generative copula-based method to accurately reidentify 99.98% of American individuals based on only 15 demographic attributes [56]. Of course, many of these variables used in this paper are not available in this dataset, but it is important to note as other models might be developed in the future those could be applied to the data shared. Overall, these risks need to be weighed against the stagnation associated with keeping these valuable data siloed. Not sharing all details pertaining to treatment efficacy and adverse drug effects are not in the best interests of general public and overall scientific and medical community. Despite these limitations, open scientific data sharing has been an enormous boon in many fields and we believe that CGT presents a proof of concept that useful medical data can be openly shared. We further demonstrated the feasibility and utility of this process in a pilot study and provide fully detailed steps for other institutions to consent and add their patients' data. The ultimate success of this platform will be determined by the flow of patient data and how it can be used to facilitate discoveries and help personalize treatment.

Conclusions

Each cancer case is unique and requires as much data as possible to inform ideal treatment decisions. The more data that exist and are released can help clinicians identify ideal personalized treatment for their patients. We found the OMOP CDM is a scalable format for dissemination, although it can be improved by better information in key data element fields such as cancer histology as compared with a prospectively collected registry format. The OHDSI Oncology Working Group [12] is currently developing an extension to OMOP to support observational cancer research that better captures and records elements we found available in the registry format but not in the current OMOP implementation. We believe such an effort is invaluable to reconcile these differences and should be integrated into the future version of CGT. Put together, we hope that the CGT framework, pilot study, and interactive visualization application furthers the ideals of the cancer Moonshot project, unleashing data trapped in silos to further cancer research and reveal patterns that can help further personalize treatment.

Acknowledgments

We acknowledge the UCSF Academic Research Systems group for making available the deidentified OMOP data. We are grateful for Foundation Medicine for supplying XML files. We thank Boris Oskotsky for help setting up the server for PatientExploreR-CGT. We thank Atul Butte and Barbara Koening for helpful comments and Max Haeussler for the early technical inspiration. We also wish to thank Rachael Liao, Bartha Knoppers, Adrian Thorogood, and the Global Alliance for Genomics and Health (GA4GH) for their collaboration and guidance. BG was affiliated with the Bakar Computational Health Sciences Institute at the University of California, San Francisco, at the time of the study and is currently affiliated with the Hasso Plattner Institute for Digital Health at Mount Sinai within the Icahn School of Medicine at Mount Sinai. This publication was supported partially by the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI [UL1 TR001872], National Cancer Institute of the National Institutes of Health under award numbers 5U24CA180951-04 and 5U24CA210974-02 to DH, as well as a Marcus Foundation Award to EC. TG graciously acknowledges support from the National Cancer Institute Oncology Model Forum project National Institutes of Health grant U24 CA195858.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials, including supplementary methods, figures, and tables.
[DOCX File , 2909 KB - jmir_v22i3e16810_app1.docx]

References

1. Schilsky RL. Personalized medicine in oncology: the future is now. Nat Rev Drug Discov 2010 May;9(5):363-366. [doi: 10.1038/nrd3181] [Medline: 20431568]
2. Iyer G, Hanrahan AJ, Milowsky MI, Al-Ahmadie H, Scott SN, Janakiraman M, et al. Genome sequencing identifies a basis for everolimus sensitivity. Science 2012 Oct 12;338(6104):221 [FREE Full text] [doi: 10.1126/science.1226344] [Medline: 22923433]
3. Sherman RE, Anderson SA, Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016 Dec 8;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
4. Glicksberg BS, Johnson KW, Dudley JT. The next generation of precision medicine: observational studies, electronic health records, biobanks and continuous monitoring. Hum Mol Genet 2018 May 1;27(R1):R56-R62. [doi: 10.1093/hmg/ddy114] [Medline: 29659828]
5. Norgeot B, Glicksberg BS, Butte AJ. A call for deep-learning healthcare. Nat Med 2019 Jan;25(1):14-15. [doi: 10.1038/s41591-018-0320-3] [Medline: 30617337]
6. Nwaru BI, Friedman C, Halamka J, Sheikh A. Can learning health systems help organisations deliver personalised care? BMC Med 2017 Oct 2;15(1):177 [FREE Full text] [doi: 10.1186/s12916-017-0935-0] [Medline: 28965492]
7. Agarwala V, Khozin S, Singal G, O'Connell C, Kuk D, Li G, et al. Real-world evidence in support of precision medicine: clinico-genomic cancer data as a case study. Health Aff (Millwood) 2018 May;37(5):765-772. [doi: 10.1377/hlthaff.2017.1579] [Medline: 29733723]
8. Adler-Milstein J, Jha AK. Sharing clinical data electronically: a critical challenge for fixing the health care system. J Am Med Assoc 2012 Apr 25;307(16):1695-1696. [doi: 10.1001/jama.2012.525] [Medline: 22535851]
9. Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. N Engl J Med 2018 Jun 7;378(23):2202-2211 [FREE Full text] [doi: 10.1056/NEJMsa1713258] [Medline: 29874542]
10. Beaulieu-Jones BK, Wu ZS, Williams C, Lee R, Bhavnani SP, Byrd JB, et al. Privacy-preserving generative deep neural networks support clinical data sharing. Circ Cardiovasc Qual Outcomes 2019 Jul;12(7):e005122. [doi: 10.1161/CIRCOUTCOMES.118.005122] [Medline: 31284738]
11. Conley RB, Dickson D, Zenklusen JC, Al Naber J, Messner DA, Atasoy A, et al. Core clinical data elements for cancer genomic repositories: a multi-stakeholder consensus. Cell 2017 Nov 16;171(5):982-986 [FREE Full text] [doi: 10.1016/j.cell.2017.10.032] [Medline: 29149611]
12. OHDSI – Observational Health Data Sciences and Informatics. OHDSI Oncology Working Group URL: https://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:oncology-sg# [accessed 2019-09-01]
13. Singal G, Miller PG, Agarwala V, Li G, Kaushik G, Backenroth D, et al. Association of patient characteristics and tumor genomics with clinical outcomes among patients with non-small cell lung cancer using a clinicogenomic database. J Am Med Assoc 2019 Apr 9;321(14):1391-1399 [FREE Full text] [doi: 10.1001/jama.2019.3241] [Medline: 30964529]
14. Jones M, Johnson M, Shervey M, Dudley JT, Zimmerman N. Privacy-preserving methods for feature engineering using blockchain: review, evaluation, and proof of concept. J Med Internet Res 2019 Aug 14;21(8):e13600 [FREE Full text] [doi: 10.2196/13600] [Medline: 31414666]
15. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: applying blockchain to securely and scalably share clinical data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
16. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and trustable electronic medical records sharing using blockchain. AMIA Annu Symp Proc 2017;2017:650-659 [FREE Full text] [Medline: 29854130]
17. Mamoshina P, Ojomoko L, Yanovich Y, Ostrovski A, Botezatu A, Prikhodko P, et al. Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget 2018 Jan 19;9(5):5665-5690 [FREE Full text] [doi: 10.18632/oncotarget.22345] [Medline: 29464026]
18. Ozercan HI, Ileri AM, Ayday E, Alkan C. Realizing the potential of blockchain technologies in genomics. Genome Res 2018 Sep;28(9):1255-1263 [FREE Full text] [doi: 10.1101/gr.207464.116] [Medline: 30076130]
19. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb 22;10(1):917 [FREE Full text] [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]
20. Kuo T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 1;26(5):462-478. [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
21. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility. J Med Internet Res 2019 Feb 8;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
22. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing blockchains for efficient health care: systematic review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]
23. Cancer Gene Trust. URL: https://www.cancergenetrust.org [accessed 2019-10-27]
24. Ethereum. URL: https://ethereum.org [accessed 2019-10-27]
25. IPFS Powers the Distributed Web. URL: https://ipfs.io [accessed 2019-10-27]
26. Cancer Gene Trust Protocol. Sharing Clinical and Genomic Data in Cancer Research URL: https://www.cancergenetrust.org/docs/cgt-ucsf-protocol.pdf [accessed 2019-10-27]

27. GitHub. Cancer Gene Trust URL: https://www.github.com/cancergenetrust [accessed 2019-10-27]
28. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013 Nov;31(11):1023-1031 [FREE Full text] [doi: 10.1038/nbt.2696] [Medline: 24142049]
29. Joseph NM, Chen Y, Nasr A, Yeh I, Talevich E, Onodera C, et al. Genomic profiling of malignant peritoneal mesothelioma reveals recurrent alterations in epigenetic regulatory genes BAP1, SETD2, and DDX3X. Mod Pathol 2017 Feb;30(2):246-254 [FREE Full text] [doi: 10.1038/modpathol.2016.188] [Medline: 27813512]
30. Foundation Medicine. URL: https://www.foundationmedicine.com/ [accessed 2019-10-27]
31. Kline CN, Joseph NM, Grenert JP, van Ziffle J, Talevich E, Onodera C, et al. Targeted next-generation sequencing of pediatric neuro-oncology patients improves diagnosis, identifies pathogenic germline mutations, and directs targeted therapy. Neuro Oncol 2017 May 1;19(5):699-709 [FREE Full text] [doi: 10.1093/neuonc/now254] [Medline: 28453743]
32. OHDSI – Observational Health Data Sciences and Informatics. URL: https://www.ohdsi.org/ [accessed 2019-09-01]
33. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann Intern Med 2010 Nov 2;153(9):600-606. [doi: 10.7326/0003-4819-153-9-201011020-00010] [Medline: 21041580]
34. Surveillance, Epidemiology, and End Results Program. SEER Program Coding and Staging Manual 2018 URL: https://seer.cancer.gov/tools/codingmanuals/ [accessed 2019-09-01]
35. NAACCR. Data Standards & Data Dictionary, Volume II URL: https://www.naaccr.org/data-standards-data-dictionary/ [accessed 2019-09-01]
36. Cancer Gene Trust. dapp URL: https://www.cancergenetrust.org/docs/about [accessed 2019-10-27]
37. Glicksberg BS, Oskotsky B, Thangaraj PM, Giangreco N, Badgeley MA, Johnson KW, et al. PatientExploreR: an extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model. Bioinformatics 2019 Nov 1;35(21):4515-4518 [FREE Full text] [doi: 10.1093/bioinformatics/btz409] [Medline: 31214700]
38. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. CRAN - R Project. 2015. Shiny: web application framework for R URL: https://cran.r-project.org/web/packages/shiny/index.html [accessed 2019-10-27]
39. Perrier V, Meyer F. CRAN - R Project. 2018. shinyWidgets: Custom Inputs Widgets for Shiny URL: https://cran.r-project.org/web/packages/shinyWidgets/shinyWidgets.pdf [accessed 2019-10-27]
40. Attali D. CRAN - R Project. 2017. shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds URL: https://cran.r-project.org/web/packages/shinyjs/index.html [accessed 2019-10-27]
41. Attali D, Edwards T. CRAN - R Project. 2018. shinyalert: Easily Create Pretty Popup Messages (Modals) in 'Shiny' URL: https://cran.r-project.org/web/packages/shinyalert/shinyalert.pdf [accessed 2019-10-27]
42. Sali A. CRAN - R Project. 2017. shinycssloaders: Add CSS Loading Animations to 'shiny' Outputs URL: https://cran.r-project.org/web/packages/shinycssloaders/index.html [accessed 2019-10-27]
43. Bailey E. CRAN - R Project. 2015. shinyBS: Twitter Bootstrap Components for Shiny URL: https://cran.r-project.org/web/packages/shinyBS/shinyBS.pdf [accessed 2019-10-27]
44. Chang W. CRAN - R Project. 2015. shinythemes: Themes for Shiny URL: https://cran.r-project.org/web/packages/shinythemes/index.html [accessed 2019-10-27]
45. Sievert C, Parmer C, Hocking T, Chamberlain S, Ram K, Corvellec M. CRAN - R Project. 2017. plotly: Create Interactive Web Graphics via 'plotly.js' URL: https://cran.r-project.org/web/packages/plotly/index.html [accessed 2019-10-27]
46. Attali D, Almende B. CRAN - R Project. 2016. timevis: Create Interactive Timeline Visualizations in R URL: https://cran.r-project.org/web/packages/timevis/index.html [accessed 2019-10-27]
47. Glicksberg BS, Oskotsky B, Giangreco N, Thangaraj P, Rudrapatna V, Datta D, et al. ROMOP: a light-weight R package for interfacing with OMOP-formatted electronic health record data. JAMIA Open 2019 Apr;2(1):10-14 [FREE Full text] [doi: 10.1093/jamiaopen/ooy059] [Medline: 31633087]
48. Dowle M, Srinivasan A, Gorecki J, Chirico M, Stetsenko P, Short T. CRAN - R Project. 2018. data.table: Extension of 'data.frame' URL: https://cran.r-project.org/web/packages/data.table/index.html [accessed 2019-10-27]
49. Xie Y, Cheng J, Allaire J, Reavis B, Gersen L, Szopka B. CRAN - R Project. 2016. DT: A Wrapper of the JavaScript Library 'DataTables' URL: https://cran.r-project.org/web/packages/DT/index.html [accessed 2019-10-27]
50. Couture-Beil A. CRAN - R Project. 2018. rjson: JSON for R URL: https://cran.r-project.org/web/packages/rjson/rjson.pdf [accessed 2019-10-27]
51. Wickham H, Francois R, Henry L, Müller K. CRAN - R Project. 2015. dplyr: A grammar of data manipulation URL: https://cran.r-project.org/web/packages/dplyr/index.html [accessed 2019-10-27]
52. PatientExploreR-CGT. URL: http://patientexplorer.cancergenetrust.org [accessed 2019-10-27]
53. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
54. Hripcsak G, Shang N, Peissig PL, Rasmussen LV, Liu C, Benoit B, et al. Facilitating phenotype transfer using a common data model. J Biomed Inform 2019 Aug;96:103253. [doi: 10.1016/j.jbi.2019.103253] [Medline: 31325501]

55. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. Br Med J 2018 Apr 30;361:k1479 [FREE Full text] [doi: 10.1136/bmj.k1479] [Medline: 29712648]
56. Rocher L, Hendrickx JM, de Montjoye Y. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 2019 Jul 23;10(1):3069 [FREE Full text] [doi: 10.1038/s41467-019-10933-3] [Medline: 31337762]

Abbreviations

CDM: common data model
CGT: Cancer Gene Trust
CT: computed tomography
EDW: Enterprise Data Warehouse
EHR: electronic health record
ETL: extract, transform, load
HIPAA: Health Insurance Portability and Accountability Act
IPFS: InterPlanetary File System
IRB: institutional review board
MRN: medical record number
OHDSI: Observational Health Data Sciences and Informatics
OMOP: Observational Medical Outcomes Partnership
PHI: protected health information
SEER: Surveillance, Epidemiology, and End Results
UCSF: University of California, San Francisco
UUID: universally unique identifier

Edited by G Eysenbach; submitted 27.10.19; peer-reviewed by L Rusu, K Yin; comments to author 15.11.19; revised version received 09.12.19; accepted 15.12.19; published 20.03.20.

Please cite as:
Glicksberg BS, Burns S, Currie R, Griffin A, Wang ZJ, Haussler D, Goldstein T, Collisson E
Blockchain-Authenticated Sharing of Genomic and Clinical Outcomes Data of Patients With Cancer: A Prospective Cohort Study
J Med Internet Res 2020;22(3):e16810
URL: http://www.jmir.org/2020/3/e16810/
doi:10.2196/16810
PMID:32196460

Corrigenda and Addenda

Correction: Associations Between Affective States and Sexual and Health Status Among Men Who Have Sex With Men in China: Exploratory Study Using Social Media Data

Zhi-Wei Zheng[1*], MS; Qing-Ling Yang[1*], MS; Zhong-Qi Liu[1], MS; Jia-Ling Qiu[1], MS; Jing Gu[1,2,3], PhD; Yuan-Tao Hao[1,2,3], PhD; Chao Song[4], PhD; Zhong-Wei Jia[5], PhD; Chun Hao[1,2,3], PhD

[1]Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, China
[2]Health Information Research Center, Guangdong Key Laboratory of Medicine, School of Public Health, Sun Yat-sen University, Guangzhou, China
[3]Sun Yat-sen Global Health Institute, Institute of State Governance, Sun Yat-sen University, Guangzhou, China
[4]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
[5]National Institute on Drug Dependence, Peking University, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Chun Hao, PhD
Department of Medical Statistics
School of Public Health
Sun Yat-sen University
No 74, Zhongshan 2nd Road
Guangzhou, 510080
China
Phone: 86 13922750299
Email: haochun@mail.sysu.edu.cn

Related Article:

Correction of: https://www.jmir.org/2020/1/e13201/

(J Med Internet Res 2020;22(3):e18135) doi:10.2196/18135

In “Associations Between Affective States and Sexual and Health Status Among Men Who Have Sex With Men in China: Exploratory Study Using Social Media Data” (J Med Internet Res 2020;22(1):e13201), there was an error which was not identified during the proofing stage.

The original published Acknowledgments section was incorrectly listed as:

> *We are grateful to the participants of this study. This project was funded by the National Natural Science Foundation of China (grant number #71974212, #71774178, and #81803334), Science and Technology Program of Guangzhou, China (grant number #201607010332), and A Major Infectious Disease Prevention and Control of the National Science and Technology Major Project (grant number #2018ZX10715004).*

The correct Acknowledgments section is:

> *We are grateful to the participants of this study. This project was funded by the National Natural Science Foundation of China (grant number #91546203, #71974212, #71774178, and #81803334), Science and Technology Program of Guangzhou, China (grant number #201607010332) and A Major Infectious Disease Prevention and Control of the National Science and Technology Major Project (grant number #2018ZX10715004).*

The correction will appear in the online version of the paper on the JMIR website on March 3, 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 05.02.20; this is a non–peer-reviewed article;accepted 17.02.20; published 03.03.20.

Please cite as:
Zheng ZW, Yang QL, Liu ZQ, Qiu JL, Gu J, Hao YT, Song C, Jia ZW, Hao C
Correction: Associations Between Affective States and Sexual and Health Status Among Men Who Have Sex With Men in China: Exploratory Study Using Social Media Data
J Med Internet Res 2020;22(3):e18135
URL: https://www.jmir.org/2020/3/e18135
doi:10.2196/18135
PMID:32130135

Corrigenda and Addenda

Correction: The Effect of Robot Attentional Behaviors on User Perceptions and Behaviors in a Simulated Health Care Interaction: Randomized Controlled Trial

Deborah L Johanson[1,2], MHealthPsych; Ho Seok Ahn[2,3], PhD; Bruce A MacDonald[2,3], PhD; Byeong Kyu Ahn[2,3], MScEng; JongYoon Lim[2,3], BComSc; Euijun Hwang[2,3], BComSc; Craig J Sutherland[2,3], PhD; Elizabeth Broadbent[1,2], PhD

[1]Department of Psychological Medicine, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand
[2]Centre for Automation and Robotic Engineering Science, The University of Auckland, Auckland, New Zealand
[3]Department of Electrical, Computer, and Software Engineering, Faculty of Engineering, The University of Auckland, Auckland, New Zealand

Corresponding Author:
Elizabeth Broadbent, PhD
Department of Psychological Medicine
Faculty of Medical and Health Sciences
The University of Auckland
Level 12 Auckland Hospital, Room 12005
2 Park Road
Auckland, 1023
New Zealand
Phone: 64 99236756
Email: e.broadbent@auckland.ac.nz

Related Article:

Correction of: https://www.jmir.org/2019/10/e13667

(J Med Internet Res 2020;22(3):e18362) doi:10.2196/18362

The authors of “The Effect of Robot Attentional Behaviors on User Perceptions and Behaviors in a Simulated Health Care Interaction: Randomized Controlled Trial” (J Med Internet Res 2019;21(10):e13667) noticed a typo in the abstract of their article. The Abstract should have detailed 4 groups rather than 3. In the Abstract, the Methods section has been revised from:

> *A parallel randomized controlled trial with a 1:1:1 allocation ration was conducted.*

to:

> *A parallel randomized controlled trial with a 1:1:1:1 allocation ratio was conducted.*

The correction will appear in the online version of the paper on the JMIR website on March 3, 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 21.02.20; this is a non–peer-reviewed article;accepted 21.02.20; published 03.03.20.

Please cite as:
Johanson DL, Ahn HS, MacDonald BA, Ahn BK, Lim J, Hwang E, Sutherland CJ, Broadbent E
Correction: The Effect of Robot Attentional Behaviors on User Perceptions and Behaviors in a Simulated Health Care Interaction: Randomized Controlled Trial
J Med Internet Res 2020;22(3):e18362
URL: https://www.jmir.org/2020/3/e18362
doi:10.2196/18362
PMID:32130132

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org