Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 20 (2018), Issue 4 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Review

Health-Related Internet Use by Children and Adolescents: Systematic Review

Eunhee Park[1], RN, PHNA-BC, PhD; Misol Kwon[1], RN, BSN

School of Nursing, University at Buffalo, Buffalo, NY, United States

Corresponding Author:
Eunhee Park, RN, PHNA-BC, PhD
School of Nursing
University at Buffalo
201E Wende Hall
Buffalo, NY, 14214
United States
Phone: 1 716 829 3701
Fax: 1 716 829 2021
Email: eunheepa@buffalo.edu

Abstract

Background: The internet is widely used by children and adolescents, who generally have a high level of competency with technology. Thus, the internet has become a great resource for supporting youth self-care and health-related services. However, few studies have explored adolescents' internet use for health-related matters.

Objective: The objective of this systematic literature review was to examine the phenomenon of children and adolescents' health-related internet use and to identify gaps in the research.

Methods: A total of 19 studies were selected from a search of major electronic databases: PubMed, Cumulative Index of Nursing and Allied Health Literature, and PsycINFO using the following search terms: "health-related internet use," "eHealth," "Internet use for health-related purpose," "Web-based resource," "health information seeking," and "online resource," combined with "child," "adolescent," "student," "youth," and "teen." The children's and adolescents' ages were limited to 24 years and younger. The search was conducted from September 2015 to October 2017. The studies identified to contain youth (<24 years) health-related internet use were all published in peer-reviewed journals in the past 10 years; these studies examined general internet use seeking health care services, resources, information, or using the internet for health promotion and self-care. Studies were excluded if they explored the role of the internet as a modality for surveys, recruitment, or searching for relevant literature without specifically aiming to study participants' health-related internet use; focused solely on quality assurance for specific websites; or were designed to test a specific internet-based intervention.

Results: Interesting patterns in adolescents' health-related internet use, such as seeking preventative health care and specific information about medical issues, were identified. Quantitative studies reported rates of the internet use and access among youth, and the purpose and patterns of health-related internet use among youth were identified. A major objective of health-related internet use is to gain information, but there are inconsistencies in adolescents' perceptions of health-related internet use.

Conclusions: This study's findings provide important information on how youth seek information and related support systems for their health care on the internet. The conceptual and methodological limitations of the identified studies, such as the lack of a theoretical background and unrepresentative samples, are discussed, and gaps within the studies are identified for future research. This review also suggests important features for potential Web-based health interventions for children and adolescents.

(J Med Internet Res 2018;20(4):e120) doi:10.2196/jmir.7731

KEYWORDS

Internet; child; adolescent; health information; health-related Internet use; eHealth

Introduction

The internet is widely used by children and adolescents, who generally exhibit a high level of competency with technology [1,2]. Its unique features and major benefits, such as highly engaging and motivating virtual components, as well as the portable, multitasking tools that give users easy and fast access to computers and mobile devices, mean that the internet has become a prevalent mode of communication and networking among youth [3,4]. Adolescents engage in many different activities on the internet, such as information searching, sharing personal information and artifacts, social media use, and recreational activities [5]; up to a quarter of their time is spent using multiple forms of media simultaneously, also known as multi-tasking [6]. As youth have a generally high level of access to the internet in their daily lives [7], it has become a major resource for them in supporting their self-care and health-related activities and services [8-10]. Although the internet is widely accessible and is well accepted by young people, there is as yet only a limited understanding of the patterns and characteristics of youth health-related internet use.

There are different patterns of the internet use by the various subgroups of this population depending on their developmental, gender, and social characteristics. As children progress to early adolescence, general internet usage increases and then levels off, presumably because of the heavier academic workload that teenagers must shoulder when they enter high school [11-13]. Similarly, research conducted on gender differences in internet use during adolescence is inconclusive [14]. Some studies have found boys (58%) to be more frequent users of the internet compared with girls (44%) [15], whereas other studies observed no significant gender difference in internet usage [16,17]. Children and adolescents also display notably different behavior in diverse regions of the globe depending on the local cultural, economic, and technological landscapes in their use of computers, mobile devices, and the internet. For example, a recent study from a cross-cultural context reported that the issue of internet addiction is not restricted to regions with high internet availability [18]. Data have shown that only 20% of African students reported spending an average of over 2 hours per day online compared with 42% and 40% of Chinese and US students, respectively [18]. However, despite the fact that access to the internet is much more limited than in either the United States or China, internet addiction is actually more prevalent in Africa [18].

The availability of high-quality health information can have a significant impact on the health outcomes of an individual. Health-related internet use is known to be associated with socioeconomic status, which is referred to as the digital divide [19]. Information obtained from interpersonal, online, or media sources facilitates the dissemination of new information, as well as influences how individuals shape their experience of health and illness [20]. This is true especially among young adults as they recognize social media as useful sources of information to supplement those received during their health care visits [21]. Online communities and social media are used to enhance access to valuable support networks, foster social inclusion, and facilitate peer-to-peer connections among adolescents with short-term or long-term diagnoses [21,22].

Young people have unique characteristics and can therefore pose special challenges for health promotion. During adolescence, teenagers undergo biological developments that involve physical, emotional, social, and pubertal maturation [23,24]. Due to these unique developmental characteristics, adolescence is also considered the most vulnerable period for engaging in various risky behaviors such as smoking, drugs, and sex [23]. However, adolescents also tend to form healthy habits and learn appropriate practices for their health concerns and management that can last for the rest of their lives [25]. Thus, youth is a critical period for the development of good health practices, highlighting the need to provide specific guidance for information and support related to their health and developmental milestones [26].

The internet offers many potential benefits for adolescent health promotion, including increasing the number of interventions for diverse topics related to the use of the internet among young people [27,28]. However, there is only a limited understanding of health-related internet use among children and adolescents. The purpose of this review is thus to conduct a systematic analysis of the research on this topic during the last 10 years and use the results to develop suggestions for important features that support effective Web-based health interventions for children and adolescents. The specific aims of this systematic review are as follows: (1) to describe the phenomenon of children and adolescents' health-related internet use, (2) to identify benefits and barriers to health-related internet use for children and adolescents, and (3) to examine conceptual and methodological issues in the current literature.

Methods

Search Overview

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses provides useful guidelines for systematic review studies [29]. This review is registered at PROSPERO (International Prospective Register of Systematic Reviews). On the basis of a careful consideration of the purposes of the study, inclusion and exclusion criteria are established to guide the subsequent search process, as shown in Figure 1.

Search Strategy

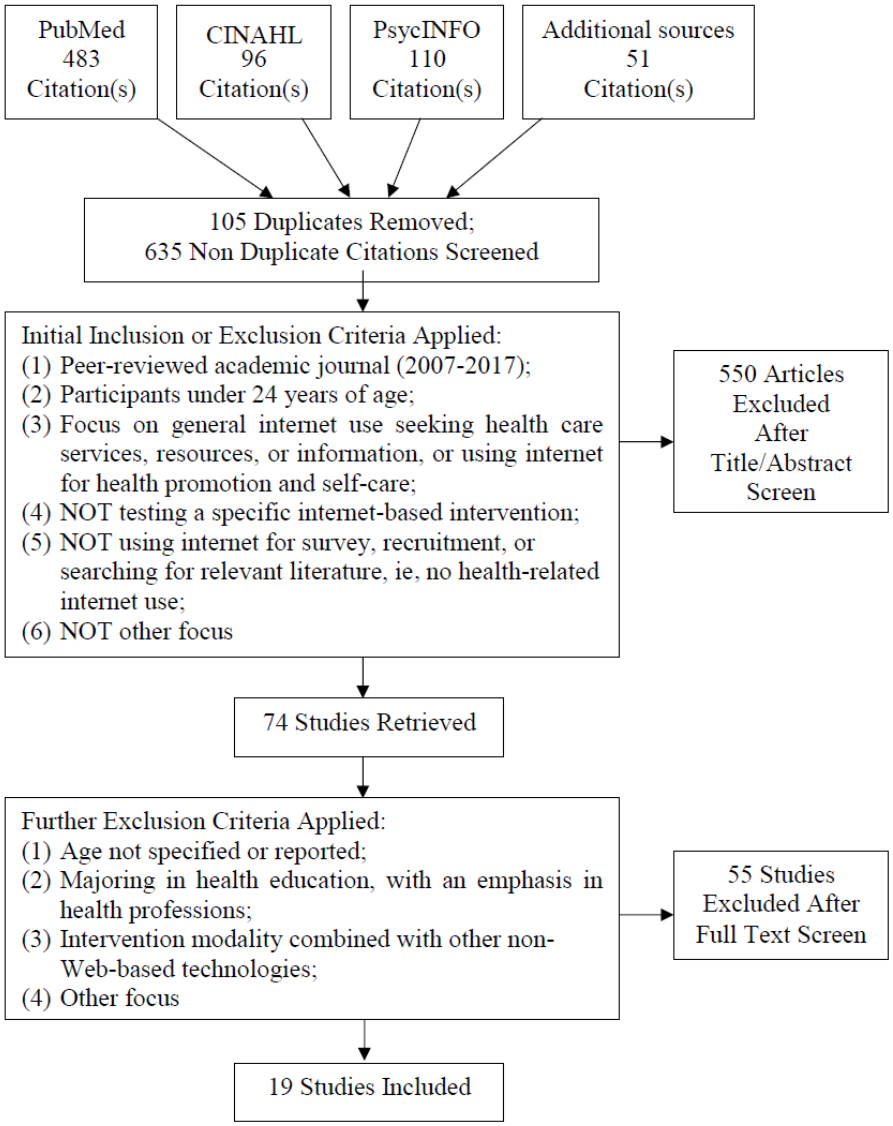
Studies were selected from a search of three major electronic databases: PubMed, the Cumulative Index of Nursing and Allied Health Literature (CINAHL), and PsycINFO. An additional search was conducted using Google Scholar. Studies were also retrieved from the reference lists of the included studies. The search terms consisted of "health-related internet use," "eHealth," "internet use for health-related purpose," "Web-based resource," "online resource," and "health information seeking," combined with "child," "adolescent," "student," "youth," and "teen." The studies were restricted to those concerning children and adolescents aged 24 years and under. The initial search was conducted from September 2015 to October 2017. Studies were included regardless of the location of the study to provide the broadest possible perspective of health-related internet use by

young people. Adopting a global perspective was expected to enable us to examine a wide range of diverse phenomena, some of which could depend on the target population and where the study was conducted.

This study includes those who are up to 24 years to gain a comprehensive picture of health-related internet use among young people. Although there is no universal definition of adolescence, it is traditionally assumed to refer to youth from 12 to 18 years of age, with those in the age range of 18 to 24 years being considered late adolescents or young adults [30]. As there has been no previous systematic study of the health-related internet use of this population, our study was intentionally adjusted to include a broader age range and thus provide a deeper understanding of the unique characteristics of health-related internet use among these subpopulations (both younger and older adolescents) irrespective of location.

The initial search identified 740 studies. After the removal of 105 duplicates, the titles and abstracts of 635 studies were reviewed to determine whether they met the inclusion criteria, resulting in a list of 74 potentially relevant studies. The full texts of these studies were then retrieved for in-depth analysis by two independent reviewers to confirm both the inclusion and exclusion criteria listed below were met, which led to 55 studies being excluded. The remaining 19 studies were included (Figure 1).

Figure 1. Flowchart of the literature search process. CINAHL: Cumulative Index of Nursing and AlliedHealth Literature.



Textbox 1. Inclusion criteria.

Studies were included if

- they were published in a peer-reviewed academic journal from 2007 to 2017
- the study participants were under 24 years of age
- the studies examined general internet use seeking health care services, resources, or information or use of the internet for health promotion and self-care
- they were written in either English or Korean

Textbox 2. Exclusion criteria.

Studies were excluded if

- the study participants were mixed with other populations aged 24 years or older
- the ages of the participants were not specified or reported
- the study participants were trained or were training to become professional health care providers (ie, physicians, nurses, or medical or nursing students)
- the intervention modality was combined with other non-Web-based technologies (such as telephones)
- the internet was simply a modality for conducting surveys or recruitment, or searching for relevant literature, rather than studying participants' health-related internet use
- the study focused solely on quality assurance for specific websites;
- the study consisted of "gray literature" such as dissertations, papers or abstracts in conference proceedings, or editorials
- the study focused on testing a specific internet-based intervention

Eligibility Criteria

The inclusion and exclusion criteria for studies are shown in Textboxes 1 and 2.

Data Extraction, Analysis, and Synthesis

One of the authors initially reviewed the titles and abstracts based on the purpose of the study and the inclusion or exclusion criteria, after which two reviewers independently reviewed the full texts of the studies that were initially selected and coded them into an analysis table. The coding scheme was developed to help identify the components relevant to the study design and to address the first two research questions. The coding scheme included the year of publication, purpose of the study, country, number of participants, participants' characteristics (eg, medical conditions and age), theoretical framework, main constructs, definition of health-related internet use provided, prevalence of health-related internet use, research design, sampling, data collection methods, instruments (including reliability and validity), data analysis, major findings, and study limitations. The coding also identified whether the findings of each of the quantitative studies indicated positive or negative perceptions about health-related internet use, as well as whether more than 50% of the participants had ever used the internet for health-related purposes. The qualitative and the quantitative studies that did not report these findings were coded as nonapplicable.

For the third research question, a coding table was created based on the guidelines suggested by the Agency for Healthcare Research and Quality criteria [29,31] that considered the research design, conceptual framework, sampling method, data collection method, instrument, analytic method, and threats to validity. The coding indicated whether the study utilized random, purposive, or convenience sampling; was quantitative or qualitative; where the data were collected, conducted at a single or multiple sites, and was an online survey, pen and pencil survey, interview, or focus group; and if the study utilized appropriate statistical analytic methods such as descriptive statistics, univariate regression and multivariate regression, or qualitative methods such as thematic analysis and content analysis. Potential threats to validity, such as self-report, a single site study, or self-selection bias were also identified and coded accordingly.

After coding tables were completed, the authors independently checked for discrepancies in the coded results to ensure accuracy. In the case of disagreement between authors, external review from experts in the area of health-related internet use would be considered. In this process, no disagreement was found. After coding was completed, authors synthesized the findings based on each research questions. The findings were then analyzed based on Eysenbach's framework and the objectives of this review [32]. The perceptions of those participating in the various studies about health-related internet use and the prevalence of participants that had ever used the internet for health-related purposes were analyzed using descriptive statistics and the chi-square test to examine the differences between the studies published from 2007 to 2012 and 2013 to 2017.

Results

Characteristics of Study Participants

A total of 19 studies met the inclusion criteria, of which the majority (n=11) were conducted in the United States. Others were conducted in the United Kingdom, Canada, Israel, Nigeria, Sweden, and Uganda. In the selected studies, the majority of the adolescent participants were not suffering from any pertinent medical conditions (n=16); the remainder were identified as having juvenile arthritis [33], type 1 diabetes mellitus [34], or undergoing orthodontic treatment [35]. Apart from 2 studies whose sample specifically consisted of female [36] and male youths [37], all the studies included mixed gender. Hispanics made up the largest group in these studies, with proportions ranging from 9% [38] to 84% [39].

The types of communities people live in serves as a partial indicator of their socioeconomic status, and participants in the studies reported in the literature covered a wide range, from living in predominantly underserved, minority community areas [37,39-42] to middle class income areas [38] and urban areas [33,43-46]. The study settings, study participants, and their characteristics, along with their main findings are summarized in Multimedia Appendix 1. Three studies included participants who were incarcerated in a juvenile detention facility [40], who had run away from home and were homeless [47], and men who had sex with men (MSM) [37] to identify the characteristics of various subgroups of youth.

In total 10,974 participants took part in the selected studies, with those enrolled in individual studies ranging from as low as 24 to 6728. The average number of participants per study was 552. All the participants were 24 years or younger.

Health-Related Internet Use

Prevalence of General Internet Use and Patterns of Health-Related Internet Use

The studies generally agreed that youths spend a large amount of time using the internet. According to the studies, 82.8% of youth in the age range of 11 to 18 years spend 1 to 4 hours/day online (Multimedia Appendix 2) [45] Interestingly, boys in the age range of 10 to 11 years reported using the internet for only 30 min/day, whereas youth in the age range of 14 to 15 years of both sexes were online for several hours/day [34]. Although the time youth in the age range of 16 to 17 years spent online drops to less than 1 hour/day [34], this is most likely related to the higher burden of academic work they are expected to accomplish at this age. Teens over 15 years reported more frequent use of the internet for searching health information than those younger than 15 years [41,42].

Researchers have also suggested that there is a difference in the frequency of daily internet usage in youths with sexual orientation differences. MSM youths exhibited significantly more frequent daily internet use (77%) than non-MSM (60%) when using it as a medium to search for their unique health information needs and to facilitate the development of their sexual identity [37]. Some of the venues used to access the internet were homes, schools, a friend's home, an internet café, or the public library [48]; many runaway and homeless youth relied primarily on public access (ie, libraries and youth services agencies) [47].

A high frequency of internet use is also widely reported, and this finding is consistent across all the study countries. Sixty-four percent of youth (aged 10-16 years) in the United Kingdom are daily users, and a further 26% use the internet at least once or twice a week [35]. In the United States, 97% use the internet at least once a month, with 87% using it at least once a week [40]. For social networking, 87% maintain a personal social networking site (SNS) profile on MySpace or Facebook [41], and 96.6% use My Yearbook, Tagged, and Bebo [49]. Even in countries where access to the internet may be more limited, SNSs were popular with young people: in Nigeria, 73% reported that they had used the internet [46].

The findings of the various studies show that a high percentage of youth have used the internet for health-related purpose [33,37,39,41,48,50,51]. Among those (n=10) that reported ever using the internet for health-related purposes, the majority (80%, 8/10) found more than 50 percent have done so [33,39,41,47,48,50,51], whereas the remaining two reported fewer than 50% in the use of internet for this purpose [35,45]. There was no difference in the percentage of participants' lifetime health-related internet use depending on the publication year when analyzed using a chi-square test (P>.05).

In a 2012 study of US teenagers, 81% reported that they had checked online health information, and 71% were very likely to search the internet for information on health; 59% sought online health information for their family's health; and 56% had heard of Medline Plus [39]. In an earlier study of youth in the age range of 18 to 19 year, 65% reported the internet to be their primary source for health-related information [51]. However, this number was not consistent across populations and depended on specific conditions. A recent study in the United States found that 91.9% of youth with juvenile arthritis used the internet for more than 5 min/day, 69.4% used it for 30 min/day, and 36.6% for more than 1 hour/day [33]. Among youth undergoing orthodontic treatment in the United Kingdom, only 8% used the internet for specific disease-related information, and 3% had seen a phone app about orthodontics. Instead, their main source of information was their health care providers (HCPs), with only 8% using the internet as a primary source of information [35]. An Israeli study that compared Jewish and Arab middle and high school students' internet access and health information–seeking behavior online found that although the two groups were similarly likely to access the internet, Arab students were far more likely to use the internet as a source of health information [48].

For the studies published from 2007 to 2012, daily users of the internet in this age group varied from 54.4% [47] to 88.2% [51], both in the United States. In studies published from 2013 to 2017, this had risen to from 64% [35] to 82.8% [45], both in the United Kingdom.

Device and Mode Used to Access the Internet

Although there has been a significant increase in the ownership of mobile phones by adolescents in recent years, many studies

did not evaluate health information seeking via internet-enabled devices. Of those that did [35,40,41], the most common means for accessing the internet were personal computers or laptops (65%), followed by cell phones or other mobile internet-enabled devices (42%), with many reporting using both [40]. Stephens et al [35] asked specifically whether their study participants accessed mobile apps, and only 3% answered in the affirmative way, whereas in another study, one-third of the Native American youth reported that the use of their cell phone (36%) was a regular mode of internet access [41].

Purpose of Health-Related Internet Use

Eysenbach's framework indicates the major types of health-related internet use as consisting of information (content), support (community), communication, and electronic commerce (e-commerce) [32]. The findings of each study were therefore coded into three categories based on this framework; support and communication were combined into a single category for the purposes of this review. These are discussed in turn below.

Information

The primary purpose for health-related internet use is seeking information. The topics that young people search for online includes information on daily health-related issues [33-35,38,39,41,43,45-47,48], physical well-being [40,41,45,48,52], sexual health [33,37,42,45-47,48-51], mental health [33,41,44,52], social problems [33,34,36,37,44,52,50,51], and culturally and religiously sensitive topics [41,48]. Daily issues that play a significant role in young peoples' lives, such as sports injuries, flu, chronic diseases, asthma, sexual health, fitness, and infections, are common areas of interest for youth on the internet [45]. This is particularly true for those suffering from particular diseases [33]. The internet also serves as a confidential source for information that may be culturally or religiously sensitive [48]; the greater likelihood of Arab youths seeking online information about mental health issues compared with their Jewish peers reflects the relative lack of mental health professionals available for Arab youth, as well as they being more culturally constrained than Jewish adolescents with regard to exposing personal concerns and problems [48]. On the other hand, there is some evidence to suggest that youth may be more likely to use the internet for less sensitive topics such as nutrition and exercise and less likely to look for sensitive topics such as violence, sexual health, bullying, tobacco, alcohol, drugs, and mental health [33]. Young people who are experiencing symptoms such as emotional difficulties often seek help for their feelings [52] and information related to their psychosocial health from peers online [44]. However, it has been reported that adolescents do not tend to use the internet for pain management [45]. Among those with diseases such as arthritis or diabetes, young people seek information related to their symptoms (52.4%) and treatment options (47.4%) [52] and may also turn to alternative sources (HCPs or peers) depending on the topic.

Support (Community) and Communication

Youths often use the internet to connect and create supportive communities on particular health issues, expressing interest in diverse online activities related to health, including messaging and connecting with others, networking, and receiving information. Intriguingly, 61.2% preferred an online support group to offline in-person groups [33], and children who were receiving hospital treatment in Sweden for a chronic disease, in this case diabetes, expressed a strong interest in using the internet for support networking, as well as for interpersonal contacts with their nondiabetic peers [34]. Youth with sexual orientation differences found the internet helpful as a way to connect to the gay community and meet partners online, as well as enabling them to discuss safe sex practices and boundaries and exchange information on HIV status before meeting prospective partners offline [37]. Interestingly, email communications with HCPs were not reported as a major purpose of health-related internet use.

Electronic Commerce

None of the studies included in this review examined young people's health-related internet use for e-commerce.

Factors Associated With Health-Related Internet Use

Gender, age, and in-school status are associated factors for the frequency of health-related internet use [34,45]. Girls tend to use the internet more often for help seeking online [41,45]. Youth of both sexes aged 16 to 17 years reported the internet to be their primary source for information, whereas those aged 10 to 11 years regarded their parents as their main source for information [34]. Similarly, youth aged 12 to 14 years regarded parents, teachers, and other adults as their primary source of health information, including sexual health [42]. Perhaps it may not be surprising that girls in Nigeria who are in school are more capable of finding information online than those who are out of school [36]. Only one study considered a potential association with race and ethnicity, reporting that among MSM in the United States, whites used the internet more frequently compared with African American and Latin American youths [37].

Notably, youths' emotional characteristics and engagement in risky behaviors are associated with internet use [33]. Young people who have a lower psychosocial quality of life tended to have higher use of the internet for health-related matters [33], although there was no association with coping skills or pain frequency [45]. Additionally, youth who engage in high risk behaviors such as smoking, less physical activity, less sun protection activity, and depression were more willing to use technology for health promotion [38].

Electronic health (eHealth) literacy level was positively associated with seeking health information online [39], as were exposure to a health course, online information seeking, exposure to MedlinePlus, parents' need for an interpreter when communicating with HCPs, upper grade in school, financial status higher health-related self-efficacy, and ethnicity (non-Hispanic), all of which are associated with a higher level of eHealth literacy [39]. An exposure to a specific website such as Medline online is known to facilitate health-related internet use; those enrolled on campuses promoting careers in the health care field and exposure to a health course are more likely to have heard of Medline Plus, and 11th graders are more likely to use Medline Plus than 9th or 10th graders [39]. Youth whose parents need interpreters to communicate between a family

member and an HCP are also more likely to have heard of Medline Plus [39]. However, no association was found between access to technology and willingness to engage in eHealth literacy [38].

Perceptions of Health-Related Internet Use

Overall, children and adolescents' perception of health-related internet use is positive. The information presented in Table 1 for the quantitative studies includes whether the findings of each study indicate positive or negative perceptions of health-related internet use. The key evidence supporting this finding is also summarized. This perception is based on participants' overall perceptions, the likelihood they will search online for health-related information, and participants' trust, preference, and interest in using the internet as their primary source for health-related purposes. Among the studies that reported the participants' perceptions on health-related internet use quantitatively (n=12), 50% (6/12) indicated that young people have generally positive perceptions about health-related internet use, with only 33.3% (4/12) reporting that children and adolescents have overall negative perceptions and 17% (2/12) reporting neutral perceptions. When we analyzed whether the perception depended on the publication year using a chi-square test, there was no statistical difference.

Table 1. Conceptual definitions and theoretical backgrounds.

Key concept related to health-related internet use	Definition and sources	Theoretical background	Authors and studies
eHealth[a] literacy	"Ability to seek, find, understand, and appraise health information from electronic resources and apply such knowledge gained to addressing or solving health problem" [53]	—[b]	Manganello et al, 2016 [43]
Health information–seeking behavior	"Purposive search for health-related information to satisfy a query" [54]	—	Stephens et al, 2013 [35]
eHealth promotion	"Web-based health education and behavior change applications" [28]	Theory of planned behavior; problem behavior theory	Tercyak et al, 2009 [38]
eHealth intervention	"Integration of information and communication technology."	—	Johnson et al, 2015 [33]
Electronic mental health	"Use of information and communication technologies to improve mental health." [55]	—	Wetterlin et al, 2014 [52]
Help seeking (help seeking online)	"Seeking assistance from mental health services, other formal services, or informal support sources for the purpose of resolving emotional or behavioral problem" [56]	Andersen behavioral model and Pescosolido's network episode model	Barman-Adhikari et al, 2011 [47]
e-patient	"Those who bring information obtained online to the medical consultation" [57]	—	Neumark et al, 2013 [48]
None	—	—	Buhi et al, 2009 [51]; Fergie et al, 2013 [44]; Gaskin et al, 2012 [40]; Ghaddar et al, 2012 [39]; Henderson et al, 2013 [45]; Magee et al, 2012 [50]; Mustanski et al, 2011 [37]; Nordfeldt et al, 2013 [34]; Nwagwu, 2007 [36]; Rushing et al, 2011 [41]; Selkie et al, 2011 [49]

[a]eHealth: electronic health.

[b]Not provided.

Perceived Benefits

Regarding the perceived importance and usefulness of the internet, 90% of the participants in one study responded that having access to health-related resources on the Web is important [40], but only 8% of those in another study stated that their preferred source of information was the internet [35]. When adolescents are asked specifically about their sexual health–related use, 48.1% reported that they are relieved or comforted by the information online [52]. This positive perception is consistent with those found in a study on youth who have been detained in a juvenile detention facility, where 90% believed that access to information on various websites was useful [40]. However, young people have also reported that they would prefer sexual health Web-based sources to contain more comprehensive [50] or broader spectrum of topics [41] rather than just sexual health information.

User-generated content is perceived as advantageous for online health content as it provides diverse views and experiential knowledge combined with anonymity [44]. Many youth with sexual orientation differences reported that the internet facilitated the development of their sexual identity by connecting them with the gay community (both online and in real life), as well as by helping them search for specific facts about HIV or sexually transmitted infections (STIs), attempt self-diagnoses

of symptoms they might be experiencing, find health centers that offer HIV or STI testing and affordable care, and learn about risk reduction techniques [37].

Perceived Barriers

In a Canadian study, 82.9% of the participants reported that they would be likely to use an information-based website at a difficult time in their life, but only 77% would be likely to use social media websites for information or to seek help [52]. The most commonly reported reason (62%-80%) for not seeking online health information was a preference for receiving information from a health professional, suggesting the use of the internet as a supplementary means rather than a replacement [48]. Only 10.9% accessed the health-related websites recommended by experts, and 10.6% sought help from social media for problems such as anxiety or depression [52].

Online privacy was a key issue for youth [34], with 87.7% stressing the importance of online privacy, which was particularly important for those with a specific health problem such as mental health issues [48]. Looking for sexual health information online was also closely linked to privacy issues as many youth felt reluctant to speak with an HCP about sensitive issues surrounding sexuality and instead use the internet to avoid embarrassment and overcome privacy issues [47]. On the other hand, lesbian, gay, bisexual, and transgender (LGBT) youth identified fear as an obstacle to online sexual health behaviors because of the perceived stigma resulting from being "caught" [50]. Although there are different perceptions in the various subgroups, 85% of the youths detained in a juvenile detention facility claimed not to be concerned about the privacy of their health information on the internet when on password-protected sites [40].

Another strong concern among youth who use the internet was the accuracy of the information [44]. When youth were asked specifically about their sexual health–related internet use, 44.4% reported that they were confused by the information they found, 25.9% were frustrated by the lack of information or an inability to find the information needed, whereas 18.5% were overwhelmed by the sheer amount of information available on the internet [51]. Some of the online experiences reported by adolescent males were not positive, with several recounting being distressed by finding information on the internet that either negatively portrayed homosexuality or described the victimization of LGBT people [37]. Those with low health literacy (28%) were more likely to rate the health information found on the internet as usually or often accurate compared with those with high health literacy (14%) [43]. Remarkably, study participants considered finding local information to be more difficult than finding general information online [49,52].

Important Features for Usability and Current Practice

Adolescents noted that they used different strategies to evaluate factual information and user-generated opinions on social media websites [44]. They highlighted the importance of the initial impression of a website and whether it made a serious and trustworthy impression on them; as they value integrity and anonymity, they were cautious about sharing their personal information [34]. Young people also stressed the importance of updating websites regularly to add value by including information such as current and recent events, facts and statistics (eg, verifiable information), as well as improving the technical aspects of websites by incorporating eye-catching design, high-quality visuals, and multimedia rather than text, although 51.9% said they never or hardly ever checked when a site was last updated or reviewed by a medical professional [51]. Furthermore, plainness (ie, clear content and layout) was another important feature that youth preferred [34]. Culturally, sexually, and religiously relevant health information targeted to specific populations, such as particular ethnic groups or sexual orientations, was preferred by minority youths and youths with sexual orientation differences [37,41,48,50]; they also preferred open access sites that did not require log-ins [34]. Regarding content, study participants wanted more information related to medications (92%), immunizations (90%), and STI test results (80%) [40].

These findings were consistent across studies examining a specific topical health (eg, sexual health) [49,51]. Regarding internet use related to sexual health, adolescents wanted sexual health education sites to be easily accessible, understandable, and user-friendly and the resources provided to be trustworthy-credible, confidential, and offered in a nonthreatening way [49]. Young people also wanted more information on specific topics and in-person resources such as local clinic resources, as these were reported as the most challenging for them to find [49,51].

When youth search for sexual health–related information, they used Google, Yahoo, and Ask most often as the first search engines [49,51], then followed sponsored links and the first three search results; another common strategy was to check for converging information across multiple websites [37]. Wikipedia and "WebMD" were the source they considered as providing the most credible sexual health information [37,49].

Conceptualization

The key concepts for health-related internet use in the studies were eHealth literacy, health information–seeking behavior, eHealth promotion, eHealth interventions, as well as electronic mental health, health seeking, and electronic patient websites (Table 1). These concepts were all based on online activities related to information seeking and understanding or communication activities for health issues, problems, and health promotion. eHealth promotion and eHealth intervention provided more nuanced definitions related to Web-based interventions and education.

Conceptual definitions were provided in only a few studies, and of these, only a few utilized a theoretical framework. In Tercyak and colleagues' study [38], the frameworks used were the theory of planned behavior and problem behavior theory, which explain the basis of the common mechanisms of multiple behavioral problems and provided frameworks that focused on individuals' motivation for eHealth promotion associated with their behavior changes. When the media influence was studied, the uses and gratifications theory [46] was applied. This theory assumed that users choose a particular medium as an avenue to actively participate while being goal-directed, rather than as mere passive recipients. This theory also considered that the medium gratifies

psychosocial needs. Another study used grounded theory [49] for its theory development.

Methodological Evaluation

Study Design

A summary of the methodological evaluation conducted for this review is shown in Multimedia Appendix 3. All the studies in the table are descriptive, with the majority being cross-sectional studies; 26% (5/19) are correlational studies. In the studies included in this review, 58% (11/19) used a quantitative study design, whereas 16 % (3/19) used a qualitative study design, and 26% (5/19) used mixed or multiple methods. For the quantitative studies, the reported rates of use and access to the internet among the study participants, as well as any associated factors related to their internet use, are identified. Generally, the qualitative and mixed-methods studies explored how youths perceived the benefits and barriers of health-related internet use.

Study Sample

Less than half of the studies 47% (9/19) used convenience sampling [33,38,40,41,43,45,47,50,52]; the remaining studies used purposive sampling strategies [34,35,44,49] and random sampling across multiple sites [39,42,46,48], with 2 studies using both convenience and purposive sampling [37,51]. Of the 11 quantitative studies, only 4 used random sampling techniques [36,39,42,48]. Over half of the studies used multiple sites for sampling (58%) or used multiple resources, for example, by recruiting from both online and offline communities. No studies specifically indicated a sample size justification.

Data Collection and Analysis

Online surveys (26%) were the most common data collection technique [33,39,45,51,52]. Most of the qualitative studies used focus groups, although a few conducted semistructured interviews. Most studies used investigator-developed questionnaires to assess health-related internet use. This poses a number of potential issues related to the validity and reliability of their questionnaires compared with existing instruments. The most common analytic technique used was descriptive, which includes descriptive statistics, univariate analyses (*t* test and chi-square test), and multivariate analyses (linear regression, logistic regression, and analysis of variance). None of the quantitative studies indicated the statistical assumptions applied, and few explained how missing data were treated. For the survey studies, the data are self-reported, which inevitably introduces bias. The analytic approaches used were generally appropriate for the level of data and measurement. For the qualitative studies, thematic analysis, content analysis, and inductive descriptive analysis were commonly used.

Discussion

Summary of Findings and Comparison With Previous Work

This review of the most recent research in this area has deepened our understanding of how young people seek information from the internet and its related support systems for their health care. Adolescents spend a great deal of time on the internet, with the majority spending more than 2 hours every day. Although there are some inconsistencies regarding the amount of time and frequency of health-related use, depending on the population and disease concerned, most young people have used the internet for health-related purposes, and it represents their most frequent source of information.

Overall, youth are positive about using the internet to search for health-related information. As their most frequently used information source, the internet is commonly used for health-related information by both healthy and nonhealthy youth. Among healthy adolescents, this information includes sensitive topics such as sexual health and violence, as well as less sensitive topics such as exercise and nutrition. For those who have been diagnosed with a medical condition, the topics searched also include finding treatment options, seeking support, and networking with fellow sufferers, which is consistent with other populations [58]. Although we found a great deal of evidence to suggest that those with specific diseases use the internet to find friends [34], this may actually be related to the unique characteristics of youth who are comfortable meeting people online. Moreover, young people tend to prefer using support groups rather than attending in-person meetings and are not particularly bound to people with similar diagnoses. These characteristics are likely to be at least partly because of the perceived benefits of internet use, as many adolescents consider the internet to be a safe space where they can share sensitive information. Young people are interested in finding information from reliable sources such as HCPs or experts, as well as user-generated information from their peers who may have experienced the same issue. Members of this generation believe that it is helpful to learn diverse views on health topics [34].

Despite their high level of health-related internet use, several perceived challenges have been reported. To ensure useful Web-based health interventions or sites available for youth, credible resources and privacy are vital for successful outcomes. Young people generally evaluate a site's credibility based upon its appearance, frequent citation, and the website's domain name such as .com, .gov, or .org, but often there is no easy way to tell [49,51]. For example, privacy and confidentiality on an SNS may indicate a lack of online help or support services in mental health [52]. Additionally, researchers have found that some adolescents have experienced difficulties when searching for specific information such as local resources, despite their competency in finding general information. User-friendly features such as sites that do not require visitors to log in are suggested as another important element that enhances usability. It is also important for sites to have good readability and be well organized. Finding the most recently updated sites or checking a website's creators are less common practices among teenagers and represent an area where education may be helpful. There is a general perception that there is a lack of useful, reliable resources for the specific information they need, such as particular disease-related information or health care topics for adolescents. Sites that can provide reliable information for youths need to be developed.

There are important findings related to the characteristics of various subgroups for health-related internet use. Youths whose

parents or older relatives are not eHealth literate, have no internet access, have low health literacy, and need interpreters have a particularly high usage of the internet and are very likely to seek health information online for their family. Interestingly, young people who are in juvenile detention facilities worry less about privacy issues and are more willing to share information on the internet, whereas the opposite is true for MSM youths, who fear stigmatization if someone finds out their search history. There are different patterns of health-related internet use depending on age, with older youth becoming more frequent users of the internet to seek information on their health. Young people who have previously taken courses or received education on internet use designed to enhance their eHealth literacy level, for example, become more competent in their health-related internet use, especially when evaluating websites, suggests the need for more extensive health literacy education [53]. No gender differences were reported for health-related internet use, except for one study that indicated girls tend to be more frequent internet users than boys for issues related to pain management. In-school education also supported youth competency for health-related internet use. Youths who have a high risk of engaging in risky behaviors tend to use the internet more often than youths with lower risk for health-related internet use, which indicates a serious need for high quality content designed specifically for preventing behavioral issues to be developed. However, the most significant gap in the research in this area is that there were no studies of children younger than 10 years. This exclusion is source for further research.

Limitations of This Review

Although this study followed evidence-based guidelines and adopted a systematic approach, chances of human error in coding are inevitable. We used a wide range of different search terms to identify relevant papers, including "health-related internet use," "eHealth," "internet use for health-related purpose," "Web-based resource," "health information seeking," and "online resource," combined with "child," "adolescent," "student," "youth," and "teen" in the databases searched; however, our choice of keywords may have resulted in missing relevant research studies eligible for inclusion. Although we used search engines most commonly used in the field of health, namely PubMed, CINAHL, and PsycINFO, this data-based selection many have created potential errors or missed relevant studies that should have been included. Furthermore, there is some potential for subjectivity in analyzing the findings, although 2 different coders carefully reviewed and coded each study independently and then discussed the results while double-checking each process. When the authors coded the methodological approaches used in each study, we tried not to assume a specific approach unless it was specifically stated in the study. For example, where no specific approach used for sampling is stated in the study, we coded these as using convenience sampling. This may have led to some potential errors regarding what the various authors actually did in their studies. Furthermore, the measures used in each study varied, and the study samples were heterogeneous, so we were unable to directly compare the outcomes for health-related internet use across all the studies examined for this review. Thus, we were not able to compare the findings based on regional differences among the samples, for example.

Implications

Although this is an emerging field of study, there have been no previous studies systematically reviewing existing research exploring the health-related internet use of teenagers and young adults. As increasing number of internet-based interventions are being developed and applied specifically to address the needs of young people, it is important to understand the characteristics of health-related internet use among youth. Although the internet is both easily accessible and widely accepted by adolescents, the so-called "digital natives," we have only a limited understanding of the patterns and characteristics of youth health-related internet use. This study therefore provides an important overview of the research findings to date related to patterns of youth health-related internet use. Although young people are generally frequent users of the internet for their health care and are positive about the practice, there remains a great need for education to support their competent and appropriate use of the internet. Additionally, there is a need for more reliable Web-based sources to be developed for this population. This study's findings include a consideration of the associated factors for health-related internet use that have an effect on adolescents' general health behaviors. A major gap identified in the review was the lack of a conceptual definition of the term "health-related internet use." Furthermore, the majority of the studies published to date have not been based on a specific theoretical framework. In addition, this review identifies several limitations of the identified studies regarding methodological issues and provides suggestions for the further rigorous research required to design efficient and effective interventions for this hard-to-reach population. HCPs and policy makers should consider how best to integrate these needs into their current practices and policies.

Recommendations for Future Research

Future research in this area needs to address several major gaps in the research, strengthen research methods, and contribute to appropriate theory development, as well as refining and conceptualizing eHealth practice and health-related internet use. The characteristics of various subpopulations need to be identified and compared with the characteristics of young people in general in this respect. In particular, internet use by younger adolescents and children who are younger than 10 years has not yet been studied. A closer examination of this younger demographic will give us a more accurate understanding of when children are first exposed to the internet and at what point its influence becomes seriously important. In this way, we will be able to identify appropriate "teachable moments" and the critical age at which to teach young people the skills they will need to become eHealth literate. Past studies have tended to focus primarily on cross-sectional studies, and it would be worthwhile to explore the longitudinal outcomes of health-related internet use. In future research in this area, studies with high-level analyses and rigorous research methods need to be conducted. For example, this review identified several studies that revealed important associated factors, and although most of the existing studies used convenience sampling, it is

important for future research to utilize randomized sampling to yield more generalizable results that are applicable to wider populations. Multivariate analyses of the factors identified in the studies reviewed here will also yield valuable information, and standard measures for health-related internet use need to be developed that are based on a clear conceptual understanding and theoretical foundation. Furthermore, nearly all the selected studies suffered from limitations when representing the diverse populations of adolescents, including their gender, race and ethnicity, socioeconomic status, and regional status, although minority populations made up over half of the study participants overall.

Implications for Health Promotion Practice

As youth are using the Web more frequently than ever before and will continue to, it is important to develop a better understanding of how they actually use the internet for health-related support and information. On the basis of in-depth understanding of youth practice, it is vital to provide health education that provides eHealth literacy skills for this population. Studies showed that youth who learned about Medline Plus are more likely higher users of the internet and more confident of using internet-based sources.

First, it is important to evaluate various online health information-seeking skills currently being taught to adolescent in schools and examine how best to help them develop the skills they will need to obtain, comprehend, and process health information, as well as online health care system information [43]. In health education for adolescents, it is necessary to include the internet as a basic component, given that so many already use the internet for their health-related needs or will do so in the near future. Studies indicate that those with low health literacy were more likely to rate the internet as usually or often accurate than those with high health literacy (28% low vs 14% high). As those with low health literacy were also more likely to use the internet daily, it is particularly important to support youth health literacy levels.

Health disparities exist, and the internet may even contribute to these, so it is important to allocate resources to the population most in need of this type of assistance, taking into account the differences identified between groups with different ethnicities reported in the research reviewed here. Internet access is one of the major factors for health-related internet use and eHealth literacy (urban vs rural). It is important to develop health education programs that focus on boosting eHealth literacy [43].

There is a great deal of room for improvement in the existing Web-based programs for teenagers and young adults. Many of the participants in the studies reviewed indicated a desire for more Web-based resources for health that are not subject to the limitations of existing websites. For example, a greater emphasis should be placed on developing an awareness of cultural values related to culturally and religiously sensitive health-related topics that may be more relevant to certain genders and youth populations, including taking into account the need to protect their privacy from parental monitoring by masking the nature of their health information-seeking [41,48], which would greatly enhance usability. Furthermore, as sensitive topics such as sexual information or mental health issues are often information that young people seek on the internet, it is important to provide reputable sources that will be accepted by the target population; more diverse content that is specifically tailored to the needs and characteristics of young people also needs to be developed. For example, people in this age group are particularly vulnerable for risky behaviors, and although they are interested in knowing more about prevention, there is a lack of good resources available to them.

Acknowledgments

This research received no financial support from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of included studies.

[PDF File (Adobe PDF File), 433KB - jmir_v20i4e120_app1.pdf]

Multimedia Appendix 2

Summary of characteristics of health-related internet use.

[PDF File (Adobe PDF File), 342KB - jmir_v20i4e120_app2.pdf]

Multimedia Appendix 3

Methodological evaluation of study quality.

[PDF File (Adobe PDF File), 293KB - jmir_v20i4e120_app3.pdf]

References

1. Prensky M. Digital natives, digital immigrants part 1. On the Horizon 2001;9(5):1-6. [doi: 10.1108/10748120110424816]
2. Selwyn N. "The digital native – myth and reality". Aslib Proceedings 2009 Jul 05;61(4):364-379. [doi: 10.1108/00012530910973776]
3. Subrahmanyam K, Smahel D. Adolescents' digital worlds. In: Digital Youth: the Role of Media in Development. New York: Springer; 2012:1-23.
4. Vanden Abeele MM. Mobile youth culture: a conceptual development. Mob Media Commun 2016 Jan;4(1):85-101. [doi: 10.1177/2050157915601455]
5. Dunahee M, Lebo H. Digitalcenter. 2016. The World Internet Project International Report (Sixth Ed) URL: http://www.digitalcenter.org/wp-content/uploads/2013/06/2015-World-Internet-Report.pdf [accessed 2017-03-20] [WebCite Cache ID 6p6s7z8FY]
6. Brown JD, Bobkowski PS. Older and newer media: patterns of use and effects on adolescents' health and well-being. J Res Adolesc 2011;21(1):95-113. [doi: 10.1111/j.1532-7795.2010.00717.x]
7. Lenhart A. Pewinternet. 2015. Teens, social media & technology overview URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [WebCite Cache ID 6p6sdaSQj]
8. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949 [FREE Full text] [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
9. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the Internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
10. Lehnhart A, Purcell K, Smith A, Zickurh K. Pewinternet. 2010. Social media & mobile Internet use among teens and young adults URL: http://www.pewinternet.org/files/old-media/Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplines.pdf [accessed 2017-03-20] [WebCite Cache ID 6p6tytPI6]
11. Park S. The association between Internet use and depressive symptoms among South Korean adolescents. J Spec Pediatr Nurs 2009 Oct;14(4):230-238. [doi: 10.1111/j.1744-6155.2009.00191.x] [Medline: 19796322]
12. Yu L, Shek DT. Internet addiction in Hong Kong adolescents: a three-year longitudinal study. J Pediatr Adolesc Gynecol 2013 Jun;26(3 Suppl):S10-S17. [doi: 10.1016/j.jpag.2013.03.010] [Medline: 23683821]
13. Yu JJ, Park SJ. Predictors and the distal outcome of general Internet use: the identification of children's developmental trajectories. Br J Dev Psychol 2017 Nov;35(4):483-497. [doi: 10.1111/bjdp.12188] [Medline: 28513971]
14. Ohannessian C. A longitudinal examination of the relationship between technology use and substance use during adolescence. In: Blair SL, Claster PN, Claster SM, editors. Technology and Youth: Growing Up in a Digital World (Sociological Studies of Children and Youth, volume 19). Bingley, UK: Emerald Group Publishing Limited; 2015:293-313.
15. Subrahmanyam K, Greenfield P, Kraut R, Gross E. The impact of computer use on children's and adolescents' development. J Appl Dev Psychol 2001 Jan;22(1):7-30. [doi: 10.1016/S0193-3973(00)00063-0]
16. Gross EF. Adolescent Internet use: what we expect, what teens report. J Appl Dev Psychol 2004 Nov;25(6):633-649. [doi: 10.1016/j.appdev.2004.09.005]
17. Lin CH, Yu SF. Adolescent Internet usage in Taiwan: exploring gender differences. Adolescence 2008;43(170):317-331. [Medline: 18689104]
18. Chen L, Nath R. Understanding the underlying factors of Internet addiction across cultures: a comparison study. Electron Commer Res Appl 2016 May;17:38-48. [doi: 10.1016/j.elerap.2016.02.003]
19. Finney Rutten LJ, Hesse BW, Moser RP, Ortiz Martinez AP, Kornfeld J, Vanderpool RC, et al. Socioeconomic and geographic disparities in health information seeking and Internet use in Puerto Rico. J Med Internet Res 2012;14(4):e104 [FREE Full text] [doi: 10.2196/jmir.2007] [Medline: 22849971]
20. Bell AV. "I think about Oprah": social class differences in sources of health information. Qual Health Res 2014 Apr;24(4):506-516. [doi: 10.1177/1049732314524637] [Medline: 24623661]
21. Reid Chassiakos YL, Radesky J, Christakis D, Moreno MA, Cross C, Council on Communications and Media. Children and adolescents and digital media. Pediatrics 2016 Nov;138(5):pii: e20162593 [FREE Full text] [doi: 10.1542/peds.2016-2593] [Medline: 27940795]
22. Council on Communications and Media. Media use in school-aged children and adolescents. Pediatrics 2016 Nov;138(5):pii: e20162592. [doi: 10.1542/peds.2016-2592] [Medline: 27940794]
23. Dahl RE. Adolescent brain development: a period of vulnerabilities and opportunities. Keynote address. Ann N Y Acad Sci 2004 Jun;1021:1-22. [doi: 10.1196/annals.1308.001] [Medline: 15251869]
24. Petersen AC. Adolescent development. Annu Rev Psychol 1988;39:583-607. [doi: 10.1146/annurev.ps.39.020188.003055] [Medline: 3278681]
25. Fadda D, Scalas LF, Meleddu M. Contribution of personal and environmental factors on positive psychological functioning in adolescents. J Adolesc 2015 Aug;43:119-131. [doi: 10.1016/j.adolescence.2015.05.019] [Medline: 26093819]
26. Kelly SA, Melnyk BM, Jacobson DL, O'Haver JA. Correlates among healthy lifestyle cognitive beliefs, healthy lifestyle choices, social support, and healthy behaviors in adolescents: implications for behavioral change strategies and future research. J Pediatr Health Care 2011;25(4):216-223. [doi: 10.1016/j.pedhc.2010.03.002] [Medline: 21700136]

27. Ritterband LM, Palermo TM. Introduction to the special issue: eHealth in pediatric psychology. J Pediatr Psychol 2009 Jun;34(5):453-456. [doi: 10.1093/jpepsy/jsp008] [Medline: 19264878]
28. Skinner HA, Maley O, Norman CD. Developing internet-based eHealth promotion programs: the Spiral Technology Action Research (STAR) model. Health Promot Pract 2006 Oct;7(4):406-417 [FREE Full text] [doi: 10.1177/1524839905278889] [Medline: 16840770]
29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
30. Hall GS. Adolescence: Its Psychology and Its Relations to Physiology, Anthropology, Sociology, Sex, Crime, Religion and Education. New York, NY: Appleton and Company; 1904.
31. West S, King V, Carey T, Lohr K, McKoy N, Sutton S. Systems to Rate the Strength of Scientific Evidence: Summary, Evidence Report/Technology Assessment: Number 47. AHRQ Publication No. 02-E015. Rockville, MD: Agency for Healthcare Research and Quality (US); 2002.
32. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371. [Medline: 15224975]
33. Johnson KR, Fuchs E, Horvath KJ, Scal P. Distressed and looking for help: Internet intervention support for arthritis self-management. J Adolesc Health 2015 Jun;56(6):666-671. [doi: 10.1016/j.jadohealth.2015.02.019] [Medline: 26003583]
34. Nordfeldt S, Angarne-Lindberg T, Nordwall M, Ekberg J, Berterö C. As facts and chats go online, what is important for adolescents with type 1 diabetes? PLoS One 2013;8(6):e67659 [FREE Full text] [doi: 10.1371/journal.pone.0067659] [Medline: 23805322]
35. Stephens R, Ryan FS, Cunningham SJ. Information-seeking behavior of adolescent orthodontic patients. Am J Orthod Dentofacial Orthop 2013 Mar;143(3):303-309. [doi: 10.1016/j.ajodo.2012.10.018] [Medline: 23452963]
36. Nwagwu WE. The Internet as a source of reproductive health information among adolescent girls in an urban city in Nigeria. BMC Public Health 2007;7:354 [FREE Full text] [doi: 10.1186/1471-2458-7-354] [Medline: 18096032]
37. Mustanski B, Lyons T, Garcia SC. Internet use and sexual health of young men who have sex with men: a mixed-methods study. Arch Sex Behav 2011 Apr;40(2):289-300 [FREE Full text] [doi: 10.1007/s10508-009-9596-1] [Medline: 20182787]
38. Tercyak KP, Abraham AA, Graham AL, Wilson LD, Walker LR. Association of multiple behavioral risk factors with adolescents' willingness to engage in eHealth promotion. J Pediatr Psychol 2009 Jun;34(5):457-469 [FREE Full text] [doi: 10.1093/jpepsy/jsn085] [Medline: 18723566]
39. Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health 2012 Jan;82(1):28-36. [doi: 10.1111/j.1746-1561.2011.00664.x] [Medline: 22142172]
40. Gaskin GL, Longhurst CA, Anoshiravani A. Internet access and attitudes toward online personal health information among detained youth. Pediatrics 2012 Nov;130(5):914-917 [FREE Full text] [doi: 10.1542/peds.2012-1653] [Medline: 23090346]
41. Rushing SC, Stephens D. Use of media technologies by Native American teens and young adults in the Pacific Northwest: exploring their utility for designing culturally appropriate technology-based health interventions. J Prim Prev 2011 Aug;32(3-4):135-145. [doi: 10.1007/s10935-011-0242-z] [Medline: 21805055]
42. Ybarra ML, Emenyonu N, Nansera D, Kiwanuka J, Bangsberg DR. Health information seeking among Mbararan adolescents: results from the Uganda Media and You survey. Health Educ Res 2008 Apr;23(2):249-258. [doi: 10.1093/her/cym026] [Medline: 17639121]
43. Manganello JA, Sojka CJ. An exploratory study of health literacy and African American adolescents. Compr Child Adolesc Nurs 2016 Jul 20;39(3):221-239. [doi: 10.1080/24694193.2016.1196264]
44. Fergie G, Hunt K, Hilton S. What young people want from health-related online resources: a focus group study. J Youth Stud 2013 Aug;16(5):579-596 [FREE Full text] [doi: 10.1080/13676261.2012.744811] [Medline: 24748849]
45. Henderson EM, Keogh E, Rosser BA, Eccleston C. Searching the internet for help with pain: adolescent search, coping, and medication behaviour. Br J Health Psychol 2013 Feb;18(1):218-232. [doi: 10.1111/bjhp.12005] [Medline: 23126577]
46. Nwagwu WE. The Internet as a source of reproductive health information among adolescent girls in an urban city in Nigeria. BMC Public Health 2007;7:354 [FREE Full text] [doi: 10.1186/1471-2458-7-354] [Medline: 18096032]
47. Barman-Adhikari A, Rice E. Sexual health information seeking online among runaway and homeless youth. J Soc Social Work Res 2011 Jun;2(2):88-103 [FREE Full text] [doi: 10.5243/jsswr.2011.5] [Medline: 22247795]
48. Wetterlin FM, Mar MY, Neilson EK, Werker GR, Krausz M. eMental health experiences and expectations: a survey of youths' Web-based resource preferences in Canada. J Med Internet Res 2014 Dec 17;16(12):e293 [FREE Full text] [doi: 10.2196/jmir.3526] [Medline: 25519847]
49. Neumark Y, Lopez-Quintero C, Feldman BS, Hirsch Allen AJ, Shtarkshall R. Online health information seeking among Jewish and Arab adolescents in Israel: results from a national school survey. J Health Commun 2013;18(9):1097-1115. [doi: 10.1080/10810730.2013.778360] [Medline: 23829662]
50. Magee JC, Bigelow L, Dehaan S, Mustanski BS. Sexual health information seeking online: a mixed-methods study among lesbian, gay, bisexual, and transgender young people. Health Educ Behav 2012 Jun;39(3):276-289. [doi: 10.1177/1090198111401384] [Medline: 21490310]
51. Selkie EM, Benson M, Moreno M. Adolescents' views regarding uses of social networking websites and text messaging for adolescent sexual health education. Am J Health Educ 2011 Jul;42(4):205-212 [FREE Full text] [Medline: 22229150]

52. Buhi ER, Daley EM, Fuhrmann HJ, Smith SA. An observational study of how young people search for online sexual health information. J Am Coll Health 2009;58(2):101-111. [doi: 10.1080/07448480903221236] [Medline: 19892646]
53. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
54. Lenz ER. Information seeking: a component of client decisions and health behavior. ANS Adv Nurs Sci 1984 Apr;6(3):59-72. [Medline: 6426379]
55. Statistics Canada 2012. Statcan.gc. 2013. Individual Internet use and e-commerce URL: http://www.statcan.gc.ca/daily-quotidien/131028/dq131028a-eng.pdf [accessed 2017-03-21] [WebCite Cache ID 6p7Uh0mBd]
56. Srebnik D, Cauce AM, Baydar N. Help-seeking pathways for children and adolescents. J Emot Behav Disord 2016 Nov 29;4(4):210-220. [doi: 10.1177/106342669600400402]
57. Giveon S, Yaphe J, Hekselman I, Mahamid S, Hermoni D. The e-patient: a survey of israeli primary care physicians' responses to patients' use of online information during the consultation. Isr Med Assoc J 2009 Sep;11(9):537-541 [FREE Full text] [Medline: 19960847]
58. Park E, Kim H, Steinhoff A. Health-related Internet use by informal caregivers of children and adolescents: an integrative literature review. J Med Internet Res 2016 Mar 03;18(3):e57 [FREE Full text] [doi: 10.2196/jmir.4124] [Medline: 26940750]

Abbreviations

CINAHL: Cumulative Index of Nursing and Allied Health Literature
e-commerce: electronic commerce
eHealth: electronic health
HCP: health care provider
LGBT: lesbian, gay, bisexual, and transgender
MSM: men who have sex with men
SNS: social networking sites
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 21.03.17; peer-reviewed by E Buhi, S Ghaddar; comments to author 22.07.17; revised version received 06.12.17; accepted 03.01.18; published 03.04.18.

Please cite as:
Park E, Kwon M
Health-Related Internet Use by Children and Adolescents: Systematic Review
J Med Internet Res 2018;20(4):e120
URL: http://www.jmir.org/2018/4/e120/
doi:10.2196/jmir.7731
PMID:29615385

Review

The Efficacy of Electronic Health–Supported Home Exercise Interventions for Patients With Osteoarthritis of the Knee: Systematic Review

Axel Georg Meender Schäfer[1], PhD; Christoff Zalpour[2], MD; Harry von Piekartz[2], PhD; Toby Maxwell Hall[3], PhD; Volker Paelke[4], PhD

[1]Course of Study Speech and Language Therapy, Occupational Therapy and Physiotherapy, Faculty of Social Work and Health, University of Applied Sciences and Arts Hildesheim, Hildesheim, Germany

[2]Institut für angewandte Physiotherapie und Osteopathie, Fakultät Wirtschafts- und Sozialwissenschaften, University of Applied Sciences Osnabrück, Osnabrück, Germany

[3]School of Physiotherapy and Exercise Science, Faculty of Health Sciences, Curtin University, Perth, Australia

[4]International Degree Programme in Media Computer Science, Electrical Engineering and Computer Science, University of Applied Sciences Bremen, Bremen, Germany

Corresponding Author:
Axel Georg Meender Schäfer, PhD
Course of Study Speech and Language Therapy, Occupational Therapy and Physiotherapy
Faculty of Social Work and Health
University of Applied Sciences and Arts Hildesheim
Gebaeude Goschentor, 2nd Floor
Goschentor 1
Hildesheim, 31141
Germany
Phone: 49 5121 881 ext 510
Email: axel.schaefer@hawk.de

Abstract

Background: Osteoarthritis of the knee is the most common cause for disability and limited mobility in the elderly, with considerable individual suffering and high direct and indirect disease-related costs. Nonsurgical interventions such as exercise, enhanced physical activity, and self-management have shown beneficial effects for pain reduction, physical function, and quality of life (QoL), but access to these treatments may be limited. Therefore, home therapy is strongly recommended. However, adherence to these programs is low. Patients report lack of motivation, feedback, and personal interaction as the main barriers to home therapy adherence. To overcome these barriers, electronic health (eHealth) is seen as a promising opportunity. Although beneficial effects have been shown in the literature for other chronic diseases such as chronic pain, cardiovascular disease, and diabetes, a systematic literature review on the efficacy of eHealth interventions for patients with osteoarthritis of knee is missing so far.

Objective: The aim of this study was to compare the efficacy of eHealth-supported home exercise interventions with no or other interventions regarding pain, physical function, and health-related QoL in patients with osteoarthritis of the knee.

Methods: MEDLINE, CENTRAL, CINAHL, and PEDro were systematically searched using the keywords osteoarthritis knee, eHealth, and exercise. An inverse variance random-effects meta-analysis was carried out pooling standardized mean differences (SMDs) of individual studies. The Cochrane tool was used to assess risk of bias in individual studies, and the quality of evidence across studies was evaluated following the Grading of Recommendations, Assessment, Development, and Evaluation approach.

Results: The literature search yielded a total of 648 results. After screening of titles, abstracts, and full-texts, seven randomized controlled trials were included. Pooling the data of individual studies demonstrated beneficial short-term (pain SMD=−0.31, 95% CI −0.58 to −0.04, low quality; QoL SMD=0.24, 95% CI 0.05-0.43, moderate quality) and long-term effects (pain −0.30, 95% CI −0.07 to −0.53, moderate quality; physical function 0.41, 95% CI 0.17-0.64, high quality; and QoL SMD=0.27, 95% CI 0.06-0.47, high quality).

Conclusions: eHealth-supported exercise interventions resulted in less pain, improved physical function, and health-related QoL compared with no or other interventions; however, these improvements were small (SMD<0.5) and may not make a meaningful

difference for individual patients. Low adherence is seen as one limiting factor of eHealth interventions. Future research should focus on participatory development of eHealth technology integrating evidence-based principles of exercise science and ways of increasing patient motivation and adherence.

(J Med Internet Res 2018;20(4):e152) doi:10.2196/jmir.9465

KEYWORDS

osteoarthritis, knee; telemedicine; exercise; treatment outcome; review; meta-analysis

Introduction

As a consequence of demographic, epidemiological, and social changes, the need for chronic care increases while health care capacities decrease [1]. This requires a change in how care is delivered [2]. As a shift from hospital care to home care is observed, self-management plays an increasingly important role to manage or improve the health of patients [3]. At the same time, home care and home therapy need to be well coordinated and consistent with quality standards [1].

Epidemiology and Consequences of Osteoarthritis of the Knee

Osteoarthritis of the knee (OAK) is an example of a chronic disease, where self-management and home therapy are an important part of health care. Following low back pain and neck pain, osteoarthritis (OA) in general is the third most common musculoskeletal disease worldwide [4]; global prevalence of OAK for persons older than 60 years is estimated at 33.6% for women and 24.3% for men [5]. Affected individuals and their families have to adapt to the disease, loss of mobility, and diminished quality of life (QoL), which are the main contributors to personal suffering. Pain, joint stiffness, instability, and decreased physical function are the major drivers for OA-related activity decline and disability [6]. As a consequence, patients with OA are at a higher risk of obesity, cardiovascular disease, and death compared with the general population [7].

Exercise for Osteoarthritis of the Knee

As mechanical factors are the main drivers for the pathogenesis of OAK, a positive response to exercise interventions and increased physical activity (PA) can be expected. A recent systematic review has shown short-term clinical meaningful improvements of pain and physical function following exercise interventions [8]. However, access to facilities offering such therapies is restricted because of the patients' mobility limitations, transport problems, and time constraints, especially in rural areas. Furthermore, the increase of chronic disease puts further strain on limited health care resources accelerating the shift toward home-based interventions and self-management.

Home exercise interventions include targeted physical activities aiming to improve muscle strength, joint range of motion, proprioception, and aerobic capacity; of these lower limb strengthening and isolated quadriceps training seem to have the largest effect on pain and physical function [8]. High intensity training results in greater beneficial effects on pain and physical function, eg, strength increase of knee extensors should be at least 30% to 40% to have a beneficial effect [9]. To achieve such a magnitude, physiologic principles of load progression need to be considered. The positive effects of increased muscle strength may be because of the positive influence on biomechanics, decreasing joint load, and focal stress on the articular cartilage [8].

Physical deconditioning and risk of obesity are closely associated with OAK [7]. Aerobic exercises may counteract these factors by leading to better overall fitness and supporting weight loss. Aerobic exercise leads to an increased peak oxygen uptake, which is inversely related to morbidity and mortality and reduces effort for submaximal daily tasks [8].

In patients with OAK, malalignment and altered kinematics may cause unequal distribution of load within the joint, which is seen as one driver for onset and progression of OAK [10]. Proprioceptive training such as stepping, standing, walking, balancing, and training of joint position sense may improve proprioceptive capacity and joint function [11].

Electronic Health Interventions

Adherence to home exercise programs is however low [12], and it seems difficult to achieve effective training intensity without adequate support and motivation. Electronic health (eHealth) technology is seen as a promising possibility to overcome these limitations [13]. eHealth-supported, home-based interventions can prevent and rehabilitate or treat many chronic conditions by providing patient education, instructions for self-management, motivation, monitoring, feedback, and enabling communication [13]. These features may enhance patient motivation and promote adherence to home exercise interventions. One example for such an eHealth intervention is the Internet-based program "join2move" [14] that includes education, exercise instruction, goal setting, time contingent exercise increase, and positive reinforcement via electronic reminders.

Although more than 43,000 health-related apps are available in the US Apple Store alone [15], the evidence base for efficacy and efficiency of many existing eHealth-assisted interventions is not sufficient. The aim of this systematic review was therefore to investigate the efficacy of eHealth-supported home-exercise interventions in the treatment of patients with OAK.

Methods

Protocol and Registration

Methods of literature search and data analysis were specified in advance and documented in a protocol. The protocol was registered under CRD42017072079 (PROSPERO CRD register). This systematic review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses recommendations [16].

Eligibility Criteria

Randomized controlled trials (RCTs) and controlled clinical trials (CCTs) investigating eHealth-supported home exercise interventions compared with no treatment or other treatments for patients with symptomatic unilateral or bilateral OAK were included. Diagnosis of OAK was based on self-report, radiography, clinical signs, or physician diagnosis. All other forms of arthritis were excluded. Studies with all types of eHealth-supported exercise interventions were included. Outcomes considered in this review were pain, function, and QoL. Studies had to be published English or German. Date limitations were not used.

Information Sources

The following databases were searched in July 2017: CENTRAL, MEDLINE via PubMed, CINAHL, PEDro, and journal websites. Additionally, reference lists of included studies were hand-searched. Date last searched was July 27, 2017.

Search Strategy

Databases were searched with the keywords Knee Osteoarthritis, Exercise AND eHealth, and RCT OR CCT and their related Medical Subject Heading and synonyms.

The terms animal OR animals and arthroplast* were used to build an exclusion filter.

The boolean operators “OR,” “AND,” and “NOT” were used to build the search strategy. Detailed search strategies for electronic databases are presented in Multimedia Appendix 1.

Study Selection

Title, keywords, abstracts, and full-texts were assessed to establish whether the study met the prespecified eligibility criteria relating to population, intervention, and study design. A checklist was used to assess eligibility criteria. Eligibility was assessed independently by two review authors (AS and CZ), and disagreements were resolved by consensus. For each selected study, the full text was retrieved for final assessment.

Data Collection Process

Data were extracted for study design, participant characteristics, intervention, control, types of outcome measures, follow-up, outcomes, and funding using a standardized form. One author (AS) extracted the data; this was checked by a second author (CZ). Disagreements were resolved by discussion. For each outcome, means, SDs, 95% CI, and *P* values were collected for each point of measurement. When necessary, SDs were calculated using available data (eg, CI) or information presented in graphical format.

Data Items

Data were retrieved for the following variables: study type; patient characteristics such as age, sex, and diagnosis; type of diagnosis (self-report, radiography, signs, and symptoms); the intervention (type of exercise intervention and eHealth technology, frequency and duration of sessions, and duration of therapy); the control intervention (type of intervention, frequency and duration of sessions, and duration of therapy); outcomes (construct, measurement instrument, length of follow-up, and points of measurement); and funding sources.

Risk of Bias in Individual Studies

The Cochrane risk of bias assessment method was used to rate the risk of bias in individual studies [17]. Two authors (AS and HvP) independently assessed the risk of bias of the included studies, and disagreements were resolved by consensus. The following bias sources were assessed: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other bias (such as recruitment bias in cluster RCTs or unbalanced groups).

Review Manager 5.3.5 (The Nordic Cochrane Centre) was used to generate a risk of bias figure.

Summary Measures

For continuous data, standard mean differences (SMDs) and 95% CIs were calculated from means and SDs using the following formula: SMD=mean difference/pooled SD.

Calculations were conducted with Review Manager 5.3.5 software (The Nordic Cochrane Centre). Not reported SDs were calculated with the calculator tool of Review Manager. SMDs of 0.2, 0.5, and 0.8 were rated as small, moderate, and large, respectively [18].

According to the guidelines for summary of findings tables [19], SMDs were translated in absolute mean differences by multiplying SMDs with a control group baseline SD extracted from one representative study and dividing it by the maximum points achievable on this measurement scale. A study was judged as representative when it represented the target population to a high degree and had a large weight within the meta-analysis. Relative differences were calculated dividing the absolute benefit by the representative control group baseline mean.

Synthesis of Results

Data from multiple studies were pooled in a meta-analysis using a random-effects model. I^2 statistic was used to assess statistical heterogeneity across pooled studies. Values of 25%, 50%, or 75% were considered as low, moderate, or high level of heterogeneity, respectively [20].

Risk of Bias Across Studies

The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach was used to evaluate the quality of evidence across studies for each outcome using the following predefined criteria [21]:

1. Inconsistency (downgraded if $I^2 \geq 50\%$)
2. Indirectness (downgraded if clinically heterogeneous)
3. Imprecision: downgraded if the pooled sample size was below the calculated sample size of an adequately powered single trial (optimal information size) [22] for each outcome. The minimal clinical important change (MCIC) is considered as delta in the power calculation. The following values are considered as MCIC for patients with OAK: pain (visual analog scale, VAS or numerical rating scale, NRS),

physical function 20% improvement [23], and QoL (36-item short form survey) 12% improvement from baseline [24]. Furthermore, Guyatt et al [22] suggest downgrading for sample sizes < 400 (200 per group) or if the CI overlaps no effect but includes an important improvement.

4. Risk of bias: downgrading should be considered when a "substantial risk of bias across most of the body of evidence" is suspected [25].

Quality was rated as high, moderate, low, or very low according to the GRADE criteria [21]. The GradePro online app [26] was used to generate GRADE evidence profiles and a summary of findings tables. Quality of evidence across studies was evaluated by one author (AS) and checked by a second author (HvP). Disagreements were resolved by consensus.

Additional Analyses

Subgroup analysis (not prespecified) was conducted for different eHealth modes of delivery (mobile apps vs telephone).

Results

Study Selection

The literature search yielded a total of 635 records. The process of study selection is presented in Figure 1. After removing duplicates and screening of titles and abstracts, 19 full-text articles were retrieved and assessed for eligibility. Of these, 12 were excluded because of inappropriate study design, intervention, population, or outcomes.

Seven articles [14,27-32] were included, and results were pooled in a meta-analysis. One study was published twice, with outcomes pain and function reported in one article [28] and QoL in the other [29].

Study Characteristics

The characteristics of included studies are presented in Multimedia Appendix 2. All studies were two-group RCTs. A total of 742 participants were randomized in intervention (n=376) or control (n=366) groups. Sample sizes in individual studies ranged from 34 to 211 participants. Sixty-four percent of the participants were female (473/742). Four studies included participants with unilateral or bilateral OAK [27-30,32]; one study included a mixed group with knee OA (67%), hip OA (21%), or both (12%) [14]; and one study included participants with chronic knee pain [31].

Interventions included exercises supported by mHealth (Internet-based programs or mobile apps) [14,27,31] and telephone-supported exercises [28-30,32]. Exercise interventions most commonly included strengthening exercises [28,29,31,32], walking [14,27-29], or PA reinforcement [14,27,30-32].

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. OAK: osteoarthritis of the knee.

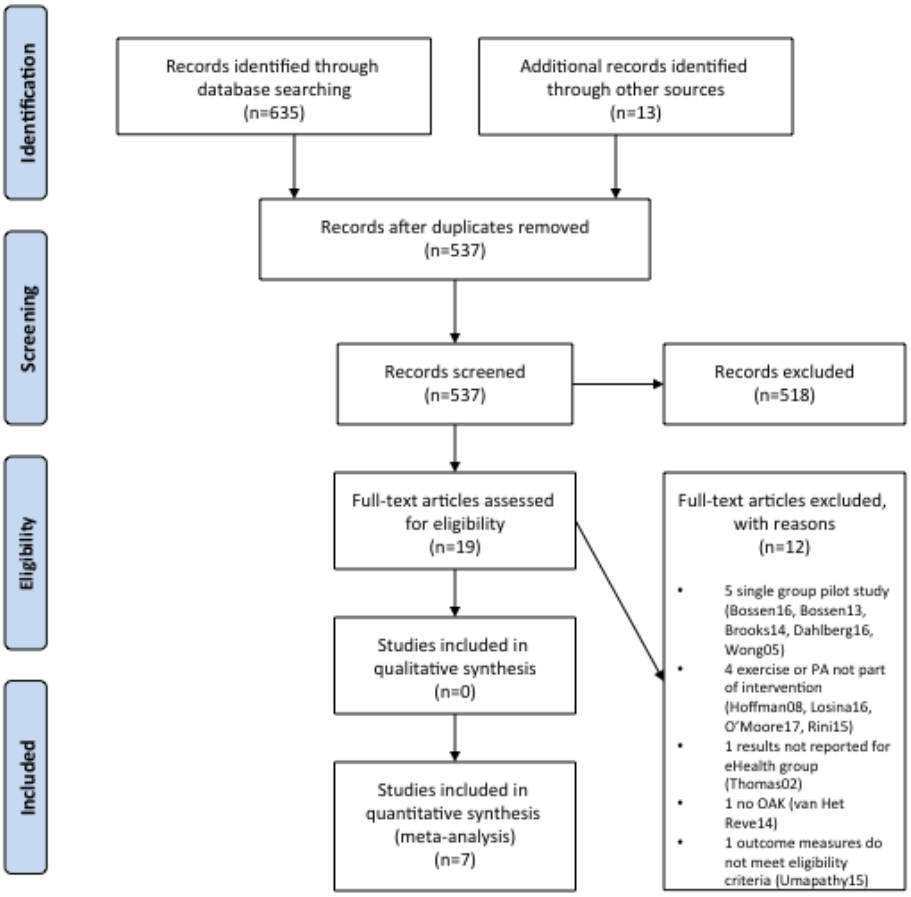
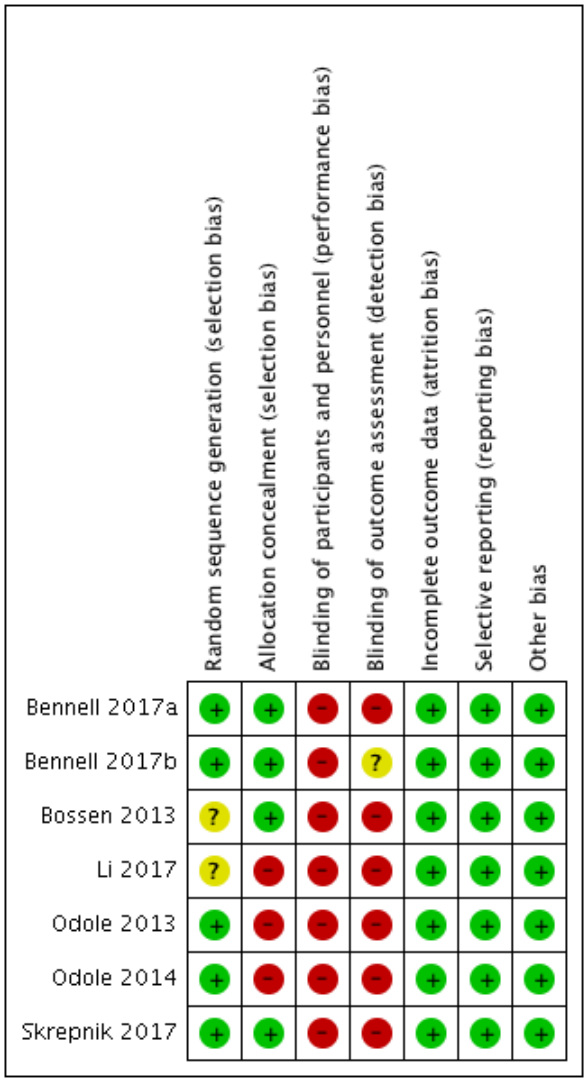


Figure 2. Risk of bias summary.



The eHealth component included education on topics such as exercise, healthy diet, pain management, and self-management. A counseling component, typically consisting of reminders, encouragement, and discussion of experienced barriers in varying proportions was also present. Modes of delivery included telephone calls [28-30,32], mobile apps [27], and Internet-based programs [14,31].

All studies reported pain and physical function as primary or secondary outcome measure. The most common measures of pain were the VAS or NRS used in certain studies [14,28,29,31,32], the pain subscales of the Western Ontario and MacMaster Universities Osteoarthritis Index (WOMAC) [33] used in one study [31], and the Knee injury and OA Outcome Score (KOOS) [34] used in another study [30]. Physical function was measured with the WOMAC in [14,31], the KOOS in [14,30], and the Ibadan Knee/Hip Osteoarthritis Outcome Measure in [28]. Health-related QoL was assessed with the Assessment of Quality of Life-Version 2 [35] used in [31,32], the KOOS in [14], and the WHO Quality of Life Assessment [36] was used in [29]. Other outcome measurements identified included global change, amount of PA, or steps walked.

Short-term follow-up time points of measurements included 1 month [30], 6 weeks [28,29], 3 months [14,27,31], and 6 months [32]. Long-term follow-up ranged from 9 months [32] to 12 months [14,32]. One study reported long-term outcomes at 18 months [32].

Risk of Bias Within Studies

Risk of bias within studies was assessed using seven criteria recommended by the Cochrane Collaboration [17] (Figure 2). None of the studies reported blinding of participants, therapists, or outcome assessors. In 2 of the studies [14,30], randomization was performed, but the method of random sequence generation was not specified. Therefore, risk of selection bias was classified as unclear for these two studies. In 2 studies [28-30], allocation concealment was not reported. Attrition, reporting, or other bias was not detected in any of the included studies.

Synthesis of Results

Pooled effect estimates including CIs are presented in this section for the outcomes pain, physical function, and health-related QoL. Calculations for absolute reduction or improvement in percentage were based on the control group baseline means (SD) from Bennell et al [32]: pain 58 (15),

physical function 45 (15), and QoL 70 (10). Quality of evidence across studies was evaluated for each outcome using the GRADE approach [21]. A summary of findings table is presented in Multimedia Appendix 3.

Pain Short Term

All 6 studies (n=742 participants) reported data for the outcome pain intensity short term (1-6 months follow-up; Figure 3). Pooled results indicate significant benefit for eHealth-supported exercise intervention (SMD=−0.31; 95% CI −0.58 to −0.04). The effect size was small according to Cohen [18] and equals a reduction of five points (95% CI 1-9) on a 0 to 100 points pain scale (0=no pain). Heterogeneity was high with I^2=67%. The quality of evidence for this outcome was low.

Pain Long Term

Three studies (n=416 participants) provided information for the outcome pain intensity long term (9-12 months follow-up; Figure 4). Pooled effect estimates showed a significant but small beneficial effect for eHealth-supported exercise (SMD=−0.30; 95% CI −0.53 to −0.07). This translates in a reduction of five points (95% CI 1-8) on a 0 to 100 points pain scale (0=no pain). Heterogeneity was low with I^2=29%. The quality of evidence for this outcome was moderate.

Physical Function Short Term

Four studies (n=479 participants) provided data for the outcome physical function short-term (1-6 months follow-up; Figure 5). Pooling of results from individual studies showed a nonsignificant, small beneficial effect (SMD=−0.30; 95% CI −0.76 to 0.17). This equals an improvement of four points (95% CI −3 to 11) on a 0 to 100 points physical function scale (100=full function). Heterogeneity was high with 83%. The quality of evidence for this outcome was low.

Physical Function Long Term

Data for the outcome physical function long term (9-12 months follow-up) were extracted from 3 studies (n=416 participants). Pooling the results of individual studies showed a small, significant beneficial effect favoring the intervention group (SMD=0.41; 95% CI 0.17-0.64; Figure 6). This equals an improvement of six points (95% CI 3-10) on a 0 to 100 points scale (higher scores indicate better function). Heterogeneity was moderate with I^2=33%. The quality of evidence for this outcome was high.

Quality of Life Short Term

Four studies (n=496 participants) provided information for the outcome QoL short term (3-6 months follow-up). Pooling results of individual studies showed a small, significant beneficial effect favoring the intervention (SMD=0.24; 95% CI 0.05-0.43; Figure 7). This translates in an improvement of three points (95% CI 1-4) on a 0 to 100 points scale (higher scores=better QoL). Heterogeneity was low with I^2=10%. The quality of evidence for this outcome was moderate.

Figure 3. Forest plot outcome pain short-term.

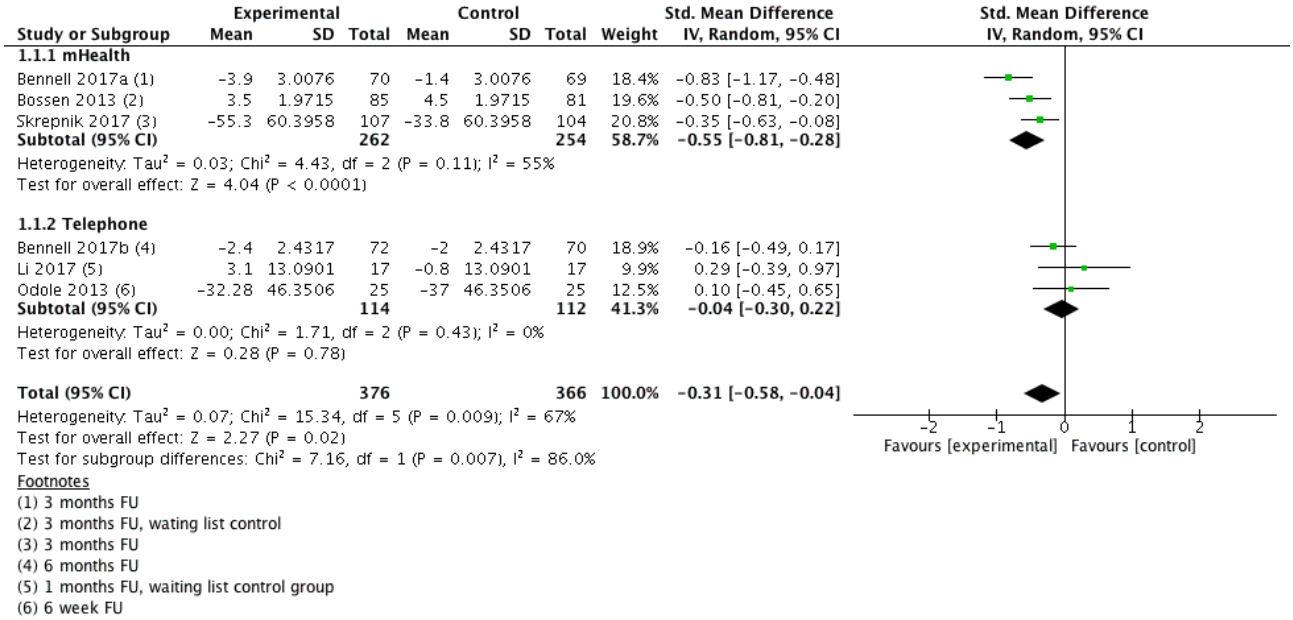


Figure 4. Forest plot outcome pain long-term.

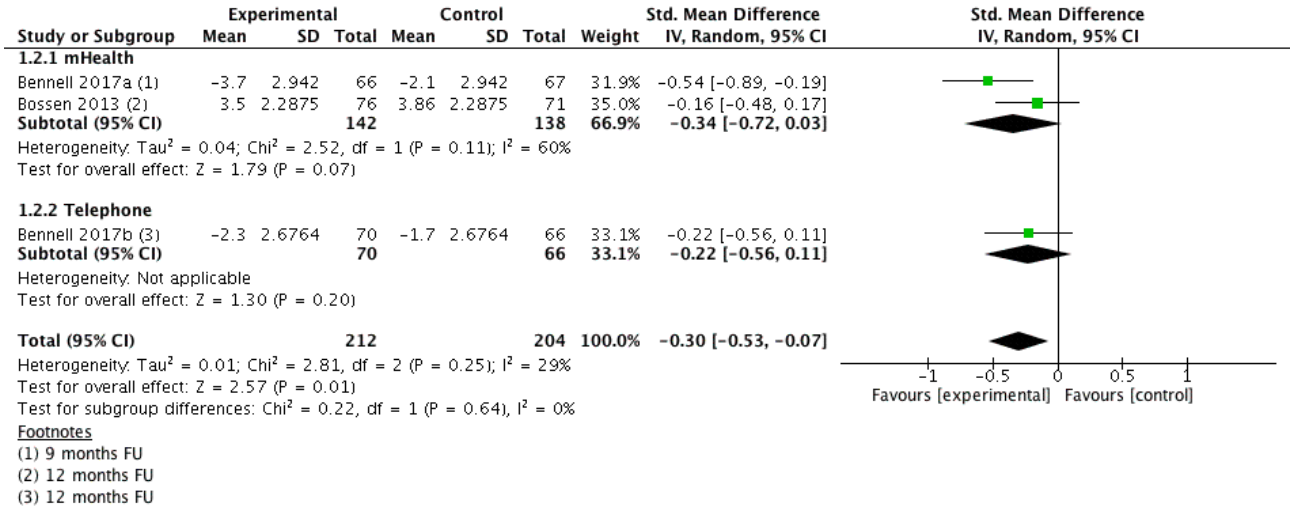


Figure 5. Forest plot outcome function short-term.

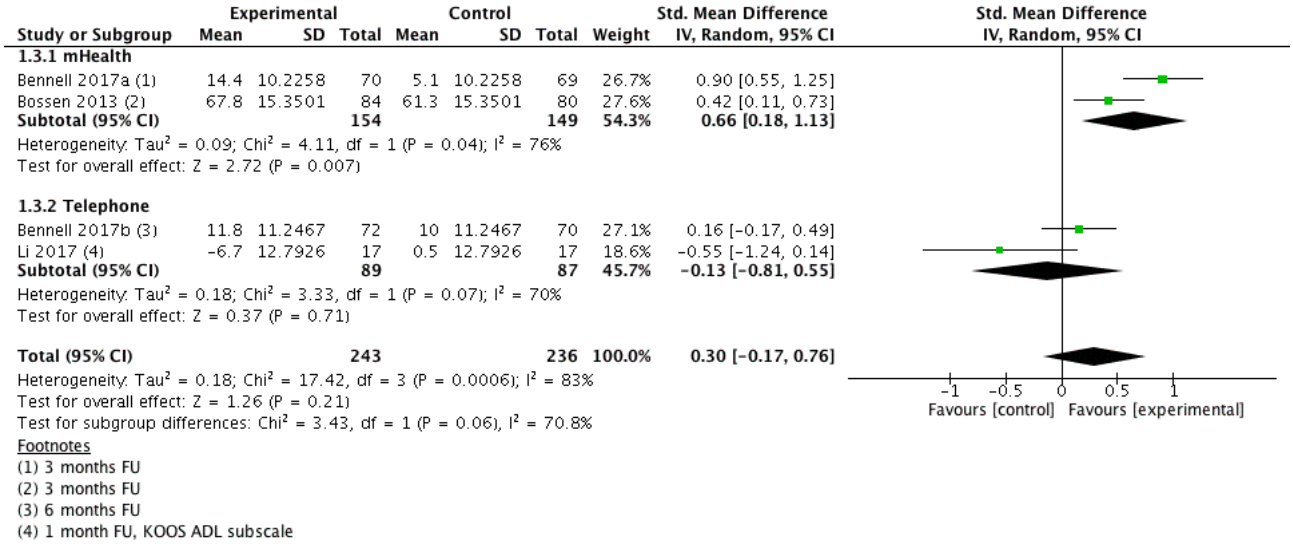
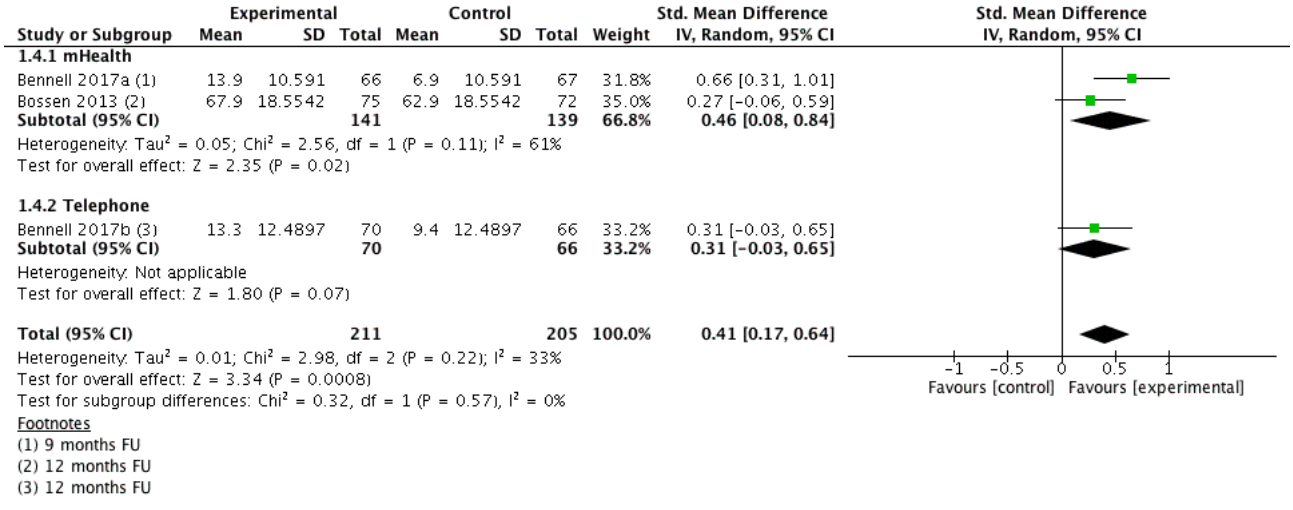


Figure 6. Forest plot outcome function long-term.



Quality of Life Long Term

Three studies (n=415 participants) provided data for the outcome QoL long term (9-12 months follow-up; Figure 8). Pooling results from individual studies yielded a small, significant beneficial effect favoring the intervention group (SMD=0.27; 95% CI 0.06-0.47). This corresponds to an improvement of three points (95% CI 1-4). Heterogeneity was low with I^2=12%. The quality of evidence for this outcome was high.

Quality of Evidence Across Studies

For each outcome, quality of evidence was assessed using the GRADE approach [21] (Tables 1 and 2). Quality of evidence for short-term outcomes were low for pain and physical function and moderate for QoL. Reasons for downgrading one level were risk of bias because of lack of blinding of therapists, patients, and outcome assessors for all short-term outcomes. Outcomes pain and physical function were further downgraded one level because of inconsistency (I^2 >50%). Quality of evidence for long-term outcomes was rated moderate for pain and high for physical function and QoL. The outcome pain long term was downgraded one level because of lack of blinding of therapists, patients, and outcome assessors.

Additional Analysis

A sensitivity analysis was conducted to assess the impact of treatment duration on heterogeneity. This was done excluding the 2 studies with the shortest intervention duration [28-30]. However, this did not substantially change the amount of observed heterogeneity between groups (pain short term: I^2 from 67%-63%; physical function short term: 83%-79%; QoL short term: 10%-0%).

Subgroup analyses were performed to investigate whether studies with different treatment delivery modes and treatment contents differed in regards to their effect size (Table 3). Studies were classified into two groups: the first group consisted of studies where treatment was delivered via mHealth technology (mobile apps), the second consisted of studies where treatment was delivered via telephone. A significant difference was found between mHealth (SMD=−0.55) and telephone (SMD=−0.04) supported exercise studies in pooled effect estimates for the outcome pain short-term (χ^2=7.2 *P*=.007). A substantial, but not significant difference was noted for the outcome physical function short-term between mHealth (SMD=−0.66) and telephone (SMD=0.13) supported exercise studies in pooled effect estimates for outcome pain short-term (χ^2=3.4 *P*=.06).

Figure 7. Forest plot outcome quality of life short-term.

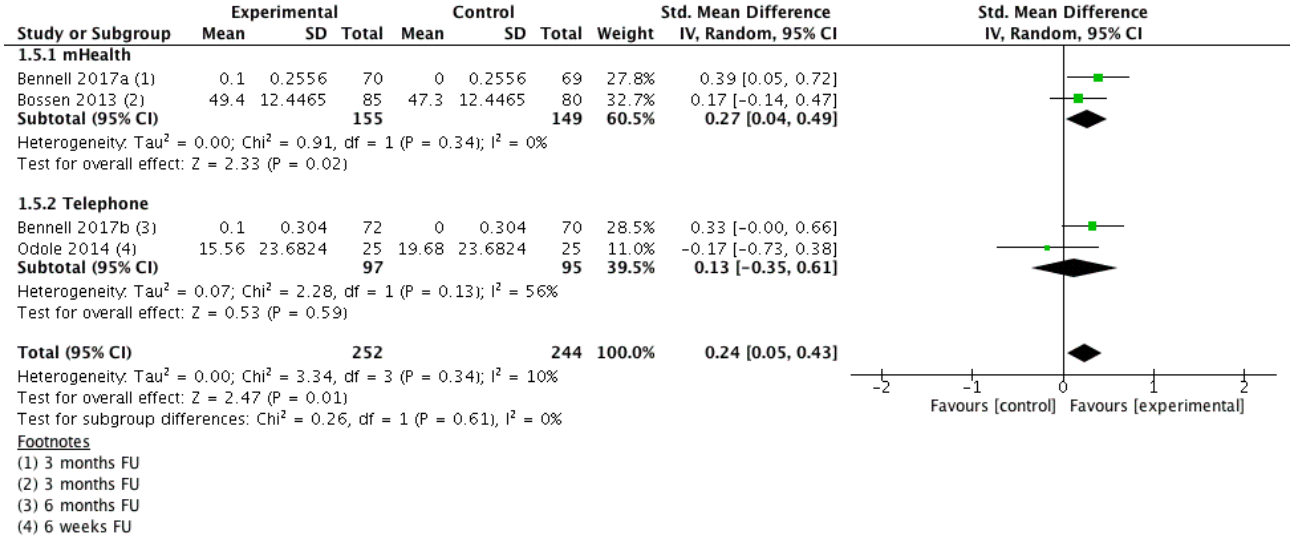


Figure 8. Forest plot outcome quality of life long-term.

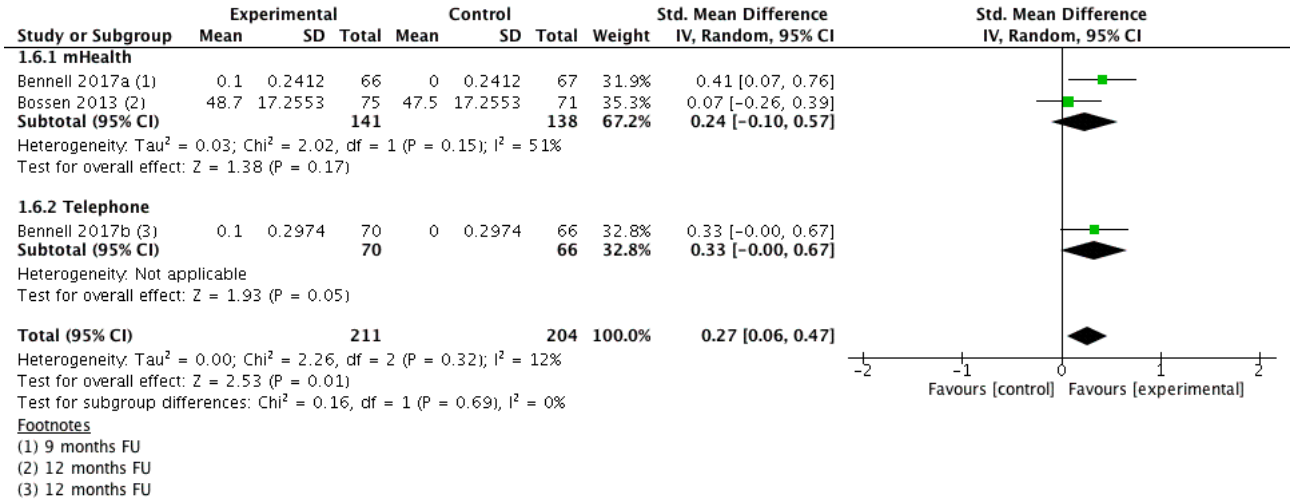


Table 1. Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) evidence profile. QoL: quality of life. RCT: randomized controlled trial.

Outcome	Quality assessment						
	Number of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations
Pain short term (follow-up: range 1 month to 6 months; assessed with self-report questionnaire 0-100 (higher numbers=more pain)	6	RCT	Serious[a]	Serious[b]	Not serious	Not serious	None
Pain long term (follow-up: range 9 months to 12 months; assessed with self-report questionnaire 0-100 (higher numbers=more pain)	3	RCT	Serious[a,c]	Not serious	Not serious	Not serious	None
Physical function short term (follow-up: range 1 month to 6 months; assessed with self-report questionnaire 0-100; higher numbers=better function)	4	RCT	Serious[a,c]	Serious[b]	Not serious	Not serious	None
Physical function long term (follow-up: range 9 months to 12 months; assessed with self-report questionnaire 0-100; higher numbers=better function)	3	RCT	Not serious	Not serious	Not serious	Not serious	None
QoL short term (follow-up: range 3 months to 6 months; assessed with self-report questionnaire 0-100; higher numbers=better QoL)	4	RCT	Serious[a]	Not serious	Not serious	Not serious	None
QoL long term (follow-up: range 9 months to 12 months; assessed with self-report questionnaire 0-100; higher numbers=better QoL)	3	RCT	Not serious	Not serious	Not serious	Not serious	None

[a]Serious risk of bias across studies because of missing blinding of therapists, patients, and outcome assessors.

[b]Heterogeneity was high with I^2 >50%.

[c]Randomization or allocation procedure unclear for some studies.

Table 2. Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) summary of findings. QoL: quality of life. SMD: standardized mean difference.

Outcome	Number of patients		Effect		Quality	Importance
	Electronic health–supported exercise	No or other intervention	Relative (95% CI)	Absolute (95% CI)		
Pain short term (follow-up: range 1 month to 6 months; assessed with self-report questionnaire 0-100 (higher numbers=more pain)	367	361	_[a]	SMD 0.31 SD lower (0.04 lower to 0.58 lower)	++oo Low	Important
Pain long term (follow-up: range 9 months to 12 months; assessed with self-report questionnaire 0-100 (higher numbers=more pain)	212	204	–	SMD 0.3 SD lower (0.07 lower to 0.53 lower)	+++o Moderate	Critical
Physical function short term (follow-up: range 1 months to 6 months; assessed with self-report questionnaire 0-100; higher numbers=better function)	243	236	–	SMD 0.3 SD higher (0.17 lower to 0.76 higher)	++oo Low	Important
Physical function long term (follow-up: range 9 months to 12 months; assessed with self-report questionnaire 0-100; higher numbers=better function)	211	205	–	SMD 0.41 SD higher (0.17 higher to 0.64 higher)	++++ High	Critical
QoL short term (follow-up: range 3 months to 6 months; assessed with self-report questionnaire 0-100; higher numbers=better QoL)	227	219	–	SMD 0.24 SD higher (0.05 higher to 0.43 higher)	+++o Moderate	Important
QoL long term (follow-up: range 9 months to 12 months; assessed with self-report questionnaire 0-100; higher numbers=better QoL)	211	204	–	SMD 0.27 SD higher (0.06 higher to 0.47 higher)	++++ High	Critical

[a]Indicates "not applicable".

Table 3. Data and analysis.

Outcome	Studies	Participants	Statistical method	Effect estimate
1.1 Pain short term	6	742	SMD[a] (IV, Random, 95% CI)	−0.31 (−0.58 to −0.04)
1.1.1 mobile health (mHealth)	3	516	SMD (IV, Random, 95% CI)	−0.55 (−0.81 to −0.28)
1.1.2 Telephone	3	226	SMD (IV, Random, 95% CI)	−0.04 (−0.30 to 0.22)
1.2 Pain long term	3	416	SMD (IV, Random, 95% CI)	−0.30 (−0.53 to −0.07)
1.2.1 mHealth	2	280	SMD (IV, Random, 95% CI)	−0.34 (−0.72 to 0.03)
1.2.2 Telephone	1	136	SMD (IV, Random, 95% CI)	−0.22 (−0.56 to 0.11)
1.3 Physical function short term	4	479	SMD (IV, Random, 95% CI)	−0.30 (−0.17 to 0.76)
1.3.1 mHealth	2	303	SMD (IV, Random, 95% CI)	0.66 (0.18 to 1.13)
1.3.2 Telephone	2	176	SMD (IV, Random, 95% CI)	—0.13 (−0.81 to 0.55)
1.4 Physical function long term	3	416	SMD (IV, Random, 95% CI)	0.41 (0.17 to 0.64)
1.4.1 mHealth	2	280	SMD (IV, Random, 95% CI)	0.46 (0.08 to 0.84)
1.4.2 Telephone	1	136	SMD (IV, Random, 95% CI)	0.31 (−0.03 to 0.65)
1.5 Quality of Life short term	4	496	SMD (IV, Random, 95% CI)	0.24 (0.05 to 0.43)
1.5.1 mHealth	2	304	SMD (IV, Random, 95% CI)	0.27 (0.04 to 0.49)
1.5.2 Telephone	2	192	SMD (IV, Random, 95% CI)	0.13 (−0.35 to 0.61)
1.6 Quality of Life long term	3	415	SMD (IV, Fixed, 95% CI)	0.27 (0.06 to 0.47)
1.6.1 mHealth	2	279	SMD (IV, Random, 95% CI)	0.24 (−0.10 to 0.57)
1.6.2 Telephone	1	136	SMD (IV, Random, 95% CI)	0.33 (0.00 to 0.67)

Discussion

Principal Findings

This systematic review included six RCTs with a total of 742 participants. Pooling the results of 6 studies demonstrated that eHealth-supported exercise interventions resulted in improved pain (SMD=−0.31, 95% CI −0.58 to −0.04) and pooled results from 4 studies (n=446 participants) indicated improvement of health-related QoL (SMD=0.24, 95% CI 0.05-0.43) immediately post intervention. These treatment effects would be considered small and translate into an absolute mean improvement of 5% (95% CI 1%-9%) for pain and 3% (95% CI 1%-4%) for health-related QoL. Improvement in pain was comparable with other interventions such as nonsteroidal antiinflammatory drugs (SMD=−0.29, 95% CI −0.35 to −0.22) or strengthening (SMD=−0.32, 95% CI −0.42 to −0.23) and were superior to aquatherapy (SMD=−0.19, 95% CI −0.35 to −0.04) [37]. Fransen et al [8] demonstrated that land-based exercise resulted in higher effect sizes of −0.49 (95% CI −0.59 to −0.39) for pain and 0.52 (95% CI 0.39 - 0.64) for physical function. Results for QoL were comparable with SMD 0.28 (95% CI 0.15-0.40).

One recent meta-analysis [38] compared exercise-based telemedicine with no intervention in patients with chronic pain and found significant mean reduction in pain (mean difference=−0.57 on a 10-point scale; 95% CI −0.81 to −0.34), which corresponds to an SMD of 0.22. Improvement in physical function (SMD=−0.20, 95% CI −0.29 to −0.12) post intervention favored the intervention group. When comparing exercise-based telemedicine with usual care or exercise-based telemedicine in addition to usual care, no significant differences were observed. These effects are smaller compared with the results identified in this review. A mixed population of chronic pain patients may respond differently to eHealth-supported exercise compared with patients with OAK. One main difference is that patients with OAK have an identifiable specific structural pathology, whereas patients with chronic pain are heterogeneous in regards to pathology and contributing factors and possibly respond to a lesser extent to exercise therapy.

One important finding of this meta-analysis was that the pooled long-term outcomes from 3 studies (n=416) showed that eHealth-supported exercise resulted in reduced pain (SMD=−0.30, 95% CI −0.53 to −0.07), improved physical function (SMD=0.41, 95% CI 0.17 - 0.64), and QoL (SMD=0.26, 95% CI 0.06 - 0.47) [27-29]. These treatment effects would be considered small and translate into absolute mean improvement of 5% (95% CI 1%-8%) for pain, 6% (95% CI 6%-10%) for physical function, and 3% (95% CI 1%-4%) for QoL. These findings indicate that the effects of eHealth-supported exercise are sustainable over a 9 to 12 months period.

Although observed improvements for most long- and short-term outcomes were statistically significant in this systematic review, they may not make a relevant difference for individual patients. Minimal clinical important changes from baseline are 20% for pain (VAS or NRS) and physical function [23] and 12% for QoL [24]. These are substantially higher than absolute changes found in this meta-analysis.

Additionally, it is important to note that 2 of the 6 studies used a waiting list control group [14,30] and another 2 studies used an education control group [27,31]. It may be possible that the

choice of the control group may have inflated the effect size. However, one study [30] with waiting list control group reported a nonsignificant effect for pain and function short term in favor of the control group.

One subgroup analysis per outcome comparing the effects of different treatment delivery modalities and treatment contents (mHealth and telephone) was conducted (Table 3). Studies in the subgroups differed in regards to mode of communication (automated in the mHealth subgroup vs personal in the telephone subgroup), access to the intervention (selfguided in the mHealth subgroup and fixed dates in the telephone subgroup), and contents of intervention. Although results have to be interpreted cautiously as the comparison is not based on randomization, a general trend for greater effect sizes in the mHealth group was observed that reached statistical significance for the outcome pain short-term. Possible reasons for the observed greater beneficial effect of mHealth interventions are stated below.

First, mHealth interventions were more complex and consisted of various elements such as information, educational material, training of self-management skills, and exercise compared with telephone interventions that typically included telephone coaching and exercises.

Second, different control group interventions may also account for a greater observed effect in studies investigating mHealth. In mHealth studies, control interventions included educational material only [31], waiting list [14], and injections plus information [27]. In comparison, control groups in the telephone studies consisted of supervised physiotherapy in 2 studies [28,29,32] and waiting list in one study [30].

Third, treatment duration was shorter in 2 of the telephone studies with 4 and 6 weeks [28-30] compared with 3 months in the mHealth studies [14,23,27,28,31]. Each of these factors alone or in combination could have contributed to the observed differences in effect size.

Limitations

In this section, limitations at study and outcome level, as well as limitations of the review process, are discussed.

Risk of Bias at Study Level

Risk of bias was low for 4 of the included studies [14,27,31,32] and moderate for the remaining 2 studies [28-30]. Lack of blinding of patients, therapists, and outcome assessors was noted in all of the studies. Although blinding of participants and therapists to treatment modality is difficult to achieve in exercise interventions, lack of blinding may nonetheless introduce overestimation of effects and should therefore be assessed [17]. Blinding of outcome assessors and statisticians would have been possible. However, only one study reported blinding of the statistician and that patients were unaware of the study hypothesis [31]. In the 2 studies with the highest risk of bias [28-30], allocation concealment was not reported; in one study [28,29], adequate random sequence generation was unclear. As the net effect of these 2 studies on the pooled effect size is in favor of the control group, the overall risk of bias is judged as low.

Overall Quality of Evidence

Quality of evidence across studies was assessed high for the outcomes physical function long term and QoL long term, moderate for pain long term and QoL short term, and low for pain short term and physical function short term.

Reasons for downgrading the quality of evidence was risk of bias and imprecision. For the outcomes pain short term, pain long term, physical function short term, and QoL short term, quality of evidence was downgraded one level because the risk of bias was assessed as serious. Reasons were lack of blinding and unclear allocation concealment across studies that may have introduced some overestimation of the results.

Quality was downgraded one level because of substantial inconsistency for the short-term outcomes pain (I^2=67%) and physical function (I^2=83%). Some reasons for inconsistency have been described above and include differences in treatment delivery modes (mHealth vs telephone), treatment duration, and control treatments. The exercise component of the eHealth intervention also varied between studies. Strengthening and reinforcement of PA was used in 3 studies [28,29,31,32], reinforcement of aerobic exercise such as walking or cycling in 2 studies [14,27], and reinforcement of general PA in one study [30].

Further reasons for inconsistency may include heterogeneity between study populations. One study [31] included patients with chronic knee pain. The proportion of patients with arthritis was probably high in this study as the inclusion criteria included age above 50 years, knee pain during walking, and more than 20 points on the WOMAC physical function subscale. Another study [14] included patients with hip and knee OA, but the majority of patients (79%) had OAK. Additionally, diverse cultural and socioeconomic backgrounds of participants coming from Australia, the Netherlands, the United States, Canada, and Nigeria may have contributed to the heterogeneity of study participants.

Limitations in the Review Process

Some limitations regarding the review process should be mentioned. These include that only studies published in English or German language were considered. Studies published in other languages could not be considered and were potentially overlooked. Studies investigating the effect of eHealth interventions are rapidly increasing; four published study protocols could be identified that matched the eligibility criteria of this systematic review [39-42]. It is possible that results from these ongoing studies may change the findings of the meta-analysis in this review.

Conclusions

Overall, eHealth-supported exercise interventions demonstrated beneficial small short- and long-term effects on pain, physical function, and QoL in patients with OAK. These effects may be too small to make a relevant difference for individual patients. The quality of evidence was low to moderate for short-term outcomes, therefore future trials are likely to change the results for short-term outcomes. The quality of evidence for long-term

outcomes were moderate to high; it seems unlikely that future studies may change results substantially.

Taking into account the balance between benefits and harm, the magnitude of effects, the importance of outcomes, the quality of evidence, the values and preferences of patients, and cost-effectiveness [25], the following recommendation is put forward:

In patients with OAK, clinicians should consider using eHealth interventions to support home exercise and self-management (weak recommendation, moderate quality of evidence).

This recommendation places a high weight on the positive balance of (small) benefits against possible adverse events and on patient's values and preferences. Less weight is placed on implementation barriers because of lack of training and financial incentives of health care providers.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies.

[PDF File (Adobe PDF File), 63KB - jmir_v20i4e152_app1.pdf]

Multimedia Appendix 2

Characteristics of included studies.

[PDF File (Adobe PDF File), 31KB - jmir_v20i4e152_app2.pdf]

Multimedia Appendix 3

Electronic health (eHealth)-supported exercise compared with no or other interventions for knee osteoarthritis GRADE summary of findings table.

[PDF File (Adobe PDF File), 40KB - jmir_v20i4e152_app3.pdf]

References

1. Robert Koch Institut, editor. Wie haben sich Angebot und Inanspruchnahme der Gesundheitsversorgung verändert? [How have the supply and use of health care changed?]. In: Gesundheit in Deutschland [Health in Germany]. Berlin: Robert Koch Institut; 2015:300-372.
2. Fischer F, Kramer A. eHealth: Hintergrund und Begriffsbestimmung [EHealth background and conceptualization]. In: Fischer F, Kramer A, editors. eHealth in Deutschland: Anforderungen und Potenziale innovativer Versorgungsstrukturen [eHealth in Germany: Requirements and Potential of Innovative Supply Structures]. Berlin: Springer Verlag; 2016:25-59.
3. Royal College of Nursing.: RCN Policy and International Department; 2014. Moving care to the community - an international perspective URL: https://matrix.rcn.org.uk/__data/assets/pdf_file/0006/523068/12.13_Moving_care_to_the_community_an_international_perspective.pdf [WebCite Cache ID 6xrX97QaJ]
4. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196. [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
5. Pereira D, Peleteiro B, Araújo J, Branco J, Santos RA, Ramos E. The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. Osteoarthritis Cartilage 2011 Nov;19(11):1270-1285 [FREE Full text] [doi: 10.1016/j.joca.2011.08.009] [Medline: 21907813]
6. Felson DT. Developments in the clinical understanding of osteoarthritis. Arthritis Res Ther 2009;11(1):203 [FREE Full text] [doi: 10.1186/ar2531] [Medline: 19232065]
7. Busija L, Bridgett L, Williams SR, Osborne RH, Buchbinder R, March L, et al. Osteoarthritis. Best Pract Res Clin Rheumatol 2010 Dec;24(6):757-768. [doi: 10.1016/j.berh.2010.11.001] [Medline: 21665124]
8. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015 Jan 09;1:CD004376. [doi: 10.1002/14651858.CD004376.pub3] [Medline: 25569281]
9. Bartholdy C, Juhl C, Christensen R, Lund H, Zhang W, Henriksen M. The role of muscle strengthening in exercise therapy for knee osteoarthritis: a systematic review and meta-regression analysis of randomized trials. Semin Arthritis Rheum 2017 Aug;47(1):9-21. [doi: 10.1016/j.semarthrit.2017.03.007] [Medline: 28438380]
10. Felson DT. Osteoarthritis as a disease of mechanics. Osteoarthritis Cartilage 2013 Jan;21(1):10-15 [FREE Full text] [doi: 10.1016/j.joca.2012.09.012] [Medline: 23041436]

11. Smith TO, King JJ, Hing CB. The effectiveness of proprioceptive-based exercise for osteoarthritis of the knee: a systematic review and meta-analysis. Rheumatol Int 2012 Nov;32(11):3339-3351. [doi: 10.1007/s00296-012-2480-7] [Medline: 22821333]
12. Pisters MF, Veenhof C, Schellevis FG, Twisk JWR, Dekker J, De Bakker DH. Exercise adherence improving long-term patient outcome in patients with osteoarthritis of the hip and/or knee. Arthritis Care Res (Hoboken) 2010 Aug;62(8):1087-1094 [FREE Full text] [doi: 10.1002/acr.20182] [Medline: 20235201]
13. Meier CA, Fitzgerald MC, Smith JM. eHealth: extending, enhancing, and evolving health care. Annu Rev Biomed Eng 2013;15:359-382. [doi: 10.1146/annurev-bioeng-071812-152350] [Medline: 23683088]
14. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
15. Dias CM, Ribeiro AG, Furtado SF. An overview about the use of healthcare applications on mobile devices. In: Miranda IM, Cruz-Cunha MM, Martinho R, Rijo R, editors. Encyclopedia of e-Health and Telemedicine. Hershey, Pennsylvania: IGI Global; 2016:285-298.
16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
17. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions.: The Cochrane Collaboration; 2011. URL: http://handbook-5-1.cochrane.org [WebCite Cache ID 6x1Xp74Zc]
18. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
19. Cochrane Musculoskeletal Group (CMSG). Summary of Findings (SoF) table guidelines for CMSG URL: http://musculoskeletal.cochrane.org/sites/musculoskeletal.cochrane.org/files/uploads/Summary [WebCite Cache ID 6v7XHc7vS]
20. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
21. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
22. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol 2011 Dec;64(12):1283-1293. [doi: 10.1016/j.jclinepi.2011.01.012] [Medline: 21839614]
23. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken) 2012 Nov;64(11):1699-1707. [doi: 10.1002/acr.21747] [Medline: 22674853]
24. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001 Aug;45(4):384-391 [FREE Full text] [doi: 10.1002/1529-0131(200108)45:4<384::AID-ART352>3.0.CO;2-0] [Medline: 11501727]
25. Schünemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook 2013. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html [WebCite Cache ID 6v7XVYoAQ]
26. GRADEpro. Gradepro.: McMaster University and Evidence Prime Inc; 2015. URL: https://gradepro.org [accessed 2017-11-20] [WebCite Cache ID 6v7XfB1wd]
27. Skrepnik N, Spitzer A, Altman R, Hoekstra J, Stewart J, Toselli R. Assessing the impact of a novel smartphone application compared with standard follow-up on mobility of patients with knee osteoarthritis following treatment with Hylan G-F 20: a randomized controlled trial. JMIR Mhealth Uhealth 2017 May 09;5(5):e64 [FREE Full text] [doi: 10.2196/mhealth.7179] [Medline: 28487266]
28. Odole AC, Ojo OD. A telephone-based physiotherapy intervention for patients with osteoarthritis of the knee. Int J Telerehabil 2013;5(2):11-20 [FREE Full text] [doi: 10.5195/ijt.2013.6125] [Medline: 25945214]
29. Odole AC, Ojo OD. Is telephysiotherapy an option for improved quality of life in patients with osteoarthritis of the knee? Int J Telemed Appl 2014;2014:1-9. [doi: 10.1155/2014/903816]
30. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A community-based physical activity counselling program for people with knee osteoarthritis: feasibility and preliminary efficacy of the track-OA study. JMIR Mhealth Uhealth 2017 Jun 26;5(6):e86 [FREE Full text] [doi: 10.2196/mhealth.7863] [Medline: 28652228]
31. 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]
32. Bennell KL, Campbell PK, Egerton T, Metcalf B, Kasza J, Forbes A, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017 Jan;69(1):84-94 [FREE Full text] [doi: 10.1002/acr.22915] [Medline: 27111441]

33. Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C, Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. J Rheumatol 1997 Dec;24(12):2413-2415. [Medline: 9415651]
34. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS)-development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88-96. [doi: 10.2519/jospt.1998.28.2.88] [Medline: 9699158]
35. Osborne RH, Hawthorne G, Lew EA, Gray LC. Quality of life assessment in the community-dwelling elderly: validation of the Assessment of Quality of Life (AQoL) Instrument and comparison with the SF-36. J Clin Epidemiol 2003 Feb;56(2):138-147. [Medline: 12654408]
36. WHO QoL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [Medline: 9626712]
37. Zhang W, Nuki G, Moskowitz R, Abramson S, Altman R, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010 Apr;18(4):476-499 [FREE Full text] [doi: 10.1016/j.joca.2010.01.013] [Medline: 20170770]
38. Adamse C, Dekker-Van Weering MG, van Etten-Jamaludin FS, Stuiver M. The effectiveness of exercise-based telemedicine on pain, physical activity and quality of life in the treatment of chronic pain: A systematic review. J Telemed Telecare 2017 Jan 01:1357633X17716576 (forthcoming). [doi: 10.1177/1357633X17716576] [Medline: 28696152]
39. Dobson F, Hinman RS, French S, Rini C, Keefe F, Nelligan R, et al. Internet-mediated physiotherapy and pain coping skills training for people with persistent knee pain (IMPACT - knee pain): a randomised controlled trial protocol. BMC Musculoskelet Disord 2014 Aug 13;15:279 [FREE Full text] [doi: 10.1186/1471-2474-15-279] [Medline: 25125068]
40. Kloek CJ, Bossen D, Veenhof C, van Dongen JM, Dekker J, de Bakker DH. Effectiveness and cost-effectiveness of a blended exercise intervention for patients with hip and/or knee osteoarthritis: study protocol of a randomized controlled trial. BMC Musculoskelet Disord 2014;15:269 [FREE Full text] [doi: 10.1186/1471-2474-15-269] [Medline: 25103686]
41. Li LC, Lineker S, Cibere J, Crooks VA, Jones CA, Kopec JA, et al. Capitalizing on the teachable moment: osteoarthritis physical activity and exercise net for improving physical activity in early knee osteoarthritis. JMIR Res Protoc 2013 May 09;2(1):e17 [FREE Full text] [doi: 10.2196/resprot.2553] [Medline: 23659903]
42. Williams QI, Gunn AH, Beaulieu JE, Benas BC, Buley B, Callahan LF, et al. Physical therapy vs. internet-based exercise training (PATH-IN) for patients with knee osteoarthritis: study protocol of a randomized controlled trial. BMC Musculoskelet Disord 2015 Sep 28;16:264 [FREE Full text] [doi: 10.1186/s12891-015-0725-9] [Medline: 26416025]

Abbreviations

eHealth: electronic health
CCT: controlled clinical trial
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
KOOS: Knee Osteoarthritis Outcome Score
MCIC: minimal clinical important change
NRS: Numerical Rating Scale
OA: osteoarthritis
OAK: osteoarthritis of the knee
QoL: quality of life
PA: physical activity
RCT: randomized controlled trial
SMD: standardized mean difference
VAS: Visual Analog Scale
WOMAC: Western Ontario and MacMaster Universities Osteoarthritis Index

Edited by G Eysenbach; submitted 27.11.17; peer-reviewed by R De Oliveira, D Bossen; comments to author 17.01.18; revised version received 07.02.18; accepted 11.02.18; published 26.04.18.

Please cite as:
Schäfer AGM, Zalpour C, von Piekartz H, Hall TM, Paelke V
The Efficacy of Electronic Health–Supported Home Exercise Interventions for Patients With Osteoarthritis of the Knee: Systematic Review
J Med Internet Res 2018;20(4):e152
URL: http://www.jmir.org/2018/4/e152/
doi:10.2196/jmir.9465
PMID:29699963

Viewpoint

Exploring the Role of In-Person Components for Online Health Behavior Change Interventions: Can a Digital Person-to-Person Component Suffice?

Sara Santarossa[1], MHK; Deborah Kane[2], PhD; Charlene Y Senn[3,4], PhD; Sarah J Woodruff[1], PhD

[1]Department of Kinesiology, University of Windsor, Windsor, ON, Canada

[2]Faculty of Nursing, University of Windsor, Windsor, ON, Canada

[3]Department of Psychology, University of Windsor, Windsor, ON, Canada

[4]Department of Women's and Gender Studies, University of Windsor, Windsor, ON, Canada

Corresponding Author:
Sara Santarossa, MHK
Department of Kinesiology
University of Windsor
401 Sunset Avenue
Windsor, ON, N9B 3P4
Canada
Phone: 1 5198198061
Email: santaros@uwindsor.ca

Abstract

The growth of the digital environment provides tremendous opportunities to revolutionize health behavior change efforts. This paper explores the use of Web-based, mobile, and social media health behavior change interventions and determines whether there is a need for a face-to-face or an in-person component. It is further argued that that although in-person components can be beneficial for online interventions, a digital person-to-person component can foster similar results while dealing with challenges faced by traditional intervention approaches. Using a digital person-to-person component is rooted in social and behavioral theories such as the theory of reasoned action, and the social cognitive theory, and further justified by the human support constructs of the model of supportive accountability. Overall, face-to-face and online behavior change interventions have their respective advantages and disadvantages and functions, yet both serve important roles. It appears that it is in fact human support that is the most important component in the effectiveness and adherence of both face-to-face and online behavior change interventions, and thoughtfully introducing a digital person-to-person component, to replace face-to-face interactions, can provide the needed human support while diminishing the barriers of in-person meetings. The digital person-to-person component must create accountability, generate opportunities for tailored feedback, and create social support to successfully create health behavior change. As the popularity of the online world grows, and the interest in using the digital environment for health behavior change interventions continues to be embraced, further research into not only the use of online interventions, but the use of a digital person-to-person component, must be explored.

(J Med Internet Res 2018;20(4):e144) doi:10.2196/jmir.8480

KEYWORDS

digital person-to-person; in-person; online intervention; behavior change; health, digital media; health care

Introduction

Background

Several aspects of the digital environment offer opportunity to support behavior change efforts, including reach, engagement, accessibility, collaboration and advocacy, and research potential [1]. Notably, there has been an increased interest from both public health organizations and those in academia, around using Web-based, mobile, and social media health behavior change interventions. It is believed that these popular digital media channels can play a valuable role in leveraging health messaging and consequently, behavior change. Although traditional face-to-face interventions or interventions with in-person components are (and continue to be) successful in health behavior change [2,3], traditional approaches can present with various barriers such as logistic problems, a challenge of keeping participants actively engaged, can be labor intensive, and

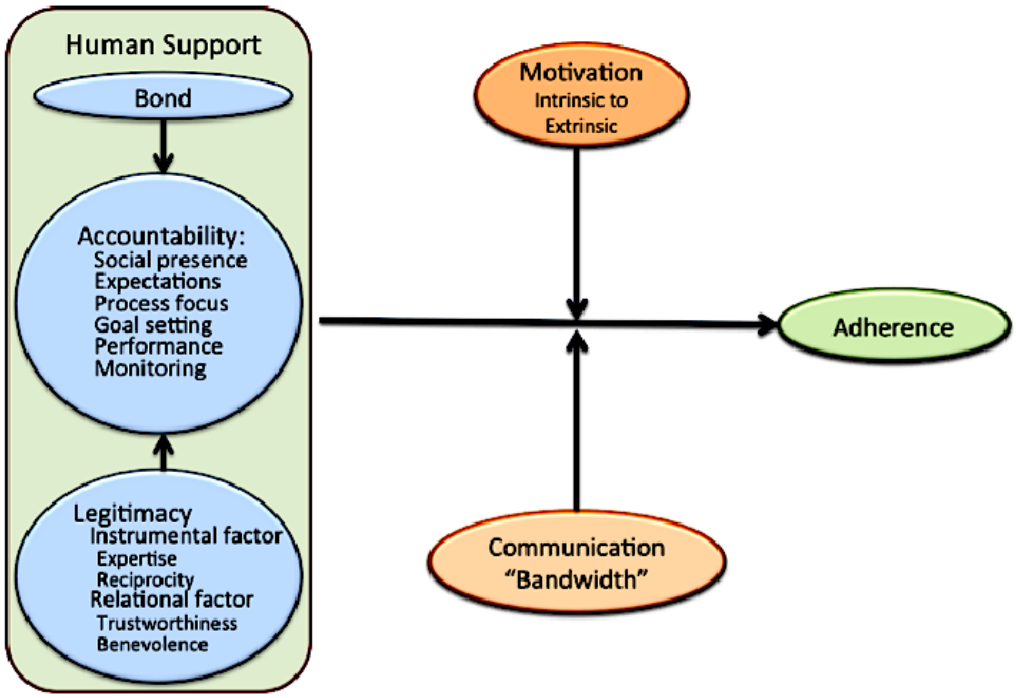
expensive to scale for larger populations. Using components of the digital environment may offer solutions to traditional challenges because of their low cost, high reach, anonymity, adaptability, and scalability [4]. Furthermore, comparisons of online interventions with traditional face-to-face interventions indicate that online treatment is generally at least as effective as conventional approaches and also possess several advantages [5-7]. Similarly, supplementary literature within the health behavior change domain has shown no significant treatment differences between the face-to-face and the online intervention groups [5-8], suggesting that online-only interventions may be just as valuable as face-to-face interventions. However, as a majority of online interventions are used in adjunct to traditional approaches [1], there is a need to understand what role in-person components play in online interventions. Furthermore, research indicates that the effectiveness of, and adherence to, online interventions is enhanced by human support [9-11]. As intervention adherence is important in predicting behavior change, the inclusion of a digital person-to-person component for an online behavior change intervention can help to combine the effectiveness and socialization opportunities of in-person meetings with the technologically enhanced active learning possibilities of the digital environment [1,5-8].

Perhaps, the dynamic, socially supportive, and interactive elements of digital media channels (ie, Web, mobile, and social media) may obviate the need for further interpersonal in-person components [6], as a digital person-to-person component can be used to cultivate a similar interpersonal connection, while overcoming the barriers of face-to-face interventions. Online human-supported interventions or digital person-to-person components have, in recent meta-analyses [12,13], obtained larger effect sizes than online self-guided programs, suggesting a need to further explore the role of the digital person-to-person relationship. For the purpose of this viewpoint paper, a "digital person-to-person" component will encompass any type of online feature that creates a sense of interpersonal connection or virtual interaction, thus, embodying qualities of a physical in-person or face-to-face components such as guidance, feedback, and support. For example, a digital person-to-person component can include human support provided through peers via online message groups or by posting or reading bulletins [14], as well as other elements of online systems that create social support such as chat forums and/or chat rooms [15]. Similarly, the digital person-to-person component can be offered on a one-to-one basis through virtual coaches, therapists, counselors, or facilitators via email, instant messaging sessions (eg, text message), and teleconferencing (eg, webcam and Skype). Digital person-to-person features create accountability, feedback, and social support, emulating traditional, physical in-person or face-to-face components and can foster motivation, encouragement, and commonality [16]. Furthermore, expectations about the digital person-to-person, such as accountability, feedback, and social support, may be grounded in social or behavioral theories including the theory of reasoned action [17] and the social cognitive theory [18], while being guided by the human support constructs of the model of supportive accountability (Figure 1) [11].

Objectives

Thus, the purpose of this viewpoint paper is to suggest that based on a comprehensive review of the literature, there is a need for a face-to-face component in Web-based, mobile, and social media health behavior change interventions but that a digital person-to-person component can foster similar results while dealing with challenges faced by traditional intervention approaches.

Figure 1. Model of supportive accountability (Mohr, Cuijpers & Lehman [11]).



Methods

To delineate the scope of the study and make it more replicable, the focus was on published references searchable through major bibliographic databases. This review adhered to the defining characteristics of Web-based, mobile, and social media health behavior change interventions, and this review excluded unpublished and untested programs. Due to the emerging state of Web 2.0 research, this viewpoint paper will not limit studies further by methodology, being inclusive of study design, participant, and setting. In particular, a wide variety of platforms were included in the search, including blogs and microblogging technologies, social networking sites, video sharing programs, and mobile health (mHealth) apps. Literature search strategies were developed using subject headings related to Web-based, mobile, and social media health behavior change interventions. Proquest Social Sciences, Web of Science, PsycINFO, Scopus, and PubMed were searched for "digital person-to-person," "in-person," "online intervention," "behavior change," "digital media," "health care," "social media," and "Web 2.0" from 2004 to April 2017. The search began with studies published since 2004 because that is when the term "Web 2.0" was coined to describe the shift to a more participatory online landscape. However, studies before 2004 were used to develop an understanding of the impact of in-person and/or traditional therapies or interventions. Finally, to ensure literature saturation, the reference lists of included studies or relevant reviews identified through the search were scanned.

Results

Using Digital Media Channels in Health Behavior Change Interventions

The digital environment consists of digital media, and although difficult to define, partly because it is ever changing, digital media in its broadest sense are content that can be transmitted over the internet or computer or phone networks [1]. Digital media channels such as the internet, mobile phones, and social media have become increasingly popular and have altered the nature of interactions around health issues. A once-passive one-way transfer of information, now it has become a network of multidirectional conversations [19]. The sense of interaction and multidirectional communication [20,21] offered by digital media channels cultivates active engagement [22-24] and information dissemination to a larger number of individuals [25]. On the basis of the popularity of these channels, an opportunity is presented to connect with individuals in their daily lives on issues concerning health and health behavior change [1].

Vast numbers of North Americans use the internet daily [26,27]. As presented in a meta-analysis of 5 papers, the literature presents substantial evidence that the use of Web-based interventions improves behavior change outcomes [28]. Web-based health interventions can be defined as "primarily self-guided intervention programs, delivered through a website, aiming to create positive change and/or improve or enhance knowledge, awareness, and understanding" [14]. Specific behavior change techniques of Web-based interventions may include real-time support, goal setting tools, alarms, reminders, and platforms to share with friends or family [1]. Particularly with health behavior change, Web-based interventions have seen several successful outcomes, including increased exercise time, increased knowledge of nutritional status, increased knowledge of asthma treatment, increased participation in health care, slower health decline, improved body shape perception, weight loss maintenance [28], and weight loss [6,29]. It has been proposed that structured (ie, lessons and activities) Web-based interventions are able to replicate health outcomes expected of a traditional, in-person intervention [30] and tend to be a more cost-effective approach [31]. However, other literature has shown that although Web-based interventions resulted in greater behavior change compared with control conditions (ie, waitlist or usual care), they had significantly less change compared with face-to-face interventions [32-34].

Particularly in North America [26,35], mobile phones are becoming a primary means of online access, as vast majority of individuals now own a mobile phone. Mobile phone are recommended as a good access point for a health behavior change intervention, since usage is high across various populations, including those considered to be underserved (ie, racial or ethnic minorities, youth, low and social economic status). [1,19,20,36]. mHealth interventions involve the use of mobile computing and communication technologies such as mobile phones, personal digital assistants, tablets, and portable media players to disseminate health information [37]. Subsequently, mHealth interventions are successful in creating health behavior change, as well as higher patient adherence, satisfaction, and acceptability than Web- or paper-based interventions [38]. Specifically, mHealth interventions have shown small but positive effects on weight loss behavior [39] and are a promising tool for decreasing risky sexual behaviors and drug use [40]. Furthermore, the use of tailored text messages as an adjunct to an in-person multidisciplinary weight management intervention resulted in improved feasibility, acceptance, and adherence [41]. The use of mobile phones offers health professionals an opportunity to engage with patients and colleagues on a scale when and where people are open to communicating and perhaps behavior change [42]. Future research on the effectiveness of text message delivery characteristics is needed to establish longer term intervention effects [43]. Moreover, the acceptance of mobile phones has helped to increase the popularity of online interactive platforms such as social media [44].

Social media are a broader concept that encompasses sites that allow users to generate and share content [21]. There are 6 main social media platforms, which include blogs, social networking sites, virtual worlds, online collaborative projects, content communities, and virtual game worlds [45]. As the use of social media continues to rise [26,46], it may indicate the potential for its role as a tool in the public health care system, specifically, health behavior change interventions. Social media platforms have been found to be successful in health behavior change interventions, with meta-analyses finding that the direction of effect for the primary outcomes favors interventions with social media components [45] and a slight positive effect of social networking site interventions on health behavior change

outcomes [47]. However, there is a lack of clear evidence of the effectiveness of social media in behavior change interventions [24,47,48], as most studies are not measuring an isolated effect of social media, thus creating a lack of ecological validity [49]. Furthermore, challenges of social media can include the spreading of misinformation and privacy breaches [45], which might suggest that using social media alone may be insufficient to promote health [48].

Role of In-Person Components in Online Health Behavior Change Interventions

It has been suggested that using a face-to-face approach is the "gold standard" in behavior change interventions [50]. Face-to-face interactions have greater bandwidth (ie, the number of communication cues a medium can convey), and this can lead to a greater ability to complete tasks, better interpersonal relations, and greater social presence [14]. Combining the verbal, nonverbal, and contextual cues of face-to-face communication could be assumed to provide the richest source of information and perhaps most positively influence behavior change. Furthermore, in face-to-face interventions, the human support created by the in-person component offers the core of the intervention while simultaneously coordinating a relationship with the participant in a way that will efficiently promote the use of the interpersonal connection to continue in the intervention [11]. In contrast, online behavior change interventions separate the content of the treatment, which is provided in a standardized manner via a website, mobile device, or social media platform, from support provided by humans, which is often intended to increase adherence [51-53]. However, Web-based, mobile, and social media interventions have shown promising results in health behavior change, but a majority of the literature focuses on these digital media platforms used in adjunct to traditional approaches [1]. The combination of online and face-to-face interventions may be reflective of the argument that it should not be necessary for online interventions to prove more effective than face-to-face treatments but rather to provide close to equivalent benefits and outcome results [50], thus, implying that online interventions are meant to provide an alternative or adjunctive component to already well-established and highly effective face-to-face interventions.

As such, some literature suggests that the idea of having a combination of online and face-to-face components within an intervention or program is ideal [1,6,54]. For example, a weight loss study that used Facebook to provide social support between monthly in-person meetings found that engagement in the Facebook support groups was significantly associated with weight loss during the 4-month maintenance period of this study, even after adjusting for face-to-face meeting attendance [55]. Furthermore, in a meta-analysis that focused on Web-based interventions and weight loss, it was reported that additional weight loss occurred when Web-based interventions were used to supplement face-to-face interventions; however, substituting face-to-face interventions with Web-based interventions resulted in significantly less weight loss [56]. These findings suggest that when digital media channels are used in conjunction with traditional approaches such as in-person behavior change interventions, they tend to be beneficial components [57] and perhaps will not be as successful if used alone. Conversely, it is important to note that some reviews have concluded that a meta-analysis could not reliably detect the effectiveness of online interventions because of the heterogeneity of designs and the small number of comparable studies [58-60].

Only a few Web-based, mobile, and social media interventions have truly measured behavior change; overall, there is a lack of comprehensive evaluation [1]. In the limited studies that investigated solely online interventions, it was recommended that having an in-person component could increase engagement and allow participants to interact and get to know each other before expecting them to interact online [61]. In addition, online behavior change interventions have had higher fidelity (ie, actual usage or intended usage of the online component) when participants knew each other before recruitment [24]. Specifically to social media, considering an in-person component has been suggested because of the "stranger phenomenon" [49]. This is based on the idea that social media are currently being used for conversations and maintenance of existing relationships and thus not being used to cultivate new acquaintances (ie, strangers). Supplementing with a face-to-face meet up (ie, more traditional way of forming acquaintances) may help overcome this particular barrier of social media. In a small pilot intervention study utilizing Facebook, which included an optional face-to-face meeting of all participants, only 3 of 8 participants attended [16]. Moreover, similar rates of participation existed in those who did not attend the in-person meeting compared with those who did [16]. Thus, further research into the need and role of in-person meetings in online interventions is warranted.

Albeit, many online interventions have supplemented with some form of in-person meeting or counseling [4,6,62]. These face-to-face interactions can be time-consuming, inconvenient, and logistically challenging. Research suggests that the use of a virtual health coach or online communication with a counselor or facilitator can be just as effective in behavior change as an in-person interaction [63,64], with implications for cost savings. Similarly, online interventions offer a promising alternative to traditional peer interventions, home visits, and/or pediatric office-based strategies to promote healthful behaviors [16], as online participants can interact frequently and at their convenience, a pattern that facilitates engagement, retention, and delivers a high intervention dose at a low cost with minimal resources [20]. These digital media channels provide a mechanism for participants to receive new information instantaneously, obtain immediate personalized automated feedback, and interact within a virtual group network, while at the same time allow for flexibility around work or school schedules and childcare responsibilities [65]. In addition, the potential anonymity of an online intervention group and its faceless quality allows participants to feel valued for the strength of their contributions rather than being evaluated on their physical appearance or disabilities [66]. Participants are likely to feel empowered, and in a safe environment, where they are able to digest the information at their own pace and better use it to enhance behavior change efficacy. However, the effectiveness of and adherence to online interventions is enhanced by human support [9,10]; and thus, the positive findings of online interventions coupled with the drawbacks of

in-person components present an opportunity for the digital person-to-person.

Using a Digital Person-to-Person Component

For a digital person-to-person component to be considered and be successful, certain adjustments need to be examined. During in-person meetings, participants are able to view nonverbal communication cues, including body language and voice qualities. These nonverbal cues may not be as obvious in online interactions. Several steps can be taken to overcome this limitation and to maintain accurate and a more complete understanding between the participant and the digital person-to-person component. Strategies to overcome nonverbal cues absent in a digital person-to-person component may include extended wording, various stylistic procedures for emphasizing text and using emoticons [14]. Extending wording and verbal expressions can help clarify messages, and the use of emoticons can enrich messages by mimicking a missing intonation or gesture [14,67,68]. In addition, participants should be well aware of the fact that messages may be misunderstood, hence a need for more probing and clarifications than in face-to-face sessions [14]. Furthermore, although possible and effective in online communications, the expression of feelings is not as automatic as in in-person meetings or relationships. This means that the digital person-to-person component must consciously consider using words and expressions that might not be used in face-to-face contact, to communicate empathy, care, concern, and warmth toward participants [14]. Again, participants have to be aware that their feelings are not as obvious and vivid as they would be in a face-to-face meeting [14]. Overall, the possibility does exist that if thoughtfully executed, a digital person-to-person component can perhaps be leveraged to substitute in-person and face-to-face components of online behavior change interventions, while overcoming traditional barriers and maintaining a sense of interpersonal connectedness. Moreover, the possibilities about the digital person-to-person relationship and opportunities for a successful alternative to in-person meetings are grounded in theory.

Research has not only shown a positive effect of grounding behavior change interventions in theory [48], but it is suggested as a necessity [1,62]. Using a digital person-to-person component in the delivery of an online behavior change intervention allows one to incorporate the best features of in-person interaction and the live instruction to personalize learning, allow thoughtful reflection, and differentiate instruction from participant to participant across a diverse group of learners. Thus, using a digital person-to-person component may be grounded in social and behavioral theories such as the theory of reasoned action [17] and the social cognitive theory [18], while being guided by the human support constructs of the model of supportive accountability (Figure 1) [11].

The theory of reasoned action [17] predicts that norms of significant people in an individual's social circles (ie, subjective norms) have a strong impact on the influence in the individual's behavioral intentions. In the digital media literature, descriptive norms, which are similar to engaging in social comparison (ie, comparing if you should or should not engage in a behavior based on what others like you are doing), are found to be more powerful in behavior change than injunctive norms [48,62]. Moreover, digital media channels such as social media thrive off social comparison and can motivate user participation by a desire to belong [69,70]. Similarly, the social cognitive theory [18] predicts social learning by observation, which can take place in both online and offline social networks. For example, the use of virtual coaches or facilitators can provide positive reinforcement for participation and model desirable behavior outcomes (ie, photographs, videos) [16]. The social cognitive perspective of social support proposes that perceived support (ie, an individual's belief that he or she is well supported) leads to better coping skills and higher self-esteem [71]. In addition, the social cognitive theory encompasses the idea of social diffusion and innovation as the way new ideas or cultural practices are transmitted or reinforced throughout a society [18]. According to Bandura [18], social innovation and diffusion can only be reinforced by media (eg, digital media) but not innovated; it is influential people who create innovation. Bandura [18] also suggested that the more media dominate people's lives, the more they will learn from it and less from people. Hence, there is a need for a digital person-to-person element in online behavior change interventions. Finally, using the human support constructs of the model of supportive accountability [11], it is suggested that the role of the digital person-to-person component can improve adherence, and consequently behavior change.

The term accountability refers to the implicit or explicit expectation that an individual may be called on to justify his or her actions or inactions [72], and being accountable to someone other than oneself enhances motivation to continue with behavioral change. Thus, adherence is an important element to consider in the development of a behavior change intervention. In the model of supportive accountability [11], human support increases adherence through accountability to a virtual coach (ie, a digital person-to-person component) who is seen as trustworthy, benevolent, and having expertise. Adherence will be further enhanced when the relationship with the virtual coach is perceived as reciprocal, clear goals and expectations are defined, and coaches are clear about the accountability process [11]. Moreover, people respond more positively to accountability demands from a coach who is perceived as legitimate [11,73]. Throughout the literature, the use of a virtual health coach has shown positive results in both weight loss [70] and physical activity adherence interventions [74]. Creating social accountability can help individuals self-monitor and follow through on their goals. Furthermore, in the model of supportive accountability [11], there are several human support constructs that are identified (see Figure 1) as integral components to how accountability is cultivated and maintained.

The existence of another human, or social presence [11], can influence accountability and subsequently adherence to behavior change interventions. For example, research suggests that although automated systems that monitor and encourage adherence, such as email reminders, can improve adherence to online interventions, digital person-to-person support enhances adherence to a significantly greater degree [13,29,75]. Expectations of the desired behavior change also play an important part in adherence [11,72]. The more people understand

and agree with the fundamental justification for the expected behavior, the greater the compliance. Such expectations need to be not only known but also clear and process, not outcome, focused [11]. Expectations should be monitored; research has shown an important feature of self-monitoring for online weight loss interventions appears to be emailing daily food intake and energy expenditure journals to a weight loss counselor rather than keeping a private record [29,64,76]. Thus, implying further rationale in need for virtual coaches (ie, digital person-to-person) support. It should be noted that the aim of performance monitoring is to provide feedback, to inform that failure to meet goals provides opportunity for self-reflection and growth, and to establish that there are no negative consequences [11]. In addition, it is suggested that supplementing online interventions with feedback and communication components can be effective in creating or generating behavior change [15].

Although all online interventions require participants to act by themselves to some extent, the type and degree of feedback that can be offered by a digital person-to-person component can vary considerably [14], from very little (ie, minimal guidance or supportive feedback mechanism provided) to high (ie, delivery of adequate amounts of tailored feedback). Moreover, immediacy of response is dependent on which communication modality is being employed. Emails and forum postings generally provide delayed feedback, whereas chat room or instant messaging sessions, Skype, and webcam calls provide participants with immediate feedback. Notably, it appears that feedback can be effective whether delivered by the internet [43] or through specific channels such as the use of text messaging [20,48,62]. Use of text messages can allow for immediate feedback on the basis of their response [25], and throughout the literature, the use of text messaging has been found to be a successful behavior change technique [20,48,62,77]. Although differentiating in their degree of direct digital person-to-person contact, feedback channels create an opportunity to foster interpersonal relationships within interactive platforms and create improvements in users' knowledge, health behavior, clinical outcomes, and social supports [78].

As previously stated, aside from accountability, a virtual coach or online counselor or facilitator aids in creating the feeling of interpersonal connectedness and can provide feedback, which tends to be effective in supportive behavior change [16,75]. Studies have found small to medium effect sizes in internet interventions that incorporate communicative functions such as online advisors [61] and use of an online counselor, compared with no counselor, resulted in greater behavior changes [15]. Similarly, in a Web-based randomized controlled trial that used no counseling, computer-automated feedback (ie, automated tailored messages), or human email counseling (ie, weekly email feedback from a counselor), results indicated that participants who had received computer-automated feedback or human email counseling had better weight loss than those with no counseling [29]. Moreover, in a 3-arm randomized controlled trial, the Facebook Plus group (ie, text messaging, personalized feedback, online support person) had significantly greater weight loss than the Facebook alone and waiting list control groups [25]. Feedback and communication components such as virtual coaches or online facilitators can also help make up an online social support system. Social support networks play an important role in determining health outcomes [79], and as more and more individuals are spending time online, research must examine the role of online social networks and their contribution to health behavior change. In addition, future research must consider the age of the virtual coaches or online facilitators, years of work experience they may have, and their accessibility to digital platforms in the workplace, as these factors play a role in the self-efficacy and utilization of digital platforms in health education organizations [80]. For optimal results, appropriate training for these platforms should be provided to those who will be providing online social support [80].

Increasing social support for a behavior change intervention can be an effective way to enhance desirable outcomes in both traditionally delivered behavioral interventions [79] as well as those delivered online [28,81]. Although utilization and seeking behaviors have been higher in women [82], research suggests that social support may be the most important aspect of online behavior change interventions, as it is the highest predictor of behavior change [48]. Online interventions that do not include some form of social support have lower utilization rates and lack of behavior change [83]. Social support can be encouraged through online social networks. Online social networks, often facilitated through social media platforms and/or a virtual coach/facilitator, have the ability to create high levels of intimacy and immediacy, meaning that support is available despite members' distance from one another. These characteristics naturally lead to high levels of social support and allow participants to provide each other with social support interactions that are present in face-to-face delivery by adding the possibility of in-the-moment posts and responses [84]. In addition, participants in a study that utilized Facebook [16] were not only successful in supporting one another in a virtual group format, but after the online intervention, the participants reported becoming Facebook "friends." Evidently, this continued peer support and gained knowledge through digital person-to-person relationships could result in further behavior change.

Online social networks can fill a void in participants, as they increase the feelings of support and connectedness. It has been found that those who reported less baseline social support had lower dropout rates, as the online social network appeared to be filling a gap [49]. Thus, it should be of no surprise that online social networks can be leveraged to foster an online community [85], as many share their personal stories, struggles, or successes [16,24,69,86], fostering a sense of interpersonal connectedness. Moreover, this sense of community can also lead to cyber worlds or communities in which people who used to feel isolated now feel a sense of togetherness [87]. The potential anonymity of online communities is particularly important in cases where health topics may be considered "taboo" or sensitive [21,49,88]. Subsequently, it is important to assess and consider the amount and/or quality of received advice or emotional support provided in online social networks as stress and stigmatization around the health topic can be induced [82]. Online social networks appear to be a predominant component in altering social norms and health behaviors on a large, often times anonymous, and cost-effective scale. Therefore, researchers should examine strategies that will further develop online social support, which

can then be used to promote continued adherence and desirable behavior change in online interventions.

Discussion

Principal Findings

The growth of the digital environment provides tremendous opportunities to revolutionize health behavior change efforts. Using digital media channels such as Web-based, mobile, or social media in online behavior change interventions can facilitate enhanced communications, research, and education and allows for the generation of multidirectional dialogs [19]. Web-based interventions have used specific behavior change techniques (eg, real-time support, reminders, and sharing platforms) [1] to produce desired behavior change outcomes [6,28,29]. Moreover, the use of mobile technology and social media in delivering health behavior change interventions also produces successful outcomes [38-40,47], and as these platforms continue to rise in popularity, continued efforts should be made to use them in the health research and health advocacy. Overall, digital media channels can be a more cost-effective approach [31] and can have a greater impact on behavior change because of high reach, anonymity, adaptability, and accessibility [1,4] than traditional face-to-face interventions. Furthermore, online technologies have been able to replicate similar results as traditional, in-person interventions [5-8,30]. Thus, online interventions now offer a real alternative, or supplement, to traditional, face-to-face interventions [14]. Although many benefits of using online behavior change interventions are documented, mixed reviews still exist on the delivery of these digital media interventions, whether or not face-to-face interventions are better, and whether in-person components are necessary [36]; thus, future research is justified.

Conclusions

Overall, face-to-face and online behavior change interventions have their respective advantages and disadvantages (ie, differing degrees of broad reach capability, anonymity, levels of treatment efficacy, and cost) and functions (ie, individual clinical treatment vs public health prevention programs), yet both serve important roles. It is suggested that perhaps the best opportunity for behavior change can be facilitated when there is a combination of face-to-face and online components [1,6,54]. It is the view of the authors that *human support* is the most important component in the effectiveness and adherence of both face-to-face and online behavior change interventions [9-11]. Thus, thoughtfully introducing a digital person-to-person component to replace face-to-face interactions can provide the needed human support while diminishing the barriers of in-person meetings. As human support in face-to-face interventions combines verbal, nonverbal, and contextual cues during in-person communications, a digital person-to-person interaction must implement strategies to overcome these challenges (ie, extended wording, various stylistic procedures for emphasizing text and using emoticons) [14]. Furthermore, using a digital person-to-person component is rooted in social and behavioral theories such as the theory of reasoned action [17] and the social cognitive theory [18] and further justified by the human support constructs of the model of supportive accountability [11]. For example, social comparison, social diffusion and innovation, accountability and adherence, feedback, and perhaps most importantly, social support and connectedness can all be accounted for using digital technology. Therefore, it must be ensured that a digital person-to-person component creates accountability [29,64,76], generates opportunities for tailored feedback [15,20,48,62], and creates social support [84,87]—all key elements in producing successful behavior change. The digital environment is ever changing, and the potential for its use in health behavior change interventions has yet to be fully harnessed [21,24,44,47]. As the popularity of the online world grows and the interest in using the digital environment for health behavior change interventions continues to be embraced, further research into not only the use of online interventions but also the use of a digital person-to-person component must be explored.

Conflicts of Interest

None declared.

References

1. McGloin AF, Eslami S. Digital and social media opportunities for dietary behaviour change. Proc Nutr Soc 2015 May;74(2):139-148. [doi: 10.1017/S0029665114001505] [Medline: 25319345]
2. Carey KB, Scott-Sheldon LA, Elliott JC, Garey L, Carey MP. Face-to-face versus computer-delivered alcohol interventions for college drinkers: a meta-analytic review, 1998 to 2010. Clin Psychol Rev 2012 Dec;32(8):690-703 [FREE Full text] [doi: 10.1016/j.cpr.2012.08.001] [Medline: 23022767]
3. Paxton SJ, McLean SA, Gollings EK, Faulkner C, Wertheim EH. Comparison of face-to-face and internet interventions for body image and eating problems in adult women: an RCT. Int J Eat Disord 2007 Dec;40(8):692-704. [doi: 10.1002/eat.20446] [Medline: 17702020]
4. Bennett GG, Herring SJ, Puleo E, Stein EK, Emmons KM, Gillman MW. Web-based weight loss in primary care: a randomized controlled trial. Obesity (Silver Spring) 2010 Feb;18(2):308-313 [FREE Full text] [doi: 10.1038/oby.2009.242] [Medline: 19696764]
5. Kaldo V, Levin S, Widarsson J, Buhrman M, Larsen HC, Andersson G. Internet versus group cognitive-behavioral treatment of distress associated with tinnitus: a randomized controlled trial. Behav Ther 2008 Dec;39(4):348-359. [doi: 10.1016/j.beth.2007.10.003] [Medline: 19027431]

6. Micco N, Gold B, Buzzell P, Leonard H, Pintauro S, Harvey-Berino J. Minimal in-person support as an adjunct to internet obesity treatment. Ann Behav Med 2007 Feb;33(1):49-56. [doi: 10.1207/s15324796abm3301_6] [Medline: 17291170]
7. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
8. 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]
9. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
10. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
11. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
12. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
13. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
14. 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]
15. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938 [FREE Full text] [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]
16. Gruver RS, Bishop-Gilyard CT, Lieberman A, Gerdes M, Virudachalam S, Suh AW, et al. A social media peer group intervention for mothers to prevent obesity and promote healthy growth from infancy: development and pilot trial. JMIR Res Protoc 2016 Aug 02;5(3):e159 [FREE Full text] [doi: 10.2196/resprot.5276] [Medline: 27485934]
17. Fishbein M, Ajzen I. Predicting Changing Behavior: The Reasoned Action Approach. New York, NY: Taylor Francis Group; 2011.
18. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
19. Burke-Garcia A, Scally G. Trending now: future directions in digital media for the public health sector. J Public Health (Oxf) 2014 Dec;36(4):527-534. [doi: 10.1093/pubmed/fdt125] [Medline: 24468840]
20. Brusse C, Gardner K, McAullay D, Dowden M. Social media and mobile apps for health promotion in Australian Indigenous populations: scoping review. J Med Internet Res 2014 Dec 10;16(12):e280 [FREE Full text] [doi: 10.2196/jmir.3614] [Medline: 25498835]
21. 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;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
22. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
23. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
24. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
25. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31 [FREE Full text] [doi: 10.1002/oby.20232] [Medline: 23505165]
26. Statistics Canada. 2016. General social survey – Canadians at work and home URL: https://www.statcan.gc.ca/eng/survey/household/5221 [accessed 2018-02-23] [WebCite Cache ID 6xSGCb8gI]
27. Pew Research Center. Internet/broadband fact sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2018-02-23] [WebCite Cache ID 6xSG14g6V]
28. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]

29. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an internet weight loss program. Arch Intern Med 2006 Aug;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
30. Wing RR. Behavioral approaches to the treatment of obesity. In: Bray GA, Bouchard C, editors. Handbook of Obesity: Clinical Applications. 2nd ed. New York, NY: Marcel Dekker, Inc; 2004:147-167.
31. Krukowski RA, Tilford JM, Harvey-Berino J, West DS. Comparing behavioral weight loss modalities: incremental cost-effectiveness of an internet-based versus an in-person condition. Obesity (Silver Spring) 2011 Aug;19(8):1629-1635 [FREE Full text] [doi: 10.1038/oby.2010.341] [Medline: 21253001]
32. Grunenberg E, Lin J, Baumeister H. [Effectiveness of web-based psychological interventions for weight loss--a systematic review]. Rehabilitation (Stuttg) 2013 Jun;52(3):182-187. [doi: 10.1055/s-0033-1343489] [Medline: 23761207]
33. Kodama S, Saito K, Tanaka S, Horikawa C, Fujiwara K, Hirasawa R, et al. Effect of web-based lifestyle modification on weight control: a meta-analysis. Int J Obes (Lond) 2012 May;36(5):675-685. [doi: 10.1038/ijo.2011.121] [Medline: 21694698]
34. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:CD007675 [FREE Full text] [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
35. Pew Research Center. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-02-23] [WebCite Cache ID 6xSGH8Oe0]
36. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
37. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
38. Carter MC, Burley VJ, Cade JE. Development of 'My Meal Mate' - A smartphone intervention for weight loss. Nutr Bull 2013 Feb 07;38(1):80-84. [doi: 10.1111/nbu.12016]
39. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
40. Reback CJ, Grant DL, Fletcher JB, Branson CM, Shoptaw S, Bowers JR, et al. Text messaging reduces HIV risk behaviors among methamphetamine-using men who have sex with men. AIDS Behav 2012 Oct;16(7):1993-2002 [FREE Full text] [doi: 10.1007/s10461-012-0200-7] [Medline: 22610370]
41. Woolford S, Clark S, Strecher V, Resnicow K. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. J Telemed Telecare 2010;16(8):458-461 [FREE Full text] [doi: 10.1258/jtt.2010.100207] [Medline: 20959393]
42. Lefebvre C. Integrating cell phones and mobile technologies into public health practice: a social marketing perspective. Health Promot Pract 2009 Oct;10(4):490-494. [doi: 10.1177/1524839909342849] [Medline: 19809002]
43. Hall AK, Cole-Lewis H, Bernhardt J. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015;36:393-415. [Medline: 25785892]
44. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016 Feb;59:299-307 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]
45. Mita G, Ni Mhurchu C, Jull A. Effectiveness of social media in reducing risk factors for noncommunicable diseases: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev 2016 Apr;74(4):237-247 [FREE Full text] [doi: 10.1093/nutrit/nuv106] [Medline: 26946250]
46. Pew Research Center. Social media fact sheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2018-03-13] [WebCite Cache ID 6xtSmDkxH]
47. Laranjo L, Arguel A, Neves A, Gallagher AM, 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]
48. Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. J Med Internet Res 2015 Jun 11;17(6):e141 [FREE Full text] [doi: 10.2196/jmir.3662] [Medline: 26068087]
49. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: systematic review. J Med Internet Res 2013 Nov;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]
50. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: in review, in use, and into the future. Prof Psychol Res 2003;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
51. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
52. Mohr DC, Duffecy J, Jin L, Ludman EJ, Lewis A, Begale M, et al. Multimodal e-mental health treatment for depression: a feasibility trial. J Med Internet Res 2010 Dec 19;12(5):e48 [FREE Full text] [doi: 10.2196/jmir.1370] [Medline: 21169164]

53. 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]
54. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. Br Med J Open 2014 Feb;4(2):e003926 [FREE Full text] [doi: 10.1136/bmjopen-2013-003926] [Medline: 24525388]
55. Hales SB, Davidson C, Turner-McGrievy GM. Varying social media post types differentially impacts engagement in a behavioral weight loss intervention. Transl Behav Med 2014 Dec;4(4):355-362 [FREE Full text] [doi: 10.1007/s13142-014-0274-z] [Medline: 25584084]
56. Reed VA, Schifferdecker KE, Rezaee ME, O'Connor S, Larson RJ. The effect of computers for weight loss: a systematic review and meta-analysis of randomized trials. J Gen Intern Med 2012 Jan;27(1):99-108 [FREE Full text] [doi: 10.1007/s11606-011-1803-9] [Medline: 21805218]
57. Kofinas JD, Varrey A, Sapra KJ, Kanj RV, Chervenak FA, Asfaw T. Adjunctive social media for more effective contraceptive counseling: a randomized controlled trial. Obstet Gynecol 2014 Apr;123(4):763-770. [doi: 10.1097/AOG.0000000000000172] [Medline: 24785602]
58. 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]
59. Manzoni GM, Pagnini F, Corti S, Molinari E, Castelnuovo G. Internet-based behavioral interventions for obesity: an updated systematic review. Clin Pract Epidemiol Ment Health 2011 Mar 04;7:19-28 [FREE Full text] [doi: 10.2174/1745017901107010019] [Medline: 21552423]
60. Neve M, Morgan PJ, Jones PR, Collins CE. 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]
61. Joseph RP, Keller C, Adams MA, Ainsworth BE. Print versus a culturally-relevant Facebook and text message delivered intervention to promote physical activity in African American women: a randomized pilot trial. BMC Womens Health 2015 Mar 27;15:30 [FREE Full text] [doi: 10.1186/s12905-015-0186-1] [Medline: 25886945]
62. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
63. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. Weight loss on the web: apilot study comparing a structured behavioral intervention to a commercial program. Obesity (Silver Spring) 2007 Jan;15(1):155-164. [doi: 10.1038/oby.2007.520] [Medline: 17228043]
64. Hunter CM, Peterson AL, Alvarez LM, Poston WC, Brundige AR, Haddock CK, et al. Weight management using the internet a randomized controlled trial. Am J Prev Med 2008 Feb;34(2):119-126. [doi: 10.1016/j.amepre.2007.09.026] [Medline: 18201641]
65. Herring SJ, Cruice JF, Bennett GG, Davey A, Foster GD. Using technology to promote postpartum weight loss in urban, low-income mothers: a pilot randomized controlled trial. J Nutr Educ Behav 2014;46(6):610-615 [FREE Full text] [doi: 10.1016/j.jneb.2014.06.002] [Medline: 25069621]
66. Wright KB. Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. J Comput Mediat Commun 2005;10(3):JCMC1034. [doi: 10.1111/j.1083-6101.2005.tb00259.x]
67. Suler J. The psychology of text relationships. In: Kraus R, Zack J, Stricker G, editors. Online Counseling: A Handbook for Mental Health Professionals. San Diego, CA: Elsevier Science; 2004:19-50.
68. Suler J. Cybertherapeutic theory and techniques. In: Barak A, editor. Psychological Aspects of Cyberspace: Theory, Research, Applications. Cambridge, UK: Cambridge University Press; 2008:102-128.
69. Chou WY, Prestin A, Lyons C, Wen KY. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
70. Merchant G, Weibel N, Pina L, Griswold WG, Fowler JH, Ayala GX, et al. Face-to-face and online networks: college students' experiences in a weight-loss trial. J Health Commun 2017 Jan;22(1):75-83. [doi: 10.1080/10810730.2016.1250847] [Medline: 28060581]
71. Lakey B, Cohen S. Social support theory and measurement. In: Cohen S, Underwood LG, Gottlieb BH, editors. Social Support Measurement and Intervention. New York, NY, US: Oxford University Press; 2000:29-52.
72. Tyler TR. The psychology of legitimacy: a relational perspective on voluntary deference to authorities. Pers Soc Psychol Rev 1997;1(4):323-345. [doi: 10.1207/s15327957pspr0104_4] [Medline: 15661667]
73. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. J Med Internet Res 2012;14(1):e1 [FREE Full text] [doi: 10.2196/jmir.1629] [Medline: 22281837]
74. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]

75. McTigue KM, Conroy MB, Hess R, Bryce CL, Fiorillo AB, Fischer GS, et al. Using the internet to translate an evidence-based lifestyle intervention into practice. Telemed J E Health 2009 Nov;15(9):851-858. [doi: 10.1089/tmj.2009.0036] [Medline: 19919191]
76. Walters S, Neighbors C. Feedback interventions for college alcohol misuse: what, why and for whom? Addict Behav 2005 Jul;30(6):1168-1182 [FREE Full text] [doi: 10.1016/j.addbeh.2004.12.005] [Medline: 15925126]
77. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
78. Hamilton N, Bhatti T. 1996. Population health promotion: an integrated model of population health and health promotion URL: https://www.canada.ca/en/public-health/services/health-promotion/population-health/population-health-promotion-integrated-model-population-health-health-promotion.html [accessed 2018-02-23] [WebCite Cache ID 6xSGisrO9]
79. Kelsey K, Earp JA, Kirkley BG. Is social support beneficial for dietary change? A review of the literature. Fam Community Health 1997;20:70-82. [doi: 10.1097/00003727-199710000-00008]
80. Alber JM, Paige S, Stellefson M, Bernhardt J. Social media self-efficacy of health education specialists: training and organizational development implications. Health Promot Pract 2016 Dec;17(6):915-921. [doi: 10.1177/1524839916652389] [Medline: 27234984]
81. Anderson-Bill ES, Winett RA, Wojcik JR, Winett SG. Web-based guide to health: relationship of theoretical variables to change in physical activity, nutrition and weight at 16-months. J Med Internet Res 2011 Mar;13(1):e27 [FREE Full text] [doi: 10.2196/jmir.1614] [Medline: 21447470]
82. Rising CJ, Bol N, Burke-Garcia A, Rains S, Wright KB. Perceived stress in online prostate cancer community participants: examining relationships with stigmatization, social support network preference, and social support seeking. J Health Comm 2017;22:469-476. [doi: 10.1080/10810730.2017.1304471]
83. McConnon A, Kirk SF, Cockroft JE, Harvey EL, Greenwood DC, Thomas JD, et al. The internet for weight control in an obese sample: results of a randomised controlled trial. BMC Health Serv Res 2007 Dec 19;7(1):206-215. [doi: 10.1186/1472-6963-7-206] [Medline: 18093289]
84. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]
85. Veale HJ, Sacks-Davis R, Weaver ER, Pedrana AE, Stoové MA, Hellard ME. The use of social networking platforms for sexual health promotion: identifying key strategies for successful user engagement. BMC Public Health 2015 Feb 06;15:85 [FREE Full text] [doi: 10.1186/s12889-015-1396-z] [Medline: 25884461]
86. Pagoto SL, Baker K, Griffith J, Oleski JL, Palumbo A, Walkosz BJ, et al. Engaging moms on teen indoor tanning through social media: protocol of a randomized controlled trial. JMIR Res Protoc 2016 Nov 29;5(4):e228 [FREE Full text] [doi: 10.2196/resprot.6624] [Medline: 27899339]
87. Duffett-Leger L. Developing a Web-Based Approach for Promoting Cervical Health in Young Women [doctoral dissertation]. New Brunswick, Canada: University of New Brunswick; 2011.
88. Olson CM. Behavioral nutrition interventions using e- and m-Health communication technologies: a narrative review. Annu Rev Nutr 2016 Jul 17;36:647-664. [doi: 10.1146/annurev-nutr-071715-050815] [Medline: 27022772]

Abbreviations

mHealth: mobile health

Edited by G Eysenbach; submitted 14.07.17; peer-reviewed by A Burke-Garcia, S Muralidharan; comments to author 16.01.18; revised version received 23.02.18; accepted 25.02.18; published 11.04.18.

Please cite as:
Santarossa S, Kane D, Senn CY, Woodruff SJ
Exploring the Role of In-Person Components for Online Health Behavior Change Interventions: Can a Digital Person-to-Person Component Suffice?
J Med Internet Res 2018;20(4):e144
URL: http://www.jmir.org/2018/4/e144/
doi:10.2196/jmir.8480
PMID:29643048

Viewpoint

Social Media as a New Vital Sign: Commentary

Sean D Young[1], MS, PhD

University of California Institute for Prediction Technology, Department of Family Medicine, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Sean D Young, MS, PhD
University of California Institute for Prediction Technology
Department of Family Medicine
University of California, Los Angeles
10880 Wilshire Blvd
Los Angeles, CA, 90024
United States
Phone: 1 310 794 8530
Email: sdyoung@mednet.ucla.edu

Abstract

Mobile technologies, such as wireless glucometers and mobile health apps, are increasingly being integrated into health and medical care. Because patients openly share real-time information about their health behaviors and outcomes on social media, social media data may also be used as a tool for monitoring patient care. This commentary describes how recent advances in computer science, psychology, and medicine enable social media data to become a new health "vital sign," as well as actionable steps that public health officials, health systems, and clinics can take to integrate social data into both public and population health as well as into individual patient care. Barriers that first need to be addressed, including privacy concerns, legal and ethical responsibilities, and infrastructure support, are discussed.

(J Med Internet Res 2018;20(4):e161) doi:10.2196/jmir.8563

KEYWORDS

social media; big data; personal health records

Introduction

Given the current discourse around social media, it might seem like "wall postings" have only existed since the turn of the millennium. However, people's psychological desire to publicly share information has existed for thousands of years. For example, scribes, messengers, and other citizens of ancient Rome shared their thoughts on a literal wall—via graffiti—and elsewhere in Europe people exchanged pamphlets featuring diary entries, quotations, and personal information [1]. These snippets of daily life entertained and informed citizens, but they also provided a window into societal well-being. That is, they were a health indicator of sorts and served as one of the earliest forms of social media.

Today, unlike antiquity, health providers and researchers have access to an endless stream of personal data via the Internet and mobile devices. In one survey, 72% of Internet users searched for health information in the past year [2], and several studies have shown that social media can have a significant positive impact on health outcomes if it is properly incorporated into health care settings [3,4]. Given the tremendous amount of publicly accessible personal data (eg, >500 million tweets per day on Twitter), organizations and research teams are already starting to integrate social media services such as Twitter and Facebook into public health surveillance by tracking influenza, sexually transmitted diseases, and predicting crime more accurately than previous models [5-10].

Patients are willing to share highly personal information on social media, including thoughts, feelings, and behaviors that they typically would not disclose because of stigma or embarrassment [11,12]. Scientists have begun to search these data for important psychological clues about patient behaviors and outcomes, such as medication adherence and adverse reactions to medications [13,14]. By combining psychological insights with data science methods such as machine learning, it is possible to develop mathematical models that mine social media data to predict people's health behaviors at a population and individual patient level. Researchers are debating the best methods to interpret and gain behavioral insights about social "big data" [15-17], but it is now feasible to monitor a host of lifestyle factors that affect a person's mental health and physical

well-being, such as stress level, physical activity, and sexual orientation.

Real-Time Interventions

In a clinic or hospital, physiological vital signs such as blood pressure and heart rate are easy to monitor, but observant physicians also take note of psychological indicators such as sadness or anxiety. If physicians could detect when a patient is anxious, or why they are scared to take medications, for example, it might be possible to provide real-time interventions and improve care delivery.

Consider depression, the leading cause of disability worldwide. Depression costs the US health care system approximately US $200 billion annually and leads to negative health outcomes such as reduced medication adherence and suicide [18]. Although real-time monitoring of depression could save money and lives, it has been difficult to track outcomes outside of a clinic. This has changed with the advent of social media. Now, researchers can work together to analyze the words and images that people organically share on social media sites for clues about depression, drug abuse, or other psychiatric disorders. Algorithms can quickly identify patterns within millions of postings and use that information to predict people's emotions and behaviors [19]. A check for suicidal ideation, for example, could be as simple as reviewing a list of keywords associated with suicide and depression [20,21]. However, despite the rapid progress in algorithm development over the years, there is still room for improvement. Algorithms implemented in a health care setting might need to apply Natural Language Processing techniques and require extensive additional research to optimize their accuracy.

While it will take time and research to develop automated algorithms capable of accurately identifying health risks within social data, we can already discuss more manual ways of implementing this approach, such as having trained staff available to review and discern whether social media posts that have been flagged by a machine for at-risk content actually do describe a patient's valid health concerns and require attention. In fact, the process of having human domain experts identify and confirm health-related content, and teach this information to a machine, is a common approach used in developing new models for monitoring and prediction [17].

With the right infrastructure and clear and informed consent from patients, it might be possible for people's real-time, organic social media data to be used as a source of information about their health and vitals. New health care tools could be developed and used based on this information to help alert health systems and their staff about potential or ongoing patient risks. How can this be done? Health systems could ask patients for permission to review their social media history, along with consent to monitor online activity when admitted to a hospital or medical facility. If a patient were to repeatedly express patterns of concern, such as talking about "pain," "being depressed," or "suicide," machines that mine this information could flag it for a health system staff member to review and intervene if deemed necessary.

Similar approaches to mine social media data can also be applied at the public health population level to help researchers and public health officials identify epidemics. This has already been the case for HIV [12,22], influenza [23], and cardiovascular disease [24]. With some refinement, social media surveillance strategies may be able to inform early warning systems and facilitate communication between doctors and public health authorities. Once a patient, population group, or geographic region of concern is identified, providers and health officials would need to follow a strict protocol to maintain privacy and confidentiality.

Roadblocks

Once equipped with these tools, providers and public health officials could have real-time access to monitor high-risk patients. However, there are a number of implementation questions that first need to be resolved. One question is whether providers and public health officials would be responsible for acting on social media data, as when providers identify a patient with high blood pressure. Because social media postings occur continuously 24 hours a day, 7 days a week, inside and outside of hospitals and clinics, this could pose a tremendous legal and economic burden on health care facilities to monitor and act on this information. For example, ethicists, lawyers, and health system management would need to discuss whether, how, and when providers should monitor and intervene. For example, if a patient posts on the social media platform of their cardiologist about a heart problem, would the cardiologist be responsible for following up? If the patient did not discuss cardiology but instead described mental health problems, an area outside of cardiology, would the responsibilities of the cardiology team change? These are some of the many unexplored questions that would need to be discussed. Before implementation of monitoring patient social data, there should be clear and ethical protocols in place to manage these situations such as having a mental health professional on-call to deal with potential crises [25].

Given that there are a limited number of studies showing a causal link between social media posts and actual health outcomes due to the novelty of this area, we suggest that it is too early to assign responsibility for monitoring social media as a health outcome. Instead, at this stage it might be used as an additional tool, instead of a standalone one, for remote monitoring of health and additional insights about patients.

Using social media to monitor and predict health risks also raises serious privacy concerns. For example, are patients willing to have their data mined to monitor their health information? Although studies on ethics suggest that people find social media to be an acceptable tool for use within public health and medicine [26,27], this is a constantly changing area. It is unclear whether people who share their data today would be willing to have their data used in the future. A recent systematic review shows that there are mixed views on the ethics of social media research, with privacy concerns being the biggest hurdle. Despite the public nature of social media, many people expressed more positive views about social media research when studies required informed consent from participants [28]. Another concern is

that some people using social media may have their profile set to public and be unaware of the privacy risks involved or have forgotten that their profiles are public. It is unclear whether these or other individuals are comfortable sharing personal health information with a health care provider. Hospitals and health care systems that implement a method of following patients on social media should therefore have consent forms that explicitly remind patients when and if they would be monitored and describe the currently known risks and benefits of this approach.

Another issue to consider is the data storage and management system. Social data could be managed within an electronic health records system or managed in a separate system, such as by clinics or health systems. Pilot testing is needed to address these questions before rolling out implementation at the larger level of electronic health records or health systems. For example, feasibility studies with small clinics will help uncover potential risks that need to be addressed before widespread implementation

Finally, there is also concern around the validity of the data. People can provide inaccurate information about their health and behaviors on social media. For instance, during the 2011-2012 flu season, flu-related Twitter models suggested that it was a typical season. However, after World Health Organization data showed that flu prevalence peaked 3 months later than a typical flu season, it was discovered that the Twitter flu model had incorrectly characterized tweets as being related or unrelated to flu. More research is therefore needed to ensure the accuracy of monitoring illnesses via social media, especially during atypical seasons [29].

The Future of Social Media and Health Care

Finally, before implementing social media as a tool for monitoring data, we need to address whether providers and health systems might not want or know how to handle this additional information about their patients. With the increasing amount of data available, it can be overwhelming to a health system to think about integrating an unfiltered, 24-hour source of information into health services and even more daunting to think about the possible risks that could occur from being one of the first to implement an approach like this. This paper is meant to initiate a discussion of the risks and benefits of this approach in order to determine whether and how social media might be integrated into health monitoring to improve public health and medical care. Although widespread implementation of social media monitoring in health care may be a number of years away, there are already case studies that use social media and other forms of big data as a tool for monitoring in other fields.

Tools to analyze social data are already successfully being used in fields like consumer behavior, education, and crime prediction [30], and there are a growing number of platforms that will soon be ready for wide-scale use by medical professionals. For example, our team at the Institute for Prediction Technology has been developing a visualization tool called Cloudberry-HIV Map [31]. We are building visualization tools like this to mine social media data and present this information to public health experts in an easy-to-understand manner so they can access “big social data” and use it to assist with the interventions and resource allocations. We have been working on applying these types of models to address HIV, natural disasters, drug addiction prevention, and suicide prevention. To date, the most extensive use of machine learning for public health in the private sector is IBM’s Watson Health, a computer that mines patient data to aid in diagnosis and treatment. In 2011, researchers started feeding the supercomputer millions of academic research papers and programmed it to understand standard treatment protocols for cancer. In one case, IBM claims that Watson Health proved better than doctors at diagnosing a rare form of leukemia [32]. Watson Health has gone through extensive testing and refinement, but it has been effective at analyzing patient medical records, identifying cancer treatment options, and providing supporting evidence for its treatment recommendations. Developing a supercomputer that can be used in a broader fashion with social media (eg, to identify early forms of mental illness) is a next step. At a minimum, it is currently possible to train computers to comb social media accounts and identify patients at risk for diseases such as depression [33]. Working with a multiplatform, rapidly evolving technology may indeed be intimidating, but research in this area suggests it could have tremendous benefits to patients and to broader public health. For this to happen, health providers, patients, and computer scientists will need to start a dialogue. The Health Insurance Portability and Accountability Act and protected health information guidelines are an important topic of discussion. Particular focus needs to be applied to whether public social media posts should be considered protected health information. In addition, infrastructure to support these tools and patient willingness to participate are a few other considerations. In some sectors this dialogue has already begun, but to make progress we need to shift our mindset on the role of social networking technologies. As health researchers and providers, being open to thinking of social media in this manner—as a potential new “vital sign”—is a wise and worthwhile change in perspective.

Conflicts of Interest

None declared.

References

1. Standage T. Writing on the Wall: Social Media - The First 2,000 Years. New York, NY: Bloomsbury; 2013.
2. Fox S. The social life of health information. 2014. URL: http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2018-01-30] [WebCite Cache ID 6wrn7f4t3]

3. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
4. Japhet G. Six ways social media can improve your health. 2014. URL: https://www.weforum.org/agenda/2014/06/social-media-health/ [accessed 2018-01-30] [WebCite Cache ID 6wrnNWc7S]
5. Young S, Mercer N, Weiss R, Torrone E, Aral S. Preventive Medicine. 2017 Dec 24. Using social media as a tool to predict syphilis URL: http://www.sciencedirect.com/science/article/pii/S0091743517305030 [accessed 2018-04-12] [WebCite Cache ID 6ydbwuTDf]
6. Gerber MS. Predicting crime using Twitter and kernel density estimation. Decision Support Systems 2014 May;61:115-125. [doi: 10.1016/j.dss.2014.02.003]
7. Bodnar T, Salathé M. Validating Models for Disease Detection Using Twitter. 2013 Presented at: WWW '13 Companion Proceedings of the 22nd International Conference on World Wide Web; May ; Rio de Janeiro, Brazil. ACM. New York, New York.?702; 2013; Brazil p. 13-17 URL: https://dl.acm.org/citation.cfm?id=2488027 [doi: 10.1145/2487788.2488027]
8. Wang X, Gerber M, Brown D. Automatic crime prediction using events extracted from twitter posts. In: Social Computing, Behavioral-Cultural Modeling and Prediction. College Park, MD: Springer; 2012:238.
9. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
10. Driscoll K, Walker S. Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data. Int J Commun 2014 Jun 16;8:1745-1764 [FREE Full text]
11. Young SD, Cumberland WG, Lee S, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. Ann Intern Med 2013 Sep 03;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]
12. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
13. Cho JHD, Gao T, Girju R. Identifying Medications that Patients Stopped Taking in Online Health Forums. 2017 Presented at: IEEE 11th International Conference on Semantic Computing (ICSC); Jan. 30-Feb. 1, 2017; San Diego, CA p. 141.
14. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
15. Boyd D, Crawford K. Critical Questions for Big Data. Information, Communication & Society 2012 Jun;15(5):662-679. [doi: 10.1080/1369118X.2012.678878]
16. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
17. Young SD. Behavioral insights on big data: using social media for predicting biomedical outcomes. Trends Microbiol 2014 Nov;22(11):601-602 [FREE Full text] [doi: 10.1016/j.tim.2014.08.004] [Medline: 25438614]
18. Depression Associated with Lower Medication Adherence Among Patients with Chronic Disease. 2011. URL: http://www.rand.org/news/press/2011/05/10.html [WebCite Cache ID 6wrrxIzMG]
19. Choudhury M, Gamon M, Counts S, Horvitz E. 2013. Predicting Depression via Social Media URL: https://www.microsoft.com/en-us/research/publication/predicting-depression-via-social-media/ [accessed 2018-04-12] [WebCite Cache ID 6ydcEarHf]
20. Burnap P, Colombo W, Scourfield J. Machine Classification and Analysis of Suicide-Related Communication on Twitter. USA: ACM; 2015 Presented at: Proceedings of the 26th ACM Conference on Hypertext & Social Media; 2015; Cyprus. [doi: 10.1145/2700171.2791023]
21. Garett R, Young S. Ethical Issues in Addressing Social Media Posts About Suicidal Intentions During an Online Study Among Youth: Case Study. JMIR Ment Health 2018 (forthcoming).
22. Young SD, Yu W, Wang W. Toward Automating HIV Identification: Machine Learning for Rapid Identification of HIV-Related Social Media Data. J Acquir Immune Defic Syndr 2017 Feb 01;74 Suppl 2:S128-S131 [FREE Full text] [doi: 10.1097/QAI.0000000000001240] [Medline: 28079723]
23. 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]
24. 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]
25. Garett R, Young S. Health Care Gamification: A Study of Game Mechanics and Elements. Tech Know Learn 2018 Jan 29:A. [doi: 10.1007/s10758-018-9353-4]
26. Fisher J, Clayton M. Who gives a tweet: assessing patients' interest in the use of social media for health care. Worldviews Evid Based Nurs 2012 Apr;9(2):100-108. [doi: 10.1111/j.1741-6787.2012.00243.x] [Medline: 22432730]
27. Chiu CJ, Menacho L, Fisher C, Young SD. Ethics issues in social media-based HIV prevention in low- and middle-income countries. Camb Q Healthc Ethics 2015 Jul;24(3):303-310. [doi: 10.1017/S0963180114000620] [Medline: 26059956]

28. Golder S, Ahmed S, Norman G, Booth A. Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review. J Med Internet Res 2017 Jun 06;19(6):e195 [FREE Full text] [doi: 10.2196/jmir.7082] [Medline: 28588006]
29. Mowery J. Twitter Influenza Surveillance: Quantifying Seasonal Misdiagnosis Patterns and their Impact on Surveillance Estimates. Online J Public Health Inform 2016;8(3):e198 [FREE Full text] [doi: 10.5210/ojphi.v8i3.7011] [Medline: 28210419]
30. Young SD. Social Media Will Help Predict Crime. 2015. URL: https://www.nytimes.com/roomfordebate/2015/11/18/can-predictive-policing-be-ethical-and-effective/social-media-will-help-predict-crime [WebCite Cache ID 6wrzR8cPp]
31. Cloudberry Demo. 2016. URL: http://predictiontechnology.ucla.edu/cloudberry-demo/ [accessed 2018-01-31] [WebCite Cache ID 6wrza6PSr]
32. Otake T. IBM big data used for rapid diagnosis of rare leukemia case in Japan. 2016. URL: http://www.japantimes.co.jp/news/2016/08/11/national/science-health/ibm-big-data-used-for-rapid-diagnosis-of-rare-leukemia-case-in-japan/#.WTCa5dy1uUm
33. Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Aug 8;6(1). [doi: 10.1140/epjds/s13688-017-0110-z]

Edited by N Collier; submitted 26.07.17; peer-reviewed by M Paul, A Sarker, J Shi, JD Sharpe; comments to author 06.12.17; revised version received 22.01.18; accepted 29.01.18; published 30.04.18.

Please cite as:
Young SD
Social Media as a New Vital Sign: Commentary
J Med Internet Res 2018;20(4):e161
URL: http://www.jmir.org/2018/4/e161/
doi:10.2196/jmir.8563
PMID:29712631

Review

Physical Activity, Sedentary Behavior, and Diet-Related eHealth and mHealth Research: Bibliometric Analysis

Andre Matthias Müller[1,2], PhD; Carol A Maher[3], PhD; Corneel Vandelanotte[4], PhD; Melanie Hingle[5], PhD; Anouk Middelweerd[6], PhD; Michael L Lopez[7], MUP; Ann DeSmet[8,9], PhD; Camille E Short[10], PhD; Nicole Nathan[11,12,13], PhD; Melinda J Hutchesson[14], PhD; Louise Poppe[8,9], MSc; Catherine B Woods[15], PhD; Susan L Williams[4], PhD; Petra A Wark[16], PhD

[1]Domain: Health Systems & Behavioural Sciences, Saw Swee Hock School of Public Healh, National University of Singapore, Singapore, Singapore
[2]Sports Centre, University of Malaya, Kuala Lumpur, Malaysia
[3]School of Health Sciences, University of South Australia, Adelaide, Australia
[4]Physical Activity Research Group, School of Health, Medical and Applied Sciences, Central Queensland University, Rockhampton, Australia
[5]Department of Nutritional Sciences, College of Agriculture and Life Sciences, The University of Arizona, Tucson, AZ, United States
[6]EMGO Institute for Health and Care Research, Department of Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, Netherlands
[7]Texas A&M AgriLife Extension Service, Texas A&M University, College Station, TX, United States
[8]Department of Movement and Sports Sciences, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium
[9]Research Foundation Flanders, Brussels, Belgium
[10]Freemasons Foundation Centre for Men's Health, Faculty of Health Sciences, University of Adelaide, Adelaide, Australia
[11]Priority Research Centre for Health Behaviour, School of Medicine and Public Health, The University of Newcastle Australia, Newcastle, Australia
[12]Hunter New England Population Health, Hunter New England Area Health Service, Newcastle, Australia
[13]Hunter Medical Research Institute, Newcastle, Australia
[14]Priority Research Centre in Physical Activity and Nutrition, School of Health Sciences, The University of Newcastle Australia, Newcastle, Australia
[15]Department of Physical Education and Sports Sciences, Faculty of Education and Health Sciences, University of Limerick, Limerick, Ireland
[16]Centre for Innovative Research Across the Life Course, Faculty of Health and Life Sciences, Coventry University, Coventry, United Kingdom

Corresponding Author:
Andre Matthias Müller, PhD
Domain: Health Systems & Behavioural Sciences
Saw Swee Hock School of Public Healh
National University of Singapore
Tahir Foundation Building
12 Science Drive, #10-01
Singapore, 117549
Singapore
Phone: 65 82997548
Fax: 65 67791489
Email: ephamm@nus.edu.sg

Abstract

Background: Electronic health (eHealth) and mobile health (mHealth) approaches to address low physical activity levels, sedentary behavior, and unhealthy diets have received significant research attention. However, attempts to systematically map the entirety of the research field are lacking. This gap can be filled with a bibliometric study, where publication-specific data such as citations, journals, authors, and keywords are used to provide a systematic overview of a specific field. Such analyses will help researchers better position their work.

Objective: The objective of this review was to use bibliometric data to provide an overview of the eHealth and mHealth research field related to physical activity, sedentary behavior, and diet.

Methods: The Web of Science (WoS) Core Collection was searched to retrieve all existing and highly cited (as defined by WoS) physical activity, sedentary behavior, and diet related eHealth and mHealth research papers published in English between January 1, 2000 and December 31, 2016. Retrieved titles were screened for eligibility, using the abstract and full-text where

needed. We described publication trends over time, which included journals, authors, and countries of eligible papers, as well as their keywords and subject categories. Citations of eligible papers were compared with those expected based on published data. Additionally, we described highly-cited papers of the field (ie, top ranked 1%).

Results: The search identified 4805 hits, of which 1712 (including 42 highly-cited papers) were included in the analyses. Publication output increased on an average of 26% per year since 2000, with 49.00% (839/1712) of papers being published between 2014 and 2016. Overall and throughout the years, eHealth and mHealth papers related to physical activity, sedentary behavior, and diet received more citations than expected compared with papers in the same WoS subject categories. The Journal of Medical Internet Research published most papers in the field (9.58%, 164/1712). Most papers originated from high-income countries (96.90%, 1659/1717), in particular the United States (48.83%, 836/1712). Most papers were trials and studied physical activity. Beginning in 2013, research on Generation 2 technologies (eg, smartphones, wearables) sharply increased, while research on Generation 1 (eg, text messages) technologies increased at a reduced pace. Reviews accounted for 20 of the 42 highly-cited papers (n=19 systematic reviews). Social media, smartphone apps, and wearable activity trackers used to encourage physical activity, less sedentary behavior, and/or healthy eating were the focus of 14 highly-cited papers.

Conclusions: This study highlighted the rapid growth of the eHealth and mHealth physical activity, sedentary behavior, and diet research field, emphasized the sizeable contribution of research from high-income countries, and pointed to the increased research interest in Generation 2 technologies. It is expected that the field will grow and diversify further and that reviews and research on most recent technologies will continue to strongly impact the field.

(J Med Internet Res 2018;20(4):e122) doi:10.2196/jmir.8954

KEYWORDS

science; telemedicine; exercise; health behavior; health resources; food; publications; movement; trends; Internet

Introduction

Being regularly active, having a less sedentary lifestyle, and consuming a healthy diet has many benefits for physical health, mental health, and well-being [1-3]. This is widely known, and the World Health Organization, the United Nations as well as many governments are committed to promoting these health behaviors [4]. Despite this, many people are not sufficiently active, are too sedentary, and/or do not adhere to dietary recommendations [5-7]. The negative consequences of the high prevalence of unhealthy behaviors are enormous for the individual, health care systems, and economies [1,6,8].

New technologies have been put forward as a cost-effective means to deliver behavioral health interventions and, as a result, prevent noncommunicable diseases (NCD) [9-12]. This is conceivable considering that the availability and personal use of information and communication technologies has increased significantly over the last two decades. Currently, 95% of the world population is covered by a mobile-cellular network and 84% is covered by a mobile-broadband network [13]. Although the Internet is still only accessible to 47% of the world population, access to the World Wide Web and smartphone usage across the globe is continuing to increase rapidly [13,14]. As such, there has been a rise in electronic health (eHealth) and mobile health (mHealth) related research for physical activity, sedentary behavior, and diet [10].

Thus far, eHealth and mHealth research related to physical activity, sedentary behavior, and diet has been summarized in several studies that focused on use and effectiveness of different technologies such as mobile phone and/or SMS (short message service) text messaging [15-18], digital games [19,20], the Internet [21-23], smartphone and/or tablet applications [24-28], social media [29], gamification features [30], and fitness trackers [31,32]. Other reviews in the field focused on specific populations such as children and adolescents [33,34], adults [35], older adults [36], overweight and obese adults [37,38], cancer survivors [39], patients with cardiovascular disease [40], and people residing in upper-middle, lower-middle, or low-income countries [41]. In addition, an international workshop addressed how eHealth and mHealth interventions should incorporate psychological theory and behavior change techniques in their design [42].

Although these studies summarized important aspects of the eHealth and mHealth research field related to physical activity, sedentary behavior, and diet, no attempt has been made to map out the entire field in a systematic manner. A bibliometric study uses publication-related information such as citations, journals, authors, and keywords to gain a bird's eye view of a field [43]. Bibliometric studies that summarized the research landscape in various fields have generated valuable insights [43-47] revealing the stage of maturity and growth of a research area, who and where the researchers are that drive the field, which journals are most prominent, and what kind of research is being conducted. This is especially useful in a relatively new research area such as eHealth and mHealth.

The purpose of this study is to examine the eHealth and mHealth research field related to physical activity, sedentary behavior and diet from its infancy until the end of 2016, and provide an overview of highly-cited papers which have considerably contributed to the maturation of the field. This will help researchers better position their work.

Methods

Search Strategy

We opted for using the Web of Science (WoS) Core Collection (Clarivate Analytics, USA) because it provides many bibliometric indicators and includes literature from most disciplines. We developed a search strategy in an iterative

manner starting from search terms used in published reviews and literature already known to us. We refined the search strategy by screening the titles of the most accessed papers listed at websites of journals that publish in the research field, and the titles of all publications of 6 researchers from different countries that are highly active in the field (see Multimedia Appendix 1).

The final search was conducted on April 26, 2017 to ensure all relevant papers that were published between January 1, 2000 and December 31, 2016 were registered in the WoS Core Collection. We used 146 search terms related to (1) physical activity, sedentary behavior, and diet and (2) use of technology (eg, smartphone, Web). Terms were combined with Boolean Operators ("OR" within the two search domains, "AND" between the two search domains). We restricted the search to publications in English and did not search for book chapters, conference proceedings, book citation indexes, and chemical indices (see Multimedia Appendix 2 for the full search strategy).

A second search using the same terms was conducted in which we only retrieved papers that WoS marked as "highly cited." WoS defines "highly cited" as being ranked within the top 1% compared with all other papers in terms of citation count in the same year and research field [48], suggesting highly-cited papers exert strong impact on the field.

The results of the two searches were exported to Microsoft Excel 2016 for screening.

Screening of Search Results

We included all journal papers on eHealth and mHealth research related to physical activity, sedentary behavior, and/or diet (including proxies, eg, weight management). They comprisedeHealth and mHealth intervention studies; papers on the components or characteristics of eHealth and mHealth (eg, use of theory in apps); papers on the relationship between technology use and the health behaviors, validation studies of consumer-based assessment tools (eg, Fitbit); and papers on the development of eHealth and mHealth interventions targeting physical activity, sedentary behavior, and/or diet. Reviews, protocols, editorials, commentaries, and original research papers were eligible to gain a comprehensive picture of the field. We excluded papers that were not related to the field (eg, biology papers); reported that technology was only used for data collection (eg, Web-based surveys) or the delivery of education without trying to change behavior (eg, nutrition science course); or were related to validation of research-grade assessment technologies (eg, ActiGraph accelerometers). The detailed screening guide is presented in Multimedia Appendix 3.

We had earlier piloted the screening procedure. Coauthors screened the same set of 20 papers (selected at random from preliminary searches) using a protocol that described the inclusion and exclusion criteria and a tutorial video. The video introduced the overall concept of a bibliometric study compared with a systematic review and detailed the inclusion and exclusion criteria with examples to illustrate how they should be applied [49]. Coauthors indicated whether they would include or exclude a paper or were unsure, while consulting the video, abstract, and full-text upon demand.

Seven trained coauthors (AMM, CAM, CV, MH, MLL, ADS, and PAW) each received a unique set of papers for title screening with optional screening of the abstract and full-text. As in the pilot phase, they chose "include," "exclude," or the option "unsure." Papers marked as unsure were screened by four of the authors (AMM, CAM, CV, and PAW) and discussed until consensus was reached.

Bibliometric Analysis

We computed the (compound) growth rate of publications over time. This was done by raising the ratio of the number of publications in 2016 over those in 2000 to the power of 1/16, after which we subtracted one and multiplied by 100:

We calculated the citation rate by dividing the number of citations per publication by the time since publication until December 2016, and expressed this per year. The citation rate does not depend on time since publication and is therefore a more precise measure of a paper's research impact than raw citation counts [50,51]. Because citation counts and citation rates are usually not normally distributed [47,52], we reported medians and interquartile ranges when studying their distributions.

Citation trends for physical activity, sedentary behavior, and diet related eHealth and mHealth research were studied between 2007 and 2016 because our analyses made use of published citation rates [53] that are only available over the most recent 10 years. We normalized the citation data for eligible papers by considering the WoS subject category and year in which a paper was published in two ways. First, we assessed the number of papers that occurred within each combination of WoS subject category and publication year among the papers included in our analysis. For each combination of WoS subject category and publication year separately, we multiplied the number of papers by the citation rate derived from the InCites Essential Science Indicator database on June 14, 2017 [53] to obtain the expected number of citations. After summing across WoS subject category, we obtained the total number of expected versus observed citations per year. Second, we compared the number of citations in each year and WoS subject category to corresponding published citation thresholds [53]. This yielded annual percentile scores that indicate the fraction of physical activity, sedentary behavior, and diet related eHealth and mHealth research articles within the top 10%, 20%, and 50% of all articles from the WoS subject categories represented by eligible papers in our search.

We explored the journals and authors who published most papers on eHealth and mHealth related to physical activity, sedentary behavior, and diet, along with the publication output of countries. We used WoS subject categories to count subject fields of papers. For the author analysis, we calculated 2 metrics using data within our dataset only: The *h*-index is the number of eligible papers of an author that were cited at least *h* times each (eg, an author with an *h*-index of 17 has at least 17 papers that were cited at least 17 times each) [54]. The *g*-index is the unique largest number of top cited eligible papers of an author that together received at least g^2 citations (eg, the 17 top cited

articles of an author with a *g*-index of 17 have at least 289 citations jointly) [55]. Countries were classified based on income as defined by the World Bank in 2017 [56].

To analyze the content of our dataset in more detail, we classified eligible papers into categories representing the studied exposure, technology, study population, setting, and methodology used. We did so by searching the title words and keywords identified by the author or by WoS editorial staff for occurrences of relevant terms. We defined the classification search terms using the agreed literature search strategy and the identified keywords in the eligible papers as starting point. The titles and keywords of the papers classified into each category were then double-checked by hand, as were those of all papers that were not classified into any or only a single category. This resulted in a refinement of the classification terms. This process was repeated until no inconsistencies were found. Using the final categorization reported in Multimedia Appendix 4, papers related to the Internet, (mobile) phone, SMS text messages, telehealth, and personal digital assistants were then classified as *Generation 1*, whereas papers on apps, wearable trackers, exergames, and social media were classified as *Generation 2*. Papers including both technologies were classified into *Generation 2*.

For the highly-cited papers, we also analyzed the papers based on their core content. These analyses were conducted independently by 2 coauthors (AMM and AM). They agreed on the descriptions of the paper's core content by also using NVivo 11 (QRS International Pty Ltd, Doncaster, Australia).

We conducted descriptive analyses using Microsoft Excel version 2016 and the Bibliometrix package version 1.7 [57] for R version 3.3.3 (Vienna, Austria) [58]. We used Stata/SE version 14.2 (College Station, TX, USA) for keyword analysis.

Results

Results of the Search

Figure 1 displays the flow of the search and screening procedure. The search resulted in 4805 hits. Of these, 336 were duplicates or conference contributions. Their exclusion led to 4469 papers to be screened. A total of 1712 papers were included in the final bibliometric analysis (Multimedia Appendix 5), 42 of which were highly cited.

Overall Trend

The number of papers on eHealth and mHealth related to physical activity, sedentary behavior, and diet increased steeply over the 17-year period (mean increase: 26% per annum). The period between 2014 and 2016 accounted for 49.00% (839/1712) of all papers (Figure 2).

The 1712 papers received 31,505 citations (median 7 per paper; interquartile range 18.5). Of the 1712 papers, 266 were not cited (15.54%), while 715 (41.67%) received 1 to 9 citations, 692 received 10 to 99 citations (40.42%), and 39 received 100 or more citations (2.28%). Overall, each paper received a median number of 2.0 citations per year (interquartile range 4.0).

Figure 1. Screening flowchart.

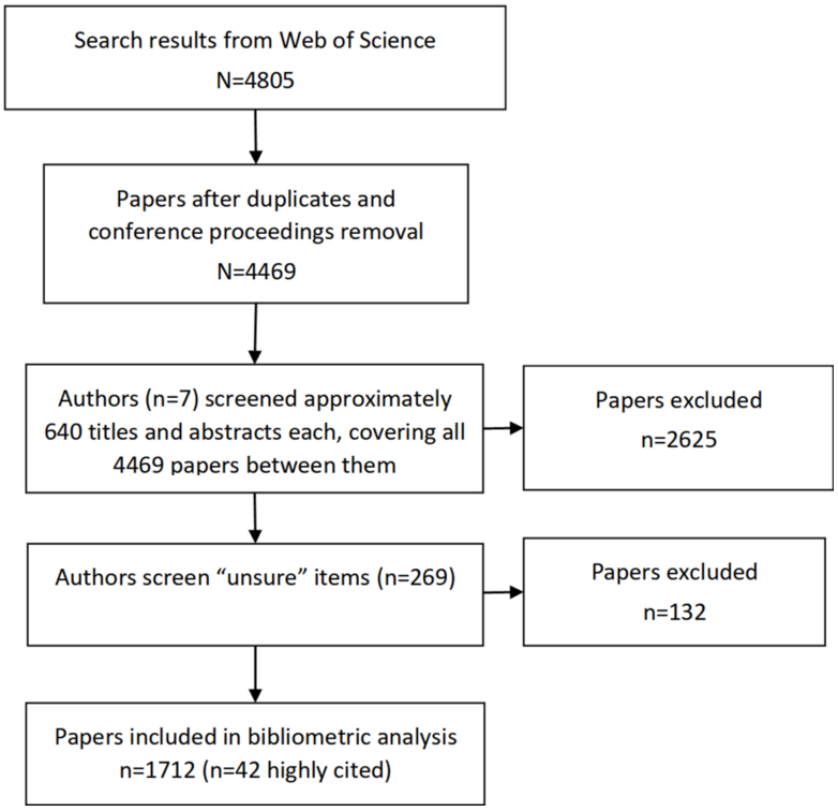
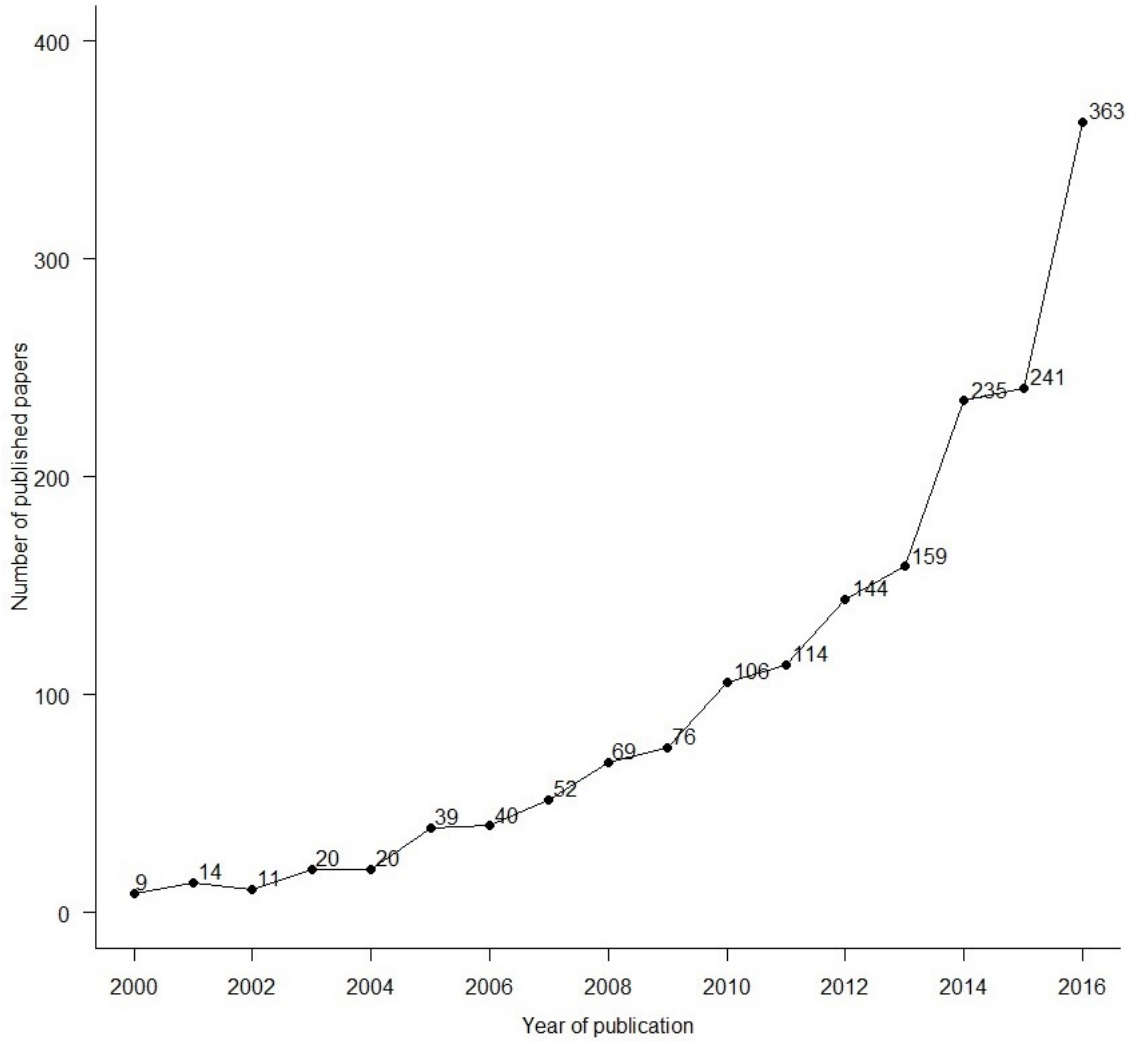


Figure 2. Publication output over time.



Compared with all papers from the same WoS subject categories, the absolute number of citations was higher than expected for included papers (see Multimedia Appendix 6 for the WoS subject categories of included papers) in all studied years (ie, between 2007 and 2016; results not shown). However, the ratio of the expected to observed citations declined from 2.6 (95% CI: 2.5-2.7) in 2007 to 1.8 (95% CI: 1.6-2.0) in 2016. Half of the eHealth and mHealth papers related to physical activity, sedentary behavior, and diet remained in the top 50% cited papers across same WoS subject categories. However, the proportion of papers in the top 20% declined from 0.6 to 0.3 between 2007 and 2016 (Figure 3).

Journals and Their Subject Categories

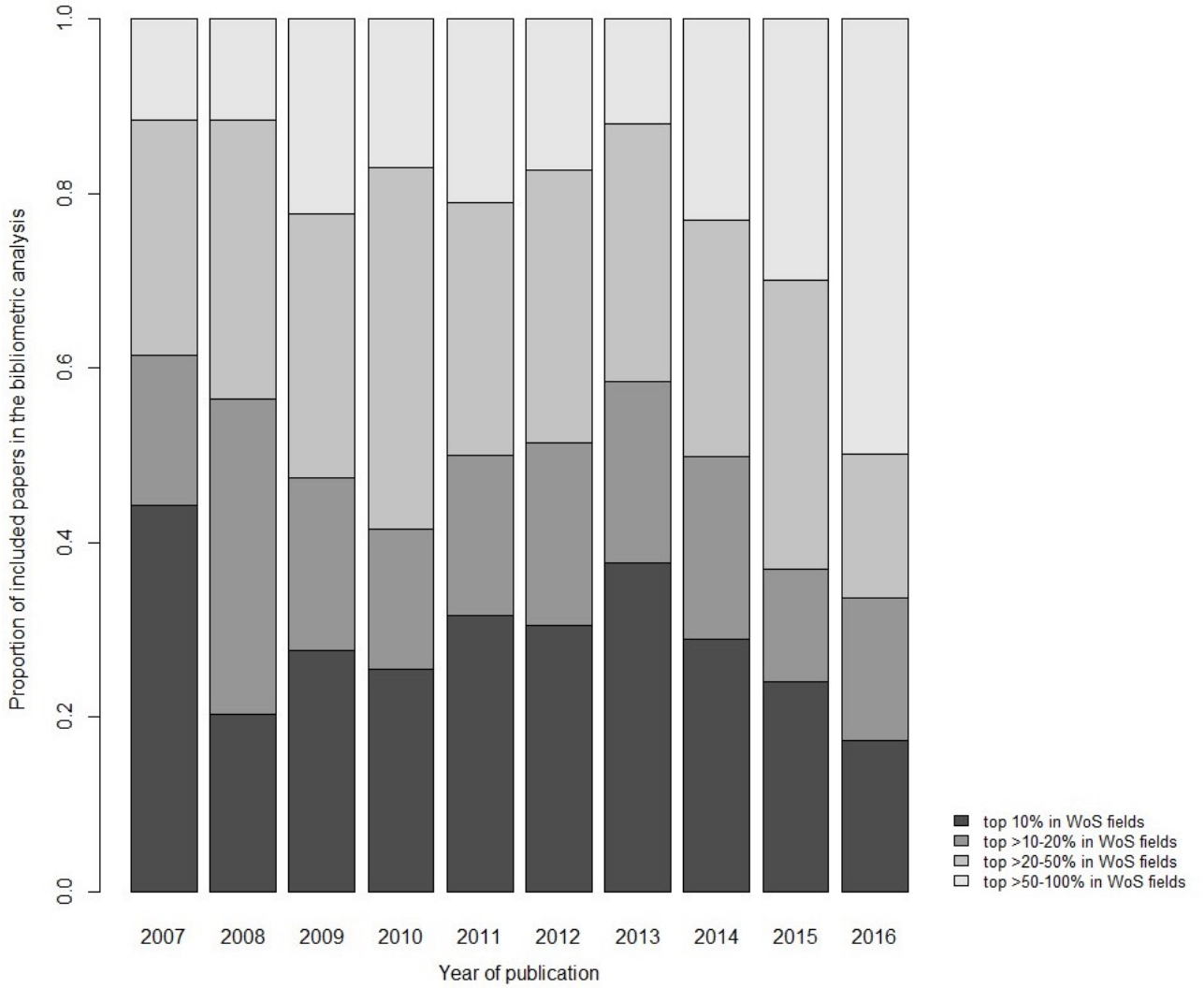
Overall, the papers were published by 471 different journals. As Table 1 shows, the *Journal of Medical Internet Research* published the most papers (9.58%, 164/1712 papers) followed by *BMC Public Health* (4.15%, 71/1712) and the *Games for Health Journal* (3.27%, 56/1712). The *Journal of Medical Internet Research* was also the highest cited journal and accounted for 13.48% of all citations in the field (n=4247 citations of 31,505 over the 17-year period). The *American Journal of Preventive Medicine* (9.37%, 2951/31,505) and *Annals of Behavioral Medicine* (4.64%, 1461/31,505) received the second and third highest number of citations, respectively.

In WoS, papers can be assigned to multiple subject categories. The papers included in this study were assigned to a total of 2797 WoS subject categories, of which 104 subject categories were unique. Table 2 shows the breakdown of subject categories present in the dataset, with only the top 10 categories shown. Throughout the years, the number of papers in journals from most fields gradually increased. However, the number of papers published in rehabilitation, health care science & services, health policy & services, as well as in medical informatics journals, increased more markedly from 2012. The number of papers published in psychology journals doubled between 2015 and 2016 (see Multimedia Appendix 6).

Authors

In total, 5654 authors contributed to the 1712 papers (median number of authors per paper 5, interquartile range 4). The top 10 authors (Table 3) contributed to 298 papers (17.41% of all papers). Vandelanotte C contributed to most papers (n=43) followed by Brug J (n=34), De Bourdeaudhuij I (n=31), and Oenema A (n=31). Vandelanotte C and Marcus B were in the top 10 of all characteristics listed in Table 3, and Brug J in all but first authorship papers.

Figure 3. Distribution of physical activity, sedentary behavior, and diet related electronic health (ehealth) and mobile health (mHealth) research papers that were in the top 10%, 20%, 50%, and 100% cited among all papers from the same Web of Science (WoS) subject categories by year of publication.



Countries

As the country of the corresponding author usually indicates where the research originated, we could also analyze the origin of the research that was published in the field. Corresponding authors were from 46 countries (see Multimedia Appendix 7). Most papers were published by authors from the United States (n=836, 48.83% of all published papers) followed by authors from Australia (n=195, 11.39%) and the Netherlands (n=125, 7.30%). Overall, 96.90% (n=1659) of all papers published were authored by researchers from high-income countries. Of the remaining 53 papers, 45 were published by authors from 9 upper-middle income countries: China (n=21), Malaysia (n=7), Iran (n=4), Brazil (n=3), Turkey (n=3), Thailand (n=2), Lebanon (n=2), Romania (n=2), and Mexico (n=1). Papers published by authors from lower-middle income countries accounted for only 0.005% of all papers and came from India (n=4), Pakistan (n=1), Nigeria (n=1), Egypt (n=1), and the Philippines (n=1). Similar patterns appeared when considering coauthorship or first authors instead of corresponding authors (data not shown).

Keywords

Authors specified a total of 2448 different keywords across papers. After also adding the keywords specified by WoS editors, this resulted in 4283 unique keywords. The number of keywords per paper varied widely (median 12, interquartile range 6). A total of 43 papers lacked specification of any keywords. The keywords that were most used reflected the exposure (eg, “physical-activity”), the general topic (ie, “health”), the study design (eg, “randomized controlled-trial”), or the population (eg, “adults”). Several of the keywords, including commonly used ones, were uninformative by themselves (eg, “risk,” “program”). We were able to use the keywords in combination with title words to classify the content of the paper into categories.

Table 1. Journals publishing most papers in physical activity, sedentary behavior, and diet electronic- and mobile health (eHealth and mHealth) research (top 20).

Journals	Rank based on total output	Papers published (N=1712), n (%)	Rank based on total citations received from any journal	Citation count (N=31,505), n (%)	Impact factor 2016[a]	5-year impact factor[a]
Journal of Medical Internet Research	1	164 (9.58)	1	4247 (13.48)	5.175	5.835
BMC Public Health	2	71 (4.15)	7	640 (2.03)	2.265	2.814
Games for Health Journal	3	56 (3.27)	21	307 (0.97)	2.019	2.242
JMIR mHealth uHealth	4	50 (2.92)	26	240 (0.76)	4.636	4.463
American Journal of Preventive Medicine	5	40 (2.34)	2	2951 (9.37)	4.020	5.412
International Journal of Behavioral Nutrition and Physical Activity	6	38 (2.22)	8	529 (1.68)	4.396	5.813
Journal of Nutrition Education and Behavior	6	38 (2.22)	10	597 (1.89)	2.491	2.439
Preventive Medicine	8	29 (1.69)	4	1038 (3.29)	3.434	3.703
Journal of Physical Activity and Health	9	27 (1.58)	36	176 (0.56)	1.946	2.400
Obesity	10	25 (1.46)	5	1032 (3.28)	3.873	4.358
Health Education Research	10	25 (1.46)	6	781 (2.48)	1.816	2.183
PLoS One	12	24 (1.40)	28	238 (0.76)	2.806	3.394
Telemedicine Journal and E-Health	12	24 (1.40)	37	174 (0.55)	2.031	2.141
Annals of Behavioral Medicine	14	23 (1.34)	3	1461 (4.64)	2.976	4.508
Computers in Human Behavior	14	23 (1.34)	46	121 (0.38)	3.435	4.252
Journal of Telemedicine and Telecare	14	23 (1.34)	17	372 (1.18)	2.008	2.371
JMIR Research Protocols	14	23 (1.34)	68	79 (0.25)	N/A	N/A
Patient Education Counseling	18	21 (1.23)	16	374 (1.19)	2.429	3.042
Translational Behavioral Medicine	19	19 (1.11)	33	221 (0.70)	2.989	2.883
American Journal of Health Promotion	20	17 (0.99)	18	361 (1.15)	2.586	2.280

[a]Obtained from InCites Journal Citation Reports (Clarivate Analytics).

Table 2. Number of papers published in journals within the top 10 leading Web of Science (WoS) subject categories. Each paper can be assigned to multiple WoS subject categories (according to the categories specified at journal level).

WoS subject category	Different journals within WoS subject category, n	Papers in journals within WoS subject category (N=2797), n (%)
Public, environmental & occupational health	70	457 (16.34)
Health care sciences & services	32	346 (12.37)
Nutrition & dietetics	46	247 (8.83)
Psychology	53	217 (7.76)
Medical informatics	19	214 (7.65)
Medicine	30	145 (5.18)
Education & education research	8	103 (3.68)
Endocrinology & metabolism	30	96 (3.43)
Rehabilitation	28	94 (3.36)
Health policy & services	12	78 (2.79)

Table 3. Top 10 most published authors in electronic health (eHealth) and mobile health (mHealth) physical activity, sedentary behavior, and diet related research in either number of papers, first authored papers, citations, *h*-or *g*-index.

Author	All papers, n (rank)	First authored papers, n (rank)	Citations, n (rank)	*h*-index[a] (rank)	*g*-index[a] (rank)
Vandelanotte C	43 (1)	13 (1)	1379 (3)	17 (2.5)	37 (1)
Brug J	34 (2)	2 (180)	1666 (1)	19 (1)	34 (2)
Oenema A	31 (3.5)	3 (71)	980 (5)	12 (9)	31 (3)
De Bourdeaudhuij I	31 (3.5)	2 (180)	788 (9)	16 (4.5)	28 (5)
Marcus B	29 (5.5)	7 (4)	1089 (4)	16 (4.5)	29 (4)
De Vries H	29 (5.5)	1 (764.5)	404 (31)	11 (13)	19 (12.5)
Thompson D	27 (7)	10 (2)	696 (13)	11 (13)	26 (6)
Collins C	25 (8.5)	4 (30)	618 (17)	11 (13)	24 (7.5)
Maddison R	25 (8.5)	5 (15)	404 (31)	10 (18.5)	19 (12.5)
Morgan P	24 (10)	3 (71)	669 (15)	13 (7)	24 (7.5)
Tate D	23 (11)	3 (71)	1512 (2)	17 (2.5)	23 (9)
Eakin E	21 (13.5)	7 (4)	869 (7)	13 (7)	21 (10.5)
Baranowski T	21 (13.5)	6 (8)	748 (10)	11 (13)	21 (10.5)
Owen N	15 (22.5)	0 (1677.5)	955 (6)	13 (7)	15 (21.5)

[a]Within our dataset only.

We classified 888 (51.87%) papers as studying *Generation 1* technologies and 742 (43.34%) papers as studying *Generation 2* technologies, with 82 (4.79%) papers being still unclassified. Before 2014, studies on *Generation 1* technologies were most common. From 2014 onwards, studies on *Generation 2* technologies were most common; their number steeply increased between 2013 and 2016. Within this period, the number of studies on *Generation 1* technologies increased less markedly (Figure 4). Vandelanotte C was the most common first author of papers on *Generation 1* technologies (n=11), followed by Harvey-Berino J (n=6). Gao Z was the most common first author of papers on *Generation 2* technologies (n=7), followed by Baranowski T (n=6).

Table 4 summaries the frequency of key study characteristics of the included papers. Physical activity was the health behavior most commonly targeted, followed by articles on weight and diet. Most studies targeted children or adolescents, while fewer focused on men and older adults. Multimedia and computer-based technologies (other than mobile apps) were most commonly studied, followed by studies focused on gamification or games, wearable technology or self-monitoring, or mobile apps or smartphones. Most studies were experimental trials, followed by reviews and/or meta-analyses. Few studies made use of creative or mixed methods.

Highly-Cited Papers

A table with all highly-cited papers can be found in Multimedia Appendix 8.

The 42 highly-cited papers received a total of 4883 citations (median 91, interquartile range 170) and were published between 2006 and 2016 across 19 journals. The *American Journal of Preventive Medicine* (n=13) and the *Journal of Medical Internet Research* (n=6) published most highly-cited papers. Corresponding authors were from nine countries, with US-based authors being the most common (n=20), followed by authors from Australia (n=8) and the United Kingdom (n=5). Overall, 226 authors contributed to the 42 highly-cited papers (mean 5.2 authors per paper) with Vandelanotte C (n=5) and Brug J (n=4) contributing to most highly-cited papers.

A systematic review and meta-analysis that reported on the link between intervention characteristics and intervention effectiveness on health behaviors (including physical activity and diet) in Internet interventions had the highest citation rate (79.5 citations per year) [59]. The authors included 85 studies and found that more extensive use of theory, a higher number of behavior change techniques and additional modes of communication (especially, text messages) were associated with larger effect sizes.

Of the 42 highly-cited papers, 20 were reviews of the literature, of which 19 used a systematic review approach, (4 of which also conducted a meta-analysis). A total of 13 papers reported data from primary studies and most of these were experimental trials such as randomized controlled trials (n=10); 8 studies reported content analyses of various smartphone apps, and 1 study introduced an eHealth and mHealth intervention development methodology. In terms of the technology used, 14 of the highly-cited papers studied *Generation 2* technologies, the majority of which related to social media, apps, or trackers (n=11). Most of the highly-cited papers were published in 2013 or later.

Figure 4. Number of published papers that studied Generation 1 technologies and Generation 2 technologies.

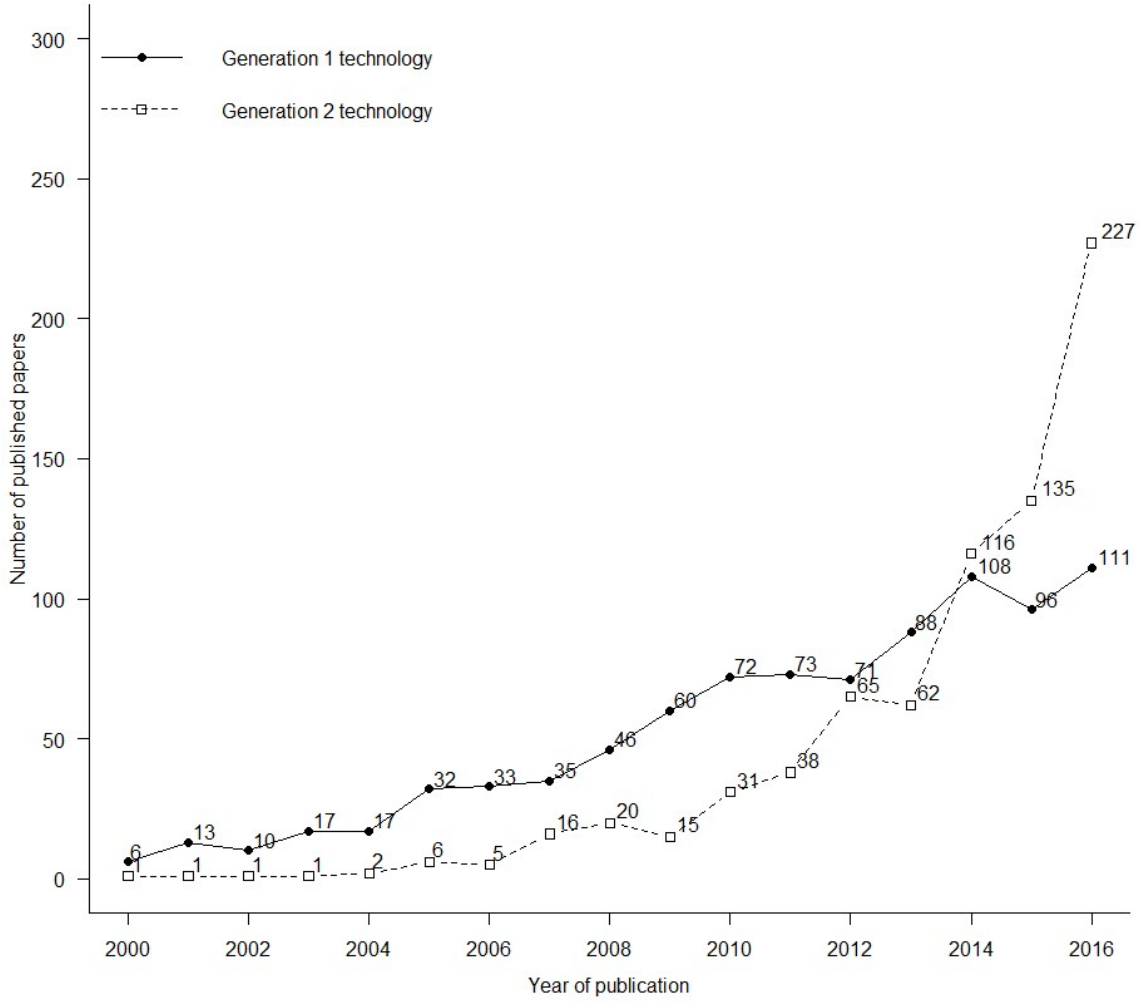


Table 4. Description of the physical activity, sedentary behavior, and diet related electronic health (eHealth) and mobile health (mHealth) research using expressions found in titles keywords of identified papers.

Category	Papers related to the category (as identified by its title or key words), n (%)[a]
Modifiable factors	
Physical activity	1236 (72.19)
Weight-related	859 (50.18)
Diet or nutrition	621 (36.27)
Sedentary behavior	169 (9.87)
Technology	
Multimedia and computer-based technologies other than mobile apps	906 (52.92)
Gamification or games	302 (17.64)
Wearable technology or self-monitoring	251 (14.66)
Mobile apps or smartphones	217 (12.68)
Telehealth	191 (11.16)
Text message	177 (10.34)
Social media or marketing	89 (5.20)
Populations	
Adults	555 (32.42)
Adolescents or youth	406 (23.71)
Children or infants	400 (23.36)
Older adults	188 (10.98)
Women	240 (14.02)
Men	12 (0.70)
Setting	
School or university	180 (10.51)
Workplace	67 (3.92)
Community	73 (4.26)
Low- or middle-income countries, or low-income settings	48 (2.80)
Family	37 (2.16)
Research methodology/focus	
Experimental trial	813 (47.49)
Review and/or meta-analysis	281 (16.41)
Qualitative study	224 (13.08)
Observational study	124 (7.24)
Costs (including cost-effectiveness and financial incentives)	77 (4.50)
Creative methods or designs	31 (1.81)
Mixed methods (including Delphi studies)	10 (0.58)

[a]Terms are not mutually exclusive. We used text search to obtain the above categorization, which resulted in a hit if any part of the title or the author-defined and WoS-defined keywords had a specific phrase or word (see methods section and Multimedia Appendix 4). The presented values should thus be used as good indicators rather than absolute values.

Discussion

Principal Findings

The purpose of this paper was to examine the entirety of the eHealth and mHealth research field related to physical activity, sedentary behavior, and diet using bibliometric data. We observed a substantial growth of research output in the field with most papers being published in recent years.

An exponential growth pattern has been observed across all research disciplines. For example, there is a 2.3% increase in scientific publications per annum leading to a doubling of the publication volume every 24 years [60]. If the overall growth rate we observed also applies to the future, we can expect the publication volume in the physical activity, sedentary behavior, and diet related to eHealth and mHealth research field to double about every 4 years. This strong growth of research output may reflect the fast development, wide availability, and increased functionality and importance of modern technology in people's daily lives. With this, opportunities to use these technologies to address behavioral health arise frequently. Across the world, researchers from various disciplines work on exploiting these new opportunities to understand and ultimately improve behavioral health. New intervention designs, methods, and analysis strategies are being developed [61-64]. We expect that these novel research initiatives will be widely disseminated in the scientific literature, which will likely lead to an increase in the research output. Compared with the entire body of research within the same WoS subject categories, eHealth and mHealth research related to physical activity, sedentary behavior, and diet was more frequently cited than expected. This likely signals an overall interest in the field.

The open access journals *Journal of Medical Internet Research* and *BMC Public Health* are the most popular outlets for researchers in this field. The *Journal of Medical Internet Research* was also the leading journal in an earlier bibliometric study that examined the overall mHealth literature [43]. The leading journals by citation count represent both open-access and nonopen access journals, with the *Journal of Medical Internet Research* receiving most and the *American Journal of Preventive Medicine* receiving the second most citations, respectively. Journals that publish open access enjoy a citation advantage in terms of speed of building up citations and overall citation count [65,66]. However, the size of this effect seems to be field-specific, which might explain why many nonopen access journals that publish physical activity, sedentary behavior, and diet eHealth and mHealth papers also accumulated a high number of citations [67,68]. The ranking of citation counts per journal is only a crude approximation of a journal's impact, as journals that publish more papers enjoy more opportunities to receive citations.

While eHealth and mHealth research related to physical activity, sedentary behavior, and diet is maturing rapidly in many high-income countries, the research output of non-high-income countries is still meager. Only 3% of all papers were from upper-middle or lower-middle income countries. None of these countries published many papers in the English language journals included in WoS. It is possible that more papers are published in a local language or in journals not included in WoS. Nonetheless, this observation is unsatisfying, considering that (1) fast globalization and urbanization in many upper-middle and lower-middle income countries is related to reduced physical activity and unhealthy diets, which is associated with an unprecedented rise in NCDs seen in many of these countries [5,69,70] and (2) the (mobile) technology infrastructure is improving rapidly, implying that technology could be used effectively in settings with limited health care resources [71]. Although research in high-income countries is important, the largest public health impacts could be generated in upper-middle, lower-middle and low-income countries where about 80% of the world population lives. There are signs that physical activity and diet research is slowly increasing in some of these countries [72]. This is promising because resulting findings can be used to address these health behaviors at scale [41,71,73]. However, barriers related to funding, prioritization, research capacity and infrastructure, and language need to be overcome.

Although research related to relatively older *Generation 1* eHealth and mHealth technologies (eg, Internet, SMS text messages) accounted for most papers in the field, its annual growth rate was 20.0% compared with 40.4% for research using *Generation 2* technologies such as smartphone apps and wearables. This might indicate that the interest in *Generation 1* technologies in the research field is slowly declining as researchers and funding agencies prioritize *Generation 2* technologies. The trend of using the newest technologies to address health behaviors is expected to continue, but whether these technologies have a meaningful and long-lasting impact on people's physical activity, sedentary behavior, and dietary habits needs to be seen [10]. Although, there are many arguments to be made for exploring very recent technologies, it is important to consider that these are currently often only available to a limited group of people. In addition, technologies may lose their appeal after a short time or are simply replaced by even newer technologies. These realities present barriers to achieving large-scale and sustainable public health impact with a specific technology.

We also found that most papers in the field were on physical activity compared with sedentary behavior and/or diet. That physical activity research is more common than research on dietary behaviors has been seen in the literature before [74,75]. Research on sedentary behavior is only a recent development as it was previously not distinguished from physical inactivity; a clear operationalization was only published recently [76]. Hence, to date, not many researchers have conducted eHealth and mHealth studies targeting sedentary behavior. Currently, only one review on eHealth and mHealth intervention studies targeting sedentary behavior exists [35].

We identified 42 highly-cited papers, which exert a strong influence on the field. These papers may be a resource for those less familiar with the field. The largest proportion of the highly-cited papers employed a systematic review approach to provide an overview of certain subfields. Systematic reviews and meta-analyses attract a high number of citations [77] as they are at the top of the evidence hierarchy in health-related subjects [78]. They are also important for researchers, policy

makers, and practitioners alike. Intervention studies, such as randomized controlled trials, were also highly cited. However, their influence may decrease when more such studies in a specific area (eg, SMS text messaging interventions) accumulate and systematic reviews become available.

Strengths and Limitations

The strength of this paper is that it provides a comprehensive overview of the eHealth and mHealth research field related to physical activity, sedentary behavior, and diet. To capture all relevant papers, we consulted the literature, discussed search terms, executed pilot searches, and refined our search before conducting the final search (with almost 150 search terms). We were similarly systematic in our paper-screening procedure. With this, we are confident that we have identified most eHealth and mHealth research papers related to physical activity, sedentary behavior, and diet.

We conducted a keyword analysis in combination with a title search to gain an insight into the studied exposures, technologies, populations, settings, and used methodologies. Because we did not confirm the obtained classification by retrieving the full-text, the number of papers for each category and subcategory we reported remains an estimate. However, this method allowed us to form an impression of the type of eHealth and mHealth research on physical activity, sedentary behavior, and diet that has been conducted between 2000 and 2016. Using keywords that are often used in a field is beneficial because others who consult databases to identify relevant research are likely to use these terms [79]. If authors use common keywords for their papers, database users are more likely to discover their work. This will also increase publication impact. Finally, in addition to providing an overview of the overall research in the field, we identified and analyzed highly-cited papers and highly published authors. This is important considering that, across all fields, only a small number of papers and authors determine the direction of a field [47,52].

Despite these strengths, limitations of our work need to be acknowledged. First, we only obtained papers from journals indexed by a single database—WoS. However, WoS is a large database that offers a wide variety of publication metrics that were vital for our analyses. Additionally, WoS only includes journals that meet certain criteria (eg, timely publishing, innovation, international diversity) to ensure high quality (more than 12,500 journals are currently indexed). Despite likely having excluded eligible papers in journals not included in WoS, we obtained papers from high-quality international journals that are the most influential source of scientific communication [52]. Using other databases such as Scopus could be explored in future studies. Second, we may have missed some papers that do not use informative keywords in the title as we did not search terminology used in abstracts. We did not review reference lists of eligible papers or their citations to identify any potentially missing papers. This, however, limited the (probably large) number of false positive results in our search. Third, we did not include gray literature (ie, conference proceedings, books, or other types of publications that are not journal papers), and we did not include papers published in languages other than English. Because of this, we may have missed relevant conference papers from fields such as human-computer interaction, computing, and engineering. Fourth, citation counts and their ranking should be interpreted with caution. Publication and citation habits vary between and even within fields [52]. The papers obtained for our bibliometric analysis were published in a variety of journals, and these journals are generally grouped under many subject categories (by WoS). Even though we considered the expected number of citations for each paper given its publication year and subject category when comparing citation trends over time, our approach remains an approximation: subject categories specified by WoS are assigned at journal level and may not reflect the field of every paper published in that journal. In addition, the citation counts we derived from WoS include self-citations, which may have influenced some of the rankings.

Finally, we have exclusively evaluated the scientific research literature and have identified trends that mainly concern scientific discovery, which will likely impact new research efforts. Therefore, the broader impact of the research outside of academia (eg, on public health, policy) cannot be deduced. Measuring and proving the societal impact of research is essential but difficult, mainly because this impact becomes usually only apparent in the far future and there are no agreed-upon measures to capture impact [80]. However, developments of measuring and analyzing impact beyond the scientific community are underway. One promising group of metrics that can be used are Altmetrics that measure the public engagement with research [81,82].

Conclusions

In this paper, we provided a bird's eye view of the research on eHealth and mHealth related to physical activity, sedentary behavior, and diet. Our analysis of 1712 papers published between January 2000 and December 2016 showed that research output is increasing rapidly; a trend that appears likely to continue. The *Journal of Medical Internet Research* was highlighted as the primary outlet for research in the field. Despite the many promising developments, research in upper-middle, lower-middle, and low-income countries is still scant. More research in such settings is needed to examine the public health impact of eHealth and mHealth interventions on physical activity, sedentary behavior, and diet where needed the most. Systematic reviews and papers that report on recent technologies (mainly smartphone apps) exert a strong impact on the field and their influence will likely remain high in the future.

Acknowledgments

CV is funded through a National Heart Foundation of Australia Future Leader Fellowship (ID 100427). CES is supported by a National Health and Medical Research Council ECR Fellowship (ID 1090517) awarded by the Australian Government. ADS

(FWO16/PDO/060, 12H6717N) and LP (11Z4716N) are supported by a Research Foundation Flanders grant. CM is supported by a National Health and Medical Research Council Career Development Fellowship awarded by the Australian Government.

Authors' Contributions

AMM, PAW, CV, and CM conceived the study, participated in its design and coordination, and drafted the manuscript. Data analyses were conducted by AMM and PAW. Articles were screened by AMM, PAW, CV, CM, MH, ADS, and MLL. All authors contributed to the writing of the manuscript, provided input on data analysis, and read and approved the final draft.

Conflicts of Interest

CV is listed in the manuscript as the author with the most publications in the field as well as most highly cited publications. His input was mostly conceptual, and he was only one of numerous contributors to the development of the search strategy and inclusion and exclusion criteria. He did not conduct any data analyses. He, therefore, did not willingly influence the results.

Multimedia Appendix 1

Journals and authors searched to refine our search strategy.

[PDF File (Adobe PDF File), 22KB - jmir_v20i4e122_app1.pdf]

Multimedia Appendix 2

Search strategy (Web of Science).

[PDF File (Adobe PDF File), 25KB - jmir_v20i4e122_app2.pdf]

Multimedia Appendix 3

Screening guide with inclusion and exclusion criteria.

[PDF File (Adobe PDF File), 32KB - jmir_v20i4e122_app3.pdf]

Multimedia Appendix 4

Classification of search terms within titles and keywords identified by the author and WoS editorial staff into categories for content analysis.

[PDF File (Adobe PDF File), 143KB - jmir_v20i4e122_app4.pdf]

Multimedia Appendix 5

File containing the included papers data from Web of Science.

[XLSX File (Microsoft Excel File), 4MB - jmir_v20i4e122_app5.xlsx]

Multimedia Appendix 6

Number of articles per subject category and publication year included in the bibliometric analysis.

[PDF File (Adobe PDF File), 109KB - jmir_v20i4e122_app6.pdf]

Multimedia Appendix 7

Publication output of countries.

[PDF File (Adobe PDF File), 33KB - jmir_v20i4e122_app7.pdf]

Multimedia Appendix 8

Highly-cited papers.

[PDF File (Adobe PDF File), 73KB - jmir_v20i4e122_app8.pdf]

References

1. Biddle SJ, Bennie JA, Bauman AE, Chau JY, Dunstan D, Owen N, et al. Too much sitting and all-cause mortality: is there a causal link? BMC Public Health 2016 Jul 26;16:635 [FREE Full text] [doi: 10.1186/s12889-016-3307-3] [Medline: 27456959]
2. Oyebode O, Gordon-Dseagu V, Walker A, Mindell JS. Fruit and vegetable consumption and all-cause, cancer and CVD mortality: analysis of Health Survey for England data. J Epidemiol Community Health 2014 Sep;68(9):856-862 [FREE Full text] [doi: 10.1136/jech-2013-203500] [Medline: 24687909]
3. Blair SN, Morris JN. Healthy hearts--and the universal benefits of being physically active: physical activity and health. Ann Epidemiol 2009 Apr;19(4):253-256. [doi: 10.1016/j.annepidem.2009.01.019] [Medline: 19344864]
4. World Heatlh Organization (WHO). Global status report on noncommunicable diseases 2014 URL: http://apps.who.int/iris/bitstream/10665/148114/1/9789241564854_eng.pdf?ua=1 [WebCite Cache ID 6tW7zibkY]
5. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, Lancet Physical Activity Series 2 Executive Committee. Progress in physical activity over the Olympic quadrennium. Lancet 2016 Sep 24;388(10051):1325-1336. [doi: 10.1016/S0140-6736(16)30581-5] [Medline: 27475270]
6. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
7. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 1;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
8. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Lancet Physical Activity Series 2 Executive Committee. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016 Sep 24;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
9. Rehman H, Kamal AK, Sayani S, Morris PB, Merchant AT, Virani SS. Using mobile health (mHealth) technology in the management of diabetes mellitus, physical inactivity, and smoking. Curr Atheroscler Rep 2017 Apr;19(4):16. [doi: 10.1007/s11883-017-0650-5] [Medline: 28243807]
10. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
11. Burke LE, Ma J, Azar KMJ, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
12. World Health Organization. Geneva: World Health Organization; 2011. mHealth new horizons for health through mobile technologies: Based on the findings of the second global survey on eHealth URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [WebCite Cache ID 6xzuC9O00]
13. International Telecommunication Union (ICT). ICT facts and figures 2016 URL: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf [WebCite Cache ID 6qaDn5Qza]
14. Statistica. Number of smartphone users worldwide from 2014 to 2020 2017 URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2017-05-19] [WebCite Cache ID 6qaFNKE7z]
15. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
16. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28 Suppl 2:1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
17. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
18. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
19. 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. [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
20. Peng W, Crouse JC, Lin JH. Using active video games for physical activity promotion: a systematic review of the current state of research. Health Educ Behav 2013 Apr;40(2):171-192. [doi: 10.1177/1090198112444956] [Medline: 22773597]
21. Bardus M, Smith JR, Samaha L, Abraham C. Mobile phone and Web 2.0 technologies for weight management: a systematic scoping review. J Med Internet Res 2015;17(11):e259 [FREE Full text] [doi: 10.2196/jmir.5129] [Medline: 26573984]

22. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
23. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
24. 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]
25. 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 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
26. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
27. 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;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
28. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: A systematic review and meta-analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
29. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
30. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS One 2017;12(3):e0173403 [FREE Full text] [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
31. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
32. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: a systematic review. BMC Public Health 2015;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]
33. Turner T, Spruijt-Metz D, Wen CK, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: a systematic review. Pediatr Obes 2015 Dec;10(6):403-409. [doi: 10.1111/ijpo.12002] [Medline: 25641770]
34. Lau PW, Lau EY, Wong DP, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011 Jul 13;13(3):e48-e49 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
35. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
36. Müller AM, Khoo S. Non-face-to-face physical activity interventions in older adults: a systematic review. Int J Behav Nutr Phys Act 2014;11(1):35 [FREE Full text] [doi: 10.1186/1479-5868-11-35] [Medline: 24612748]
37. 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]
38. Neve M, Morgan PJ, Jones PR, Collins CE. 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]
39. Kopp LM, Gastelum Z, Guerrero CH, Howe CL, Hingorani P, Hingle M. Lifestyle behavior interventions delivered using technology in childhood, adolescent, and young adult cancer survivors: a systematic review. Pediatr Blood Cancer 2017 Jan;64(1):13-17. [doi: 10.1002/pbc.26166] [Medline: 27468131]
40. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity ehealth interventions for people with cardiovascular disease: systematic review. J Med Internet Res 2017 Aug 02;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]
41. Müller AM, Alley S, Schoeppe S, Vandelanotte C. The effectiveness of e-& mHealth interventions to promote physical activity and healthy diets in developing countries: a systematic review. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109 [FREE Full text] [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]

42. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
43. Sweileh WM, Al-Jabi SW, AbuTaha AS, Zyoud SH, Anayah FM, Sawalha AF. Bibliometric analysis of worldwide scientific literature in mobile - health: 2006-2016. BMC Med Inform Decis Mak 2017 May 30;17(1):72 [FREE Full text] [doi: 10.1186/s12911-017-0476-7] [Medline: 28558687]
44. Andrade A, Dominski FH, Coimbra DR. Scientific production on indoor air quality of environments used for physical exercise and sports practice: bibliometric analysis. J Environ Manage 2017 Jul 01;196:188-200. [doi: 10.1016/j.jenvman.2017.03.001] [Medline: 28284941]
45. Müller AM, Ansari P, Ebrahim NA, Khoo S. Physical activity and aging research: a bibliometric analysis. J Aging Phys Act 2016 Dec;24(3):476-483. [doi: 10.1123/japa.2015-0188] [Medline: 26671908]
46. Şenel E, Demir E. A global productivity and bibliometric analysis of telemedicine and teledermatology publication trends during 1980–2013. Dermatologica Sin 2015 Mar;33(1):16-20. [doi: 10.1016/j.dsi.2014.10.003]
47. van Noorden R, Maher B, Nuzzo R. The top 100 papers. Nature 2014 Oct 30;514(7524):550-553. [doi: 10.1038/514550a] [Medline: 25355343]
48. Web of Science. InCites Essential Science Indicators URL: http://ipscience-help.thomsonreuters.com/incitesLiveESI/ESIGroup/indicatorsGroup/citationThresholds/thresholdHighlyCited.html [WebCite Cache ID 6sNew7gQp]
49. Müller AM. Youtube.: Study screening guide. Youtube; 2016. e- & mHealth in physical activity and diet research URL: https://www.youtube.com/watch?v=7p6w0zaTGH8 [accessed 2017-05-21] [WebCite Cache ID 6qce9rsiw]
50. Ho YS. Classic articles on social work field in Social Science Citation Index: a bibliometric analysis. Scientometrics 2013 Apr 13;98(1):137-155. [doi: 10.1007/s11192-013-1014-8]
51. Chuang KY, Wang MH, Ho YS. High-impact papers presented in the subject category of water resources in the essential science indicators database of the institute for scientific information. Scientometrics 2011 Mar 5;87(3):551-562. [doi: 10.1007/s11192-011-0365-2]
52. Bornmann L, Mutz R, Neuhaus C, Daniel HD. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. Ethics Sci Environ Polit 2008 Jun 03;8:93-102. [doi: 10.3354/esep00084]
53. Clarivate A. Clarivate.libguides. 2017. Essential Science Indicators: Learn the Basics URL: http://clarivate.libguides.com/esi [accessed 2017-08-04] [WebCite Cache ID 6sSIeyJm1]
54. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA 2015 Nov;102(46):16569-16572. [Medline: 16275915]
55. Egghe L. Theory and practise of the g-index. Scientometrics 2013 Jun 20;69(1):131-152. [doi: 10.1007/s11192-006-0144-7]
56. World Bank. Country and Lending Groups URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [WebCite Cache ID 6r4c6w6BP]
57. Aria M, Cuccurullo C. Bibliometrix. 2016. Bibliometrix: R package for bibliometric and co-citation analysis URL: http://www.bibliometrix.org [accessed 2017-07-28] [WebCite Cache ID 6sHyJezk4]
58. R-project. Vienna, Austria; 2017. R: A language and environment for statistical computing URL: https://www.r-project.org/ [accessed 2017-07-28] [WebCite Cache ID 6sHygUtqx]
59. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
60. Bornmann L, Mutz R. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J Assn Inf Sci Tec 2015 Apr 29;66(11):2215-2222. [doi: 10.1002/asi.23329]
61. Farooq M, McCrory MA, Sazonov E. Reduction of energy intake using just-in-time feedback from a wearable sensor system. Obesity (Silver Spring) 2017 Apr;25(4):676-681. [doi: 10.1002/oby.21788] [Medline: 28233942]
62. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832. [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
63. 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. [doi: 10.1007/s10865-016-9797-8] [Medline: 27722907]
64. Müller AM, Blandford A, Yardley L. The conceptualization of a Just-In-Time Adaptive Intervention (JITAI) for the reduction of sedentary behavior in older adults. Mhealth 2017 Sep;3:37 [FREE Full text] [doi: 10.21037/mhealth.2017.08.05] [Medline: 29184889]
65. Ottaviani J. The post-embargo open access citation advantage: it exists (probably), its modest (usually), and the rich get richer (of course). PLoS One 2016;11(8):e0159614 [FREE Full text] [doi: 10.1371/journal.pone.0159614] [Medline: 27548723]
66. Eysenbach G. Citation advantage of open access articles. PLoS Biol 2006 May;4(5):e157 [FREE Full text] [doi: 10.1371/journal.pbio.0040157] [Medline: 16683865]

67. Wray KB. No new evidence for a citation benefit for Author-Pay Open Access Publications in the social sciences and humanities. Scientometrics 2016 Jan 12;106(3):1031-1035. [doi: 10.1007/s11192-016-1833-5]
68. Sotudeh H, Ghasempour Z, Yaghtin M. The citation advantage of author-pays model: the case of Springer and Elsevier OA journals. Scientometrics 2015 Jun 3;104(2):581-608. [doi: 10.1007/s11192-015-1607-5]
69. Goryakin Y, Lobstein T, James WP, Suhrcke M. The impact of economic, political and social globalization on overweight and obesity in the 56 low and middle income countries. Soc Sci Med 2015 May;133:67-76 [FREE Full text] [doi: 10.1016/j.socscimed.2015.03.030] [Medline: 25841097]
70. Popkin BM. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. Am J Clin Nutr 2006 Aug;84(2):289-298 [FREE Full text] [Medline: 16895874]
71. Hyder AA, Wosu AC, Gibson DG, Labrique AB, Ali J, Pariyo GW. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. J Med Internet Res 2017 May 05;19(5):e133 [FREE Full text] [doi: 10.2196/jmir.7246] [Medline: 28476722]
72. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC, Lancet Physical Activity Series 2 Executive Committee. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. Lancet 2016 Sep 24;388(10051):1337-1348. [doi: 10.1016/S0140-6736(16)30728-0] [Medline: 27475273]
73. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
74. Fjeldsoe B, Neuhaus M, Winkler E, Eakin E. Systematic review of maintenance of behavior change following physical activity and dietary interventions. Health Psychol 2011 Jan;30(1):99-109. [doi: 10.1037/a0021974] [Medline: 21299298]
75. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
76. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
77. Uthman OA, Okwundu CI, Wiysonge CS, Young T, Clarke A. Citation classics in systematic reviews and meta-analyses: who wrote the top 100 most cited articles? PLoS One 2013;8(10):e78517. [doi: 10.1371/journal.pone.0078517] [Medline: 24155987]
78. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. J Clin Nurs 2003 Jan;12(1):77-84. [Medline: 12519253]
79. Blank I, Rokach L, Shani G. Leveraging metadata to recommend keywords for academic papers. J Assn Inf Sci Tec 2016 Mar 15;67(12):3073-3091. [doi: 10.1002/asi.23571]
80. Bornmann L. What is societal impact of research and how can it be assessed? A literature survey. J Am Soc Inf Sci Tec 2012 Dec 17;64(2):217-233. [doi: 10.1002/asi.22803]
81. Bornmann L. Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. J Informetr 2014 Oct;8(4):895-903. [doi: 10.1016/j.joi.2014.09.005]
82. Piwowar H. Altmetrics: Value all research products. Nature 2013 Jan 10;493(7431):159. [doi: 10.1038/493159a] [Medline: 23302843]

Abbreviations

eHealth: electronic health
eHealth: electronic health
NCD: Noncommunicable disease
SMS: short message service
WoS: Web of Science

Edited by G Eysenbach; submitted 26.09.17; peer-reviewed by M Bardus, R Hand; comments to author 08.11.17; revised version received 17.12.17; accepted 17.12.17; published 18.04.18.

Please cite as:
Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, DeSmet A, Short CE, Nathan N, Hutchesson MJ, Poppe L, Woods CB, Williams SL, Wark PA
Physical Activity, Sedentary Behavior, and Diet-Related eHealth and mHealth Research: Bibliometric Analysis
J Med Internet Res 2018;20(4):e122
URL: http://www.jmir.org/2018/4/e122/
doi:10.2196/jmir.8954
PMID:29669703

Original Paper

Development of an Internet-Administered Cognitive Behavior Therapy Program (ENGAGE) for Parents of Children Previously Treated for Cancer: Participatory Action Research Approach

Anna Wikman[1], PhD; Laura Kukkola[1], MSc; Helene Börjesson[1], MA; Martin Cernvall[1], PhD; Joanne Woodford[1], PhD; Helena Grönqvist[1], PhD; Louise von Essen[1], PhD

Clinical Psychology in Healthcare, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

Corresponding Author:
Anna Wikman, PhD
Clinical Psychology in Healthcare
Department of Women's and Children's Health
Uppsala University
Uppsala University Hospital
Uppsala, 75185
Sweden
Phone: 46 186110000
Fax: 46 186115583
Email: anna.wikman@kbh.uu.se

Abstract

Background: Parenting a child through cancer is a distressing experience, and a subgroup of parents report negative long-term psychological consequences years after treatment completion. However, there is a lack of evidence-based psychological interventions for parents who experience distress in relation to a child's cancer disease after end of treatment.

Objective: One aim of this study was to develop an internet-administered, cognitive behavior therapy–based, psychological, guided, self-help intervention (ENGAGE) for parents of children previously treated for cancer. Another aim was to identify acceptable procedures for future feasibility and efficacy studies testing and evaluating the intervention.

Methods: Participatory action research methodology was used. The study included face-to-face workshops and related Web-based exercises. A total of 6 parents (4 mothers, 2 fathers) of children previously treated for cancer were involved as parent research partners. Moreover, 2 clinical psychologists were involved as expert research partners. Research partners and research group members worked collaboratively throughout the study. Data were analyzed iteratively using written summaries of the workshops and Web-based exercises parallel to data collection.

Results: A 10-week, internet-administered, cognitive behavior therapy–based, psychological, guided, self-help intervention (ENGAGE) was developed in collaboration with parent research partners and expert research partners. The content of the intervention, mode and frequency of e-therapist support, and the individualized approach for feedback were modified based on the research partner input. Shared solutions were reached regarding the type and timing of support from an e-therapist (eg, initial video or telephone call, multiple methods of e-therapist contact), duration and timing of intervention (eg, 10 weeks, 30-min assessments), and the removal of unnecessary support functions (eg, removal of chat and forum functions). Preferences for study procedures in future studies testing and evaluating the intervention were discussed; consensus was not reached for all aspects.

Conclusions: To the best of our knowledge, this study is the first use of a participatory action research approach to develop a psychological intervention for parents of children previously treated for cancer and to identify acceptable study procedures. Involvement of parents with lived experience was vital in the development of a potentially relevant and acceptable intervention for this population.

(J Med Internet Res 2018;20(4):e133) doi:10.2196/jmir.9457

KEYWORDS

cognitive therapy; psychology, clinical; e-therapy; community participation; Sweden

Introduction

Background

Although majority of the children diagnosed with cancer survive their disease [1], childhood cancer impacts the entire family from cancer diagnosis to survivorship [2]. For parents, a child's treatment completion represents an important milestone, but it also represents a period of psychological vulnerability [3,4]. Indeed, a subgroup of parents report negative long-term psychological consequences years after completion of treatment [4-6]. However, currently, there is a lack of evidence-based psychological interventions for parents who experience distress in relation to a child's cancer disease after end of treatment. Recently published clinical guidelines, outlining how children diagnosed with cancer and their family members should be cared for, recommend referrals to appropriate psychosocial and therapeutic support into long-term survivorship [7]. Despite these recommendations, we have recently shown that subgroups of parents report an unmet need of psychological support after end of treatment [8]. Furthermore, although face-to-face cognitive behavior therapy (CBT) shows promise in decreasing post-traumatic stress symptoms (PTSS), depression, and anxiety among parents of children previously treated for cancer (personal communication by L Ljungman, 2017-10-23), challenges remain regarding provision of psychological support to those parents of children previously treated for cancer who need such support. Indeed, our recent findings showing an unmet need of psychological support among parents of children previously treated for cancer [8] are in line with findings from one study in Australia showing that formal psychological support was difficult to access and rarely received by parents after cancer treatment completion [9]. This study concluded that factors related to staff availability, models of assessment and delivery of services, and size and location of pediatric cancer centers may hinder the provision of support.

Provision of CBT via the internet may increase access to psychological support and may be an alternative for parents of children previously treated for cancer. One previous study has shown high acceptability and feasibility of a Web-based, group-based, CBT intervention, delivered "live" by a psychologist, for parents following cancer treatment completion [10]. In a randomized controlled trial (RCT), we have shown a Web-based psychological self-help intervention to be effective in reducing PTSS, depression, and anxiety among parents of children recently diagnosed with cancer [11], with improvements maintained at 1-year follow-up [12]. This RCT was initiated in 2010, and the intervention, although Web-based, utilized few technological features, as it provided material in the form of PDF files and secure written communication between the participant and therapist. During this trial, challenges, such as low enrolment rate and considerable attrition, were identified [11,12]. Factors to enhance inclusion and retention rates such as end-user involvement when developing interventions and study procedures to test and evaluate interventions for parents of children with cancer have been suggested as essential next steps for intervention research [13]. Furthermore, others have put forward that internet-administered, CBT-based, self-help interventions should be developed with the target population in mind [14]. Indeed, poorer levels of acceptability have been found for internet-administered interventions not developed for and tailored toward specific populations [15]. Additionally, research suggests recruitment and adherence rates may be improved if the perspective of the population is adopted [16].

Objective

Due to previous findings [4-6,8] as well as challenges with recruitment and attrition when offering parents psychological support during the child's treatment, it was decided to focus subsequent research on parents later during their child's disease trajectory, specifically after the end of treatment. Building on our group's previous work with this population (personal communication by L Ljungman, 2017-10-23) [4,5], representing the Medical Research Council (United Kingdom) phase I (development) research [17], this study adopted a participatory action research (PAR) approach [18]. One aim of this study was to develop an internet-administered, CBT-based, psychological, guided, self-help intervention (ENGAGE) for parents of children previously treated for cancer. Another aim was to identify acceptable procedures for future feasibility and efficacy studies testing and evaluating the intervention.

Methods

Design and Setting

The study was carried out according to PAR, which is a collaborative process of knowledge production and colearning, placing people with lived experience at the center of the process [18]. A group of people with lived experience of parenting a child previously treated for cancer were involved as parent research partners (PRPs) and took part in 8 workshops and related Web-based exercises. The study was carried out at Uppsala University, Uppsala, Sweden, and was conducted over an 8-month period during 2016 on weekday evenings based on PRPs' preferences. An additional workshop was carried out during this period with 2 expert research partners (ERPs), both clinical psychologists and experts in internet-administered psychological interventions. Ethical approval was granted by the regional ethical review board in Uppsala, Sweden (Dnr: 2015/426).

Research Partners

Parent Research Partners

Eligible PRPs were the ones who (1) lived near Uppsala (≤100 km), (2) spoke Swedish, (3) were a parent of a child previously treated for cancer, and (4) had experienced or were experiencing psychological distress related to the child's cancer disease. Current severe psychological distress (eg, symptoms of a severe and enduring mental health difficulty, misuse of alcohol or drugs, acutely suicidal) excluded parents from participation. Parents of children previously treated for cancer who had participated in our group's previous intervention research [11,12] (personal communication by L Ljungman, 2017-10-23) were invited strategically via letter and a telephone call, considering variation in gender and socioeconomic status. If 2 parents of 1 child had participated in a previous study, then only 1 parent was invited. This was to avoid overrepresentation of 1 family's experience in the group, and to achieve an open

discussion environment, which may have been influenced by the presence of one's partner. In addition, information about the study was posted on the Swedish Childhood Cancer Foundation Website and the Swedish Childhood Cancer Association Middle-Sweden Facebook page to increase awareness of the study, including an open invitation to participate in the study. The sample consisted of 6 parents, 5 of whom had participated in our previous intervention research. PRPs participated outside their regular working hours and were reimbursed for time (fixed amount for workshops and Web-based assignments completed) and travel expenses at the end of the study.

Parent Research Partner Characteristics

All eligible parents who provided informed written consent were included as PRPs (4 mothers, 2 fathers). All PRPs reported living with a partner and the majority (n=4) reported having completed a university degree. Their mean age was 50 years (SD 3), the child's average age when diagnosed was 9.1 years (SD 4), and mean time since end of treatment was 5 years (SD 3). On average, the PRPs completed an average of 5 workshops (of 8) and 7.8 Web-based exercises (of 8).

Expert Research Partners

Two clinical psychologists with expertise in development and clinical use of internet-administered CBT programs were involved as ERPs. One of the ERPs had extensive experience of working with children and their family members. The ERPs were reimbursed for time spent reviewing the materials and participating in the ERP workshop and travel expenses.

Aspects of the Intervention and Procedures for Future Studies Set at the Start of the Study

Some aspects regarding the intervention and procedures were preset before the start of the study by the research group and communicated to the PRPs and ERPs. First, the intervention should be informed by evidence-based knowledge concerning psychological distress experienced by parents of children treated for cancer and conceptualization and treatment of this distress [4,5,11,12] (personal communication by L Ljungman, 2017-10-23). Second, the intervention should be delivered via the U-CARE portal (Portal). The Portal is a secure infrastructure developed by our research group and includes functions such as log-in via bank-issued electronic identification; provision of internet-administered, guided, self-help material; communication between participants and e-therapists via internal messages and homework reports; video calls; collection of questionnaire data at predefined observation points; logging of participant behavior; Web-based library; participant chat; forum; diary; and a question and answer (Q&A) function. Third, the procedures should be possible to carry out considering the available resources and should follow national ethical research regulations.

Procedure

Workshops and Web-Based Exercises With Parent Research Partners

Table 1 shows an overview of the PAR process. The workshops with PRPs were facilitated by one parent of a child previously treated for cancer with professional experience of teaching, MSc in English (coauthor HB), and one PhD student, MSc in Clinical Psychology (coauthor LK). The facilitators were responsible for constructing the Web-based exercises and materials used, taking meeting notes during the workshops, facilitating discussions, and reviewing study materials. After each workshop, HB and LK reflected upon the process with members of the research group (coauthors MC, LvE, HG, and AW). Collaboration with PRPs included 8 workshops and related Web-based exercises. Workshops were carried out in person, and Web-based exercises were carried out individually via the Portal. Each workshop had a predefined theme to focus the discussions.

Table 1. Collaboration process, workshop overview, and research partner activity carried out during May to December 2016. CBT: cognitive behavior therapy; PRP: parent research partner.

Content	Workshop 1	Workshop 2	Workshops 3 and 4	Workshops 5 and 6	Expert research partner workshop	Workshop 7	Workshop 8
Topic	Welcome	Initial engagement	Intervention	Study procedures	Treatment manual	Prototype	Evaluation
Activities	Establishing contact; Presenting the study context	Discussion of opportunities and barriers for initial engagement	Discussion of modes of delivery and support functions; Refining a CBT model	Discussion of pros and cons of randomization and acceptable study procedures	Reviewing and refining a treatment manual; Discussion of manual and draft module	Reviewing and refining a treatment prototype and materials	Summary and evaluation of participatory action research process
Web-based exercises	None	Completed by 6 PRPs	Completed by 6 PRPs	Completed by 6 PRPs	None	Completed by 5 PRPs	Completed by 6 PRPs
Present at workshop	2 workshop facilitators; 2 researchers; 4 parent research partners (PRPs)	2 workshop facilitators; 3 PRPs	Workshop 3: 2 workshop facilitators; 2 PRPs Workshop 4: 2 workshop facilitators; 4 PRPs	Workshop 5: 2 workshop facilitators; 3 PRPs Workshop 6: 2 workshop facilitators; 4 PRPs	2 expert research partners; 5 researchers	2 workshop facilitators; 5 PRPs	2 workshop facilitators; 2 researchers; 6 PRPs
Time-frame	May 2016	May 2016	May 2016	June 2016	September 2016	November 2016	December 2016

Workshops 2 to 7 and Web-based exercises were conducted in an iterative manner, including the presentation of repeated written process summaries from the previous workshop to be reviewed and discussed with PRPs at the beginning of each subsequent workshop. During the first workshop with PRPs, the primary aim was to establish contact to facilitate the group process and present the study context. In the final workshop, the overall PAR process was summarized and evaluated. Workshops 2 to 7 with PRPs were carried out according to the following 2-stage process:

First, PRPs were asked to complete individual Web-based exercises before each workshop. These generally consisted of a summary from the previous Web-based exercise and workshop (workshops 2 to 7), educational or intervention materials (videos, PowerPoint lectures, or PDF texts), and related open- and close-ended questions. At the end of each Web-based exercise, PRPs were asked to provide feedback on the materials for discussion in the subsequent workshop. PRPs could indicate if they did not wish for their feedback to be discussed at the subsequent workshop and were encouraged to suggest topics they wished to reflect on in the group. PRPs were asked to provide responses to the Web-based exercises the day before the respective workshop at the latest. On the day of each workshop, the facilitators reviewed and summarized the PRPs' individual answers to the Web-based exercises. On the basis of this summary, areas for further investigation for each PRP workshop were decided at a research group meeting, and added to the workshop agenda.

Second, each workshop began with a round of brief reflection, providing PRPs with the possibility to share thoughts and feelings relating to the Web-based exercise and previous workshop. Then, the agenda for the workshop was presented and PRPs were encouraged to add items to the agenda. Each workshop lasted for 2 hours, including a brief refreshment break, and consisted of individual reflection and group discussions. At the end of each meeting, an individual reflection practice was carried out where PRPs individually answered some open-ended questions. PRPs were encouraged to provide suggestions for upcoming workshops and identify the most valuable topics of the workshop. PRPs spent approximately 5 min on the task. Following this, PRPs were offered the opportunity to stay for another hour to freely reflect upon their experiences with the facilitators or to socialize with other parents, without any documentation being carried out and refreshments were provided. Figure 1 illustrates a workshop in progress.

Workshop With Expert Research Partners

The written intervention material (consisting of 109 A4 pages), including 1 draft module (eg, text, video, and audio materials), was presented to ERPs following the 6 workshops with PRPs. The material was written in parallel to the PRP workshops, with workshops informing the content included. During the ERP workshop, ERPs provided their expert perspectives on the material and draft module. Members of the research group (coauthors MC, LvE, HG, and AW) were present at the workshop, which was facilitated by LK. Following the workshop, the intervention material was modified, and a prototype of the intervention was presented to PRPs at workshop 7.

Figure 1. Parent research partners (PRPs) and facilitators during a workshop. Note that consent was obtained from all PRPs present to use this photo.

Data Collection

Workshops with PRPs were documented using written meeting notes (coauthors HB and LK). Materials used for discussion practices, such as post-it notes and summaries by PRPs, were saved. At the end of each workshop (with exception of workshops 1 and 8), PRPs completed written individual process evaluations (ie, feedback on the workshops) to guide the continuation of the collaboration process. The PRPs' responses to the Web-based exercises were documented on the Portal to ensure secure communication. During the Web-based exercises, PRPs provided feedback on extracts of the intervention and education materials presented by answering questions via the Portal. Questions focused on PRPs' experiences of and views on aspects such as the design of the intervention materials, the Portal interface, preferences for optional support functions, and suggestions on how to improve aspects of the intervention. The ERP workshop was documented using written meeting notes from the workshop (coauthor LK, with assistance from a research assistant), with ERPs providing their written reflections on the materials after the workshop.

Data Analysis

Data were analyzed iteratively parallel to data collection. Pragmatism was adopted as the underlying research paradigm, selecting an approach to data analysis most appropriate to the study aims [19]. Specifically, as per other studies utilizing PAR [20,21], a thematic analysis approach was adopted [22]. PRPs' responses to the Web-based exercises were read and summarized by HB and LK and reported back to the PRPs in the subsequent workshop. Areas of further exploration in subsequent workshops were identified and discussed by the research group consisting of the research group leader (professor, PhD in clinical psychology, and clinical psychologist, coauthor LvE), 2 researchers with a PhD in psychology (coauthors HG and AW), and 2 clinical psychologists (coauthors MC and LK). Possible solutions to parents' questions and further areas of interest were identified by the research group before each subsequent workshop. Although each workshop had a predefined theme, the agenda for each subsequent PRP workshop partially emerged from the PRPs' Web-based exercises and reflections by the research group, with these reflections reported back to the PRPs to establish trustworthiness of the interpretations made by the research group. Furthermore, additional agenda items could be suggested by PRPs. This iterative process was used throughout the study and continued until no new themes related to study aims were identified. Over the course of the PAR process, discussions during workshops continued until data saturation was reached, that is, until no new data emerged. Following standard approaches to thematic analysis [22], data were then synthesized by members of the research group (coauthors MC, LvE and AW) firstly into descriptive and topic codes, followed by identification of themes relating to the study aims. To enhance trustworthiness, identified themes were then presented to and further discussed by the wider research group to ensure agreement on identified themes and that all data were included in the identified themes.

Results

Parent Research Partners' Views on the Intervention

An overview of the results from the PAR process is shown in Table 2.

Duration, Content, and Presentation

PRPs stressed the importance of the intervention not being too burdensome. As such, parents suggested exercises to be shortened and provided over a longer period than originally suggested. Spending 1 hour per week working with the intervention was seen as optimal. Some suggested a duration of 7 weeks, whereas others considered a 12-week intervention better. Although consensus was not reached, a 10-week intervention was deemed acceptable.

PRPs stressed the importance of communication suggesting that the researchers should communicate clearly to potential participants in the future feasibility study that the intervention was developed to fulfill parents' needs for psychological support, and provide clear treatment goals (eg, reducing PTSS and depression). One aspect discussed with PRPs concerned the extent to which the intervention content could be individualized, given the majority of PRPs highlighted an interest in a tailored intervention, for example, choosing specific topics to meet an individual's particular needs. However, for the purpose of testing the feasibility and acceptability of the intervention in the upcoming feasibility study, it is important that its content remains the same for all participants.

Collaboration with PRPs resulted in including a video or telephone support call, the use of case vignettes, and replacing one of the actors in the vignettes presented to PRPs. Furthermore, the language used to describe suffering and content in modules was modified based on PRP preferences. For example, in terms of the conceptualization of distress presented during workshops 3 and 4, PRPs preferred terms such as changed life situation to depressive inactivity and difficult or painful emotions and memories rather than traumatic stress. The final version of the intervention consists of 1 introductory module followed by 10 internet-administered, guided, self-help modules.

PRPs appreciated a combination of text, audio, and video materials of high quality, including an option to print materials. The inclusion of case vignettes was valued. However, the importance of authenticity was discussed, with modifications to the vignettes made based on PRPs' feedback on the prototype presented in workshop 7, to improve relevance and authenticity. PRPs suggested that the inclusion of an introductory video of the intervention, presented by an e-therapist and a parent, would increase trustworthiness and motivation to participate. Furthermore, an optional support function including an Web-based library containing information about CBT, self-help, literature suggestions, links to relevant Websites, as well as CBT exercises from the intervention were seen as advantageous.

Table 2. Overview of results from the participatory action research (PAR) process, summarizing research partners' views on the intervention ENGAGE and acceptable procedures for future feasibility and effectiveness studies. CBT: cognitive behavior therapy.

Research partner	Views on intervention		Views on procedures	
	Theme	Description	Theme	Description
Parent research partners	Duration, content and presentation of the intervention	Not too burdensome Short Web-based exercises One hour per week 10-week program Text, audio, and video materials Language used to describe distress Case vignettes, actor changes Introductory video Web-based library including, for example, information about CBT, self-help, literature suggestions, links to relevant Websites as well as CBT exercises from the intervention	Information about study participation	Engaging and interesting information, highlighting benefits of guided self-help Information video about the study before consenting
Parent research partners	Support and contact during the interventions	Initial video or telephone session Booster video or telephone session at half-time Multiple methods of e-therapist contact (video call, telephone call, and written communication) Written feedback throughout intervention Single-item mood assessment using 5-point Likert scale to communicate changes in mood with the e-therapist	Time aspects	30 min acceptable time to complete assessments at each observation point (ie, baseline, post treatment, 6-month follow-up) 10 min acceptable time to complete weekly assessments
Expert research partners		Clear Web-based exercises Evidence-based CBT exercises Reviewing treatment goals throughout the intervention "Less is more": intervention content reduced CBT exercises available in the Web-based library Professionally designed materials	No views provided on study procedures	

The design aspects emerging as important during the PAR process were subsequently incorporated into the design of the intervention (Table 2). Optional support functions planned for inclusion before the collaborative process included a chat function, a forum, and a Q&A function. However, these were in general viewed as unnecessary and subsequently removed from the intervention.

A written prototype of the intervention was presented to PRPs during workshop 7 and was perceived positively, with one parent stating that "parents could benefit from this."

Support and Contact

At the start of the study, it was planned for the intervention to include weekly written feedback from an e-therapist trained in supporting the intervention. An e-therapist is a mental health care professional who provides support electronically, for example, via email or videoconferencing [23]. E-therapists will be psychology program students, in at least their 4th year of study, having completed a minimum of their first term of advanced studies in CBT, but will have not yet begun their prescribed practical service (ie, praktisk tjänstgöring för psykologer/PTP). PRPs highlighted the importance of personal contact via a video or telephone call with an e-therapist at the start of the intervention for parents to be able to "tell their story" and midway through the intervention to increase motivation. However, all PRPs raised the importance of parents being able to "tell their story" in their preferred way (eg, contact with the e-therapist via video or telephone call, or via written communication). As such, personal contact at the start and midway through the intervention with an e-therapist via a video or telephone call has been incorporated (Table 2). However, it was not deemed possible to completely individualize the procedure for parents "telling their story" as it is important the intervention is delivered in the same way to all participants for testing purposes in the forthcoming feasibility study.

A further issue discussed related to PRPs' views on receiving feedback on changes in mood during the intervention. Although some considered weekly charts illustrating changes in mood as an important aspect of the intervention, the majority preferred personal feedback from their e-therapist. In the final version of the intervention, written feedback from the e-therapist is the primary method of communication as it was considered most important by PRPs. However, in line with the PRPs' requests, a single-item mood assessment, using a 5-point Likert scale, has been incorporated as a means for participants to inform their e-therapist about their mood, in addition to the written communication.

Expert Research Partners' Views on the Intervention

Following the ERPs' input, the intervention was further refined (Table 2). Changes included clarifying the key learning components included in each module. For example,

evidence-based CBT exercises were included in each module, and weekly action plans were designed based on the content of the modules to carry out as homework. ERPs highlighted the importance of revisiting treatment goals during the intervention to maintain focus, which was subsequently more clearly incorporated in the intervention. In line with PRPs' views, one take-home message was "less is more." Consequently, the content in the intervention was reduced. ERPs agreed with PRPs that CBT exercises should be easily available, for example, in the Web-based library, to enable parents to return to previous exercises based on individual preferences and needs. To make the user experience more appealing, presenting a professional look, it was decided that materials and illustrations used in the intervention should be created by professionals in graphic design as suggested by the ERPs. Figure 2 shows an example of an internet-administered intervention module screen.

Parent Research Partners' Views on What Procedures Should be Used in Future Studies Testing and Evaluating the Intervention

Information About Study Participation

One key procedural aspect discussed concerned recruitment to future feasibility and effectiveness studies. PRPs stressed the importance of how study information should be provided to engage and motivate participation, highlighting the potential benefits of CBT-based, internet-administered, guided, self-help interventions. A phone call, followed by written information, was the preferred method of receiving study information. It was positively perceived to receive a study-code to access materials on the Web with further information about the study, including an informational film before consenting to participation, both of which have been included in the intervention (Table 2). PRPs were of the opinion that study information should be communicated by health care personnel, or representatives from relevant organizations, to increase trustworthiness. PRPs indicated a preference for recruitment via the hospital departments in connection with end of the child's cancer treatment. However, as recruitment is planned for up to 5 years after end of treatment, it would not be feasible to recruit via hospital departments as parents are likely no longer in contact with the hospital clinics at that time.

Time Aspects

PRPs were unable to reach consensus regarding the ideal timing for the invitation to participate in a controlled study of an internet-administered, guided, self-help intervention for parents of children previously treated for cancer. Although some indicated that 3 to 6 months following end of the child's cancer treatment would be preferable, others argued that more time must pass as parents may not be aware of their needs for psychological support so soon after end of treatment. However, all PRPs considered it important that parents are offered the participation, rather than having to seek out participation. Views differed regarding how long parents were willing to wait to receive the intervention in a hypothetical controlled study with a wait-list control group. Although a waiting period of 1 to 3 months was generally accepted, parents raised concerns about waiting when suffering emotionally following the end of their child's treatment.

The time required to complete pre- and postintervention assessments as part of a controlled study was discussed, and PRPs considered 30 min to be an acceptable length of time for completing these assessments. In addition, spending 10 min weekly for completing the questionnaires while receiving the intervention was an acceptable amount of time according to PRPs.

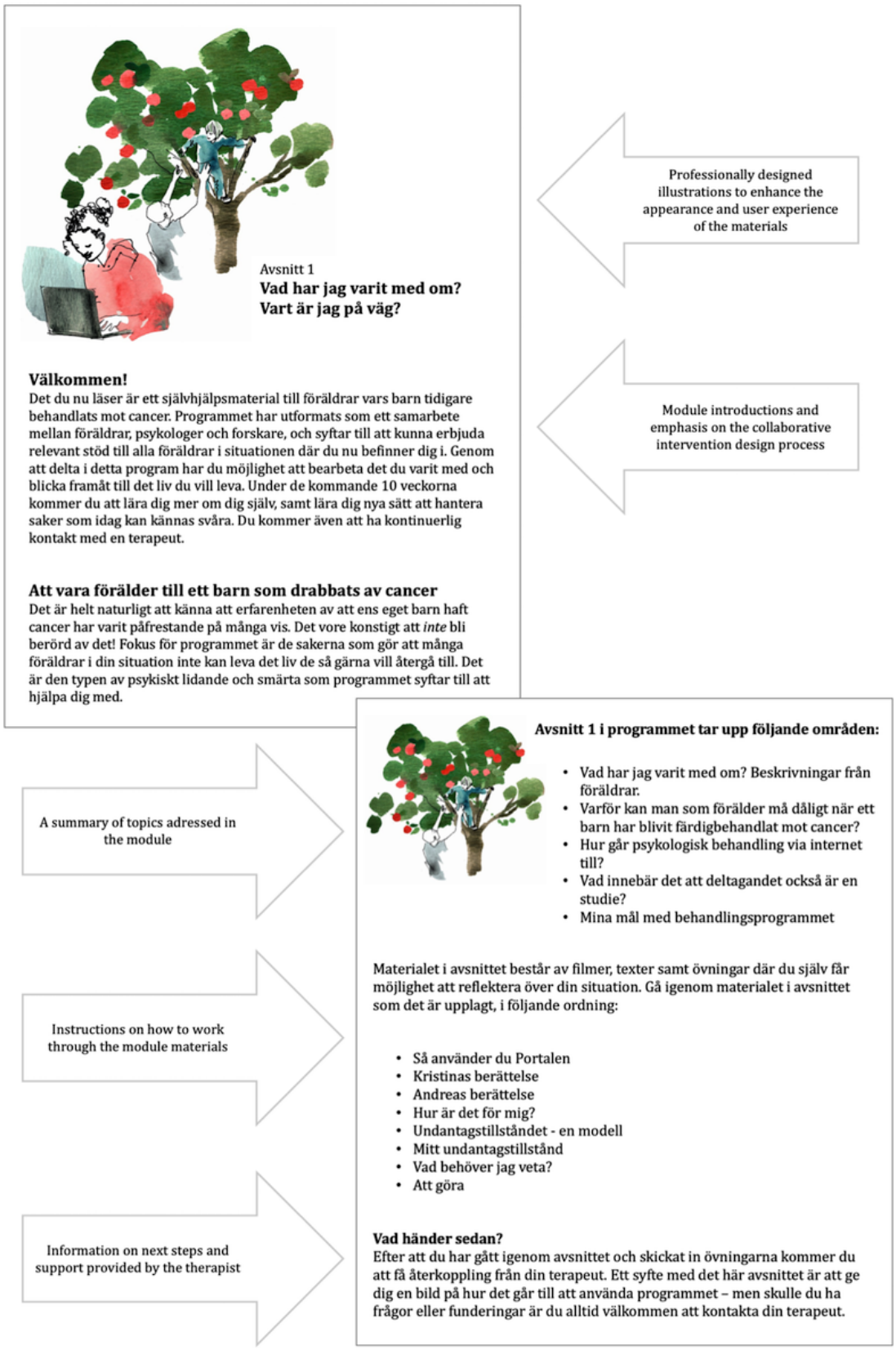
Although PRPs agreed on many procedural aspects for future studies testing and evaluating the intervention, some issues remained undecided following this process. To further address parents' preferences regarding study procedures, a cross-sectional, Web-based study has been conducted (results forthcoming) to examine their attitudes and preferences regarding, for example, the mode of study invitation, how study information should be presented and by whom, type of control conditions in a controlled study, and the acceptable waiting time in a controlled study with a wait-list control condition.

Process Evaluation

To facilitate the collaboration, continuous process evaluations were carried out. PRPs suggested some clarifications, for example, with regard to the time period of interest regarding their lived experience (eg, during treatment, at treatment completion, or at present) and according to which role they should provide feedback on the material (as hypothetical participants in a future study using the material as part of the intervention, or as research partners in the ongoing PAR process). PRPs were encouraged to approach the materials as research partners, but to give their opinion, they also needed to test the materials as hypothetical participants in a future study. Instructions were modified according to clarifications.

PRPs expressed concerns about working so closely with the facilitators who also created the materials for the workshops and Web-based exercises and whether this might influence their perspective as PRPs. To reduce potential respondent bias, the research group adapted the PAR process to include the presentation of materials by members of the research group not otherwise involved in the workshops. Furthermore, PRPs highlighted the psychological and emotional demands of taking part in the PAR process. However, having received CBT within our previous research [11,12] (personal communication by L Ljungman, 2017-10-23) was perceived as helpful in coping with these demands.

Figure 2. Example of an internet-administered intervention module screen (illustrations by Annika Carlsson).



Discussion

Principal Findings

A 10-week internet-administered, CBT-based, psychological, guided, self-help intervention (ENGAGE) for parents of children previously treated for cancer, alongside procedures for future studies testing and evaluating the intervention, were developed in collaboration with PRPs and ERPs. Specifically, the content of the intervention, mode and frequency of e-therapist support, and the individualized approach for feedback were modified based on input from the PRPs or ERPs. Shared solutions were reached regarding the type and timing of support from an e-therapist, duration and timing of intervention, and the removal of "unnecessary" support functions. The ENGAGE intervention will include written, audio, and video materials as well as an initial support session via video or telephone in which individual problem analyses and idiographic goals will be formulated. The Web-based intervention will be delivered via the Portal with weekly e-therapist support. A "booster" session will be provided via video or telephone midway through the intervention. Participants will complete 1 module per week over a 10-week treatment period. Guidance from e-therapists will consist of 1 video- or telephone-assessment session, weekly Web-based written support, and a mid-treatment video- or telephone "booster" session. Module content will include, for example, psychoeducation, case vignettes, exercises based on specific CBT techniques, action plans, and questionnaires to assess symptoms. Preferences for study procedures in future studies testing and evaluating the intervention were discussed, but consensus was not reached. Overall, collaborative work added significantly to the study as a whole, with PRPs' feedback continuously informing the research process, highlighting the value of working closely with the target population when conducting intervention research.

PRPs' preferences regarding the delivery and content of the intervention are largely consistent with the existing, related literature for other study populations. For example, including patient vignettes in the intervention materials to facilitate normalization [24] and enhance identification with the intervention content [25]. Furthermore, PRPs' concerns regarding limitations of not being able to personalize the content of the intervention are consistent with wider research [14]. Moreover, the lack of personal interaction with an e-therapist has been highlighted [26], with participants reporting difficulties developing a relationship with an e-therapist in the absence of face-to-face contact [27]. However, the inclusion of a video call at the beginning of the intervention may overcome these concerns. Indeed, the delivery of CBT via videoconference is as effective as face-to-face CBT [28], allowing similar communication to that in face-to-face therapy [29]. The inclusion of some form of face-to-face contact with an e-therapist may help develop a therapeutic alliance [30]. Furthermore, supported by the wider literature, was the suggestion to enhance the acceptance of the internet-administered intervention, and thus potentially facilitate recruitment, by presenting an informative video about internet-administered CBT to potential participants [31-33].

Although some findings were in line with existing related literature concerning the development of internet-administered interventions, this study further extends our understanding of acceptable intervention adaptation for the population. First, before the study, the research group posited that the inclusion of support functions (eg, a chat function, a forum, and a Q&A function) would be desirable and enhance intervention interactivity. However, PRPs felt such functions were unnecessary, and these were thus not included in the intervention. Second, PRPs were introduced to the conceptualization of distress experienced by parents of children previously treated for cancer, as informed by previous research conducted by our group [4,5,11,12] (personal communication by L Ljungman, 2017-10-23). Although PRPs reported that the conceptualization of distress was acceptable, and reflected their own experiences, a number of modifications were suggested concerning the adoption of "lay language," as opposed to the psychological terminology used in the intervention material. As such, collaboration with PRPs further contributed to our understanding of how to communicate psychological principles to the population using more acceptable and accessible terminology.

The key elements of a PAR approach include understanding, mutual involvement, change, and a process that promotes personal growth [34]. Fulfillment of these elements could be identified from the PRPs' responses to workshop evaluations. Establishing participants as equal research partners can be difficult, especially if participants experience low self-efficacy regarding their ability to participate in the process as equal partners [35]. Future PAR research may look to include training for research partners without research experience [36]. However, PRPs reported personal growth and helping others as significant motivators for contributing to the study.

Strengths and Limitations

A strength of the study was the involvement of PRPs and ERPs via a PAR approach in developing the intervention ENGAGE and study procedures for future studies testing and evaluating the intervention. As such, the development was informed by previous research [4,5,11,12] (personal communication by L Ljungman, 2017-10-23), PRPs' lived experience, and ERPs' expert knowledge. We expect the importance of this approach to be reflected in the acceptability and feasibility of the intervention and study procedures in a forthcoming feasibility study. However, it should be noted that PRPs had received psychological support in previous intervention research. This means that they might have a more positive attitude toward research in comparison with the wider population and that the current findings may not describe the experiences of parents of children previously treated for cancer who have not accessed psychological support. It should also be considered that although PRPs of both genders and from different socioeconomic backgrounds were included, the sample was small and recruited from a small geographical area. In addition to inviting parents who had participated in our group's previous intervention research, an open invitation to participate in the study was posted through the Swedish Childhood Cancer Foundation Website and the Swedish Childhood Cancer Association Middle-Sweden Facebook page. However, this recruitment strategy did not result

in any further potential participants registering interest in participating in the study. As such, future research may benefit from adopting more assertive recruitment methods to identify researcher partners from the wider community of parents of children previously treated for cancer [36]. Although consensus was not always reached, in general, acceptable alternatives were agreed upon by PRPs. Ideas were discussed by PRPs and facilitators to generate shared solutions. In some cases, PRPs' preferences were deemed unfeasible by the research group. For example, PRPs mentioned a preference for recruitment via hospital departments. However, as recruitment to the forthcoming feasibility study is planned for up to 5 years after end of a child's treatment and parents are likely no longer in contact with the hospital department at that time, such a recruitment strategy is not possible to adopt.

Conclusions

To the best of our knowledge, this study is the first use of a PAR approach to develop a CBT-based, internet-administered, guided, self-help intervention for parents of children previously treated for cancer and acceptable procedures for future studies testing and evaluating the intervention. We believe involvement of parents with lived experience and experts with expert knowledge have been vital in the development of a potentially relevant and acceptable intervention for this population. Specifically, the PAR process informed intervention content, including language, duration, mode, and frequency of e-therapist support. Furthermore, planned recruitment strategies for use in a planned feasibility study were modified, which may enhance participation, for example, via the inclusion of an informative video about internet-administered interventions. In addition, this planned feasibility study will further examine the relevance and acceptability of the developed intervention. The PAR process adopted in this study may inform the future use of PAR techniques to adapt and tailor internet-administered, psychological, guided, self-help interventions for other populations.

Acknowledgments

This work was supported by funding to LvE from The Swedish Research Council (grant number K2015-99X-20836-08-4) and the Swedish Cancer Society (grant number 140790). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors gratefully acknowledge the contributions of the 6 parents of children previously treated for cancer, and the 2 clinical psychologists, who acted as highly valued research partners. The authors would like to thank Anna Liisa Aho, PhD; Gustaf Ljungman, PhD, MD; and Lisa Ljungman, PhD, for their valuable contributions and collaboration during the undertaking of this study.

Authors' Contributions

LvE, MC, HG, LK, AW, and JW were involved in the study conception and design. HB and LK were involved in workshop facilitation and data collection. MC, LvE, and AW were involved in data analysis. LvE, LK, AW, and JW were involved in interpretation of the results and in manuscript writing. HB, MC, LvE, HG, LK, AW, and JW were involved in critical review of the manuscript.

Conflicts of Interest

None declared.

References

1. Gustafsson G, Kogner P, Heyman M. 2013. Childhood cancer incidence and survival in Sweden 1984-2010 URL: https://www.cceg.ki.se/documents/ChildhoodCancerIncidenceandSurvivalinSweden1984_2010.pdf [accessed 2018-01-22] [WebCite Cache ID 6wf0tuN6Y]
2. Pai AL, Greenley RN, Lewandowski A, Drotar D, Youngstrom E, Peterson CC. A meta-analytic review of the influence of pediatric cancer on parent and family functioning. J Fam Psychol 2007 Sep;21(3):407-415. [doi: 10.1037/0893-3200.21.3.407] [Medline: 17874926]
3. Wakefield CE, McLoone JK, Butow P, Lenthen K, Cohn RJ. Parental adjustment to the completion of their child's cancer treatment. Pediatr Blood Cancer 2011 Apr;56(4):524-531. [doi: 10.1002/pbc.22725] [Medline: 21298736]
4. Ljungman L, Hovén E, Ljungman G, Cernvall M, von Essen L. Does time heal all wounds? A longitudinal study of the development of posttraumatic stress symptoms in parents of survivors of childhood cancer and bereaved parents. Psychooncology 2015 Dec;24(12):1792-1798 [FREE Full text] [doi: 10.1002/pon.3856] [Medline: 26042579]
5. Ljungman L, Cernvall M, Grönqvist H, Ljótsson B, Ljungman G, von Essen L. Long-term positive and negative psychological late effects for parents of childhood cancer survivors: a systematic review. PLoS One 2014;9(7):e103340 [FREE Full text] [doi: 10.1371/journal.pone.0103340] [Medline: 25058607]
6. Bruce M. A systematic and conceptual review of posttraumatic stress in childhood cancer survivors and their parents. Clin Psychol Rev 2006 May;26(3):233-256. [doi: 10.1016/j.cpr.2005.10.002] [Medline: 16412542]

7. Steele AC, Mullins LL, Mullins AJ, Muriel AC. Psychosocial interventions and therapeutic support as a standard of care in pediatric oncology. Pediatr Blood Cancer 2015 Dec;62(Suppl 5):S585-S618. [doi: 10.1002/pbc.25701] [Medline: 26700919]
8. Kukkola L, Hovén E, Cernvall M, von Essen L, Grönqvist H. Perceptions of support among Swedish parents of children after end of successful cancer treatment: a prospective, longitudinal study. Acta Oncol 2017 Oct 03;56(12):1705-1711. [doi: 10.1080/0284186X.2017.1374554] [Medline: 28971717]
9. Wakefield CE, McLoone J, Butow P, Lenthen K, Cohn RJ. Support after the completion of cancer treatment: perspectives of Australian adolescents and their families. Eur J Cancer Care (Engl) 2013 Jul;22(4):530-539. [doi: 10.1111/ecc.12059] [Medline: 23730980]
10. Wakefield CE, Sansom-Daly UM, McGill BC, Ellis SJ, Doolan EL, Robertson EG, et al. Acceptability and feasibility of an e-mental health intervention for parents of childhood cancer survivors:"Cascade". Support Care Cancer 2016 Dec;24(6):2685-2694. [doi: 10.1007/s00520-016-3077-6] [Medline: 26781620]
11. Cernvall M, Carlbring P, Ljungman L, Ljungman G, von Essen L. Internet-based guided self-help for parents of children on cancer treatment: a randomized controlled trial. Psychooncology 2015 Sep;24(9):1152-1158. [doi: 10.1002/pon.3788] [Medline: 25728688]
12. Cernvall M, Carlbring P, Wikman A, Ljungman L, Ljungman G, von Essen L. Twelve-month follow-up of a randomized controlled trial of internet-based guided self-help for parents of children on cancer treatment. J Med Internet Res 2017 Jul 27;19(7):e273 [FREE Full text] [doi: 10.2196/jmir.6852] [Medline: 28751300]
13. Stehl ML, Kazak AE, Alderfer MA, Rodriguez A, Hwang W, Pai AL, et al. Conducting a randomized clinical trial of an psychological intervention for parents/caregivers of children with cancer shortly after diagnosis. J Pediatr Psychol 2009 Sep;34(8):803-816 [FREE Full text] [doi: 10.1093/jpepsy/jsn130] [Medline: 19091806]
14. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323 [FREE Full text] [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
15. Scott K, Beatty L. Feasibility study of a self-guided cognitive behaviour therapy internet intervention for cancer carers. Aust J Prim Health 2013;19(4):270-274. [doi: 10.1071/PY13025] [Medline: 23796231]
16. Ferwerda M, van Beugen S, van Burik A, van Middendorp H, de Jong EM, van de Kerkhof PC, et al. What patients think about E-health: patients' perspective on internet-based cognitive behavioral treatment for patients with rheumatoid arthritis and psoriasis. Clin Rheumatol 2013 Jun;32(6):869-873. [doi: 10.1007/s10067-013-2175-9] [Medline: 23354514]
17. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
18. Kindon S, Pain R, Kesby M, editors. Participatory action research approaches and methods: connecting people, participation and place. New York: Routhledge; 2008.
19. Feilzer MY. Doing mixed methods research pragmatically: implications for the rediscovery of pragmatism as a research paradigm. J Mix Methods Res 2009;4(1):6-16. [doi: 10.1177/1558689809349691]
20. Goeman D, Michael J, King J, Luu H, Emmanuel C, Koch S. Partnering with consumers to develop and evaluate a Vietnamese Dementia Talking-Book to support low health literacy: a qualitative study incorporating codesign and participatory action research. BMJ Open 2016 Sep 26;6(9):e011451 [FREE Full text] [doi: 10.1136/bmjopen-2016-011451] [Medline: 27670516]
21. O'Brien N, Heaven B, Teal G, Evans EH, Cleland C, Moffatt S, et al. Integrating evidence from systematic reviews, qualitative research, and expert knowledge using co-design techniques to develop a web-based intervention for people in the retirement transition. J Med Internet Res 2016 Aug 03;18(8):e210 [FREE Full text] [doi: 10.2196/jmir.5790] [Medline: 27489143]
22. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. Br Med J 2000 Jan 8;320(7227):114-116 [FREE Full text] [Medline: 10625273]
23. Manhal-Baugus M. E-therapy: practical, ethical, and legal issues. Cyberpsychol Behav 2001 Oct;4(5):551-563. [doi: 10.1089/109493101753235142] [Medline: 11725648]
24. Karageorge A, Murphy MJ, Newby JM, Kirsten L, Andrews G, Allison K, et al. Acceptability of an internet cognitive behavioural therapy program for people with early-stage cancer and cancer survivors with depression and/or anxiety: thematic findings from focus groups. Support Care Cancer 2017 Jul;25(7):2129-2136. [doi: 10.1007/s00520-017-3617-8] [Medline: 28213818]
25. Gerhards SA, Abma TA, Arntz A, de Graaf LE, Evers SM, Huibers MJ, et al. Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: a qualitative study on patient experiences. J Affect Disord 2011 Mar;129(1-3):117-125. [doi: 10.1016/j.jad.2010.09.012] [Medline: 20889214]
26. Wang J, Lam RW, Ho K, Attridge M, Lashewicz BM, Patten SB, et al. Preferred features of E-mental health programs for prevention of major depression in male workers: results from a Canadian national survey. J Med Internet Res 2016;18(6):e132 [FREE Full text] [doi: 10.2196/jmir.5685] [Medline: 27267782]

27. 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]
28. Stubbings DR, Rees CS, Roberts LD, Kane RT. Comparing in-person to videoconference-based cognitive behavioral therapy for mood and anxiety disorders: randomized controlled trial. J Med Internet Res 2013;15(11):e258 [FREE Full text] [doi: 10.2196/jmir.2564] [Medline: 24252663]
29. Ciuca AM, Berger T, Crişan LG, Miclea M. Internet-based treatment for Romanian adults with panic disorder: protocol of a randomized controlled trial comparing a Skype-guided with an unguided self-help intervention (the PAXPD study). BMC Psychiatry 2016 Jan 14;16:6 [FREE Full text] [doi: 10.1186/s12888-016-0709-9] [Medline: 26769021]
30. Holst A, Nejati S, Björkelund C, Eriksson MC, Hange D, Kivi M, et al. Patients' experiences of a computerised self-help program for treating depression - a qualitative study of Internet mediated cognitive behavioural therapy in primary care. Scand J Prim Health Care 2017 Mar;35(1):46-53 [FREE Full text] [doi: 10.1080/02813432.2017.1288813] [Medline: 28277055]
31. Baumeister H, Seifferth H, Lin J, Nowoczin L, Lüking M, Ebert D. Impact of an acceptance facilitating intervention on patients? acceptance of Internet-based pain interventions: a randomized controlled trial. Clin J Pain 2015 Jun;31(6):528-535. [doi: 10.1097/AJP.0000000000000118] [Medline: 24866854]
32. Mitchell N, Gordon PK. Attitudes towards computerized CBT for depression amongst a student population. Behav Cognit Psychother 2007 May 14;35(04):421. [doi: 10.1017/s1352465807003700]
33. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 1;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
34. Kidd SA, Kral MJ. Practicing participatory action research. J Counsel Psychol 2005;52(2):187-195. [doi: 10.1037/0022-0167.52.2.187]
35. Munns A, Toye C, Hegney D, Kickett M, Marriott R, Walker R. Peer-led Aboriginal parent support: program development for vulnerable populations with participatory action research. Contemp Nurse 2017 Aug 02;53(5):558-575. [doi: 10.1080/10376178.2017.1358649] [Medline: 28738761]
36. Soong CS, Wang MP, Mui M, Viswanath K, Lam TH, Chan SS. A “Community Fit” community-based participatory research program for family health, happiness, and harmony: design and implementation. JMIR Res Protoc 2015;4(4):e126 [FREE Full text] [doi: 10.2196/resprot.4369] [Medline: 26510960]

Abbreviations

CBT: cognitive behavior therapy
ERP: expert research partner
PAR: participatory action research
PRP: parent research partner
PTSS: post-traumatic stress symptoms
Q&A: question and answer
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.11.17; peer-reviewed by C Wakefield, H Miller, M Lucassen; comments to author 27.12.17; accepted 05.02.18; published 18.04.18.

Please cite as:
Wikman A, Kukkola L, Börjesson H, Cernvall M, Woodford J, Grönqvist H, von Essen L
Development of an Internet-Administered Cognitive Behavior Therapy Program (ENGAGE) for Parents of Children Previously Treated for Cancer: Participatory Action Research Approach
J Med Internet Res 2018;20(4):e133
URL: http://www.jmir.org/2018/4/e133/
doi:10.2196/jmir.9457
PMID:29669710

Original Paper

Assessing Therapeutic Alliance in the Context of mHealth Interventions for Mental Health Problems: Development of the Mobile Agnew Relationship Measure (mARM) Questionnaire

Katherine Berry[1*], BSc, MSc, ClinPsyD, PhD; Amy Salter[1*], BSc, MSc; Rohan Morris[1*], BSc, PhD; Susannah James[1*], BSc, MSc; Sandra Bucci[1*], BSc (Hons), ClinPsyD

Division of Psychology and Mental Health, School of Health Sciences, University of Manchester, Manchester, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Katherine Berry, BSc, MSc, ClinPsyD, PhD
Division of Psychology and Mental Health
School of Health Sciences
University of Manchester
Oxford Road
Manchester, M139PL
United Kingdom
Phone: 44 1613060422
Fax: 44 1613060400
Email: katherine.berry@manchester.ac.uk

Abstract

Background: Digital health interventions in the form of smartphone apps aim to improve mental health and enable people access to support as and when needed without having to face the stigma they may experience in accessing services. If we are to evaluate mobile health (mHealth) apps and advance scientific understanding, we also need tools to help us understand in what ways mHealth interventions are effective or not. The concept of therapeutic alliance, a measure of the quality of the relationship between a health care provider and a service user, is a key factor in explaining the effects of mental health interventions. The Agnew Relationship Measure (ARM) is a well-validated measure of therapeutic alliance in face-to-face therapy.

Objective: This study presented the first attempt to (1) explore service users' views of the concept of relationship within mHealth mental health interventions and (2) adapt a well-validated face-to-face measure of therapeutic alliance, the Agnew Relationship Measure (ARM), for use with mHealth interventions.

Methods: In stage 1, we interviewed 9 mental health service users about the concept of therapeutic alliance in the context of a digital health intervention and derived key themes from interview transcripts using thematic analysis. In stage 2, we used rating scales and open-ended questions to elicit views from 14 service users and 10 mental health staff about the content and face validity of the scale, which replaced the word "therapist" with the word "app." In stage 3, we used the findings from stages 1 and 2 to adapt the measure with the support of a decision-making algorithm about which items to drop, retain, or adapt.

Results: Findings suggested that service users do identify relationship concepts when thinking about mHealth interventions, including forming a bond with an app and the ability to be open with an app. However, there were key differences between relationships with health professionals and relationships with apps. For example, apps were not as tailored and responsive to each person's unique needs. Furthermore, apps were not capable of portraying uniquely human-like qualities such as friendliness, collaboration, and agreement. We made a number of changes to the ARM that included revising 16 items; removing 4 items due to lack of suitable alternatives; and adding 1 item to capture a key theme derived from stage 1 of the study ("The app is like having a member of my care team in my pocket").

Conclusions: This study introduces the mHealth version of the ARM, the mARM, that has good face and content validity. We encourage researchers to include this easy-to-use tool in digital health intervention studies to gather further data about its psychometric properties and advance our understanding of how therapeutic alliance influences the efficacy of mHealth interventions.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 34966555; http://www.isrctn.com/ISRCTN34966555 (Archived by WebCite at http://www.webcitation.org/6ymBVwKif)

(J Med Internet Res 2018;20(4):e90) doi:10.2196/jmir.8252

KEYWORDS

mobile health; health care provider; digital interventions; therapeutic alliance; mental health; measure development

Introduction

Smartphone technology is constantly evolving. Many individuals use their phones and internet regularly, with 89% of adults over the age of 16 owning and using a smartphone [1]. Health care providers are capitalizing on this societal development by exploring innovative ways of using smartphone technology to support the delivery of digital health care interventions (DHIs). Indeed, the National Health Service in the United Kingdom has a digital strategy and mobile technology tool kit that seeks to exploit the capacity of smartphones to improve the efficiency and timeliness of health care interventions [2].

There are now a variety of health-related software applications (apps) available freely to be downloaded for a whole range of health care problems, including mental health problems. The growth in apps for mental health problems is an important development. The near-constant connectivity of smartphones means that people can access support as and when needed without having to overcome the stigma or barriers they might experience in accessing traditional face-to-face mental health services [3]. Accessing face-to-face support can be particularly difficult for those with severe mental health difficulties, such as psychosis, which is characterized by a mistrust of others, impaired social functioning, and difficulties in developing relationships with others [4]. A recent systematic review of Web-based, social media, and mobile technologies for severe mental health problems found that as many as 75% to 95% of service users reported technology-based interventions to be positive and useful for their mental health [5]. Furthermore, DHIs for psychosis have the potential to reduce hospital admissions, improve symptom outcomes, and improve medication adherence [5]. However, research examining the efficacy of mobile health (mHealth) interventions is lagging behind their production. The majority of mHealth interventions are not theory-driven or evidence-based. As with face-to-face delivered interventions, there is an imperative need to evaluate the efficacy of DHIs to ensure that mHealth interventions are actually providing a beneficial treatment.

The concept of therapeutic alliance, a measure of the quality of the relationship between a health care provider and a service user, is a key factor in explaining the effects of face-to-face interventions [6,7]. However, this concept has received little empirical attention in the mHealth field. mHealth interventions present a challenge to the importance of the concepts of alliance and therapeutic relationships, as relationships with health professionals might be diminished, or, in some cases, completely absent. Studies that have investigated the concept of alliance in relation to internet-delivered mental health interventions more generally, including computerized programs, have either assessed therapeutic relationship with a therapist supporting the person to access the technology, or assessed the relationship with the technological device or program itself [8]. These studies have suggested that the concept of alliance may be a less robust predictor of outcomes than in traditional face-to-face interventions for mental health problems, particularly when the direct role of the therapist is minimal and service users are asked to comment on their relationship with a computer or mobile device rather than a therapist assisting with a computer-based intervention (eg, [9-11]). However, there is some evidence that higher scores on these alliance measures are associated with more engagement with interventions [11].

It has been suggested that the relative lack of attention that developers of computerized interventions pay to the relationship-building qualities of self-help technologies for mental health interventions may partly account for the smaller effect sizes compared with face-to-face therapies and higher rates of attrition. For example, the responsivity to data entered, the degree of individually tailored responses, the consistency of advice, and the use of illustrative characters could all enhance the sense of relationship with the device or program in question [8]. There are also problems with how existing studies have measured the concept of alliance within this emerging field. Thus far, existing studies have used measures of alliance developed for use in face-to-face interventions, substituting the word “therapist” with “program.”

In this paper, we described 3 stages of a research process that ultimately aimed to develop a measure of alliance within DHIs. We aimed to build on previous research by exploring service users’ and mental health professionals’ concept of relationship with mHealth interventions and investigated how we might enhance the relationship element of existing mHealth interventions (stage 1). We also used this knowledge from stage 1 of the study, combined with an assessment of the face and content validity of a face-to-face alliance measure (stage 2) to more rigorously adapt the measure for use within the digital health context (stage 3).

Methods

Overall Context of the Research

The service user participants from all phases of this study were identified through the Actissist trial [12] (trial registration: ISRCTN34966555), a proof-of-concept trial investigating the feasibility and acceptability of a theory-driven, smartphone-delivered psychological intervention targeting areas of distress in early psychosis. Service users who participated in this research were approached on the basis that they were aged 16 years or older and had been registered under early intervention services across the North West of England, United Kingdom. Consent to participate in this study was obtained through the Actissist trial protocol, which was approved by the relevant ethics and research governance committees. This mixed-methods study occurred across two stages as described below and consisted of 33 participants (n=23 service users; n 10 mental health staff).

Stage 1: Qualitative Study

Nine service users (7 males, 2 females) participated in one-to-one, face-to-face interviews. A researcher interviewed participants about their views and experience of the Actissist app as well as the concept of therapeutic alliance related to DHIs. Specific questions that pertained most directly to the concept of therapeutic alliance were as follows:

1. Did you feel as though you were in a partnership with the app, like you were working with the app towards a shared goal or purpose?
   Prompt: What are your thoughts about the partnership? Can you tell me more about this experience?
2. Did you think that the app was supportive?
3. Prompt: Why? In what way? Examples?
4. Did you find yourself looking to the app for solutions to your difficulties?
5. Prompt: Can you tell me of an example? How did you feel about looking to an app for solutions?'

Interviews lasted from 45 to 60 min, were recorded, and transcribed verbatim. Interviews were analyzed using thematic analysis [13]. Throughout the analysis, pertinent excerpts of the concept of *relationship* were deductively extracted and labeled with codes. Codes were then organized together to develop themes. The analysis was undertaken by AS under the guidance of KB and SB during regular coding meetings where transcripts were read and coded by all team members to check for consistency in ratings. The organization of the final set of codes into themes was discussed and agreed by the research team.

AS is a postgraduate student studying for a Master's degree in Clinical and Health Psychology. She has limited direct experience of working with people with mental health problems or in psychological therapies without or without digital components. KB and SB are both clinical psychologists and academics who are experienced in developing therapeutic relationships and carrying out therapy with people with mental health problems. KB has published widely on the concept of therapeutic alliance in people with psychosis. SB has also published on the therapeutic alliance and is the lead investigator on two major trials of DHI for people with psychosis, which includes the Actissist study described in this paper.

Stage 2: Assessment of Content and Face Validity of the Agnew Relationship Measure

The Agnew Relationship Measure (ARM) is a well-validated measure of therapeutic alliance in face-to-face therapy [14]. The ARM was used as a platform to help develop a measure specifically to assess the concept of alliance in mHealth interventions, which we term the mobile Agnew Relationship Measure (mARM). We chose to adapt the ARM, as this measure has been used in previous studies assessing therapeutic alliance in the context of computerized and mobile-based mental health interventions (e, [9,11,15]). The ARM assesses five concepts thought to comprise therapeutic alliance: bond, partnership, confidence, openness, and client initiative. Bond encompasses feelings of positive regard from and toward the therapist; partnership concerns the collaboration between the client and the therapist; confidence pertains to the competency of the therapist; openness is characterized by the freedom of personal disclosure; and client initiative is associated with feelings of control and empowerment. Consistent with previous studies, the first iteration of the measure that we shared with participants simply replaced the word "therapist" with "app."

In addition to the ARM, a questionnaire was developed to explore individuals' opinions about the ARM as a measure for assessing therapeutic alliance in mHealth interventions for psychosis. This included a relevancy scale for each item whereby participants were asked to assess each question on how relevant they believed the item to be to the concept of therapeutic alliance with a DHI on a scale of 1 to 4 (1=not relevant, 2=somewhat relevant, 3=quite relevant, 4=highly relevant). This scale also investigated the wording and format of the questionnaire using open-ended questions that prompted participants to expand on their reasoning or suggest alternative wording of items.

Fourteen service users (none of whom participated in the first part of the study), and 10 members of staff took part in the assessment of the content and face validity of the measure. All participants were presented with the version of the ARM and relevancy scale described above. For convenience, this was done through a range of methods. Eleven participants were interviewed over the phone, 11 were interviewed as part of a small group discussion, and 2 members of staff returned the questionnaire via email. For each method, participants were given at least 24 hours to consider participating in this study. Initially, the participants were contacted over the phone and given information about the study, including the aims of the research. Verbal consent was obtained and a phone interview or face-to-face meeting date was arranged. The relevancy scale (alongside the ARM) was sent to people via email so they were able to familiarize themselves with the measure before the interview.

Stage 3: Development of the Mobile Agnew Relationship Measure (mARM)

We developed an algorithm for scrutinizing and adapting ARM's items. The rationale for the algorithm was to support our decision-making process about which items to retain, drop, or reword. The algorithm stated that if at least one service user or staff member rated an item "not relevant" or one participant suggested an alternative wording, then the item was discussed by the research team. For each of these items, an alternative wording was debated, with reference to comments made by staff or service users. Themes and excerpts from stage 1 of the research were also drawn upon in considering the relevance of the item or alternative wording. If the alliance concept of an item did not translate to an mHealth context and no suitable alternative wording could be identified, then the item was removed from the measure.

Results

Stage 1: Qualitative Study

Two overarching alliance-related themes were identified from the interviews: (1) forming a bond and relationship with the app; and (2) preference for an app instead of a human therapist.

Table 1. Summary of themes from stage 1 with illustrative quotes.

Theme	Illustrative quote
Forming a bond and relationship with the app	
Building a supportive relationship with the app	"I actually miss doing it cause it does help on different days. Not doing it, I miss it." [Participant 6]
	"Sometimes, it would have been better for me, I would have preferred to have the app, even now, even though I probably wouldn't use it for 2 or 3 weeks, and still be relying on the information. There comes a time where you mind goes blank to the situation and you're in complete panic mode and to be able to say 'I need someone. I can't get [Care Coordinator] on the phone, I can't get my care-worker on the phone. What do I do?'" [Participant 1]
Mimic human support	"It's another tool in the arsenal...your CPN, he's only there once a week, there's a long period of time when you're on your own, you've got nobody. You feel like you're not wanted, not needed, and that app, it says 'you've got a CPN in your pocket, I've got a care provider in my pocket that I can go out quite freely now without my CPN.'" [Participant 1]
	"It did start to feel like part of my normal routine...it was sort of like having a buddy um so yeah, every time it sort of asked you to check in, it was quite a good feeling...it started to seem natural very early..." [Participant 1]
Barriers to bond	"There's the reassuring side of I'm not going insane, but theres also the cons to it as well...It's only a phone telling me I'm not going insane, it's not actually a Dr saying well actually, you're not that bad." [Participant 9]
	"I did wonder if it'd be able to portray information as well as a person. Which obviously it can't do because you can't ask it questions." [Participant 5]
Preference for an app instead of a human therapist	
Apps provide greater freedom than face-to-face	"If you are feeling criticised, you can actually be honest with it...Say, I'm feeling criticised that day and say, be honest with it, it's - it, yeah, so that's what's vital I think. It can create somewhere, instead of having whoever you live with... instead of opening up and having a [an argument], releasing it that way, you can release it into [the app]." [Participant 1]
Apps promote self exploration	"With the app, you do find yourself opening up a little bit more, personally, on your own privately with your app...Not with your psychologist, not with a CPN, but with yourself and that's when you find out your being a bit more honest with yourself, you're able to articulate that with your psychologist 'this is how I'm feeling, I know because I've had it on the app'...I've got the app there to be, alone, on my own. I can be an emotional wreck if I want to be, I can be a blubbery, crying baby again if I want to be cons I'm on my own, I'm with my app and it's getting me through this and that's what I got from it." [Participant 1]
	"...You were able to tell it what was worrying you uh - you didn't feel like you couldn't express the worries so um that was good." [Participant 2]

Themes are summarized in Table 1, along with illustrative quotes supporting the themes generated.

Theme 1: Forming a Bond and Relationship With the App

The accounts given by the service users demonstrated that they felt a strong sense of support from the app and consequently, when the app was no longer available for use, people reported missing the support that it had provided. This perceived loss suggested that participants had potentially formed a relational bond with the app. In addition, participants reported that the app provided a sense of security by giving instant support when required.

Several service users indicated that using the app was akin to always having a member of their care team available. Some described it as having a "therapist" in their pocket. Furthermore, a number of participants reported feeling as though they had a friend in the app that would offer encouragement and reassurance. However, there were also accounts where some participants felt that the app was not tailored to their individual needs; the information provided was either repetitive and/or was too generic. This notion was deemed relevant to the concept of developing a relationship with the app, as participants perceived that it was these "robotic" features that hindered the development of a relational bond.

Theme 2: Preference for an App Instead of a Human Therapist

In some accounts, participants alluded to a freedom to be honest and open with the app, which was noted as a key difference in the app compared with face-to-face therapy. For example, there seemed to be a reduced risk of embarrassment from confiding in the app as opposed to confiding in a person. This theme was deemed relevant to the concept of relationship, as embarrassment or perceptions of therapist judgment may be a factor in the rupture in face-to-face relationships. Several participants discussed how the app helped them to be honest with themselves and that further down the line, this improved their ability to be open with members of their care team. In this respect, communicating with the app served as a bridge that helped participants to communicate with other people in their social world.

Table 2. Service user and staff relevancy rating frequency count. There were 24 respondents overall, but not all respondents provided data for each item.

Modified ARM[a] items	Not relevant n (%)	Somewhat relevant n (%)	Quite relevant n (%)	Highly relevant n (%)
I feel free to express the things that worry me	1 (5)	0 (0)	9 (45)	10 (50)
I feel friendly towards the app	1 (5)	4 (21)	8 (42)	6 (32)
I am worried about embarrassing myself when using the app	6 (32)	5 (26)	5 (26)	3 (16)
I take the lead when using the app	0 (0)	6 (32)	9 (47)	4 (21)
I keep some important things to myself and don't share them with the app	5 (26)	2 (11)	5 (26)	7 (37)
I have confidence in the app and its techniques	0 (0)	0 (0)	10 (53)	9 (47)
I feel optimistic about my progress	0 (0)	1 (5)	13 (68)	5 (26)
I feel I can openly express my thoughts and feelings when using the app	0 (0)	2 (11)	8 (42)	9 (47)
I feel critical or disappointed in the app	5 (26)	6 (32)	4 (21)	4 (21)
I can share personal matters I am ordinarily ashamed or afraid to reveal	2 (11)	4 (21)	6 (32)	7 (37)
I look to the app for solutions to my problems	0 (0)	6 (32)	8 (42)	5 (26)
The app's skills are impressive	1 (5)	8 (42)	8 (42)	2 (11)
The app accepts me no matter how I respond	2 (11)	0 (0)	10 (53)	7 (37)
I feel influenced by the app in ways that are not beneficial to me	6 (32)	3 (16)	6 (32)	4 (21)
The app finds it hard to understand me	7 (37)	3 (16)	5 (26)	4 (21)
The app's approach is warm and friendly with me	1 (5)	5 (26)	5 (26)	8 (42)
The app does not give me the guidance I would like	5 (26)	4 (21)	4 (21)	6 (32)
The app feels persuasive	4 (21)	7 (36)	5 (26)	3 (16)
The app is supportive	1 (5)	1 (5)	11 (58)	6 (32)
The app follows its own plans, ignoring my views on how to proceed	6 (32)	3 (16)	5 (26)	5 (26)
The app is confident in its messages and techniques	3 (16)	3 (16)	7 (37)	6 (32)
The app seems bored or impatient with me	6 (32)	5 (26)	6 (32)	2 (11)
The app expects me to take responsibility rather than be dependent on it	4 (21)	6 (32)	7 (37)	2 (11)
The app and I are willing to work hard together	3 (16)	6 (32)	8 (42)	1 (11)
I take the lead and the app expects it of me	2 (11)	7 (37)	7 (37)	3 (16)
The app and I agree about how to work together	3 (16)	3 (16)	9 (47)	4 (21)
The app and I have difficulty working jointly as a partnership	5 (26)	6 (32)	6 (32)	2 (11)
The app and I are clear about our role and responsibilities when we interact	2 (11)	6 (32)	8 (42)	3 (16)

[a]ARM: Agnew Relationship Measure.

Stage 2: Assessment of Face and Content Validity of the Agnew Relationship Measure

Stage 2 of the study aimed to assess the face and content validity of the first iteration of the mARM, which involved simply replacing the word "therapist" with the word "app." The number of participants who rated each item as not relevant, somewhat relevant, quite relevant, and highly relevant is presented in Table 2. An overarching theme within the relevancy scale and open-ended questions was that some items inappropriately anthropomorphized the app. Both staff and service users reported finding some items hard to answer or relate to the app due to the human nature of the item. For example, the partnership subscales of the ARM, such as agreeing a goal with an app or the concept of the app and the person working together, were particularly criticized. Several staff highlighted the need to qualify items with a perception rather than a statement of fact in instances where human qualities were referenced. For example, replacing phrases, such as "the app is supportive" with "the app seems supportive." However, staff and service users were more likely to endorse relationship items which reflected their own thoughts and feelings about the app, suggesting that the concept of relationship was important even if it was a less reciprocal relationship. For example, the majority of staff and service users reported that the item, "I feel friendly towards the app" was relevant. Items which referred to techniques proposed by the app or confidence in the app's capacity to help, which were not uniquely human qualities were also generally less problematic for participants. For example, items such as, "I have confidence in the app and its techniques" or "I feel

optimistic about my progress" were endorsed by most people as being relevant. Similarly, items that involved expressing thoughts and feelings, such as "I feel I can openly express my thoughts and feelings when using the app" were felt to be relevant.

Stage 3: Development of the Mobile Agnew Relationship Measure

Stage 3 of the study aimed to use the results of stages 1 and 2 to reach a consensus about which items to include in the mARM. Eight of the original items were retained, 16 were revised, 4 were removed due to a lack of suitable alternatives and one was added to capture a key theme from the interviews in stage 1 that we felt was not adequately captured by other existing items or other reworded items ("The app is like having a member of my care team in my pocket"). The changes we made to the measure along with the rationale for changing, not changing, or removing items are displayed in Multimedia Appendix 1. The finalized version of the mARM can be found in Multimedia Appendix 2.

Discussion

Summary of Findings

This study aimed to explore service users' views of the concept of *relationship* within mental health DHIs and to examine and adapt a well-validated face-to-face measure of therapeutic alliance for use with mHealth interventions. Stage 1 of this mixed-methods project suggested that service users were able to form a bond and relationship with the app; although, the generic nature of the app's messages that at times did not feel personalized could hinder the development of this therapeutic relationship. Stage 1 of this study also revealed that some people preferred a relationship with an app as opposed to a relationship with a human therapist. For example, people could be open and honest with the app and this had the potential to promote future communication with other people. Stage 2 of this project suggested that simply replacing the word "therapist" with "the app" in traditional measures of alliance is insufficient to capture the nuances of alliance in mHealth. The findings of stages 1 and 2 were used to inform the development of an mHealth measure of alliance with good content and face validity as determined by both service users and mental health professionals.

Contextualizing Findings Within Existing Literature

The majority of existing studies have explored alliance in the context of interventions with some degree of overt therapist input into a client's treatment, such as a therapist assisting the person to use an internet-based self-help program. In these instances, it is difficult to conclude whether existing findings reflected the quality of clients' working relationships with the therapists involved or with the computer programs themselves. Our study is one of the few studies that have examined the relationship concept in the complete absence of a therapist. Other similar studies have focused on alliance ratings to computer programs for mild to moderate mental health problems [9-11]. These studies suggested that alliance ratings in relation to computer programs were similar to therapist alliance ratings; although, alliance ratings were slightly lower than typical ratings of therapists and alliance was a less important predictor of outcome. However, all of these studies reworded traditional alliance substituting the word "therapist" with "program" (or equivalent term). Future research needs to assess the extent to which the mARM, which was specifically adapted to assess alliance in the mHealth context, is a better predictor of therapy outcomes. However, our finding that some participants felt that the app was not tailored to their individual needs suggested that apps may not yet be sophisticated to mimic human qualities of responsivity and sensitivity to a sufficient degree. At present, truly dialogic engagement may present a challenge for even the most intelligent automated approaches due to its highly contextual and interactive requirements [16].

The findings that people could be more open with the app due to reduced fears about embarrassing themselves compared with talking to a person might suggest that the nonhuman element to the app could prove helpful for some people. In this respect, apps serve a slightly different function to relationships with therapists and consequently perhaps researchers should not be aspiring to fully mimic human interactions when developing a DHI for mental health. Indeed, our adaptions to the ARM items focused on reducing the humanization of the app, which many of the service users and staff in our study found hard to relate to. In making these changes, it was, however, important to hold in mind that we still needed to capture the concept of relationship within our measures rather than developing a measure that purely assesses satisfaction with an app or program.

We are aware that there are existing measures of the quality and functionality of apps [17]. For example, the user Mobile Application Rating Scale ([18]) assesses factors such as an app's customization, interactivity, ease of use, and the quality and credibility of information. Although the two concepts of quality and alliance are undoubtedly related (eg, if an individual is not satisfied with the quality of the app, it is unlikely that s/he will develop a positive alliance with the app), it is important that measures of alliance in mHealth are sufficiently focused on concepts directly relevant to alliance such as bond, responsiveness to need, and communication, as opposed to other more general indices of app quality.

Our finding that people could be open with the app and empowered by the app reflected Clarke et al's [11] finding that alliance subscales measuring perceived empowerment and perceived freedom to self-disclose were significantly positively correlated with self-monitoring frequency, suggesting better engagement with the program. Moreover, participants' ratings of the quality of their emotional connection with the program were positively correlated with program log-ins, frequency of self-monitoring, and number of treatment modules completed. These findings are particularly important as one of the concerns about mHealth interventions is that people are more likely to disengage from them compared with face-to-face therapies [8].

Limitations

There are some limitations that need to be accounted for when interpreting our findings and considering future uses of the revised mARM. While this research has developed a user-informed measure of alliance for mHealth interventions,

the measure requires more rigorous validation. Arguably, this research may not have captured a truly representative range of views about how to measure alliance in mHealth interventions, in particular as participants were recruited from a single clinical population, first episode psychosis.

We attempted to capture the themes derived from stage 1 of the research by rephrasing items in the ARM, using information from the interviews. We also explored an additional item to tap the concept of the app being like a therapist in the pocket. However, arguably the process of basing the measure on the ARM, as opposed to generating items solely from the interviews may have resulted in us missing some potentially important concepts that are uniquely relevant to the assessment of alliance for apps, including the degree to which participants perceive the app's responses as automated and robotic.

Summary of Implications

The mARM needs to be subjected to further empirical testing with a wider range of clinical groups. As part of this process, we need to explore whether the mARM is capable of predicting outcomes in DHIs in the same way that therapeutic alliance measures predict outcomes in face-to-face therapy with human therapists. It would also be important to compare the predictive validity of the ARM, which is specifically designed to assess the relationship aspect of therapy with more generic measures of app satisfaction.

Conclusions

The mARM has attempted to capture unique elements of a digital therapeutic relationship from user feedback. Feedback was obtained from both service users who had first-hand experience with an mHealth intervention and from experienced mental health professionals. The suggestions of these participants and the consequent adaptations we made to an existing, well-established measure of alliance have resulted in an alliance measure with good content and face validity that can assess alliance in mHealth contexts. Simply replacing "therapist" with "the app" in an established measure of alliance is insufficient to capture the nuances of therapeutic alliance in mHealth. The next crucial step in this program of research is to carry out a more comprehensive assessment of the psychometric properties of the scale, ideally within the context of a large mHealth trial, so that the association with outcomes can be systematically explored.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Modification and justification of the amended Agnew Relationship Measure (ARM).

[PDF File (Adobe PDF File), 32KB - jmir_v20i4e90_app1.pdf]

Multimedia Appendix 2

The mARM—the amended Agnew Relationship Measure for use in mHealth interventions.

[PDF File (Adobe PDF File), 23KB - jmir_v20i4e90_app2.pdf]

References

1. Ofcom. 2015. The Communications Market Report 2015 URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0017/12833/cmr-2015-analyst_briefing.pdf [accessed 2018-02-23] [WebCite Cache ID 6xRaYuVxp]
2. Mental Health Taskforce. 2016. The Five Year Forward View for Mental Health URL: https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf [accessed 2018-02-20] [WebCite Cache ID 6xNGFfycc]
3. Proudfoot J. The future is in our hands: the role of mobile phones in the prevention and management of mental disorders. Aust N Z J Psychiatry 2013 Feb;47(2):111-113. [doi: 10.1177/0004867412471441] [Medline: 23382507]
4. The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.
5. 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]
6. 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]
7. Priebe S, Richardson M, Cooney M, Adedeji O, McCabe R. Does the therapeutic relationship predict outcomes of psychiatric treatment in patients with psychosis? A systematic review. Psychother Psychosom 2011;80(2):70-77. [doi: 10.1159/000320976] [Medline: 21196804]
8. 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]

9. 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]
10. 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]
11. Clarke J, Proudfoot J, Whitton A, Birch MR, Boyd M, Parker G, et al. Therapeutic alliance with a fully automated mobile phone and web-based intervention: secondary analysis of a randomized controlled trial. JMIR Ment Health 2016 Feb 25;3(1):e10 [FREE Full text] [doi: 10.2196/mental.4656] [Medline: 26917096]
12. Bucci S, Barrowclough C, Ainsworth J, Morris R, Berry K, Machin M, et al. Using mobile technology to deliver a cognitive behaviour therapy-informed intervention in early psychosis (Actissist): study protocol for a randomised controlled trial. Trials 2015;16(1):404 [FREE Full text] [doi: 10.1186/s13063-015-0943-3] [Medline: 26357943]
13. Braun V, Clarke V. Successful Qualitative Research: A practical guide for beginners. London: SAGE Publications Ltd; 2013.
14. Agnew-Davies R, Stiles WB, Hardy GE, Barkham M, Shapiro DA. Alliance structure assessed by the Agnew Relationship Measure (ARM). Br J Clin Psychol 1998 May;37(Pt 2):155-172. [Medline: 9631204]
15. Kay-Lambkin F, Baker A, Lewin T, Carr V. Acceptability of a clinician-assisted computerized psychological intervention for comorbid mental health and substance use problems: treatment adherence data from a randomized controlled trial. J Med Internet Res 2011;13(1):e11 [FREE Full text] [doi: 10.2196/jmir.1522] [Medline: 21273184]
16. Mason RB. Student engagement with, and participation in, an e-forum. J Educ Techno Soc 2011;14(2):258-268.
17. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
18. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]

Abbreviations

ARM: Agnew Relationship Measure
DHI: digital health interventions
mARM: mobile Agnew Relationship Measure
mHealth: mobile health

Edited by G Eysenbach; submitted 04.10.17; peer-reviewed by M Ashford, S Stoyanov, E Scherer, M Puttaroo, L Odom; comments to author 29.11.17; revised version received 08.12.17; accepted 08.12.17; published 19.04.18.

Please cite as:
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;20(4):e90
URL: http://www.jmir.org/2018/4/e90/
doi:10.2196/jmir.8252
PMID:29674307

Original Paper

Trajectories of 12-Month Usage Patterns for Two Smoking Cessation Websites: Exploring How Users Engage Over Time

Jonathan B Bricker[1,2], PhD; Vasundhara Sridharan[1,2], MS; Yifan Zhu[1], PhD; Kristin E Mull[1], MS; Jaimee L Heffner[1], PhD; Noreen L Watson[1], PhD; Jennifer B McClure[3], PhD; Chongzhi Di[1], PhD

[1]Fred Hutchinson Cancer Research Center, Seattle, WA, United States

[2]Department of Psychology, University of Washington, Seattle, WA, United States

[3]Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States

Corresponding Author:
Jonathan B Bricker, PhD
Fred Hutchinson Cancer Research Center
1100 Fairview Ave N
Seattle, WA, 98109
United States
Phone: 1 2066675074
Email: jbricker@fredhutch.org

Abstract

Background: Little is known about how individuals engage with electronic health (eHealth) interventions over time and whether this engagement predicts health outcomes.

Objective: The objectives of this study, by using the example of a specific type of eHealth intervention (ie, websites for smoking cessation), were to determine (1) distinct groups of log-in trajectories over a 12-month period, (2) their association with smoking cessation, and (3) baseline user characteristics that predict trajectory group membership.

Methods: We conducted a functional clustering analysis of 365 consecutive days of log-in data from both arms of a large (N=2637) randomized trial of 2 website interventions for smoking cessation (WebQuit and Smokefree), with a primary outcome of 30-day point prevalence smoking abstinence at 12 months. We conducted analyses for each website separately.

Results: A total of 3 distinct trajectory groups emerged for each website. For WebQuit, participants were clustered into 3 groups: 1-week users (682/1240, 55.00% of the sample), 5-week users (399/1240, 32.18%), and 52-week users (159/1240, 12.82%). Compared with the 1-week users, the 5- and 52-week users had 57% higher odds (odds ratio [OR] 1.57, 95% CI 1.13-2.17; *P*=.007) and 124% higher odds (OR 2.24, 95% CI 1.45-3.43; *P*<.001), respectively, of being abstinent at 12 months. Smokefree users were clustered into 3 groups: 1-week users (645/1309, 49.27% of the sample), 4-week users (395/1309, 30.18%), and 5-week users (269/1309, 20.55%). Compared with the 1-week users, 5-week users (but not 4-week users; *P*=.99) had 48% higher odds (OR 1.48, 95% CI 1.05-2.07; *P*=.02) of being abstinent at 12 months. In general, the WebQuit intervention had a greater number of weekly log-ins within each of the 3 trajectory groups as compared with those of the Smokefree intervention. Baseline characteristics associated with trajectory group membership varied between websites.

Conclusions: Patterns of 1-, 4-, and 5-week usage of websites may be common for how people engage in eHealth interventions. The 5-week usage of either website, and 52-week usage only of WebQuit, predicted a higher odds of quitting smoking. Strategies to increase eHealth intervention engagement for 4 more weeks (ie, from 1 week to 5 weeks) could be highly cost effective.

Trial Registration: ClinicalTrials.gov NCT01812278; https://www.clinicaltrials.gov/ct2/show/NCT01812278 (Archived by WebCite at http://www.webcitation.org/6yPO2OIKR)

(J Med Internet Res 2018;20(4):e10143) doi:10.2196/10143

KEYWORDS

engagement; trajectories; eHealth; websites; tobacco; smoking; acceptance and commitment therapy; smokefree.gov; patient participation; telemedicine; tobacco use cessation; smoking cessation

Introduction

Electronically delivered health interventions (or eHealth interventions), such as websites and mobile apps, have been successful methods of health behavior change [1-4]. In this body of research, people who engage more with eHealth interventions tend to have better treatment outcomes [5]. However, while eHealth intervention engagement is usually measured with simple counts of the number of log-ins and modules completed [5], little is known about how users engage with eHealth interventions *over time* and whether those temporal patterns predict better treatment outcomes. In the educational literature, a well-documented finding is that learning new material becomes more effective when it occurs over a longer period of time as opposed to over a short period of time [6]. This process, called spaced practice, works by way of increasing variability in learning and remembering new information [7].

Websites and mobile apps for health behavior change are usually available for participants to use at will, which results in high variations of individual usage patterns, or usage trajectories, over time. For example, some users may follow a trajectory of logging in several times within the first few days of starting an intervention and then never return. Others may follow a trajectory where they log in consistently and then gradually taper off. And other users may follow a trajectory where they consistently log in over the course of many months. It is possible that some groups of individuals follow unique usage trajectories over time that are associated with differential health outcomes. For example, people who log in consistently over the course of many months might have positive health outcomes because they have consistently benefited from the information and skills presented in the intervention. Alternatively, consistent log-ins may be a marker of ongoing challenges and struggles to change a health behavior, and thus may indicate poorer treatment outcomes. Since we do not know which trajectories of use predict successful behavior change, studying distinct groups of usage trajectories that people follow can help us identify which usage patterns are beneficial and make recommendations for future program use. This will help inform the design of eHealth interventions to improve successful behavior change.

Within the social and behavioral sciences, identifying usage trajectories has been applied for several decades to understanding behavior patterns over time [8-11]. More recently, a few studies have analyzed usage trajectories for eHealth interventions. One study examined 8-week usage trajectories of a diabetes management mobile app. The study found 3 distinct trajectories of usage and described the clusters of people following these trajectories as minimal users, intermittent waning users, and consistent users [12]. However, the study was limited by a small sample size (N=84), as well as short duration (8 weeks), and whether the trajectories predicted health outcomes was not reported. Other research identified 5 distinct usage trajectories of a short message service (SMS) text-messaging–based smoking cessation program over 5 weeks, namely high engagement, increasing engagement, rapid decrease, delayed decrease, and low engagement [13]. The study found that the high engagement and increasing engagement groups were more likely than the other groups to be abstinent over the course of 5 weeks.

If eHealth intervention usage trajectories that predict health outcomes can be identified, understanding the groups of individuals who tend to follow more or less successful trajectories is an important next step. This would reveal the qualities of individuals who are likely to have engagement patterns that are related to successful and unsuccessful outcomes. Knowing these baseline characteristics might allow researchers and intervention designers to tailor eHealth interventions to users' unique challenges, needs, and limitations. While studies have found that being a woman, being older, and having a higher education are generally consistent predictors of greater eHealth intervention usage [14-17], very little is known about the user characteristics that are associated with different patterns of use over time. To our knowledge, only 1 study has examined this question [12] and found that being female and having higher baseline motivation were associated with more consistent log-in trajectories.

Using the example of smoking cessation websites, in this study we aimed to determine (1) distinct groups of log-in trajectories, (2) their prediction of the smoking cessation outcome, and (3) baseline user characteristics that are associated with different usage trajectory groups. The overall goal was to advance the study of analytic methods of user engagement and, ultimately, the design of more effective interventions that are tailored to users and their longitudinal patterns of engagement. To accomplish these aims, in this study we analyzed 365 consecutive days of log-in data from both arms of a large (N=2637), 2-arm randomized trial of website interventions for smoking cessation (NCT01812278).

Methods

Participants

As described in the main outcome article for the trial [18], we recruited participants (N=2637) from across the United States to participate in a study comparing 2 Web-delivered smoking cessation programs. Participants were recruited between March 24, 2014 and August 11, 2015. To be eligible for the study, participants had to be adult smokers in the United States (≥18 years of age), smoking at least 5 cigarettes daily, motivated to quit in the next 30 days, and have internet access. The 2637 participants were assigned to 1 of 2 Web-based smoking cessation interventions using stratified black randomization (on smoking frequency, education, and sex): WebQuit (n=1319; experimental arm) [18] or Smokefree (n=1318; control arm) [19].

Smoking Cessation Interventions

Participants accessed their assigned website with a unique username and password. For the first 4 weeks, all participants in both programs could opt to receive up to 4 short daily tips via SMS text messaging or email, which were designed to increase engagement. Participants were free to use their assigned program as they wished for 1 year from the date of enrollment.

The WebQuit program was based on acceptance and commitment therapy (ACT) [20], an approach that teaches skills

to smokers to let their urges pass without smoking. The program had 4 parts. Step 1, Make a Plan, enabled users to develop a personalized quit plan, identify smoking triggers, learn about US Food and Drug Administration (FDA)-approved cessation medications, and upload a photo of their inspiration to quit (ACT processes: Values and Committed Action). Step 2, Be Aware, contained 3 exercises to illustrate the problems with trying to control thoughts, feelings, and physical sensations rather than allowing them to come and go (ACT process: Creative Hopelessness). Step 3, Be Willing, contained 8 exercises to help users practice allowing thoughts, feelings, and physical sensations that trigger smoking (ACT processes: Willingness, Being Present, and Cognitive Defusion). Step 4, Be Inspired, contained 15 exercises to help participants identify deeply held values inspiring them to quit smoking and to exercise self-compassion in response to smoking lapses (ACT processes: Values and Self-as-Context). The program also prompted users to track smoking, cessation medications, and practice of ACT skills. Tracking results were displayed graphically along with the user's inspiration for quitting and badges earned for program use. Participants could log in and use the program as much as they liked.

For the control arm, we hosted a secured private version of the US National Cancer Institute's Smokefree.gov site. This intervention was also named WebQuit so that participants would be blinded to group assignment. Smokefree follows the US clinical practice guidelines [21] and provides standard treatment that teaches skills to smokers to avoid urges. Users were able to navigate through all pages of the website at any time, and there were no restrictions on the order in which they could view the content. Smokefree had 3 main sections: Quit Today, Preparing to Quit, and Smoking Issues. The Quit Today section had 7 pages of content that provided tips for the quit day, staying smoke-free, and dealing with cravings. The section also provided information on withdrawal, benefits of quitting, and FDA-approved cessation medications. The Prepare to Quit section had 7 content pages providing information on various reasons to quit, what makes quitting difficult, how to make a quit plan, and using social support during a quit attempt. The Smoking Issues section provided 5 pages on health effects of smoking and quitting, depression, stress, secondhand smoke, and coping with the challenges of quitting smoking for the lesbian, gay, bisexual, and transgender community. The section also contained 5 quizzes that provided feedback about level of depression, stress, nicotine dependence, nicotine withdrawal, and secondhand smoke, as well as tips for coping with them.

Measures

Baseline Characteristics

At baseline, participants reported on demographics, alcohol use, smoking history, and whether they had a partner and friends who smoked. We measured nicotine dependence with all 6 items of the Fagerström Test for Nicotine Dependence (FTND) [22]. Participants also filled out the Commitment to Quitting Scale [23], which has 8 items measuring participants' motivation to stay abstinent (example item, "I'm willing to put up with whatever discomfort I have to in order to quit smoking."). The scale, which has been used in multiple smoking cessation trials [18,24], has been shown to have good reliability and validity [23]. We screened participants for mental health conditions including depression (Center for Epidemiologic Studies Depression scale) [25], generalized anxiety (Generalized Anxiety Disorder 7-item scale) [26], panic disorder (Autonomic Nervous System Questionnaire) [27], posttraumatic stress disorder (PTSD; PTSD Checklist) [28], and social anxiety (mini-Social Phobia Inventory) [29]. We included the results as covariates and predictors, since prior research has shown that mental health symptoms are a predictor of engagement in eHealth interventions [30,31].

Engagement

For each participant, we recorded time- and date-stamped log file records of each page opening. For this analysis, we used a binary measure indicating whether each participant logged in at least once each day (ie, had at least one page opening recorded in the log file data). Using this method, we obtained for each participant a 0/1 code for each day for 365 days from the date of randomization.

Cessation Outcome

The primary outcome of the study was self-reported 30-day point prevalence abstinence (ie, no smoking at all in the past 30 days) at 12-month follow-up. Self-reported smoking or abstinence is a standard method for assessing the efficacy of Web-delivered interventions [32]. The Society for Research on Nicotine and Tobacco Subcommittee on Biochemical Verification has suggested that biochemical confirmation is not necessary in population-based studies with no face-to-face contact and in studies where data are collected through the Web, telephone, or mail because of low demand characteristics of these studies [33,34].

Statistical Analyses

To determine distinct groups of log-in trajectories for each website, we used a functional clustering approach consisting of 3 steps: (1) presmoothing the binary daily engagement time series; (2) conducting functional principal component analysis [35], a dimension reduction procedure to summarize each participant's log-in trajectory by low-dimensional functional principal component scores; and (3) applying the clustering large applications algorithm [36] to the derived functional principal component scores. This procedure does not rely on any assumptions on the shapes of trajectories and is capable of handling large datasets and complex missing data patterns. We determined the total number of trajectories for each website using predictive strength [37], which is a statistical criterion to assess how many groups can be predicted from the data and how well. We obtained each study participant's log-in trajectory by transforming longitudinal sequences of log-in time stamps into a binary time series indicating log-in occurrence each day. Note that we chose not to use latent class growth curve approaches that have been used in other eHealth intervention engagement studies [12,13] because these methods do not handle very densely recorded longitudinal data without substantial data reduction (eg, reducing data into weekly or monthly log-in counts per participant) and often rely on restrictive assumptions on the shapes of trajectories.

After determining distinct trajectory clusters, we applied logistic regression models to investigate the associations between the trajectory clusters and the smoking cessation outcome. Both unadjusted and covariate-adjusted regression models were fitted. For covariate-adjusted models, we selected variables by stepwise Akaike information criterion (AIC) in both backward and forward directions. Covariates considered for adjustment were the baseline characteristics described above in the Measures subsection, including commitment to quit smoking, to control for participant characteristics that may confound any association with cessation outcomes. Finally, to identify baseline user characteristics associated with trajectory membership, we applied multinomial logistic regression models with baseline covariates as predictors and the log-in trajectory clusters as outcome. We selected variables in the final multivariate model via a stepwise AIC procedure from a pool of candidate baseline covariates that had a univariate association with log-in trajectory clusters.

Results

Description of Sample

Table 1 shows the baseline demographics and participant characteristics in both the WebQuit and Smokefree arms. Overall, participants were on average 46 years old, about 80% were female, about 80% were white, about 52% were employed, and about 72% had greater than high school education.

Table 1. Summary of baseline characteristics of participants from both WebQuit and Smokefree arms, by log-in trajectories. LGBT: lesbian, gay, bisexual, and transgender; FTND: Fagerström Test for Nicotine Dependence; PTSD: posttraumatic stress disorder.

Participant characteristics	WebQuit (n=1240)				Smokefree (n=1309)			
	1-week users (n=682)	5-week users (n=399)	52-week users (n=159)	Overall	1-week users (n=645)	4-week users (n=395)	5-week users (n=269)	Overall
Age (years), mean (SD)	44.6 (13.6)	47.4 (12.5)	51.4 (12.6)	46.4 (13.3)	45.4 (13.2)	46.0 (13.8)	48.1 (12.8)	46.2 (13.3)
Male, n (%)	149 (21.8)	67 (16.8)	32 (20.1)	248 (20.0)	133 (20.6)	90 (22.8)	49 (18.2)	272 (20.8)
Married, n (%)	263 (38.6)	174 (43.6)	56 (35.2)	493 (39.8)	234 (36.3)	140 (35.4)	89 (33.1)	463 (35.4)
Working, n (%)	354 (51.9)	212 (53.1)	83 (52.2)	649 (52.3)	362 (56.2)	196 (49.6)	118 (43.9)	676 (51.6)
High school or less, n (%)	204 (29.9)	98 (24.6)	42 (26.4)	344 (27.7)	185 (28.7)	107 (27.1)	70 (26.0)	362 (27.7)
LGBT, n (%)	63 (9.2)	32 (8.0)	14 (8.8)	109 (8.8)	63 (9.8)	42 (10.6)	30 (11.2)	135 (10.3)
White, n (%)	558 (81.2)	323 (81.0)	123 (77.4)	1004 (81.0)	530 (82.2)	321 (81.3)	218 (81.0)	1069 (81.7)
Hispanic, n (%)	52 (7.6)	34 (8.5)	6 (3.8)	92 (7.42)	52 (8.1)	42 (10.6)	27 (10.0)	121 (9.2)
Any quit attempt in last 12 months, n (%)	269 (42.0)	165 (43.9)	70 (46.1)	504 (43.1)	285 (45.7)	169 (44.5)	121 (46.9)	575 (45.6)
FTND score, mean (SD)	5.68 (2.19)	5.54 (2.18)	5.58 (2.21)	5.62 (2.19)	5.70 (2.10)	5.71 (2.17)	5.33 (2.32)	5.63 (2.17)
Smoking characteristics								
Half a pack or more, n (%)	539 (79.0)	313 (78.4)	128 (80.5)	980 (79.0)	523 (81.1)	313 (79.2)	195 (72.5)	1031 (78.8)
>10 years, n (%)	530 (77.7)	327 (82.0)	139 (87.4)	996 (80.3)	509 (78.9)	310 (78.5)	223 (82.9)	1042 (79.6)
Partner smokes, n (%)	459 (67.3)	279 (69.9)	115 (72.3)	853 (68.8)	454 (70.4)	270 (68.4)	200 (74.3)	924 (70.6)
No. of friends who smoke, mean (SD)	2.2 (1.6)	2.1 (1.6)	2.1 (1.7)	2.2 (1.6)	2.3 (1.6)	2.1 (1.6)	2.3 (1.7)	2.2 (1.6)
Commitment score, mean (SD)	4.01 (0.74)	3.97 (0.75)	3.94 (0.77)	3.99 (0.75)	4.01 (0.77)	3.96 (0.73)	4.04 (0.82)	4.00 (0.77)
Mental health measures, n (%)								
Depression	400 (59.0)	208 (52.3)	80 (50.3)	688 (55.7)	374 (58.4)	208 (52.9)	149 (55.8)	731 (56.2)
Anxiety	246 (36.4)	123 (31.1)	41 (25.8)	410 (33.3)	238 (37.0)	130 (33.0)	93 (34.6)	461 (35.3)
Social anxiety	197 (29.0)	123 (30.8)	37 (23.4)	357 (28.9)	191 (29.7)	125 (31.7)	88 (32.8)	404 (30.9)
Panic	304 (49.0)	167 (47.6)	58 (42.3)	529 (47.7)	292 (49.7)	173 (49.3)	117 (48.3)	582 (49.3)
PTSD	368 (54.3)	207 (52.1)	72 (45.3)	647 (52.4)	357 (55.4)	187 (47.5)	143 (53.2)	687 (52.6)
Hazardous alcohol use, n (%)	83 (12.4)	38 (9.7)	13 (8.6)	134 (11.0)	82 (13.0)	36 (9.5)	23 (8.7)	141 (11.1)
Alcohol or drug abuse, n (%)	36 (5.3)	20 (5.0)	10 (6.3)	66 (5.32)	37 (5.7)	31 (7.8)	14 (5.2)	82 (6.3)

There were no baseline differences between treatment arms on how often participants had used the internet in the last 30 days ($\chi^2_{2,\ n=2495}$=2.3, *P*=.32). Fewer than half (about 42%) of the participants had made a quit attempt in the last year, and about 80% of the sample had been smoking for more than 10 years, with an average FTND score of 5.6 (moderate nicotine dependence). The data retention rate was 87.56% (2309/2637) and did not differ between arms.

Description of Distinct Groups of Trajectories

The functional clustering analysis of 52 weeks of log-ins revealed 3 distinct groups of trajectories for each of the intervention websites. Figure 1 shows log-in patterns for the first 16 weeks for WebQuit (left) and for Smokefree (right). The trajectories were easiest to visualize for the first 16 weeks of use. However, Multimedia Appendix 1 shows the full 52 weeks for reference. For the WebQuit website (Figure 1, left), the first trajectory group (682/1240, 55.00% of sample) logged at least one day in the first week and then had almost no log-ins after that. They were termed 1-week users. The second trajectory group (399/1240, 32.18% of sample) logged in an average of 1.8 days in the first week, 0.8 days in the second week, once every 3 weeks until week 5, and had very sporadic log-ins in week 6 and beyond. They were termed 5-week users. The third trajectory group (159/1240, 12.82% of sample) logged in an average of 3.7 days in the first week, 3.3 days in the second week, 2.7 days in the third week, 2.4 days in the fourth week, 1.6 days in week 5, once in week 6, and then on average once every month starting in week 7 and continuing in this pattern until week 52. They were termed 52-week users.

For the Smokefree website (Figure 1, right), the first trajectory group (645/1309, 49.27% of sample) logged in less than once on average in the first week and then had almost no log-ins after that. As with WebQuit, they were termed 1-week users. The second trajectory group (395/1309, 30.18% of sample) logged in once in week 1, every other week until week 4, and then had almost no log-ins after that. They were termed 4-week users. The third trajectory group (269/1309, 20.55% of sample) logged in an average of 1.5 days in weeks 1 and 2, once in week 3, every other week over the period of weeks 4 to 5, and then had almost no log-ins after that. They were termed 5-week users. Note also that in both intervention arms, there was a pattern of a spike in log-ins at week 12, corresponding to the invitation to complete the 12-week outcome survey that, while completely independent of the interventions, likely triggered some users to engage with their assigned intervention website.

Trajectory Membership Prediction of Smoking Cessation Outcome

Table 2 shows each intervention arm's trajectory group membership as a predictor of 30-day point prevalence abstinence at the 12-month follow-up. For WebQuit, abstinence rates for these 3 trajectory groups were 116/562 (20.6%) for 1-week users, 100/370 (27.0%) for 5-week users, and 51/149 (34.2%) for 52-week users. Compared with 1-week users, 5-week users had 57% higher odds (OR 1.57, 95% CI 1.13-2.17; *P*=.007) of being abstinent at 12 months, and 52-week users had 124% higher odds (OR 2.24, 95% CI 1.45-3.43; *P*<.001) of being abstinent at 12 months. These models adjusted for the baseline covariates selected as outlined in the Methods section and included smoking half a pack or more, the commitment to quitting score, and screening positive for panic disorder. Descriptively, for Smokefree, abstinence rates for the 3 trajectory groups were 139/562 (24.7%) for 1-week users, 85/349 (24.4%) for 4-week users, and 81/252 (32.1%) for 5-week users.

Figure 1. Average weekly log-in trajectory for each cluster from the (left) WebQuit (n=1240) arm and (right) Smokefree (n=1309) arm for first 16 weeks of use.

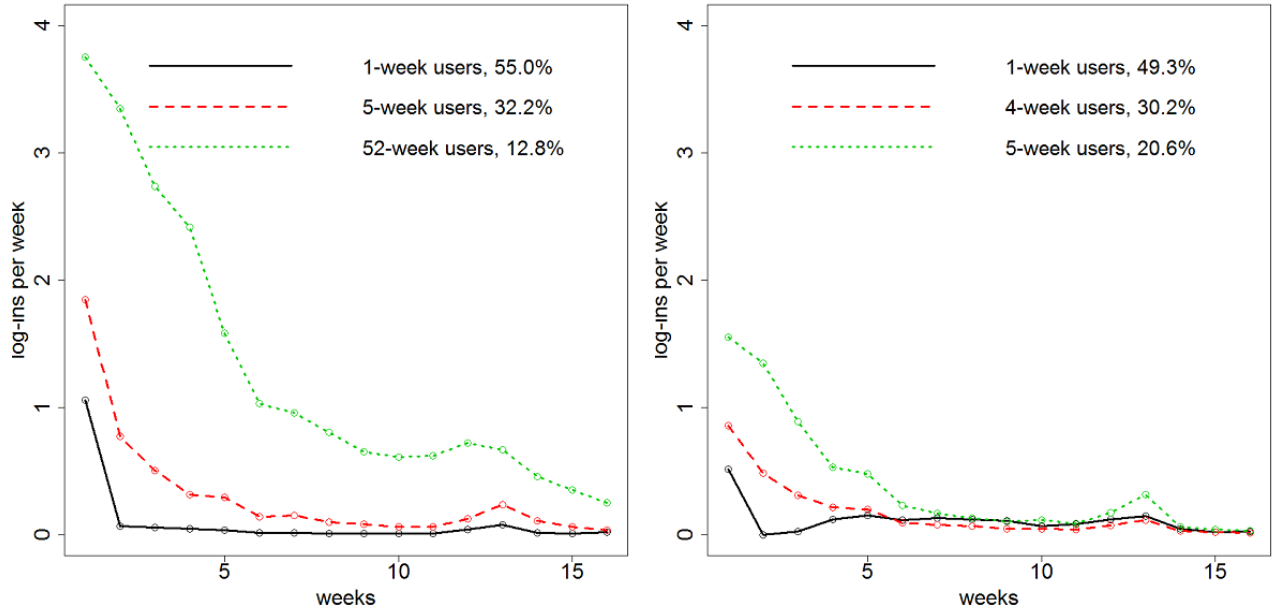


Table 2. Logistic regression models predicting 12-month smoking cessation outcome by groups of engagement trajectories and other covariates[a].

Arm and covariate	Odds ratio	95% CI	*P* value
WebQuit			
5-week users	1.57	1.13-2.17	.007
52-week users	2.24	1.45-3.43	<.001
Half a pack or more	0.58	0.41-0.82	.002
Commitment	1.69	1.37-2.10	<.001
Panic[b]	0.75	0.55-1.01	.06
Smokefree			
4-week users	1.00	0.73-1.37	.99
5-week users	1.48	1.05-2.07	.02
High school or less	0.69	0.50-0.94	.02
Smoking >10 years	0.76	0.55-1.06	.10
Smokes within 5 minutes of waking	0.75	0.57-0.99	.05
Commitment	1.71	1.41-2.07	<.001
Partner smokes	0.62	0.47-0.84	.001

[a]Reference group: 1-week users. Only significant predictors have been included in this table for ease of reading.

[b]Refers to whether participants screened positive for panic disorder.

Table 3. Multinomial logistic regression results predicting log-in trajectory cluster membership from baseline characteristics[a].

Arm, cluster, and characteristic	Odds ratio	95% CI
WebQuit		
5-week users		
Smoking >10 years	1.27	0.93-1.75
Not anxious[b]	1.25	0.96-1.64
52-week users		
Smoking >10 years	1.90	1.14-3.14
Not anxious	1.56	1.06-2.33
Smokefree		
4-week users		
Less than half a pack	1.16	0.85-1.61
Unemployed	1.33	1.03-1.72
No posttraumatic stress disorder[c]	1.43	1.11-1.85
5-week users		
Less than half a pack	1.72	1.23-2.44
Unemployed	1.79	1.33-2.38
No posttraumatic stress disorder	1.16	0.88-1.56

[a]Reference group: 1-week users. Only significant predictors have been included in this table for ease of reading.

[b]Refers to screening negative for generalized anxiety disorder.

[c]Refers to screening negative for posttraumatic stress disorder.

Compared with 1-week users, 4-week users were not more likely to be abstinent at 12 months (OR 1.00, 95% CI 0.73-1.37; *P*=.99), but 5-week users had 48% higher odds of being abstinent (OR 1.48, 95% CI 1.05-2.07; *P*=.02). This analysis adjusted for selected baseline covariates of education, smoking more than 10 years, smoking within 5 minutes of waking, commitment to quitting, and whether one has a partner who smokes.

Baseline Characteristics Predicting Trajectory Membership

Since the groups of trajectories were different across the 2 arms, we explored the baseline characteristics predicting membership in the groups for the 2 arms separately. For WebQuit, baseline characteristics associated with trajectory membership were age, smoking for at least the past 10 years, screening positive for depression, and screening positive for anxiety (all $P<.05$; results not shown). Controlling for the impact of related covariates, the adjusted multivariate regression model selected by stepwise AIC procedure showed that smoking for at least the past 10 years and screening negative for anxiety each, respectively, predicted a 90% higher odds (OR 1.90, 95% CI 1.14-3.14) and a 56% higher odds (OR 1.56, 95% CI 1.06-2.33) of being a 52-week user (compared with being a 1-week user) (Table 3). Since smoking history is partly a reflection of one's age, and the variables age, smoking history, and anxiety were correlated with each other, when we calculated a model containing age (categorized by decade), only age emerged as a significant predictor (see Multimedia Appendix 2).

For Smokefree, the baseline characteristics associated with trajectory membership in univariate analysis were being unemployed, smoking less than half a pack per day, and screening as not having PTSD (all $P<.05$; results not shown). Controlling for the impact of related covariates, the multivariate regression model showed that smoking less than half a pack per day predicted a 72% higher odds (OR 1.72, 95% CI 1.23-2.44) of being a member of the 5-week group, compared with the 1-week user group (Table 3). Being unemployed predicted a 79% higher odds (OR 1.79, 95% CI 1.33-2.38) of being a member of the 5-week user group relative to the 1-week group. Screening negative for PTSD predicted 43% higher odds (OR 1.43, 95% CI 1.11-1.85) of being a member of the 4-week user group relative to the 1-week user group. There was no evidence in either sample that sex predicted trajectory membership (all $P>.05$).

Discussion

Principal Findings

To our knowledge, this was one of few studies to analyze usage trajectories of eHealth interventions and examine the association between trajectory group membership and health outcomes [12,13]. The study found (1) 3 distinct groups of log-in trajectories for 2 Web-delivered interventions for smoking cessation, (2) that these trajectory groups differentially predicted smoking outcomes at 12 months, and (3) that certain user characteristics are associated with membership in certain trajectory groups. A 5-week usage of either website, and 52-week usage only of WebQuit, predicted a higher odds of quitting smoking. In general, the WebQuit intervention had a greater number of weekly log-ins within each of the 3 trajectory groups as compared with those of the Smokefree intervention. These major results are synthesized and interpreted in greater detail in this discussion.

Usage Trajectories and Health Outcomes

Regarding the first trajectory group, half the participants in both arms were 1-week users, which is a significant concern because they were the least likely to abstain from smoking at 12 months. Thus, it is imperative to learn why a participant would have almost no log-ins after a single week of use. User-centered design research, including laboratory observations and diary studies, could help elucidate the qualities of the intervention that cause an individual to discontinue use of the website. These individuals might benefit from a more intensive intervention, an eHealth intervention that uses a different treatment model, or one that is *not* eHealth (eg, individual telephone coaching). Regarding the second trajectory group, 5-week users were more likely to quit smoking in the WebQuit intervention (as well as for Smokefree, which had 5-week users as its third trajectory group). These results suggest that strategies to increase eHealth intervention engagement for 4 more weeks (ie, from 1 week to 5 weeks) could be highly cost effective. Example strategies worth testing include (1) proactive check-ins (via text message or phone calls) from staff about progress with the website, (2) daily automated text messages notifying the user of new content now available on the website, (3) rewards for each day's use of the website with badges or redeemable prizes, and (4) a 5-week challenge that shows other users' daily log-in progress toward the goal of 5 weeks of usage.

Regarding the third trajectory group, each intervention website had distinct log-in patterns that are likely explained by differing website structures. For Smokefree, this group was the 5-week users. The fact that they had almost no log-ins at 5 weeks and beyond is likely a reflection of Smokefree's structure—an informational resource for users, functioning like reference material. Thus, 5 weeks may be sufficient time for a user to glean all needed information from Smokefree and apply it appropriately to quitting smoking, as they had 48% higher odds of quitting smoking (compared with 1-week users). For WebQuit, this group was the 52-week users, who had 124% higher odds of quitting smoking (compared with 1-week users). Their much longer-term engagement is likely a reflection of WebQuit's structure—a step-by-step skills-based program that includes tracking progress with urges and smoke-free days. This program structure may have encouraged long-term, spaced skills practice [6], which may have contributed to the 34% 12-month quit rates observed in WebQuit's third trajectory group. In general, the findings for both websites' third trajectory group suggest that consistent use of each program over time is prognostic of a better health outcome, which is contrary to the notion that consistent log-ins may be a marker of ongoing challenges and struggles to change a health behavior. E-intervention design should thus focus on methods to encourage engagement over time, which may include strategies similar to those suggested above.

Personal Characteristics and Usage Trajectories

The impact of personal characteristics on usage trajectories appeared to vary by intervention. Specifically, WebQuit users who had smoked for at least 10 years were more likely to be 5-week users and nearly twice as likely to be 52-week users than 1-week users. However, smoking history differences may

be a reflection of age: users aged 50 years and over were over 8 times more likely to be 52-week users. This finding is consistent with past research showing that being older is a predictor of higher eHealth use [14-17], even though it was found only for WebQuit, not Smokefree, in this analysis. On the other hand, participants who screened positive for a mental health condition in either website (PTSD in Smokefree, and anxiety or depression in WebQuit) were more likely to be 1-week users, which suggests the need develop strategies to promote longer-term engagement for people with mental health disorders. There was no evidence in this study that sex predicted trajectory membership. Nonetheless, we recommend that future research examine many subgroup differences (eg, sex, race, age) in eHealth intervention trajectories as research on this model methodology expands to a wide variety of populations. Overall, these analyses suggest a need for further research on what baseline factors might predict different usage trajectories, and therefore inform the development of tailored interventions that facilitate long-term, consistent engagement, based on an individual's specific baseline characteristics.

Limitations and Future Directions

The study had several key limitations. First, we tested only 2 websites, and both were focused on smoking cessation; thus, future research should examine the extent to which results generalize to other behaviors and to other types of eHealth interventions. Second, cessation outcome data were self-reported for reasons stated in the Methods. Remote biochemical validation of smoking cessation would have introduced biases, including low response rates, prohibitive cost, challenges with confirming the identity of the person providing the sample, and inability to confirm abstinence beyond 24 hours [33,34].

Conclusions

In general, the WebQuit intervention had a greater number of weekly log-ins within each of the 3 trajectory groups as compared with those of the Smokefree intervention. The 1-, 4-, and 5-week usage of websites may be common patterns of how people engage in eHealth interventions over time. The 5-week usage of either website, and 52-week usage only of WebQuit, predicted a higher odds of quitting smoking. Strategies to increase eHealth intervention engagement for 4 more weeks (ie, from 1 week to 5 weeks) could be highly cost effective.

Acknowledgments

The authors would like to acknowledge funding by the US National Cancer Institute (R01 CA166646; R01CA192849), and the US National Institute on Drug Abuse (R01 DA038411). Funders had no role in the trial conduct or interpretation of results.

Conflicts of Interest

In July 2016, JBB was a consultant to GlaxoSmithKline, the makers of a nicotine replacement therapy. He now serves on the Scientific Advisory Board of Chrono Therapeutics, the makers of a nicotine replacement therapy device. JLH has received research support from Pfizer, the makers of a smoking cessation medication.

Multimedia Appendix 1

Average weekly log-in trajectory each for cluster from the WebQuit (n=1240) arm and Smokefree (n=1309) arm for 52 weeks of use.

[PNG File, 72KB - jmir_v20i4e10143_app1.png]

Multimedia Appendix 2

Age as dichotomous variable (reference: <30, age not selected in SmokeFree arm), odds ratios reported with 1-week users as comparison group.

[PDF File (Adobe PDF File), 16KB - jmir_v20i4e10143_app2.pdf]

References

1. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
2. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
3. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
4. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

5. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
6. Shaughnessy JJ. Long-term retention and the spacing effect in free-recall and frequency judgments. Am J Psychol 1977 Dec;90(4):587. [doi: 10.2307/1421733]
7. Pyc MA, Rawson KA. Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? J Mem Lang 2009 May;60(4):437-447. [doi: 10.1016/j.jml.2009.01.004]
8. Chassin L, Presson CC, Pitts SC, Sherman SJ. The natural history of cigarette smoking from adolescence to adulthood in a midwestern community sample: multiple trajectories and their psychosocial correlates. Health Psychol 2000;19(3):223-231. [doi: 10.1037/0278-6133.19.3.223]
9. Donovan JE. Adolescent alcohol initiation: a review of psychosocial risk factors. J Adolesc Health 2004 Dec;35(6):529.e7-529.18. [doi: 10.1016/j.jadohealth.2004.02.003] [Medline: 15581536]
10. Duncan SC, Duncan TE. A multivariate latent growth curve analysis of adolescent substance use. Struct Equ Model 1996 Jan;3(4):323-347. [doi: 10.1080/10705519609540050]
11. Macleod J, Oakes R, Copello A, Crome I, Egger M, Hickman M, et al. Psychological and social sequelae of cannabis and other illicit drug use by young people: a systematic review of longitudinal, general population studies. Lancet 2004 May 15;363(9421):1579-1588. [doi: 10.1016/S0140-6736(04)16200-4] [Medline: 15145631]
12. Goh G, Tan NC, Malhotra R, Padmanabhan U, Barbier S, Allen JC, et al. Short-term trajectories of use of a caloric-monitoring mobile phone app among patients with type 2 diabetes mellitus in a primary care setting. J Med Internet Res 2015;17(2):e33 [FREE Full text] [doi: 10.2196/jmir.3938] [Medline: 25648130]
13. Christofferson DE, Hertzberg JS, Beckham JC, Dennis PA, Hamlett-Berry K. Engagement and abstinence among users of a smoking cessation text message program for veterans. Addict Behav 2016 Nov;62:47-53. [doi: 10.1016/j.addbeh.2016.06.016] [Medline: 27318948]
14. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
15. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
16. Balmford J, Borland R, Benda P. Patterns of use of an automated interactive personalized coaching program for smoking cessation. J Med Internet Res 2008;10(5):e54 [FREE Full text] [doi: 10.2196/jmir.1016] [Medline: 19097975]
17. Zbikowski SM, Hapgood J, Smucker BS, 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;10(5):e41 [FREE Full text] [doi: 10.2196/jmir.999] [Medline: 19017583]
18. Bricker JB, Mull KE, McClure JB, Watson NL, Heffner JL. Improving quit rates of web-delivered interventions for smoking cessation: full-scale randomized trial of WebQuit.org versus Smokefree.gov. Addiction 2017 Dec 13. [doi: 10.1111/add.14127] [Medline: 29235186]
19. US Department of Health and Human Services, National Instutues of Health, National Cancer Institute. Smokefree.gov.: Tobacco Control Research Branch, National Cancer Institute URL: https://smokefree.gov/ [accessed 2018-04-06] [WebCite Cache ID 6yTxFOwLI]
20. Hayes SC, Levin ME, Plumb-Vilardaga J, Villatte JL, Pistorello J. Acceptance and commitment therapy and contextual behavioral science: examining the progress of a distinctive model of behavioral and cognitive therapy. Behav Ther 2013 Jun;44(2):180-198 [FREE Full text] [doi: 10.1016/j.beth.2009.08.002] [Medline: 23611068]
21. Fiore MC, Jaen CR, Baker T, Bailey WC, Benowitz NL, Curry SE, et al. Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: US Department of Health and Human Services; 2008.
22. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [Medline: 1932883]
23. Kahler CW, Lachance HR, Strong DR, Ramsey SE, Monti PM, Brown RA. The Commitment to Quitting Smoking Scale: initial validation in a smoking cessation trial for heavy social drinkers. Addict Behav 2007 Oct;32(10):2420-2424 [FREE Full text] [doi: 10.1016/j.addbeh.2007.04.002] [Medline: 17478057]
24. Kahler CW, Metrik J, LaChance HR, Ramsey SE, Abrams DB, Monti PM, et al. Addressing heavy drinking in smoking cessation treatment: a randomized clinical trial. J Consult Clin Psychol 2008 Oct;76(5):852-862 [FREE Full text] [doi: 10.1037/a0012717] [Medline: 18837602]
25. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
26. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
27. Stein MB, Roy-Byrne PP, McQuaid JR, Laffaye C, Russo J, McCahill ME, et al. Development of a brief diagnostic screen for panic disorder in primary care. Psychosom Med 1999;61(3):359-364. [Medline: 10367617]

28. Lang AJ, Stein MB. An abbreviated PTSD checklist for use as a screening instrument in primary care. Behav Res Ther 2005 May;43(5):585-594. [doi: 10.1016/j.brat.2004.04.005] [Medline: 15865914]
29. Connor KM, Kobak KA, Churchill LE, Katzelnick D, Davidson JR. Mini-SPIN: a brief screening assessment for generalized social anxiety disorder. Depress Anxiety 2001;14(2):137-140. [Medline: 11668666]
30. Rotondi AJ, Eack SM, Hanusa BH, Spring MB, Haas GL. Critical design elements of e-health applications for users with severe mental illness: singular focus, simple architecture, prominent contents, explicit navigation, and inclusive hyperlinks. Schizophr Bull 2015 Mar;41(2):440-448. [doi: 10.1093/schbul/sbt194] [Medline: 24375458]
31. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
32. Civljak M, Stead LF, Hartmann-Boyce J, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2013;7:CD007078. [doi: 10.1002/14651858.CD007078.pub4] [Medline: 23839868]
33. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]
34. Cha S, Ganz O, Cohn AM, Ehlke SJ, Graham AL. Feasibility of biochemical verification in a web-based smoking cessation study. Addict Behav 2017 Oct;73:204-208. [doi: 10.1016/j.addbeh.2017.05.020] [Medline: 28551588]
35. Ramsay JO, Silverman BW. Applied Functional Data Analysis: Methods and Case Studies. New York, NY: Springer; 2007.
36. Kaufman L, Rousseeuw PJ. Finding Groups in Data. Hoboken, NJ: Wiley Online Library; 1990.
37. Tibshirani R, Walther G. Cluster validation by prediction strength. J Comput Graph Stat 2005 Sep;14(3):511-528. [doi: 10.1198/106186005X59243]

Abbreviations

ACT: acceptance and commitment therapy
AIC: Akaike information criterion
eHealth: electronic health
FDA: Food and Drug Administration
FTND: Fagerström Test for Nicotine Dependence
PTSD: posttraumatic stress disorder
SMS: short message service

Edited by G Eysenbach; submitted 15.02.18; peer-reviewed by N Chavannes, JC Chen, S Bhatlekar; comments to author 14.03.18; revised version received 23.03.18; accepted 24.03.18; published 20.04.18.

Please cite as:
Bricker JB, Sridharan V, Zhu Y, Mull KE, Heffner JL, Watson NL, McClure JB, Di C
Trajectories of 12-Month Usage Patterns for Two Smoking Cessation Websites: Exploring How Users Engage Over Time
J Med Internet Res 2018;20(4):e10143
URL: http://www.jmir.org/2018/4/e10143/
doi:10.2196/10143
PMID:29678799

Original Paper

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

Shinyi Wu[1,2,3,4], PhD; Kathleen Ell[1], DSW; Haomiao Jin[1,2,3], PhD; Irene Vidyanti[3,5], PhD; Chih-Ping Chou[6], PhD; Pey-Jiuan Lee[1], MS; Sandra Gross-Schulman[7], MD, MPH, RN; Laura Myerchin Sklaroff[7,8], MA; David Belson[3], PhD; Arthur M Nezu[9], PhD; Joel Hay[4], PhD; Chien-Ju Wang[7], MS; Geoffrey Scheib[7], RN, BS; Paul Di Capua[10,11], MBA, MD; Caitlin Hawkins[3], PhD; Pai Liu[3], PhD; Magaly Ramirez[3,12], PhD; Brian W Wu[6], MD, PhD; Mark Richman[13], MD; Caitlin Myers[7], RN, MN, FNP; Davin Agustines[7], DO; Robert Dasher[7], MD; Alex Kopelowicz[7], MD; Joseph Allevato[7], MD; Mike Roybal[7], MD; Eli Ipp[14,15,16], MD; Uzma Haider[15], MD; Sharon Graham[7], MD; Vahid Mahabadi[7], MD; Jeffrey Guterman[7,14], MD

[1]Suzanne Dworak-Peck School of Social Work, University of Southern California, Los Angeles, CA, United States

[2]Roybal Institute on Aging, University of Southern California, Los Angeles, CA, United States

[3]Daniel J. Epstein Department of Industrial and Systems Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States

[4]Schaeffer Center for Health Policy and Economics, University of Southern California, Los Angeles, CA, United States

[5]Policy Analysis Unit, Los Angeles County Department of Public Health, Los Angeles, CA, United States

[6]Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[7]Los Angeles County Department of Health Services, Los Angeles, CA, United States

[8]College of Social and Behavioral Sciences, California State University, Northridge, Los Angeles, CA, United States

[9]Department of Psychology, Drexel University, Philadelphia, PA, United States

[10]Caremore Medical Group, East Haven, CT, United States

[11]Herbert Wertheim College of Medicine, Florida International University, Miami, FL, United States

[12]Department of Health Policy and Management, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, United States

[13]Department of Emergency Medicine, Northwell Health Long Island Jewish Medical Center, New Hyde Park, NY, United States

[14]David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States

[15]Harbor-UCLA Medical Center, University of California Los Angeles, Los Angeles, CA, United States

[16]Los Angeles Biomedical Research Institute, Los Angeles, CA, United States

Corresponding Author:
Shinyi Wu, 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 2138216442
Fax: 1 2137401120
Email: shinyiwu@usc.edu

Abstract

Background: Comorbid depression is a significant challenge for safety-net primary care systems. Team-based collaborative depression care is effective, but complex system factors in safety-net organizations impede adoption and result in persistent disparities in outcomes. Diabetes-Depression Care-management Adoption Trial (DCAT) evaluated whether depression care could

be significantly improved by harnessing information and communication technologies to automate routine screening and monitoring of patient symptoms and treatment adherence and allow timely communication with providers.

Objective: The aim of this study was to compare 6-month outcomes of a technology-facilitated care model with a usual care model and a supported care model that involved team-based collaborative depression care for safety-net primary care adult patients with type 2 diabetes.

Methods: DCAT is a translational study in collaboration with Los Angeles County Department of Health Services, the second largest safety-net care system in the United States. A comparative effectiveness study with quasi-experimental design was conducted in three groups of adult patients with type 2 diabetes to compare three delivery models: usual care, supported care, and technology-facilitated care. Six-month outcomes included depression and diabetes care measures and patient-reported outcomes. Comparative treatment effects were estimated by linear or logistic regression models that used generalized propensity scores to adjust for sampling bias inherent in the nonrandomized design.

Results: DCAT enrolled 1406 patients (484 in usual care, 480 in supported care, and 442 in technology-facilitated care), most of whom were Hispanic or Latino and female. Compared with usual care, both the supported care and technology-facilitated care groups were associated with significant reduction in depressive symptoms measured by scores on the 9-item Patient Health Questionnaire (least squares estimate, LSE: usual care=6.35, supported care=5.05, technology-facilitated care=5.16; *P* value: supported care vs usual care=.02, technology-facilitated care vs usual care=.02); decreased prevalence of major depression (odds ratio, OR: supported care vs usual care=0.45, technology-facilitated care vs usual care=0.33; *P* value: supported care vs usual care=.02, technology-facilitated care vs usual care=.007); and reduced functional disability as measured by Sheehan Disability Scale scores (LSE: usual care=3.21, supported care=2.61, technology-facilitated care=2.59; *P* value: supported care vs usual care=.04, technology-facilitated care vs usual care=.03). Technology-facilitated care was significantly associated with depression remission (technology-facilitated care vs usual care: OR=2.98, *P*=.04); increased satisfaction with care for emotional problems among depressed patients (LSE: usual care=3.20, technology-facilitated care=3.70; *P*=.05); reduced total cholesterol level (LSE: usual care=176.40, technology-facilitated care=160.46; *P*=.01); improved satisfaction with diabetes care (LSE: usual care=4.01, technology-facilitated care=4.20; *P*=.05); and increased odds of taking an glycated hemoglobin test (technology-facilitated care vs usual care: OR=3.40, *P*<.001).

Conclusions: Both the technology-facilitated care and supported care delivery models showed potential to improve 6-month depression and functional disability outcomes. The technology-facilitated care model has a greater likelihood to improve depression remission, patient satisfaction, and diabetes care quality.

(J Med Internet Res 2018;20(4):e147) doi:10.2196/jmir.7692

KEYWORDS

primary care; disease management; depression; diabetes mellitus; health information technology; telemedicine; comparative effectiveness research; propensity score; population health; patient reported outcome measures

Introduction

Depression Care in Underserved Populations

Depression, an underdiagnosed comorbidity for those with chronic illness [1], impairs functional status and worsens clinical outcomes, including morbidity and mortality; it also increases cost [2-5]. Diabetes doubles the risk of depression relative to the general population; 10% to 15% of adults with diabetes also have major depressive disorder [6,7]. The relationship between diabetes and depression might be bidirectional [8,9]. Depression with diabetes may significantly worsen the course of both disorders, leading to reduced functioning and quality of life [8-11].

Low-income, culturally diverse populations with chronic illnesses have an even higher risk of depression [7,12,13]. Hispanics and Latinos have a higher prevalence of diabetes compared with non-Hispanic whites [14] and are less likely to meet glycated hemoglobin (HBA_{1c}) and serum cholesterol goals [15]. Racial and ethnic minority populations also experience disparities in terms of mental health care, including appropriate mental health diagnosis, counseling, antidepressant medication prescriptions, and depression care follow-up [16-23]. Hispanics and Latinos are less than half as likely as non-Hispanic whites to receive guideline-level depression care [20].

Research has shown that there are effective ways to reduce these disparities. For example, there is growing evidence that a team-based collaborative depression care model is effective in improving care for low-income patients (including racial and ethnic minority populations) with chronic illnesses [24-27]. The US Preventive Services Task Force recommends depression screening, and an adaptive treatment approach has been shown to be effective in helping patients find successful antidepressant options or psychotherapy [28,29].

Safety-net primary care clinics are the preferred venue for underserved patients to access depression care because it is a common point of service delivery [30-32]. However, these settings encounter a complex mix of patient, provider, and health system factors that can impede the adoption of evidence-based collaborative depression care and result in persistent disparities in depression outcomes [33].

Safety-net systems organize and deliver a significant level of health care and other related services to uninsured, Medicaid, and other vulnerable populations [34]. Safety-net primary care providers often find it challenging to engage patients with major

depression, particularly when it is accompanied by chronic illnesses, because the patient must participate in active and ongoing depression symptom assessment and management in addition to managing the other medical conditions [16,33,35-37]. Concurrently, minority patients are less likely to voluntarily report depressive symptoms, 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 [33,38]. Therefore, low-income, predominantly minority patients in safety-net care systems often miss out on depression diagnosis and follow-up assessments [33,39].

Diabetes-Depression Care-Management Adoption Trial

The Diabetes-Depression Care-management Adoption Trial (DCAT) is a translational study in a large safety-net system of primary care settings. The study compared a technology-facilitated care (TC) model with a usual care (UC) model and a supported care (SC) model to assess whether technology could facilitate the adoption of collaborative depression care for patients with type 2 diabetes. The DCAT TC model is an information and communication technology (ICT)–facilitated care management approach that harnesses automated telephone assessment (ATA) technology and integrates it with a disease management registry (DMR) system to automate key aspects of depression care. The UC model is standard care in a safety-net system in which primary care physicians (PCPs) develop individualized treatment plans for depression and diabetes care. The SC model is a team-based collaborative care management approach that involves care team staff members to provide depression and diabetes care.

The DCAT study is expected to fill two important gaps in current collaborative depression care implementation research. First, existing studies largely rely on labor-intensive, team-based SC approaches to implement collaborative depression care [24-27,40-43]. There is evidence that this SC model is effective and can be cost-effective compared with UC [24-27,44-47]. However, it is challenging for SC teams to integrate depression comorbidity care when patients are presented with other medical conditions because of the intensive labor and time needed to proactively screen for depression, follow up on treatment, and monitor and manage long-term care [33,41,48]. By relieving providers in resource-constrained safety-net clinics from many labor-intensive tasks such as collecting, summarizing, and reviewing individual or aggregate patient data to facilitate care, automation can facilitate the adoption of collaborative depression care. Therefore, DCAT tested ICT that automated critical information collection and processing for depression care tasks, including (1) Depression assessments and symptom monitoring, (2) Patient self-management behavior prompting, (3) Optimization of treatment follow-up, and (4) provider collaborative communication.

Second, existing research has not fully addressed ways to develop a patient-centered approach to implement collaborative depression care for low-income, predominantly minority safety-net patient populations. Prior studies have demonstrated that the team-based approach can effectively implement collaborative depression care in safety-net environments [24-27]; DCAT built on this evidence by applying the ICT to further address language, time, and stigma barriers [33,41,48] affecting safety-net patients. DCAT accomplished this by customizing calls with the patient's preferred language and call time, making multiple attempts (if needed) to connect with the patient, and establishing a private and machine-only venue to report sensitive depression measures to reduce social desirability bias [49-51]. About 80% of the patients agreed or strongly agreed that the DCAT-tested ICT was easy to use (86.2%, 94/109), nonintrusive (87.1%, 95/109), and private and secure (75.9%, 82/108) [51].

Paper Purpose and Hypothesis

DCAT-related improvements have been reported in several publications, including trial design [33], TC technology design and evaluation [52], patient acceptance of the technology [51], patient engagement [50], provider implementation reactions [53], depression risk profiling and prediction [54,55], and cost-effectiveness analysis (unpublished data, 2018, [56]). This paper reports DCAT-related depression symptom and diabetes care outcomes after 6 months. In addition, to provide the patients' perspective on treatment benefits, this paper includes patient-reported outcomes, including physical and mental well-being, functional impairment, and satisfaction with care [57].

The main hypothesis of the paper is that, compared with the UC group, both the TC and SC groups will have statistically significant greater improvement in depression symptoms, diabetes care processes and outcomes, and patient-reported outcomes. TC uses an ATA system to ease the adoption of collaborative depression care rather than direct clinical intervention with patients; therefore, although the researchers have no hypothesis on how TC outcomes will compare with SC outcomes, this paper also explores whether the TC group will have better outcomes than the SC group.

Methods

Diabetes-Depression Care-Management Adoption Trial Study Design

DCAT is a translational study conducted in collaboration with Los Angeles County Department of Health Services (LACDHS), the second largest safety-net care system in the United States. Institutional review board approval was obtained from the University of Southern California, the Olive View–University of California Los Angeles Medical Center, and the Los Angeles Biomedical Research Institute.

Study Sites and Intervention Period

The study used a quasi-experimental comparative effectiveness design to compare three delivery models in three groups: UC, SC, and TC. Eight clinics were selected to participate in the study based on criteria that reflected geographic and diabetes care model diversity. These clinics were matched by geographic location and patient sociodemographics to form the three study groups. The patients were not randomly assigned; each patient was assigned to a study group based on the clinic from which he or she was recruited.

The UC group featured two community clinics and represents the status quo of clinical practice, wherein the translation and

adoption of depression care evidence is performed by PCPs and their staffs. The SC and TC groups each featured two care teams from an LACDHS diabetes disease management program (DMP) to incorporate depression care. In both the SC and TC groups, one of the two teams practiced in both a hospital-based outpatient clinic and a satellite community clinic; the other team practiced in a community clinic in a different geographic area.

The intervention period was 12 months, and the study occurred from 2011 to 2013. During the first 6 months, the UC group received usual primary care, whereas the SC group received DMP-supported depression care, and the TC group received the ATA application in the DMP setting. After 6 months, all SC and TC patients were transferred back to their usual primary care, although the ATA calls were continued for the full 12 months. This paper reports the 6-month outcomes.

Intervention Description

Table 1 shows the intervention elements of the UC, SC, and TC models, described below.

Table 1. Intervention elements of the usual care (UC), supported care (SC), and technology-facilitated care (TC) models. ATA: automated telephone assessment; DMP: disease management program; DMR: disease management registry; LACDHS: Los Angeles County Department of Health Service; PCP: primary care physicians; PHQ-9: 9-item Patient Health Questionnaire; PST: problem-solving therapy.

Elements	Usual care	Supported care	Technology-facilitated care
Care paradigm	Standard primary care; optional PST	Diabetes DMP-supported care; PST; Interactive DMR system	Diabetes DMP-supported care; PST; ATA linked to DMR enhanced with clinical decision support software
Clinic setting	Two community non-DMP clinics	Two diabetes DMP teams in safety-net clinics: one serving both hospital-based outpatient clinic and a satellite community clinic, and other serving in a community clinic in a different geographic area	Two diabetes DMP teams in safety-net clinics: one serving both hospital-based outpatient clinic and a satellite community clinic, and other serving in a community clinic in a different geographic area
Patient education and care resources	Depression educational pamphlets (in English) or *fotonovella* (in Spanish) adapted for diabetes patients; Standard provider contact and community resource information	Depression educational pamphlets (in English) or *fotonovella* (in Spanish) adapted for diabetes patients; Standard provider contact and community resource information	Depression educational pamphlets (in English) or *fotonovella* (in Spanish) adapted for diabetes patients; Standard provider contact and community resource information
Physician education	Psychiatrist expert conducts webinars about collaborative depression care evidence, offers PCP depression screening and treatment didactic, and provides personal copy of the Los Angeles County Department of Health Services depression care protocol	Psychiatrist expert conducts webinars about collaborative depression care evidence, offers PCP depression screening and treatment didactic, and provides personal copy of the Los Angeles County Department of Health Services depression care protocol	Psychiatrist expert conducts webinars about collaborative depression care evidence, offers PCP depression screening and treatment didactic, and provides personal copy of the Los Angeles County Department of Health Services depression care protocol
Clinician training for PST	Optional for UC clinicians	Mandatory for DMP nurses, nurse practitioners, and social workers; conducted by psychology and social work experts	Mandatory for DMP nurses, nurse practitioners, and social workers; conducted by psychology and social work experts
Depression screen and ongoing symptom monitoring	Standard care determined by PCP practice	Performed by DMP clinical social worker: PHQ-9 screening when patients join the DMP; Ongoing symptom monitoring per clinical judgment based on LACDHS depression care protocol and treatment guideline	Performed by the ATA system and enhanced DMR: Quarterly depression screening (PHQ-2) for nondepressed patients; Monthly symptom monitoring (PHQ-2, -9, other tailored questions) for depressed patients
Depression treatment	Standard care: Antidepressant medication; Referral to clinic social worker or community mental health care	DMP based on the LACDHS protocol and treatment guideline: Antidepressant with optional PST; Option of community referrals	DMP based on the LACDHS protocol and treatment guideline with ATA responses and DMR data: identify at-risk patients, determine treatment, and promptly follow up on treatment adherence issues.
Provider collaborative communication	LACDHS standard clinic collaboration	DMP nurse initiates communication with medication prescriber; Refers patient to social worker if patient refuses medication or needs PST	DMP plus enhanced team care collaboration facilitated by DMR: Reminders prompt designated responders to follow up; Responders include DMP nurse, social worker, medication prescriber or PCP, or psychiatrist
Patient relapse prevention	Standard care	Monthly telephone calls by nurse or social worker	Monthly or quarterly automated telephone calls prompt for relapse prevention.

Usual Care Model

The UC model was standard primary care. UC clinicians were offered an optional training opportunity (described in *Provider Training and Depression Treatment Protocol* section below).

Supported Care Model

The SC model used the diabetes DMP team (comprising nurse care managers, nurse practitioners, and a consulting or supervising physician) to deliver depression care. SC diabetes care management was designed to proactively identify, risk stratify, and treat patients using clinical protocols that emphasized patient empowerment. The DMP was nurse driven and physician supervised and used structured approaches and protocols; in these programs, nurses delivered the majority of the diabetes care. The approaches included a patient-signed commitment to take an active role in his or her diabetes care, case management, PCP designation, group patient education, self-management support, and care coordination. The diabetes-specific management was provided initially via in-person visits, with follow-up primarily via telephone visits. The DMP included a homegrown, Web-based, interactive chronic DMR system to support clinical assessment and decisions. The DMP was designed for limited-time care management (typically 6 months), after which patients were transferred back to their primary medical providers.

During the study, the SC team supplemented diabetes management with two periodic depression symptom screening and monitoring tools: (1) the 9-item Patient Health Questionnaire (PHQ-9) [58], a standard tool in each clinic's disease registry and (2) the LACDHS depression care protocol and treatment guideline (see "Provider Training and Depression Treatment Protocol"). In PHQ-9, the patient scores each of the nine Diagnostic and Statistical Manual of Mental Disorders, 4th edition criteria as "0" (not at all) to "3" (nearly every day) to provide both a dichotomous diagnosis of probable major depression and a continuous severity score [58]. The SC program also designated a social worker to provide problem-solving therapy (PST), an evidence-based depression treatment [59].

Technology-Facilitated Care Model

The TC model also operated in a DMP clinic setting with a DMR and supplemental depression care based on the LACDHS depression care protocol and treatment guideline. The TC model, however, was designed to assist time-pressured clinical social workers and medical and nursing providers by using an ATA system to routinely screen and monitor patient depression symptoms and treatment adherence and communicate the results to providers. As described elsewhere [33,52], the ATA system was linked with the DMR to automatically trigger depression care management calls on a predetermined calendar schedule. The call contents were individually tailored, driven by a preprogrammed algorithm that scanned patient medical records and call histories to determine applicable questions. The ATA used a persona, "Amy," who spoke in a natural voice rather than a system-generated text-to-speech robotic voice to administer the assessment questions. During study enrollment, patients selected their preferred language (English or Spanish) and preferred call time. The DCAT ATA built in both automated speech recognition and interactive voice response technologies [60] that allowed patients to either speak their responses to Amy's questions or punch numbers on a phone keypad. Automated speech recognition has the advantage of eliminating number-punching errors, which are a concern for diabetes patients with sensing or vision problems. Unfortunately, automated speech recognition was only available in English, not Spanish, because of suboptimal recognition accuracy in different Spanish accents.

There were two main ATA call scripts: one for screening and one for monitoring. The screening calls were for people who had no prior history of depression or who had been clear of a depression diagnosis for at least 6 months. The ATA collected information in four categories: (1) depressive symptoms; (2) pain; (3) self-management activities, including regular physical and fun activities; and (4) patient request to be contacted by a provider. PHQ-2, the first two items of the PHQ-9, were used for screening; if a patient score exceeded the cut-off of 3 points out of the possible 6 points, the ATA automatically asked the remaining PHQ-9 items. The monitoring calls were for depressed patients; the monitoring calls addressed all four categories and administered PHQ-9. If the patient had been prescribed an antidepressant, the call asked questions about medication adherence and side effects. If the patient was receiving psychotherapy, the call asked questions about problem-solving skills practice. Depending on the questions asked and patient response time, each ATA call lasted about 2 to 5 min. If a patient did not answer the call, the ATA system repeated the calling attempt multiple times a day for up to 1 week [52]. If a patient did not pick up the call within a week, the scheduled call was forfeited, and the patient was contacted again for the next scheduled call.

The telephone was selected as the communication platform because phones were the most accessible technology among safety-net patients at the time of the study. The calls were low intensity (ie, one call every month for monitoring or every 3 months for screening based on each patient's depression condition) to balance information need and patient burden. Clinic officials have reported that patients who are depressed seem more likely to miss their scheduled visit appointments and often delay or forgo calling for help when symptoms fail to improve or worsen. The ATA system mitigated this dilemma by contacting patients rather than relying on them to initiate calls, proactively reaching patients and identifying their care needs. The TC model did accommodate patients who preferred a personal call over an automated call; in those cases, staff members made the calls according to the patient's language and schedule preferences (25/366 or 6.8% of the patients made the request).

The patient-reported ATA data were tethered to the DMR, which in the TC model was enhanced by clinical decision support software for provider collaborative communication. The decision support software automatically generated task reminders and alerts based on the patient records in the DMR and the assessment data; the reminders and alerts prompted DMP providers to follow up with specific patients in need of care. For example, the automatically generated provider tasks in the DMR would remind a care manager to follow up with a patient

who self-reported an antidepressant adherence issue, task a social worker to follow up with a patient with major depression symptoms, or task a nonclinical assistant to address patient callback requests. Task reminders included structured, radio-button lists of potential care management actions with the option of free text to support evidence-based practices and to ease providers' documentation burden.

If a patient expressed an inclination toward self-harm or suicide (ie, responded to PHQ-9 question 9 with a score of 2=more than half the days or 3=nearly every day), the ATA system automatically initiated contact (via mobile phone SMS [short message service] text message and email) with an emergency response physician to get help for the patient. If the physician did not respond within 15 min, the ATA system initiated contact with the next physician on the emergency response team. This process repeated up to the fifth physician (the first three were psychiatrists and the last two were emergency medicine physicians) to ensure the patient received attention. During the study, the ATA was able to reach an emergency response physician in every instance.

Provider Training and Depression Treatment Protocol

DMP depression care in both SC and TC was based on the LACDHS depression care protocol and treatment guideline, which was developed by the DCAT team and described in the study design paper [33]. All SC and TC care providers were trained by an expert psychiatrist in the collaborative depression care model and adaptive treatment approach via one of three webinars (each approximately 2 hours). They were also offered training in PST via a 1-day (6-hour) workshop conducted by an academic psychologist and a social worker faculty member who are PST experts. UC providers were also invited to participate in these training opportunities, but they were not given time off from clinical duties to participate in the trainings.

While TC used technology to support providers for depression symptom and treatment adherence monitoring and to facilitate care coordination, SC providers monitored patients in the traditional way by calling patients and coordinating care among themselves. All patients in the SC and TC groups received support from a nurse care manager by telephone or in the clinic; in the TC group, patients also received the ongoing follow-up ATA calls in English or Spanish. If a patient in either group was confirmed for depression, weeks 1 to 8 of the depression care protocol included first-line treatment with antidepressant medication prescribed according to the protocol by the treating physician or nurse practitioner. If the patient refused medication, the care manager referred the patient to a social worker for six to eight PST sessions.

During weeks 9 to 12, the care manager would refer patients with a partial response (reduction in PHQ-9 scores) or nonresponse back to the treating physician or nurse practitioner, who would adjust antidepressant medication dosage (or encourage nonmedicated patients to begin medication) and the addition of PST. Patients with a full response (PHQ-9 score less than 8) received monthly treatment maintenance and relapse-prevention behavioral activation.

Consistent with the depression care protocol, patients with persistent PHQ-9 scores of 10 or higher were offered additional PST booster sessions; augmentation with low-dose trazodone, an antidepressant medication that also helps treat anxiety and insomnia; or referrals to specialty mental health care.

Participant Eligibility and Recruitment

Textboxes 1 and 2 show the eligibility and ineligibility criteria for patients.

Every enrollee received a set of educational and community resource materials in Spanish or English.

The enrollment period was from April 2011 to May 2012. Patients with type 2 diabetes were identified for recruitment from the DMR database and clinic records. Patients provided verbal consent to bilingual research assistants during study eligibility screening.

Textbox 1. Eligibility criteria for patients.

Eligibility criteria

- 18 years or older
- Had been diagnosed with type 2 diabetes
- Had a working phone number
- Spoke English or Spanish
- Could read and understand the consent form

Textbox 2. Ineligibility criteria for patients.

Ineligibility criteria

- Patients with baseline possible suicidal ideation
- Patients with cognitive impairment
- Patients with alcohol abuse
- Patients with recent lithium or antipsychotic medication

Table 2. Primary outcome measures in the Diabetes-Depression Care-management Adoption Trial (DCAT), Los Angeles, 2011 to 2013. HBA_{1c}: glycated hemoglobin; PHQ-9: 9-item Patient Health Questionnaire; PST: problem-solving therapy; SF-12: 12-item Short Form Survey.

Variables	Description
Depression, measured at baseline and 6 months post intervention	
PHQ-9 [58]	Continuous variable assessing severity of depression. Scoring: PHQ-9 of 5-9=mild depression; PHQ-9 of 10-14=moderate depression; PHQ-9 of 15-19=major depression; PHQ-9 of 20-27=severe depression. For purposes of this study, PHQ-9 ≥10 indicated depression serious enough to consider pharmacologic or PST treatment.
Depression remission	Dichotomous variable assessing effectiveness of treating patients with major depression. Depression remission defined as baseline PHQ-9 ≥10 and 6-month PHQ-9 ≤8 with a reduction ≥50%.
Diabetes, measured at baseline and 6 months post intervention if not otherwise indicated	
HBA_{1c} value[a]	Continuous variable assessing severity of diabetes. HBA_{1c} value indicates average plasma glucose concentration over prolonged periods.
HBA_{1c} tested[a]	Dichotomous variable assessing diabetes care process.
Total cholesterol[a]	Continuous variable assessing cholesterol levels and severity of dyslipidemia
Diabetes self-care [61]	Days per week of diabetes self-care. Treated as a continuous variable.
Exercise	Days of participating in at least 30 min of exercise during previous week.
Patient reported outcomes, measured at baseline and 6 months post intervention	
SF-12 physical score [62]	Continuous variables assessing functional health and well-being
SF-12 mental score [62]	
Sheehan Disability Scale [63,64]	Self-reported tool assessing functional impairment in work or school, social, and family life.
Satisfaction with diabetes care	Five-level score assessing diabetes care satisfaction. Treated as a continuous variable.
Satisfaction with care for emotional problems	Five-level score assessing mental health care satisfaction. Treated as a continuous variable.
Satisfaction with care for emotional problems, baseline PHQ-9 ≥10	Five-level score assessing mental care satisfaction of patients with major depression. Treated as a continuous variable.

[a]The HBA_{1c} value, HBA_{1c} tested, and total cholesterol value were obtained from the LACDHS electronic medical record system. The measurement periods were within 3 months of baseline and 6-month post intervention. If more than one value was available, the values closest to the baseline and the 6-month follow-up period were chosen.

Outcome Measures

Measures were taken at baseline and at 6, 12, and 18 months by independent English-Spanish bilingual interviewers. Primary outcomes included three depression outcomes, five diabetes care measures, and six patient-reported outcomes measuring physical and mental well-being, functional impairment, and satisfaction with care (see Table 2).

Sample Size Calculation

The target sample size was based on power analysis for two primary outcomes: reduction of depressive symptoms (measured by PHQ-9 score) and depression remission. Power analyses were conducted using nQuery (Statistical Solutions Ltd, Boston MA) [65] to estimate effect sizes of the treatment with nonrandomized pre- and postintervention comparisons and longitudinal statistical approaches for repeated measures to compare trends in depression-related outcomes. The calculations assumed an alpha of .05, power of 0.80, attrition rates less than 20% at each 6-month follow-up assessment up to 18 months, and 25% to 30% depression prevalence among patients with diabetes [25]. For pre- and postintervention comparisons across all three program conditions, a sample size of approximately 500 in each study group would allow detection of a small effect size of 0.1.

Statistical Analysis

All analyses were carried out according to the intention-to-treat rule consistent with standard practice in most clinical trials. The propensity score method has proved to be an effective approach to analyzing observational or quasi-experimental studies [66-70]. A propensity score is defined as the probability that a patient is likely to receive treatment or control given the patient's baseline characteristics. Patients with the same propensity scores are like those in a randomized controlled trial.

The classical propensity score method is only applicable to two-way comparisons. Thus, we used a generalized propensity score (GPS) method designed for comparing two or more interventions versus one comparison group [71,72], wherein the GPS is defined as the conditional probability that a patient

is likely to be in a specific group given this patient's baseline characteristics. As recommended [72], a multinomial logistic regression was used to estimate GPSs. The model used study group as the dependent variable and the measured baseline characteristics shown in Table 3 (see "Results") as the independent variables. We subsequently checked the distribution of the estimated GPSs because comparisons between groups are suspect if substantial separation occurs between study groups [72-74].

Comparative treatment effects were estimated by linear or logistic regression models that used outcomes at 6 months as the dependent variable and study group, care team, outcome variables at baseline, two of the three estimated GPSs, insulin use, HBA_{1c}, age, gender, and preferred language as the independent variables. Regression that includes estimated GPSs as covariates has been shown to be an effective tool to adjust sample biases in observational or quasi-experimental studies [71,72]. Three care team variables were used to adjust for differences among providers. Two of the three estimated GPSs adjusted for imbalance in baseline characteristics. Insulin use, HBA_{1c}, age, gender, and preferred language were included because their effects on outcomes were of clinical interest; and their inclusion is consistent with prior findings in behavioral and clinical factors associated with depression in patients with diabetes [75]. The coefficients of study group predicted comparative treatment effects while controlling for other covariates. All statistical analyses were conducted at 0.05 significance level (two-tailed) using SAS (SAS Institute Inc, Cary NC) software, version 9.3.

Results

Baseline Characteristics and Participant Flow

A total of 1704 patients were screened, of which 101 patients met the exclusion criteria, 128 patients refused to participate, 12 patients did not sign the Health Insurance Portability and Accountability Act agreement, and 57 patients were excluded after not completing the baseline assessment. Men had a significantly lower enrollment rate than women (84.0% [536/638] vs 89.02% [949/1066], respectively; *P*=.003), which was partly associated with poor alcohol use scores (4.9% [31/638] for men vs 0.56% [6/1066] for women).

Among the 1406 patients enrolled in DCAT (484 in UC, 480 in SC, and 442 in TC), 1309 patients (416 in UC, 461 in SC, and 432 in TC) had complete data in the measures used in estimating the GPSs after interviews at baseline. As shown in Table 3, there were no significant differences in baseline depressive symptoms measured by the PHQ-9 score, anxiety symptoms measured by the Brief Symptom Inventory score, functional disability measured by the Sheehan Disability Scale (SDS), and the overall mental status measured by the 12-item Short Form Survey mental score. SC and TC patients had higher HBA_{1c} compared with UC patients because the SC and TC patients were enrolled from the DMP program designed for patients with severe diabetes. Other significant differences were diabetes self-care score and psychological stress measures (economic stress, number of stressors, sum of stress level, diabetes emotional burden, and diabetes regimen stress). The unbalanced samples were expected because of the quasi-experimental design. A Consolidated Standards of Reporting Trials diagram outlining participant flow is shown in Figure 1. See Multimedia Appendix 1 for comparison of baseline characteristics of samples included in versus excluded from the regression analysis. No significant differences were identified.

Six-Month Outcomes

With the final sample size of 1087 (341 in UC, 380 in SC, and 366 in TC) to evaluate intervention effects, the study has the statistical power of 0.80 to detect an effect size of Cohen *d*=0.12, a small effect size. Regression analysis with GPS adjustment results regarding 6-month outcomes in DCAT are shown in Tables 4 and 5. The GPSs were estimated by the previously described multinomial logistic regression model. The distributions of the estimated GPSs across study groups were similar; thus, treatment effects can be predicted based on the estimated GPSs instead of actual group assignment.

Compared with UC, both SC and TC were significantly associated with decreased PHQ-9 scores (least squares estimate, LSE: UC=6.35, SC=5.05, TC=5.16; *P* value: SC vs UC=.02, TC vs UC=.02) and reduced prevalence of depression as measured by PHQ-9 ≥10 (SC vs UC: adjusted odds ratio, AOR=0.45, 95% CI 0.23-0.88, *P*=.02; TC vs UC: AOR=0.33, 95% CI 0.17-0.65, *P*=.007). Only TC was significantly associated with improved depression remission relative to UC (AOR=2.98, 95% CI 1.08-8.25, *P*=.04), although SC came close. There were no significant differences in depression outcomes between the SC and TC groups.

Regarding diabetes care measures, no significant differences existed between SC and UC. However, TC was significantly associated with reduced total cholesterol level (LSE: UC=176.40, TC=160.46; *P*=.01) and increased odds that the patient would have an HBA_{1c} test (TC vs UC: AOR=3.40, 95% CI 1.58-7.31, *P*<.001). The latter was positively correlated with depression remission (AOR=2.67, 95% CI 1.15-4.17, *P*=.004). There were no significant differences in diabetes care measures between the SC and TC groups.

Table 3. Descriptive of baseline measures used in estimating the generalized propensity scores. PHQ-9: 9-item Patient Health Questionnaire; SC: supported care; SF-12: 12-item Short Form Survey; TC: technology-facilitated care; UC: usual care.

Baseline characteristic	Usual care (n=416)[a]	Supported care (n=461)[a]	Technology-facilitated care (n=432)[a]	SC vs UC (*P* value)	TC vs UC (*P* value)	TC vs SC (*P* value)
Age in years, mean (SD)	55.15 (9.21)	51.92 (9.29)	52.63 (8.74)	<.001	<.001	.47
Female, n (%)	293 (70.4)	271 (58.8)	266 (61.6)	<.001	.02	.66
Latino, n (%)	389 (94.0)	386 (83.7)	390 (90.5)	<.001	.23	.003
Spanish as preferred language, n (%)	366 (88.0)	360 (78.1)	352 (81.5)	<.001	.03	.38
Body mass index, mean (SD)	32.55 (7.04)	32.73 (7.64)	33.11 (7.16)	.92	.50	.72
Less than high school education, n (%)	310 (74.5)	287 (62.3)	306 (70.8)	<.001	.47	.02
Unemployed, n (%)	275 (66.1)	30 (67.0)	286 (66.2)	.96	.99	.96
Economic distress[b], mean (SD)	3.91 (2.44)	3.76 (1.98)	4.35 (2.10)	.57	.009	<.001
Number of stressors[c], mean (SD)	2.16 (2.20)	2.57 (2.30)	2.54 (2.11)	.02	.03	.98
Sum of stress level[d], mean (SD)	14.50 (16.23)	19.26 (19.49)	17.16 (16.87)	<.001	.07	.18
Predicted future health cost[e], mean (SD)	6711.47 (3347.32)	6839.82 (3854.07)	6376.52 (3930.77)	.87	.39	.15
Age at onset of diabetes, mean (SD)	45.20 (10.52)	41.84 (10.19)	42.32 (9.84)	<.001	<.001	.76
Insulin use, n (%)	114 (27.4)	310 (67.2)	282 (65.3)	<.001	<.001	.80
SF-12 physical, mean (SD)	43.24 (11.19)	45.83 (10.91)	43.96 (10.89)	.002	.61	.03
SF-12 mental, mean (SD)	50.09 (12.12)	49.33 (14.16)	50.39 (12.44)	.66	.94	.45
Number of diabetes complications[f], mean (SD)	0.71 (0.45)	0.73 (0.44)	0.65 (0.48)	.76	.15	.02
Whitty-9 diabetes symptoms scale[g], mean (SD)	1.67 (0.63)	1.71 (0.63)	1.56 (0.53)	.54	.03	<.001
Diabetes emotional burden[h], mean (SD)	2.75 (1.96)	3.69 (2.08)	2.53 (1.88)	<.001	.22	<.001
Diabetes regimen stress[h], mean (SD)	2.61 (1.91)	3.61 (2.11)	2.40 (1.85)	<.001	.27	<.001
Diabetes self-care[i], mean (SD)	4.04 (1.34)	4.76 (1.24)	4.23 (1.23)	<.001	.07	<.001
PHQ-9[j], mean (SD)	6.55 (5.51)	6.80 (6.43)	6.44 (5.97)	.81	.96	.65
Brief Symptom Inventory[k], mean (SD)	1.32 (3.02)	1.27 (3.24)	0.98 (2.72)	.96	.22	.32
Sheehan Disability Scale[l], mean (SD)	2.19 (2.80)	2.10 (3.00)	2.06 (2.87)	.87	.78	.98
Dysthymia, n (%)	55 (13.2)	116 (25.2)	64 (14.8)	<.001	.81	<.001
Previous diagnosis of major depressive disorder, n (%)	23 (5.5)	75 (16.3)	17 (3.9)	<.001	.68	<.001
Chronic pain, n (%)	127 (30.5)	129 (28.0)	71 (16.4)	.65	<.001	<.001
Satisfaction with diabetes care, mean (SD)	4.61 (0.74)	4.81 (0.50)	4.67 (0.53)	<.001	.33	<.001
Satisfaction with care for emotional problems, mean (SD)	4.22 (0.99)	4.70 (0.63)	4.52 (0.66)	<.001	<.001	<.001
HBA_{1c} value, mean (SD)	8.37 (1.93)	9.57 (2.20)	9.73 (1.93)	<.001	<.001	.46

[a]Values are numbers (column percentage) for categorical variables and mean (SD) for continuous variables.

[b]Assessed by 12 general and health-related economic distresses, scored 0-12; higher scores indicate a higher level of economic distress.

[c]Assessed by 12 stressors related to work, family, social, and legal problems, scored 0-12; higher scores indicate a larger number of stressors.

[d]Assessed by 12 stressors related to work, family, social, and legal problems, each rated by level of stress from 0-10; therefore, total scores range from 0-120, with higher scores indicating a higher level of stress.

[e]Prediction of future health cost using the RxRisk model [76].

[f]Assessed by 7 diabetes complications: vision problems, loss of feeling in feet or legs, kidney problems, foot ulcer, amputation, sexual impairment, and heart attack, scored 0-7; higher scores indicate a larger number of diabetes complications.

[g]Assessed by the 9-item diabetes symptoms scale [77], scored 1-5; higher scores indicate more severe diabetes.

[h]Assessed by the 2-item Diabetes Distress Scale [78], scored 1-6; higher scores indicate a higher level of diabetes distress.

[i]Assessed by the Toobert Diabetes Selfcare Scale [61], scored 0-7; higher scores indicate better diabetes self-care.

[j]Assessed by the 9-item Patient Health Questionnaire [58], scored 0-27; higher scores indicate worse depressive symptoms.

[k]Assessed by the Brief Symptoms Inventory [79], scored 0-24; higher scores indicate worse anxiety.

[l]Assessed by the Sheehan Disability Scale [63,64], scored 0-30; higher scores indicate more significant functional impairment.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram: participant flow of Diabetes-Depression Care-Management Adoption Trial (DCAT).

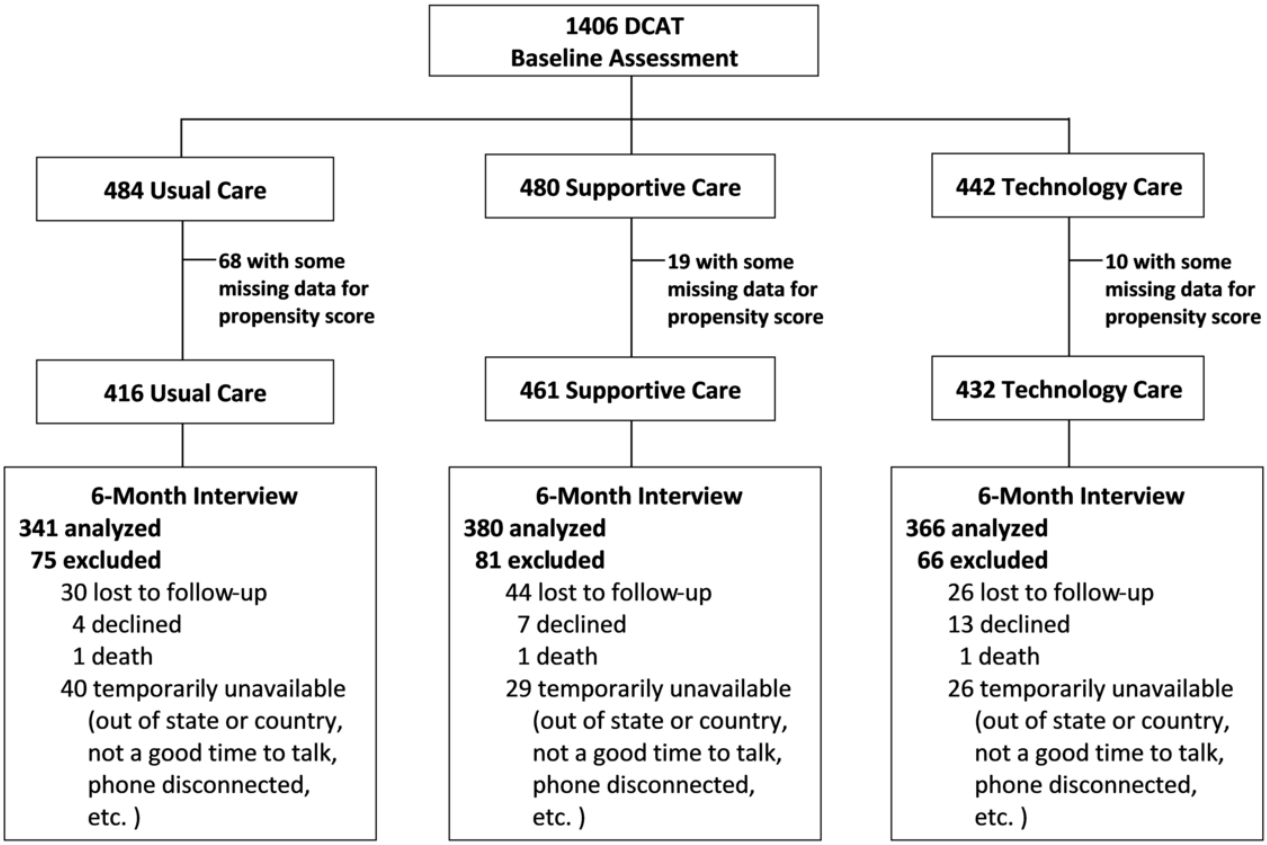


Table 4. Regression analysis of continuous 6-month outcomes adjusted for baseline characteristics and propensity scores in the Diabetes-Depression Care-management Adoption Trial (DCAT), Los Angeles, 2011 to 2013. Linear regression models are adjusted for care team, outcome variable at baseline, two of the three estimated generalized propensity scores, insulin use, glycated hemoglobin (HBA_{1c}), age, gender, and preferred language. Least squares means and SE reported for continuous outcomes. LSE: least squares estimate; PHQ-9: 9-item Patient Health Questionnaire; SC: standard care; SF-12: 12-item Short Form Survey; TC: technology-facilitated care; UC: usual care.

Continuous outcome	Usual care (n=341), LSE (SE)	Supported care (n=380), LSE (SE)	Technology-facilitated care (n=366), LSE (SE)	SC vs UC (*P* value)	TC vs UC (*P* value)	TC vs SC (*P* value)
PHQ-9	6.35 (0.49)	5.05 (0.47)	5.16 (0.48)	.02	.02	.81
HBA_{1c} value	7.95 (0.17)	7.79 (0.16)	8.05 (0.16)	.41	.57	.10
Total cholesterol	176.40 (5.27)	166.90 (4.96)	160.46 (5.04)	.12	.01	.16
Diabetes self-care	4.66 (0.13)	4.70 (0.12)	4.78 (0.12)	.80	.38	.52
Exercise	4.73 (0.28)	4.90 (0.26)	4.86 (0.27)	.59	.66	.88
SF-12 physical score	42.99 (0.97)	42.46 (0.95)	41.87 (0.95)	.63	.27	.55
SF-12 mental score	48.38 (1.04)	50.07 (1.01)	49.87 (1.02)	.16	.17	.85
Sheehan Disability Scale	3.21 (0.26)	2.61 (0.25)	2.59 (0.25)	.04	.03	.95
Satisfaction with diabetes care	4.01 (0.09)	4.15 (0.09)	4.20 (0.09)	.17	.05	.58
Satisfaction with care for emotional problems	3.25 (0.10)	3.64 (0.10)	3.46 (0.10)	.01	.07	.06
Satisfaction with care for emotional problems, among patients with baseline PHQ-9 ≥10	3.20 (0.22)	3.58 (0.21)	3.70 (0.21)	.16	.05	.56

Table 5. Regression analysis of binary 6-month outcomes adjusted for baseline characteristics and propensity scores in the DCAT, Los Angeles, 2011-2013. Logistic regression models are adjusted for care team, outcome variable at baseline, two of the three estimated generalized propensity scores, insulin use, glycated hemoglobin (HBA_{1c}), age, gender, and preferred language. Adjusted odds ratio (AOR) and 95% CIs reported for binary outcomes. PHQ-9: 9-item Patient Health Questionnaire

Binary outcome	Supported care vs usual care		Technology-facilitated care vs usual care		Technology-facilitated care vs supported care	
	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value
PHQ-9≥10	0.45 (0.23-0.88)	.02	0.33 (0.17-0.65)	.007	0.75 (0.39-1.41)	.37
Depression remission	2.86 (0.98-8.40)	.06	2.98 (1.08-8.25)	.04	1.04 (0.47-2.31)	.92
HBA_{1c} tested[a]	1.82 (0.89-3.71)	.10	3.40 (1.58-7.31)	<.001	1.87 (0.82-4.27)	.14

[a]Adjusted relative risk for HBA_{1c} tested, supported care vs usual care=1.13 (0.97-1.23), technology-facilitated care vs usual care=1.22 (1.10-1.29), technology-facilitated care vs supported care=1.12 (0.95-1.21).

Both SC and TC were significantly associated with improved SDS scores relative to UC (LSE: UC=3.21, SC=2.61, TC=2.59; *P* value: SC vs UC=.04, TC vs UC=.03). SC was significantly associated with improved satisfaction with care for emotional problems compared with UC (LSE: UC=3.25, SC=3.64; *P*=.01), but only TC was significantly associated with improved satisfaction with diabetes care (LSE: UC=4.01, TC=4.20; *P*=.05) and satisfaction with care for emotional problems among patients with depression, as measured by PHQ-9 ≥10 at baseline (LSE: UC=3.20, TC=3.70; *P*=.05). There were no significant differences in patient-reported outcomes between the SC and TC groups.

Discussion

Principal Findings

Analysis of 6-month DCAT outcomes revealed that both the TC and SC groups were significantly associated with better outcomes compared with UC in terms of depressive symptoms reduction. Using the PHQ-9 score, which ranges from 0 to 27, the 1.3 (SC group) and 1.2 (TC group) points of improvements compared with the UC group are clinically meaningful given that the baseline PHQ-9 score is only about 6.5 points. The magnitude of improvements is consistent with a recent collaborative depression care study that included both depressed and nondepressed patients [80]. This finding supports the hypothesis that the two intervention groups would be associated with better depression care outcomes.

It was not surprising to find the ATA technology did not improve depression outcomes of the TC group over the SC group in this case because it was designed to facilitate the adoption of collaborative depression care rather than direct clinical intervention with patients. Clinically, DMP care teams in both groups were trained in and practiced the LACDHS depression protocol and treatment guideline. The SC DMP providers monitored patients using traditional mechanisms

(specifically by calling patients), and they coordinated care among themselves; therefore, the study-related depression care resulted in additional new workload for the SC providers. For the TC care team, the technology, albeit at a low-intensity of contact, helped alleviate the workload for depression symptom and treatment adherence monitoring.

The ATA technology also prompted providers to follow up and alerted emergency responders to immediately contact patients with suicidal ideation. That could be the reason why only the TC group was significantly associated with depression remission and increased patient satisfaction among depressed patients. The results are encouraging that a well-designed technology can be an effective aide to the adoption of collaborative depression care. Full justification of the TC model via a complete cost-effectiveness analysis is presented elsewhere [56]. The cost-effectiveness analysis revealed that the intervention models improved quality-adjusted life years, depression-free days, and medical costs. The TC model was cost-effective compared with SC and cost-saving compared with UC. The 6-month and cost-effectiveness results suggest the TC model is promising in facilitating better and cost-effective care for depressed patients.

Moreover, only the TC model was significantly associated with improved diabetes care processes, indicated by reduced total cholesterol level and increased odds that the patient would have an HBA_{1c} test. One possible explanation for these improvements in diabetes care is that as depressive symptoms are increasingly monitored and timelier addressed, patients may become more willing to take active care of their diabetes; this explanation is supported by the significant correlation between the odds of having an HBA_{1c} test and the improvement in depression remission. Another possible explanation is that providers may address patients' diabetes care needs in addition to depression care needs when they respond to the task reminders generated by the technology.

Compared with the UC group, both the SC and TC models were significantly associated with better improvement in 6-month patient-reported functioning in family, work, and social life, as measured by the 3-item SDS. The 0.6 points of improvements in SC and TC groups compared with the UC group are meaningful as most patients at baseline had only minimal functional impairment (average baseline SDS score was 2.1 points; SDS score >6 indicates functional disability [63,64]), which implies that the room for improvement is small. This finding suggests that the two enhanced care delivery models not only improved depressive symptoms but also translated such symptom improvement into better perceived life functioning.

While the TC model delivered positive results, and most patients in the TC groups reported high acceptance of the ICT tested in DCAT [51], there is significant room for improvement in using the ATA technology. Specifically, only half of the scheduled calls were answered successfully because of phone connectivity issues or lack of time for the patient to answer the calls [50,52]. One challenge may have been that during the study, most patients in LACDHS did not use cellphones. Now that cellphones are more readily available, attention should be turned to other ICT (such as SMS text messages and smartphone apps) to improve patient contact and to capture patient-reported outcomes. Such technology has greater portability and versatility, may extend the ATA capabilities in reaching and engaging patients, may potentially increase the model effectiveness, and may reduce costs.

Researchers can expect unpredictable consequences after making changes (such as the DCAT implementation) in complex systems such as LACDHS. As discussed in another DCAT study [53], two particularly important implications that emerged were the strengthened role of social workers (in both SC and TC) and the importance of suicide-alert responders (in TC). Every DMP site had a colocated clinical social worker, an evidence-based method of improving quality of depression care [81,82]. The clinical social workers were an underutilized resource before the study; during the DCAT trial, they proved instrumental in adopting depression care. Furthermore, the suicide-alert responders appeared to play a much larger role than anticipated. Providers facing typical barriers in mental health care (including lack of familiarity with guidelines, lack of self-efficacy, and lack of outcome expectancy [83]) were reassured by the availability of an organizational resource for the patient to fall back on in the "worst case scenario," namely severe suicidal ideation. Taken together with the strengthened role of the social worker, the interventions seemed to have leveraged the available mental health resources into a more cohesive, integrated model of mental health care in a primary setting. In other words, the SC and TC models used existing diabetes disease management teams and leveraged available mental health resources to implement depression symptom monitoring and treatment protocols, provider collaborative communication, and patient relapse prevention.

In summary, the 6-month DCAT findings suggest that both the TC and SC delivery models are significantly associated with improved depression outcomes and life functioning, and that the TC model offers additional promise in terms of improved depression remission, diabetes care processes, and patient satisfaction. Given the rapid rise of diabetes during the past several decades—especially among low-income, minority populations—and the immense opportunity to improve diabetes-related measures and outcomes, a growing number of health plans and health care organizations are trying to manage their diabetes population through disease management programs. The SC and TC models demonstrated that an important and valuable way to support providers is to add evidence-based collaborative depression care and facilitate adoption of ICT in diabetes DMPs designed to reduce disparities in commonly comorbid diabetes and depression care. When enhanced by ICT, DMPs may be able to greatly improve overall care, cost, and effectiveness of health care delivery for underserved patients. DCAT SC and TC models improved diabetes and depression outcomes in the second largest US safety-net health system; other resource-constrained programs may replicate these models to improve comorbid diabetes and depression outcomes.

Limitations

The main limitation of this study is its employment of a quasi-experimental design, which introduces bias because of

the differences in both patient characteristics and care teams at each facility. To mitigate the bias, in the regression analysis we adjusted for patient differences through propensity scores and the assignment of the six care teams at eight facilities (each facility was staffed by only one team; two of the six teams served two facilities). Although care facilities were matched by geographic location and patient sociodemographics among the three study groups, the quality of care can vary from facility to facility; therefore, the regression analysis included a check in which care team assignments were replaced for each facility. This analysis did not change the direction and significance of intervention effects; however, the adjustment may not be sufficient. Differences in the facilities, the DMP care teams, and the unmeasured patient sociodemographics, diabetes and comorbid conditions, and psychological stress measures may differentiate diabetes and depression care needs and outcomes [84]. Providers should consider these differences when applying the technology.

Another limitation may be the predominantly Latino sample, which raises concern about the generalizability of findings. Applying the DCAT TC model to other groups should be done cautiously and with further evaluation.

The third limitation is the focus in this paper on clinical outcomes. However, the DCAT TC model was designed to accelerate the adoption of evidence-based collaborative care to improve the overall care process. Full analyses of 18-month clinical outcomes and the cost-effectiveness of the TC model will be reported elsewhere.

The fourth limitation is lack of data to understand the practical mechanism by which the SC DMP and the technology-enhanced TC DMP led to improvements in depression. Possible improvement mechanisms for future research include better treatment initiation or adjustment, receipt of PST, greater patient adherence, or referrals and visits to other mental health providers.

Comparison With Prior Work

DCAT adds to the growing number of telehealth studies that are employing technology to improve depression care in primary care settings for patients with chronic diseases [85-92]. A key strength of the DCAT TC model over earlier studies is that it used automated calls, which reduces provider depression monitoring workload and allows more time for clinical encounters such as timely adjustment of treatment. The TC model is especially effective in a resource-constrained environment such as safety-net care systems, improving care for predominantly minority and low-income patients.

Applying automated remote monitoring ICT, electronic clinical decision support, and even artificial intelligence to facilitate chronic disease management is an emerging research topic. Prior studies revealed that ATA is valid in conducting depression screening and suggested the technology can be incorporated into the care management model [93,94]. However, evidence is limited regarding the comparative effectiveness of the technology. Kroenke et al [95] tested ATA with care management for pain and depression in patients with cancer; Ratanawongsa et al [96] and Handley et al [97] tested ATA-facilitated diabetes management for low-income Medicaid and safety-net patients. The DCAT study uniquely addressed depression care for patients with type 2 diabetes in a safety-net setting where comorbidity of the two diseases is common and where many barriers such as culture diversity, financial stress, and limited provider resources impede the adoption of evidence-based depression interventions. Results from DCAT are consistent with prior studies [91-97] and support ATA as a promising technology to facilitate care management, even for sensitive conditions such as depression, for diverse populations and in primary care settings.

Conclusions

Both SC and TC models are associated with improved 6-month depression outcomes and reduced functional disability among adult patients with type 2 diabetes. However, the TC model is more likely to achieve greater improvements in depression remission, as well as measures of patient satisfaction and diabetes care quality. This paper provides encouraging evidence that a well-designed automated ICT system is an effective facilitator that can support delivery of evidence-based collaborative depression care to patients with type 2 diabetes in a resource-constrained urban safety-net primary care setting. This is a promising solution to reduce health disparities, improve patient experience of care, and improve the health of low-income, minority populations.

Acknowledgments

This project was supported with a grant from the US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation (Grant No. 1R18AE000054-01). SW was the principal investigator. The authors acknowledge the LACDHS clinics, providers, DMP staff, and patients who assisted with or participated in DCAT; the research team for its contribution to this study; and 4PatientCare for assisting with the development of the ATA calls. The authors also posthumously acknowledge Dr Wayne Katon for providing training on the collaborative depression care model and stepped care depression treatment approach to the participating clinics.

Conflicts of Interest

JG reports a proprietary or commercial interest in the ATA system discussed in this paper. He is the chief medical officer of 4PatientCare. He receives consulting remuneration and owns a significant equity position in the company.

Multimedia Appendix 1

Comparison of baseline characteristics of included vs excluded samples for the regression analysis.

[PDF File (Adobe PDF File), 29KB - jmir_v20i4e147_app1.pdf]

References

1. Li C, Ford ES, Zhao G, Ahluwalia IB, Pearson WS, Mokdad AH. Prevalence and correlates of undiagnosed depression among U.S. 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. [doi: 10.1007/s11606-016-3650-1] [Medline: 26969312]
6. Ducat L, Philipson LH, Anderson BJ. The mental health comorbidities of diabetes. J Am Med Assoc 2014 Aug 20;312(7):691-692 [FREE Full text] [doi: 10.1001/jama.2014.8040] [Medline: 25010529]
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. Chen PC, Chan YT, Chen HF, Ko MC, Li CY. Population-based cohort analyses of the bidirectional relationship between type 2 diabetes and depression. Diabetes Care 2013 Feb;36(2):376-382 [FREE Full text] [doi: 10.2337/dc12-0473] [Medline: 23150281]
9. Golden SH, Lazo M, Carnethon M, Bertoni AG, Schreiner PJ, Diez Roux AV, et al. Examining a bidirectional association between depressive symptoms and diabetes. J Am Med Assoc 2008 Jun 18;299(23):2751-2759 [FREE Full text] [doi: 10.1001/jama.299.23.2751] [Medline: 18560002]
10. de Groot M, Anderson R, Freedland KE, Clouse RE, Lustman PJ. Association of depression and diabetes complications: a meta-analysis. Psychosom Med 2001;63(4):619-630. [Medline: 11485116]
11. Egede LE. Diabetes, major depression, and functional disability among U.S. adults. Diabetes Care 2004 Feb;27(2):421-428. [Medline: 14747223]
12. 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. [Medline: 14514586]
13. Olfson M, Shea S, Feder A, Fuentes M, Nomura Y, Gameroff M, et al. Prevalence of anxiety, depression, and substance use disorders in an urban general medicine practice. Arch Fam Med 2000;9(9):876-883. [Medline: 11031395]
14. Schneiderman N, Llabre M, Cowie CC, Barnhart J, Carnethon M, Gallo LC, et al. Prevalence of diabetes among Hispanics/Latinos from diverse backgrounds: the Hispanic community health study/study of Latinos (HCHS/SOL). Diabetes Care 2014 Aug;37(8):2233-2239 [FREE Full text] [doi: 10.2337/dc13-2939] [Medline: 25061138]
15. Stark Casagrande S, Fradkin JE, Saydah SH, Rust KF, Cowie CC. The prevalence of meeting HBA 1c, blood pressure, and LDL goals among people with diabetes, 1988-2010. Diabetes Care 2013 Aug;36(8):2271-2279 [FREE Full text] [doi: 10.2337/dc12-2258] [Medline: 23418368]
16. 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]
17. Atdjian S, Vega WA. Disparities in mental health treatment in U.S. racial and ethnic minority groups: implications for psychiatrists. Psychiatr Serv 2005 Dec;56(12):1600-1602. [doi: 10.1176/appi.ps.56.12.1600] [Medline: 16339626]
18. Chow JC, Jaffee K, Snowden L. Racial/ethnic disparities in the use of mental health services in poverty areas. Am J Public Health 2003 May;93(5):792-797. [Medline: 12721146]
19. Jimenez DE, Cook B, Bartels SJ, Alegría M. Disparities in mental health service use of racial and ethnic minority elderly adults. J Am Geriatr Soc 2013 Jan;61(1):18-25 [FREE Full text] [doi: 10.1111/jgs.12063] [Medline: 23252464]
20. 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]
21. McGuire TG, Miranda J. New evidence regarding racial and ethnic disparities in mental health: policy implications. Health Aff (Millwood) 2008;27(2):393-403 [FREE Full text] [doi: 10.1377/hlthaff.27.2.393] [Medline: 18332495]

22. Miranda J, McGuire TG, Williams DR, Wang P. Mental health in the context of health disparities. Am J Psychiatry 2008 Sep;165(9):1102-1108. [doi: 10.1176/appi.ajp.2008.08030333] [Medline: 18765491]
23. Wells K, Klap R, Koike A, Sherbourne C. Ethnic disparities in unmet need for alcoholism, drug abuse, and mental health care. Am J Psychiatry 2001 Dec;158(12):2027-2032. [doi: 10.1176/appi.ajp.158.12.2027] [Medline: 11729020]
24. Ell K, Katon W, Cabassa LJ, Xie B, Lee P, Kapetanovic S, et al. Depression and diabetes among low-income Hispanics: design elements of a socioculturally adapted collaborative care model randomized controlled trial. Int J Psychiatry Med 2009;39(2):113-132 [FREE Full text] [doi: 10.2190/PM.39.2.a] [Medline: 19860071]
25. Ell K, Katon W, Xie B, Lee P, Kapetanovic S, Guterman J, et al. Collaborative care management of major depression among low-income, predominantly Hispanic subjects with diabetes: a randomized controlled trial. Diabetes Care 2010 Apr;33(4):706-713 [FREE Full text] [doi: 10.2337/dc09-1711] [Medline: 20097780]
26. Ell K, Xie B, Quon B, Quinn DI, Dwight-Johnson M, Lee PJ. Randomized controlled trial of collaborative care management of depression among low-income patients with cancer. J Clin Oncol 2008 Sep 20;26(27):4488-4496 [FREE Full text] [doi: 10.1200/JCO.2008.16.6371] [Medline: 18802161]
27. Ell K, Xie B, Kapetanovic S, Quinn DI, Lee PJ, Wells A, et al. One-year follow-up of collaborative depression care for low-income, predominantly Hispanic patients with cancer. Psychiatr Serv 2011 Feb;62(2):162-170. [doi: 10.1176/ps.62.2.pss6202_0162] [Medline: 21285094]
28. Eisenberg Center at Oregon Health & Science University. Choosing antidepressants for adults: clinician's guide. In: Comparative Effectiveness Review Summary Guides for Clinicians. Rockville, MD: Agency for Healthcare Research and Quality; 2007.
29. Anderson RJ, Gott BM, Sayuk GS, Freedland KE, Lustman PJ. Antidepressant pharmacotherapy in adults with type 2 diabetes: rates and predictors of initial response. Diabetes Care 2010 Mar;33(3):485-489 [FREE Full text] [doi: 10.2337/dc09-1466] [Medline: 20032276]
30. Ell K, Aranda MP, Xie B, Lee PJ, Chou CP. Collaborative depression treatment in older and younger adults with physical illness: pooled comparative analysis of three randomized clinical trials. Am J Geriatr Psychiatry 2010 Jun;18(6):520-530 [FREE Full text] [doi: 10.1097/JGP.0b013e3181cc0350] [Medline: 20220588]
31. González HM, Vega WA, Williams DR, Tarraf W, West BT, Neighbors HW. Depression care in the United States: too little for too few. Arch Gen Psychiatry 2010 Jan;67(1):37-46 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.168] [Medline: 20048221]
32. Bartels SJ, Coakley EH, Zubritsky C, Ware JH, Miles KM, Areán PA, PRISM-E Investigators. Improving access to geriatric mental health services: a randomized trial comparing treatment engagement with integrated versus enhanced referral care for depression, anxiety, and at-risk alcohol use. Am J Psychiatry 2004 Aug;161(8):1455-1462. [doi: 10.1176/appi.ajp.161.8.1455] [Medline: 15285973]
33. 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]
34. Institute Of Medicine. America's Health Care Safety Net: Intact But Endangered. Washington, DC: The National Academies Press; 2000.
35. 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]
36. 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]
37. Heithoff K. Does the ECA underestimate the prevalence of late-life depression? J Am Geriatr Soc 1995 Jan;43(1):2-6. [Medline: 7806734]
38. 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. [Medline: 7884106]
39. Richards D. Prevalence and clinical course of depression: a review. Clin Psychol Rev 2011 Nov;31(7):1117-1125. [doi: 10.1016/j.cpr.2011.07.004] [Medline: 21820991]
40. Katon WJ, Von Korff M, Lin EH, Simon G, Ludman E, Russo J, et al. The pathways Study: a randomized trial of collaborative care in patients with diabetes and depression. Arch Gen Psychiatry 2004 Oct;61(10):1042-1049. [doi: 10.1001/archpsyc.61.10.1042] [Medline: 15466678]
41. 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]
42. Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. N Engl J Med 2010 Dec 30;363(27):2611-2620 [FREE Full text] [doi: 10.1056/NEJMoa1003955] [Medline: 21190455]

43. Unützer J, Katon W, Callahan CM, Williams Jr JW, Hunkeler E, Harpole L, IMPACT Investigators. Improving Mood-Promoting Access to Collaborative Treatment. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. J Am Med Assoc 2002 Dec 11;288(22):2836-2845. [Medline: 12472325]
44. Katon W, Unützer J, Fan MY, Williams Jr JW, Schoenbaum M, Lin EH, et al. Cost-effectiveness and net benefit of enhanced treatment of depression for older adults with diabetes and depression. Diabetes Care 2006 Feb;29(2):265-270. [Medline: 16443871]
45. Katon W, Russo J, Lin EH, Schmittdiel J, Ciechanowski P, Ludman E, et al. Cost-effectiveness of a multicondition collaborative care intervention: a randomized controlled trial. Arch Gen Psychiatry 2012 May;69(5):506-514 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2011.1548] [Medline: 22566583]
46. Von Korff M, Katon W, Bush T, Lin EH, Simon GE, Saunders K, et al. Treatment costs, cost offset, and cost-effectiveness of collaborative management of depression. Psychosom Med 1998;60(2):143-149. [Medline: 9560861]
47. Simon GE, Katon WJ, VonKorff M, Unützer J, Lin EH, Walker EA, et al. Cost-effectiveness of a collaborative care program for primary care patients with persistent depression. Am J Psychiatry 2001 Oct;158(10):1638-1644. [doi: 10.1176/appi.ajp.158.10.1638] [Medline: 11578996]
48. Katon W, Unützer J. Collaborative care models for depression: time to move from evidence to practice. Arch Intern Med 2006 Nov 27;166(21):2304-2306. [doi: 10.1001/archinte.166.21.2304] [Medline: 17130381]
49. Logan DE, Claar RL, Scharff L. Social desirability response bias and self-report of psychological distress in pediatric chronic pain patients. Pain 2008 Jun;136(3):366-372. [doi: 10.1016/j.pain.2007.07.015] [Medline: 17822848]
50. Vidyanti I, Wu B, Wu S. Low-income minority patient engagement with automated telephonic depression assessment and impact on health outcomes. Qual Life Res 2015 May;24(5):1119-1129 [FREE Full text] [doi: 10.1007/s11136-014-0900-8] [Medline: 25543270]
51. Ramirez M, Wu S, Jin H, Ell K, Gross-Schulman S, Myerchin Sklaroff L, 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]
52. 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]
53. Di Capua P, Wu B, Sednew R, Ryan G, Wu S. Complexity in redesigning depression care: comparing intention versus implementation of an automated depression screening and monitoring program. Popul Health Manag 2016 Dec;19(5):349-356. [doi: 10.1089/pop.2015.0084] [Medline: 27028043]
54. 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 03;12:E142 [FREE Full text] [doi: 10.5888/pcd12.150047] [Medline: 26334714]
55. 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]
56. Hay JW, Lee PJ, 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 2018:1-8 (forthcoming). [doi: 10.1016/j.jval.2017.11.005]
57. Patrick DL, Guyatt GH, Acquadro C. Patient-reported outcomes. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons; 2008.
58. 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] [Medline: 11556941]
59. Nezu AM, Nezu CM, D'Zurilla T. Problem-Solving Therapy: A Treatment Manual. New York, NY: Springer Publishing Company; 2013.
60. Rabiner L, Juang BH. Fundamentals of Speech Recognition. Upper Saddle River, NJ: Prentice-Hall; 1993.
61. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950 [FREE Full text] [Medline: 10895844]
62. Stewart AL, Ware Jr JE. Measuring Function and Well-Being: The Medical Outcomes Study Approach. Durham, NC: Duke University Press; 1992.
63. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol 1996 Jun;11 Suppl 3:89-95. [Medline: 8923116]
64. Sheehan KH, Sheehan DV. Assessing treatment effects in clinical trials with the discan metric of the Sheehan disability scale. Int Clin Psychopharmacol 2008 Mar;23(2):70-83. [doi: 10.1097/YIC.0b013e3282f2b4d6] [Medline: 18301121]
65. Elashoff JD. Hubspot. Boston, MA: Statistical Solutions; 2007. nQuery Advisor Version 7.0 User's Guide URL: http://cdn2.hubspot.net/hubfs/488764/nQ70_Manual.pdf [accessed 2018-03-28] [WebCite Cache ID 6yFiaYoPB]
66. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984 Sep;79(387):516-524. [doi: 10.2307/2288398]
67. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55. [doi: 10.1093/biomet/70.1.41]

68. van den Eeden SK, Glasser M, Mathias SD, Colwell HH, Pasta DJ, Kunz K. Quality of life, health care utilization, and costs among women undergoing hysterectomy in a managed-care setting. Am J Obstet Gynecol 1998 Jan;178(1 Pt 1):91-100. [Medline: 9465810]
69. Mitra N, Indurkhya A. A propensity score approach to estimating the cost-effectiveness of medical therapies from observational data. Health Econ 2005 Aug;14(8):805-815. [doi: 10.1002/hec.987] [Medline: 15791679]
70. Hammond GC, Croudace TJ, Radhakrishnan M, Lafortune L, Watson A, McMillan-Shields F, et al. Comparative effectiveness of cognitive therapies delivered face-to-face or over the telephone: an observational study using propensity methods. PLoS One 2012;7(9):e42916 [FREE Full text] [doi: 10.1371/journal.pone.0042916] [Medline: 23028436]
71. Imbens GW. The role of the propensity score in estimating dose-response functions. Biometrika 2000 Sep 01;87(3):706-710. [doi: 10.1093/biomet/87.3.706]
72. Spreeuwenberg MD, Bartak A, Croon MA, Hagenaars JA, Busschbach JJ, Andrea H, et al. The multiple propensity score as control for bias in the comparison of more than two treatment arms: an introduction from a case study in mental health. Med Care 2010 Feb;48(2):166-174. [doi: 10.1097/MLR.0b013e3181c1328f] [Medline: 20068488]
73. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol 2006 Jun 15;163(12):1149-1156 [FREE Full text] [doi: 10.1093/aje/kwj149] [Medline: 16624967]
74. Pasta DJ. SAS. San Francisco, CA: Lewin Group; 2000. Using propensity scores to adjust for group differences: examples comparing alternative surgical methods URL: http://www2.sas.com/proceedings/sugi25/25/st/25p261.pdf [accessed 2018-03-28] [WebCite Cache ID 6yFix5jQI]
75. Katon W, Von Korff M, Ciechanowski P, Russo J, Lin E, Simon G, et al. Behavioral and clinical factors associated with depression among individuals with diabetes. Diabetes Care 2004 Apr;27(4):914-920. [Medline: 15047648]
76. Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'Keeffe Rosetti MC. Risk adjustment using automated ambulatory pharmacy data: the RxRisk model. Med Care 2003 Jan;41(1):84-99. [doi: 10.1097/01.MLR.0000039830.19812.29] [Medline: 12544546]
77. Whitty P, Steen N, Eccles M, McColl E, Hewison J, Meadows K, et al. A new self-completion outcome measure for diabetes: is it responsive to change? Qual Life Res 1997 Jul;6(5):407-413. [Medline: 9290307]
78. Fisher L, Glasgow RE, Mullan JT, Skaff MM, Polonsky WH. Development of a brief diabetes distress screening instrument. Ann Fam Med 2008;6(3):246-252 [FREE Full text] [doi: 10.1370/afm.842] [Medline: 18474888]
79. Derogatis LR, Savitz KL. The SCL-90-R brief symptom inventory (BSI) in primary care. In: Marvish ME, editor. Handbook of Psychological Assessment in Primary Care Settings. Mahwah, NJ: Erlbaum; 2000.
80. Gilbody S, Lewis H, Adamson J, Atherton K, Bailey D, Birtwistle J, et al. Effect of collaborative care vs usual care on depressive symptoms in older adults with subthreshold depression: the CASPER randomized clinical trial. J Am Med Assoc 2017 Dec 21;317(7):728-737. [doi: 10.1001/jama.2017.0130] [Medline: 28241357]
81. Pincus HA, Pechura CM, Elinson L, Pettit AR. Depression in primary care: linking clinical and systems strategies. Gen Hosp Psychiatry 2001;23(6):311-318. [Medline: 11738461]
82. Menear M, Duhoux A, Roberge P, Fournier L. Primary care practice characteristics associated with the quality of care received by patients with depression and comorbid chronic conditions. Gen Hosp Psychiatry 2014;36(3):302-309. [doi: 10.1016/j.genhosppsych.2014.01.013] [Medline: 24629824]
83. Maxwell M, Harris F, Hibberd C, Donaghy E, Pratt R, Williams C, et al. A qualitative study of primary care professionals' views of case finding for depression in patients with diabetes or coronary heart disease in the UK. BMC Fam Pract 2013 Apr 04;14:46 [FREE Full text] [doi: 10.1186/1471-2296-14-46] [Medline: 23557512]
84. Eckel RH, Kahn SE, Ferrannini E, Goldfine AB, Nathan DM, Schwartz MW, et al. Obesity and type 2 diabetes: what can be unified and what needs to be individualized? J Clin Endocrinol Metab 2011 Jun;96(6):1654-1663 [FREE Full text] [doi: 10.1210/jc.2011-0585] [Medline: 21602457]
85. Hunkeler EM, Meresman JF, Hargreaves WA, Fireman B, Berman WH, Kirsch AJ, et al. Efficacy of nurse telehealth care and peer support in augmenting treatment of depression in primary care. Arch Fam Med 2000 Aug;9(8):700-708. [Medline: 10927707]
86. Gellis ZD, Kenaley BL, Ten Have T. Integrated telehealth care for chronic illness and depression in geriatric home care patients: the integrated telehealth education and activation of mood (I-TEAM) study. J Am Geriatr Soc 2014 May;62(5):889-895. [doi: 10.1111/jgs.12776] [Medline: 24655228]
87. Choi NG, Marti CN, Bruce ML, Hegel MT, Wilson NL, Kunik ME. Six-month postintervention depression and disability outcomes of in-home telehealth problem-solving therapy for depressed, low-income homebound older adults. Depress Anxiety 2014 Aug;31(8):653-661 [FREE Full text] [doi: 10.1002/da.22242] [Medline: 24501015]
88. Lichstein KL, Scogin F, Thomas SJ, DiNapoli EA, Dillon HR, McFadden A. Telehealth cognitive behavior therapy for co-occurring insomnia and depression symptoms in older adults. J Clin Psychol 2013 Oct;69(10):1056-1065 [FREE Full text] [doi: 10.1002/jclp.22030] [Medline: 24014056]
89. Gilbody S, Whitty P, Grimshaw J, Thomas R. Educational and organizational interventions to improve the management of depression in primary care: a systematic review. J Am Med Assoc 2003 Jun 18;289(23):3145-3151. [doi: 10.1001/jama.289.23.3145] [Medline: 12813120]

90. Simon GE, Ludman EJ, Tutty S, Operskalski B, Von Korff KM. Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment: a randomized controlled trial. J Am Med Assoc 2004 Aug 25;292(8):935-942. [doi: 10.1001/jama.292.8.935] [Medline: 15328325]
91. Dietrich AJ, Oxman TE, Williams Jr JW, Schulberg HC, Bruce ML, Lee PW, et al. Re-engineering systems for the treatment of depression in primary care: cluster randomised controlled trial. Br Med J 2004 Sep 11;329(7466):602 [FREE Full text] [doi: 10.1136/bmj.38219.481250.55] [Medline: 15345600]
92. Sheldon C, Waxmonsky JA, Meir R, Morris C, Finkelstein L, Sosa M, et al. Telephone assessment, support, and counseling for depression in primary care medical clinics. Cogn Behav Pract 2014 Aug;21(3):282-295. [doi: 10.1016/j.cbpra.2014.04.005]
93. Turvey CL, Willyard D, Hickman DH, Klein DM, Kukoyi O. Telehealth screen for depression in a chronic illness care management program. Telemed J E Health 2007 Feb;13(1):51-56. [doi: 10.1089/tmj.2006.0036] [Medline: 17309355]
94. Sheeran T, Rabinowitz T, Lotterman J, Reilly CF, Brown S, Donehower P, et al. Feasibility and impact of telemonitor-based depression care management for geriatric homecare patients. Telemed J E Health 2011 Oct;17(8):620-626 [FREE Full text] [doi: 10.1089/tmj.2011.0011] [Medline: 21780942]
95. Kroenke K, Theobald D, Wu J, Norton K, Morrison G, Carpenter J, et al. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. J Am Med Assoc 2010 Jul 14;304(2):163-171 [FREE Full text] [doi: 10.1001/jama.2010.944] [Medline: 20628129]
96. Ratanawongsa N, Handley MA, Quan J, Sarkar U, Pfeifer K, Soria C, et al. Quasi-experimental trial of diabetes Self-Management Automated and Real-Time Telephonic Support (SMARTSteps) in a medicaid managed care plan: study protocol. BMC Health Serv Res 2012;12:22 [FREE Full text] [doi: 10.1186/1472-6963-12-22] [Medline: 22280514]
97. Handley MA, Shumway M, Schillinger D. Cost-effectiveness of automated telephone self-management support with nurse care management among patients with diabetes. Ann Fam Med 2008;6(6):512-518 [FREE Full text] [doi: 10.1370/afm.889] [Medline: 19001303]

Abbreviations

HBA $_{1c}$: glycated hemoglobin
AOR: adjusted odds ratio
ATA: automatic telephone assessment
DCAT: Diabetes-Depression Care-Management Adoption Trial
DMP: disease management program
DMR: disease management registry
GPS: generalized propensity score
ICT: information and communication technology
LACDHS: Los Angeles County Department of Health Services
LSE: least squares estimate
OR: odds ratio
PCP: primary care physician
PHQ-9: 9-item Patient Health Questionnaire
PST: problem-solving therapy
SC: supported care
SDS: Sheehan Disability Scale
SF-12: 12-item Short Form Survey
TC: technology-facilitated care
UC: usual care

Edited by G Eysenbach; submitted 25.04.17; peer-reviewed by O Demasi, N Choi, A Centi, N Ratanawongsa, T Bose; comments to author 15.07.17; revised version received 10.12.17; accepted 13.01.18; published 23.04.18.

Please cite as:
Wu S, Ell K, Jin H, Vidyanti I, Chou CP, Lee PJ, Gross-Schulman S, Sklaroff LM, Belson D, Nezu AM, Hay J, Wang CJ, Scheib G, Di Capua P, Hawkins C, Liu P, Ramirez M, Wu BW, Richman M, Myers C, Agustines D, Dasher R, Kopelowicz A, Allevato J, Roybal M, Ipp E, Haider U, Graham S, Mahabadi V, Guterman J
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;20(4):e147
URL: http://www.jmir.org/2018/4/e147/
doi:10.2196/jmir.7692
PMID:29685872

Original Paper

Effectiveness of an Internet- and App-Based Intervention for College Students With Elevated Stress: Randomized Controlled Trial

Mathias Harrer[1], BSc; Sophia Helen Adam[1], MSc; Rebecca Jessica Fleischmann[1], BSc; Harald Baumeister[2], PhD; Randy Auerbach[3], PhD, ABPP; Ronny Bruffaerts[4], PhD; Pim Cuijpers[5], PhD; Ronald C Kessler[6], PhD; Matthias Berking[1], PhD; Dirk Lehr[7], PhD; David Daniel Ebert[1], PhD

[1]Clinical Psychology and Psychotherapy, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

[2]Clinical Psychology and Psychotherapy, University of Ulm, Ulm, Germany

[3]Department of Psychiatry, Columbia University, New York, NY, United States

[4]Universitair Psychiatrisch Centrum, Katholieke Universiteit Leuven, Leuven, Belgium

[5]Department of Clinical, Neuro, and Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[6]Department of Health Care Policy, Harvard Medical School, Boston, MA, United States

[7]Division of Online Health Trainings, Innovation Incubator, Leuphana University, Lüneburg, Germany

Corresponding Author:
Mathias Harrer, BSc
Clinical Psychology and Psychotherapy
Friedrich-Alexander-University Erlangen-Nuremberg
Nägelsbachstraße 25a
Erlangen, 91052
Germany
Phone: 49 1708237654
Email: mathias.harrer@fau.de

Abstract

Background: Mental health problems are highly prevalent among college students. Most students with poor mental health, however, do not receive professional help. Internet-based self-help formats may increase the utilization of treatment.

Objective: The aim of this randomized controlled trial was to evaluate the efficacy of an internet-based, app-supported stress management intervention for college students.

Methods: College students (n=150) with elevated levels of stress (Perceived Stress Scale 4-item version, PSS-4 ≥8) were randomly assigned to either an internet- and mobile-based stress intervention group with feedback on demand or a waitlist control group. Self-report data were assessed at baseline, posttreatment (7 weeks), and 3-month follow-up. The primary outcome was perceived stress posttreatment (PSS-4). Secondary outcomes included mental health outcomes, modifiable risk and protective factors, and college-related outcomes. Subgroup analyses were conducted in students with clinically relevant symptoms of depression (Center for Epidemiological Studies' Depression Scale >17).

Results: A total of 106 participants (76.8%) indicated that they were first-time help-seekers, and 77.3% (intervention group: 58/75; waitlist control group: 58/75) showed clinically relevant depressive symptoms at baseline. Findings indicated significant effects of the intervention compared with the waitlist control group for stress (d=0.69; 95% CI 0.36-1.02), anxiety (d=0.76; 95% CI 0.43-1.09), depression (d=0.63; 95% CI 0.30-0.96), college-related productivity (d=0.33; 95% CI 0.01-0.65), academic work impairment (d=0.34; 95% CI 0.01-0.66), and other outcomes after 7 weeks (posttreatment). Response rates for stress symptoms were significantly higher for the intervention group (69%, 52/75) compared with the waitlist control group (35%, 26/75, P<.001; number needed to treat=2.89, 95% CI 2.01-5.08) at posttest (7 weeks). Effects were sustained at 3-month follow-up, and similar findings emerged in students with symptoms of depression.

Conclusions: Internet- and mobile-based interventions could be an effective and cost-effective approach to reduce consequences of college-related stress and might potentially attract students with clinically relevant depression who would not otherwise seek help.

Trial Registration: German Clinical Trial Register DRKS00010212; http://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00010212 (Archived by WebCite at http://www.webcitation.org/6w55Ewhjd)

(J Med Internet Res 2018;20(4):e136) doi:10.2196/jmir.9293

KEYWORDS

randomized controlled trial; stress, psychological; depression; telemedicine; students; help-seeking behavior

Introduction

Background

Between 25% and 50% of college students meet the criteria for at least one mental health disorder in a given year [1,2]. Data suggest that mental disorders account for about half the disease burden of young adults in developed countries [3] and are associated with a range of negative consequences, including lowered academic performance [4] and college attrition [5].

Despite the availability of effective treatment [6], only 1 in 5 students with mental disorders receives minimally adequate treatment [1]. Reasons for this treatment gap include attitudinal barriers such as stigma and a preference for self-help [7].

Internet- and mobile-based interventions [8] might help to increase the utilization of psychological interventions, as they can be easily accessed, allow for high scalability, and can be provided at a low cost [9,10]. Internet-based interventions may also be suitable for college student populations [11], with research indicating that preference for help-seeking through the internet is higher among younger and well-educated individuals [12]. There is meta-analytic evidence suggesting the efficacy of internet interventions for a range of conditions and populations [6,13-16], including college students [17], with effect sizes of technology-delivered interventions ranging from standardized mean difference (SMD) of 0.42 to 0.43 for depression, 0.30 to 0.56 for anxiety, and 0.73 to 0.82 for stress [17,18]. However, the few results for internet and mobile-based interventions targeting stress in students are conflicting in terms of their effectiveness [19,20], warranting further research.

A recent meta-analysis suggests that intervention effects are considerably higher in indicated compared with general student populations [18], stressing the importance of developing suitable intervention approaches for at-risk students. Internet-based interventions which are labeled to improve stress coping skills, as opposed to focusing on reducing symptoms of mental disorders, could represent a promising way to reach such burdened individuals. In an Australian investigation among severely distressed college students, 55.7% indicated that they were quite or very likely to use an internet-delivered program to seek help [21]. A significant association between heightened stress levels and positive attitudes toward internet intervention usage has also been found in a German general population sample [22].

If proven to be effective, internet- and mobile-based approaches could provide a feasible instrument to help avert the onset of more severe stress-related mental health concerns in at-risk college students [8]. More research is therefore required to corroborate results on the effectiveness of internet- and mobile-based stress interventions and assess the potential of such interventions to reach and be effective in burdened students who already show symptoms of mental illness such as depression. Facing the deleterious effects of poor mental health on academic functioning, it is also important to assess whether such interventions may have an impact on important college-related outcomes such as academic self-efficacy and impairment [17].

Objectives

The aim of this study is thus to evaluate the effectiveness of an internet- and mobile-based intervention targeting university students with heightened stress levels. We hypothesized the internet intervention to be more effective in reducing symptoms of stress compared with a waitlist control group (WCG). It was furthermore assumed that more students participating in the intervention compared with the WCG would show a reliable change in perceived stress outcomes and attain close to symptom-free status. The second objective of this study was to investigate the hypothesized positive effect of the intervention on further mental health outcomes, modifiable risk and protective factors, and college-related outcomes compared with the WCG. Finally, our aim was to explore intervention participants' adherence to, and acceptance of, the intervention.

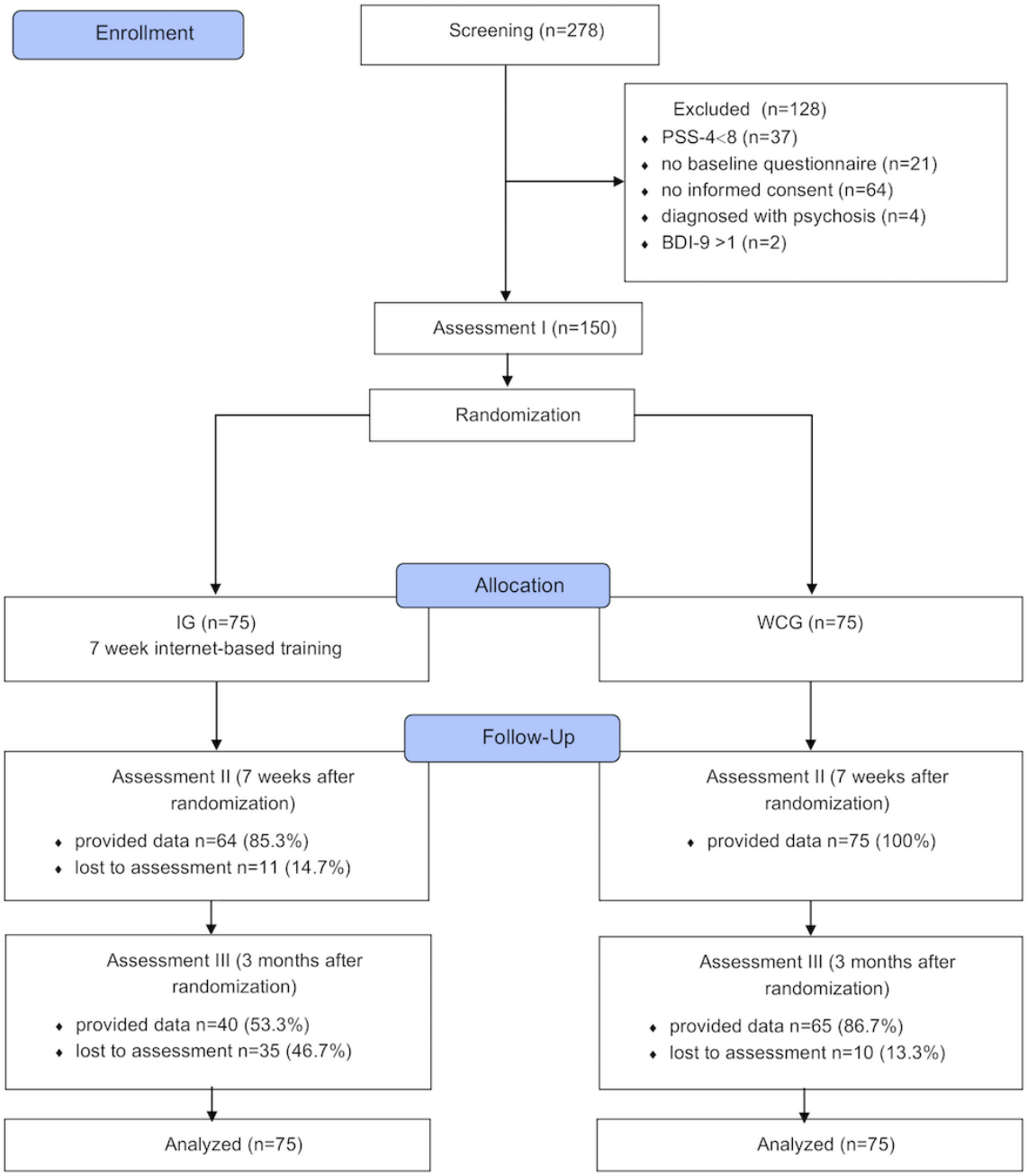
Methods

This study was carried out as part of the WHO World Mental Health International College Student project [23]. The WHO World Mental Health International College Student project aims to obtain accurate cross-national information on the prevalence, incidence, and correlates of mental, substance, and behavioral problems among college students worldwide, to describe patterns of service use and unmet need for treatment, to investigate the associations of these disorders with academic functioning, and to evaluate the effects of a wide range of preventive and clinical interventions on student mental health, functioning, and academic performance.

Design

A 2-armed randomized controlled trial was conducted with 150 participants, comparing an internet and app-based intervention with feedback on demand (*StudiCare Stress*) to a waitlist control group (WCG). Both conditions had full access to treatment as usual (TAU). The sample size allowed to detect effect sizes of d=0.41 with a power ($1-\beta$) of 0.80 with alpha of .05 and was based on a meta-analysis on internet-based interventions for college students, which reported an SMD of 0.73 for stress but lower effects for depression outcomes (SMD=0.43) [17]. A sample size of 150 was therefore chosen to also detect significant changes for secondary outcomes in this study such as depression.

Figure 1. Flow of participants (CONSORT flow chart). BDI: Beck Depression Inventory; IG: intervention group; WCG: waitlist control group; CONSORT: Consolidated Standards of Reporting Trials.



Assessments took place at baseline (T1), posttreatment (T2; 7 weeks), and 3 months after baseline (T3; see Figure 1). Self-report data were collected using a Web-based assessment tool (Advanced Encryption Standard, 256-bit encryption). All procedures involved in the study were consistent with the generally accepted standards of ethical practice. The study was approved by the University of Erlangen-Nuremberg ethics committee (Erlangen, Germany; 322_15 B). The trial is registered in the German Clinical Trials Register (DRKS00010212 [24]). This study describes the main effectiveness analysis of the intervention; we also assessed moderator and mediator variables, which are listed in the trial registration (see Multimedia Appendix 1) and will be analyzed and reported in due length elsewhere.

Participants

Inclusion criteria were (1) elevated levels of perceived stress (Perceived Stress Scale 4-item version, PSS-4≥8 [25]; representing a level of stress one *SD*=2.92 above the mean of 4.49 in a large student sample [25]), (2) enrollment in a German-speaking university at the beginning of the training, (3) age ≥18 years, (4) internet access, (5) willingness to provide self-report data at all assessment points, and (6) informed consent. Exclusion criteria were (1) self-reported diagnosis of dissociative symptoms or psychosis in the past or (2)

considerable risk for suicide (Beck Depression Inventory item 9 >1; "I feel I would be better off dead" or "I would kill myself if I had the chance"). Individuals showing an elevated risk for suicide were given detailed information about treatment options and were asked to see a physician or psychiatrist as soon as possible.

Recruitment

Participants were recruited via university press reports, student counseling services, and social media platforms. Potential participants declared interest in partaking in the study by filling out a Web-based registration form on the study website.

Assessment of Eligibility and Randomization

Individuals who declared interest in participating received an information letter along with an informed consent sheet and were asked to provide an email address for their intervention platform profile. Applicants were informed that withdrawal from the study was possible at any time, did not go along with any negative consequences, and all collected case data could be deleted on request during the study. Interested participants were asked to complete the written informed consent form and fill out the Web-based screening questionnaire.

Individuals meeting all the inclusion and none of the exclusion criteria were invited to fill out the baseline assessment. After completion, individuals were randomly allocated to either the IG or the WCG. Randomization took place at a ratio of 1:1 and a block size of 2 using an automated computer-based random integer generator (*Randlist*, Datinf GmbH, Tübingen, Germany) and was performed by a researcher not otherwise involved in the study. Participants could not be blinded to study conditions; yet, during the randomization process, the allocation was concealed from participants, researchers involved in recruitment, and e-coaches.

Study Conditions

Intervention Condition

The framework for *StudiCare Stress* was derived from *GET.ON Stress*, a Web-based stress management intervention for employees [26]. Changes in form and therapeutic content were made to tailor the intervention to university students' needs.

The intervention is based on cognitive-behavioral and third-wave techniques and aligns with Lazarus' transactional model of stress [27] in differentiating between problem-focused and emotion regulation–focused coping. For problem-focused coping, cognitive-behavioral problem-solving strategies are applied to reduce and eliminate modifiable stressors. Emotion regulation refers to the processes through which individuals monitor, evaluate, modify, and thus control emotions to reach relevant needs or goals and has been shown to be influential in reducing various symptoms of mental illness [28]. Elective modules integrated at the end of session 2 to 7 could be chosen based on individual need and interest, covering student-specific topics: social support, rumination and worrying, time management, procrastination, test anxiety, sleep, motivation, nutrition and exercise, and dealing with writer's block and concentration.

The intervention comprised 8 main modules. Completing 1 module took 30 to 90 min, and participants were advised to complete at least one and a maximum of 2 modules per week. Thus, the intervention was intended to be completed in about 5 to 7 weeks (see Multimedia Appendix 2 for a detailed description of the modules).

Strong emphasis was put on the transfer of acquired knowledge, strategies, and techniques into the students' daily life through homework assignments. A personal diary app could be downloaded by participants to keep track of mood fluctuations, monitor factors contributing to their stress levels and reflect on intervention elements they could implement into their daily life. The diary app was introduced in module 1 as an adjunct to the main sessions and contained standardized free-text fields, rating scales, and gave the opportunity to add a photo to the entry (see Textbox 1). Participants were also provided with a PDF version of the diary and were instructed to monitor their mood 2 to 3 times each week, using either the app or a printout of the PDF for their entries.

In addition, before beginning with the intervention, participants could request automatic daily messages containing short motivational prompts and ultrabrief training exercises via SMS (short message service), aimed at facilitating transfer of learned strategies into daily life routine. Messages were prescheduled to roughly mirror content and exercises provided through the progression of the intervention.

Participants were guided by an eCoach, a trained student in a master's program in psychology. Contact between the eCoach and intervention participants was solely established online, and there were no face-to-face meetings. An adherence-focused guidance concept in accordance with the human accountability model [29,30] was applied, which has been shown to be noninferior to intensive guidance while minimizing human resources ([31]; for a detailed description see [26]). Guidance consisted of 3 parts: (1) monitoring adherence (sending up to 3 reminders when a module was not completed during 1 week through the internal platform messaging system and via email), (2) checking the intervention platform back-end for participants who had completed a new module to unlock the next module and send standardized motivational messages through the platform, and (3) providing feedback on demand. When requesting help, participants received feedback within 48 hours.

The feedback reflected the participants' individual questions and problems and gave positive reinforcement. Feedback on demand was available for each participant from module 1 until completion of the booster session and was given via the internal messaging system of the training platform. Only few participants (5%, 4/75) requested individual feedback, resulting in 5 content feedbacks for the entire sample. In total, the eCoach sent 289 reminders (3.85 reminders per participant).

Control Condition

Students assigned to the waitlist control condition (WCG) completed the same assessments at T1, T2, and T3 as the intervention condition, but were not given access to the intervention until 3 months after randomization. Yet, they had full access to TAU offered by routine health care.

Textbox 1. General structure of the app-based diary entries.

1. How do you feel today? (Emoticons: Happy–Sad–Anxious–Angry)
2. How stressed out do you feel today? (Rating scale 1-10)
3. Describe what happened today. (Free text)
4. Were you able to identify any things contributing to your stress levels today? (Free text)
5. Are there any techniques you previously learned that you may be able to apply? (Free text)
6. Do you want to add a photo to your entry? (Upload button)

Primary Outcome Measure

The primary outcome was perceived stress as measured by the PSS-4 [25]. The PSS-4 assesses the degree to which individuals evaluate their lives as stressful, especially regarding how uncontrollable and overloading relevant aspects of life are perceived. The PSS-4 comprises 4 items (Item 1: "How often have you felt you were unable to control the important things in your life?"; Item 2: "How often have you felt confident about your ability to handle your personal problems?"; Item 3: "How often have you felt that things were going your way?"; Item 4: "How often have you felt difficulties were piling up so high that you could not overcome them?"), yielding a score between 0 and 16. Participants rated their level of perceived stress within the last 2 weeks on a 5-point Likert scale (0= *never*; 4= *very often*). A two-factor structure has been commonly found for the PSS [32-35], with positively framed items representing perceived coping self-efficacy and negative items reflecting hopelessness, the latter being a strong predictor for depression [36]. Higher scores on the PSS have shown to have good predictive validity for several adverse health outcomes [36-38]. Despite its brevity, the PSS-4 has been found to have acceptable to good psychometric properties [39,40]. The scale has a good level of internal consistency in this study as indicated by a Cronbach alpha of .83.

Secondary Outcome Measures

Unless otherwise specified, all outcomes were measured for a retrospective time frame of 2 weeks. All measures were administered in German. When no German translation was available, scales were translated independently by two of the researchers (MH and SHA), who then compared and discussed the translations to resolve disagreement.

Mental Health

To examine effects of the stress intervention on symptoms of common mental disorders, we included mental health outcomes associated with elevated distress in college students, including depression (short German form of the Center for Epidemiological Studies' Depression Scale, CES-D [41]; 15 items, scale 0-3, range 0-45) and state anxiety (Spielberger State-Trait Anxiety Inventory [42]; 6 items, scale 1-4, range 6-24; *at the moment*) [43]. General well-being as an overall marker of mental health was assessed by the WHO-Five Well-Being Index (WHO-5 [44]; 5 items, scale 0-6, range 0-30), and emotional exhaustion using the Maslach Burnout Inventory-student version ([45]; 5 items, scale 1-6, range 5-30).

Risk and Protective Factors

Following measures for established risk and protective factors were assessed to investigate the intervention's effect on individual resources and vulnerabilities related to the development and proliferation of mental illness: dysfunctional perfectionism [46] (Revised Almost Perfect Scale [47]; translated; 8 items, scale 1-7, range 8-56), resilience [48] (Connor-Davidson Resilience Scale short form [49]; translated; 2 items, scale 0-4, range 0-8), self-compassion [50,51] (Self-Compassion Scale [52]; 12 items, scale 1-5, range 12-60), and self-esteem [53] (Rosenberg Self-Esteem Scale [54]; 10 items, scale 1-4, range 10-40).

College-Related Outcomes

To evaluate presenteeism and loss of productivity, the Presenteeism Scale for Students'([55]; translated) subscale for work impairment (Work Impairment Scale; 10 items, scale 1-5, range 10-50) was administered. Productivity losses were assessed by an adaption of the Presenteeism Scale for Students' work output scale, investigating the current percentage to which participants were able to reach their usual academic productivity. Productivity could be rated on a visual analog scale ranging from 0%= *completely unproductive* to 100%= *full productivity*. Academic self-efficacy was measured by the academic self-efficacy scale (Wirkstud [56]; 7 items, scale 1-4, range 7-28), and academic worrying using the Academic Worrying Questionnaire ([57]; translated; 10 items, scale 0-4, range 0-40).

Additional Measures

Additional questionnaires assessed demographic variables, prior contact with professional health providers, and satisfaction with the intervention (IG only; Client Satisfaction Questionnaire, adapted to the web context, CSQ-8 [58]; 8 items, scale 1-4). Treatment credibility and expectancies were measured at baseline by the Credibility and Expectancy Questionnaire ([59]; translated; 4 items, scale 1-5, range 4-20, 2 items, 0%-100%). Participants in the IG could give feedback on each modules' usefulness (1= *not useful at all*, 5= *very useful*), complexity (1= *very complex*, 5= *very easy*), and duration until termination (1= *less than ½ hour*, 5= *more than 1½ hours*) on a 5-point Likert or 4-point scale, respectively.

Statistical Analyses

Main Effectiveness Evaluation

All results are reported according to the Consolidated Standards of Reporting Trials statement ([60]; see Multimedia Appendix 3). Analyses based on the intention-to-treat (ITT) principle were conducted, with missing data imputed using a Markov chain

Monte Carlo multivariate imputation algorithm (multiple imputation functions in IBM SPSS 23; IBM Corp, Armonk, NY, USA) with 100 estimations per missing and all variables set as predictors for imputation. Imputed datasets were then aggregated to obtain 1 imputed dataset.

The hypothesized superiority of the internet intervention was tested with regard to (1) change in participants' perceived stress and secondary outcomes from baseline (T1) to post intervention (T2) and 3-month follow-up (T3), (2) the number of participants with treatment response, (3) the number of students achieving close to symptom-free status, and (4) the amount of participants who experienced symptom deterioration.

Differences in change of perceived stress between study arms were assessed using univariate analysis of covariance (ANCOVA) with scores at baseline as covariate to control for varying degrees of baseline scores. Effect sizes (Cohen's *d*) were calculated based on the imputed dataset for between-group differences, using the pooled IG and WCG SD [61]. To calculate 95% CIs, the formula by Rosnow and Rosenthal [62] was used. According to Cohen [63], *d*=0.2 can be considered a small effect, *d*=0.5 a medium and *d*=0.8 a large effect. A significance level of .05 (2-sided) was used for all analyses.

To ascertain the number of participants attaining a reliable improvement in stress symptomatology, participants were coded as responders or nonresponders according to the Reliable Change Index [64]. Accordingly, response was attained when participants' scores on the PSS-4 differed more than −2.17 points from baseline to T2 and T3, respectively. Furthermore, the numbers needed to treat (NNT) to achieve 1 additional treatment response were calculated. Negative effects of the intervention were evaluated by the number of participants with reliable symptom deterioration concerning perceived stress through the Reliable Change Index. Participants were defined as symptom-free when scoring more than 2 SDs below the mean at baseline for the full sample (PSS-4 ≤7.29).

Subgroup Analysis

To estimate the interventions' efficacy in a clinical population, a subgroup analysis was conducted including only participants with a score of >17 on the CES-D short form at baseline and following the same procedure as the main analysis. A score of 18 has been shown to be a valid cut-off in indicating a high probability of clinical depression [65]. Participants were classified as responders if they showed reliable change in depressive symptoms according to the Reliable Change Index.

Study Completer Analysis

Completer analysis based on the sample of participants who provided data at all 3 assessment points was conducted additionally as a sensitivity analysis.

Process Evaluation

Descriptive statistics were used for process evaluation. To assess overall user satisfaction across various domains, item data provided by the CSQ-8 was examined individually. Acceptance of intervention modules was analyzed using the module feedback of the IG. Adherence was assessed by analyzing intervention completion rates tracked within the intervention platform. Finally, we analyzed the proportion of participants who accessed the diary app and requested automated short messages via SMS.

Results

Recruitment for the study started on May 9, 2016. The last follow-ups were completed on January 30, 2017.

Participants

The study flow can be found in Figure 1. Participants who were lost to follow-up at T2, T3, or both assessments did not differ significantly from participants who adhered to the protocol on any baseline characteristic (all *P*>.05). Table 1 summarizes detailed baseline characteristics of study participants. The majority (76.8%, 106/138) of the participants indicated that they had not consulted a physician, psychotherapist, or counselor for their health-related problems and may thus be considered first-time help-seekers. Descriptive data including all 3 assessment points for all outcomes is depicted in Table 2. Both study arms did not differ significantly (all *P*>.05) on any characteristic at baseline.

Main Effectiveness Analysis

Changes in Perceived Stress

As hypothesized, the ANCOVA controlling for baseline scores revealed a significant effect for perceived stress at posttest ($F_{1,147}$=19.70, *P*<.001) and at 3-month follow-up ($F_{1,147}$=15.10, *P*<.001; see Table 3), with moderate to large effect sizes at both T2 (*d*=0.69; 95% CI 0.36-1.02) and T3 (*d*=0.57; 95% CI 0.24-0.89).

Treatment Response for Perceived Stress

Chi-squared tests revealed that significantly more participants in the IG (69%, 52/75) were classified as responders compared with the WCG (35%, 26/75) at posttest (χ^2_1=18.1, *P*<.001), resulting in an NNT of 2.89 (95% CI 2.01-5.08). At 3-month follow-up, 55 of 75 participants in the IG (73%) and 33 of 75 in the WCG (44%) were coded as responders (χ^2_1=13.3, *P*<.001) which equals an NNT of 3.41 (95% CI 2.25-7.00).

Symptom-Free Status for Perceived Stress

Symptom-free status was achieved by significantly (χ^2_1=6.7, *P*=.01) more participants in the IG (44%, 33/75) compared with the WCG (24%, 18/75) at T2, and at T3 (IG: 53%, 40/75; WCG: 35%, 26/75) with χ^2_1=5.3 (*P*=.02), resulting in an NNT of 5 (T2; 95% CI 2.87-19.31) and 5.36 (T3; 95% CI 2.92-32.66), respectively.

Symptom Deterioration for Perceived Stress

Only a small proportion of participants experienced symptom deterioration. Fewer participants' stress symptomatology deteriorated in the IG (0%, 0/75) compared with the WCG, where 7 of 75 (9%) participants' symptoms deteriorated (χ^2_1=7.3, *P*<.001; NNT=10.58, 95% CI 6.19-44.18) at T2. Symptom deterioration did not differ at T3, with 1 case of 75

participants (1%) in the IG and 3 of 75 (4%) in the WCG (χ^2_1=1.0, *P*=.31).

Table 1. Baseline characteristics.

Characteristics	All participants (N=150)	Intervention (N=75)	Control (N=75)
Sociodemographics			
Age in years, mean (SD)	24.1 (4.1)	24.0 (4.6)	24.2 (3.6)
Gender, female, n (%)	112 (74.7)	54 (72)	58 (77)
In a relationship, n (%)	79 (52.7)	39 (52)	40 (53)
Married, n (%)	6 (4.0)	4 (5)	2 (3)
Major			
Business & Economics, n (%)	33 (22.0)	16 (21)	17 (23)
Computer Science & Engineering, n (%)	13 (8.7)	9 (12)	4 (5)
Education, n (%)	17 (11.3)	7 (9)	10 (13)
Humanities, n (%)	12 (8.0)	5 (8)	7 (9)
Law, n (%)	6 (3.3)	2 (3)	4 (5)
Medicine, n (%)	15 (10.0)	7 (9)	8 (11)
Natural Sciences, n (%)	20 (13.3)	11 (15)	9 (12)
Social Sciences, n (%)	34 (22.7)	18 (24)	16 (21)
Number of semesters (previous studies included), mean (SD)	6.7 (3.6)	6.4 (3)	7.07 (3.9)
Type of tertiary education facility, n (%)			
College	119 (79.3)	56 (74)	63 (84)
University of Applied Sciences	31 (20.6)	19 (25)	12 (16)
Housing situation, n (%)			
Alone	31 (20.7)	18 (24)	13 (17)
Flat share	95 (63.3)	48 (64)	47 (63)
With parents	24 (16.0)	9 (12)	15 (20)
Main source of funding, n (%)			
Parents	64 (42.7)	33 (44)	31 (41)
Job	48 (32.02)	25 (33)	23 (31)
Loan	34 (22.7)	15 (20)	19 (25)
Partner	2 (1.3)	1 (1)	1 (1)
Scholarship	2 (1.3)	1 (1)	1 (1)

Secondary Outcome Analysis

Table 3 summarizes the results of the ITT analyses for the secondary outcomes. ANCOVAs revealed significant effects (*P*<.05) in favor of the IG for the majority of outcomes at both assessment points, with effect sizes ranging from *d*=0.33 (95% CI 0.01-0.65) for productivity (T2) to *d*=0.82 (95% CI 0.49-1.15) for emotional exhaustion (T2). No statistically significant effect was found for perfectionism ($F_{1,147}$=0.38, *P*=.53) at T2, but at T3 (*P*<.001). Resilience (T2: $F_{1,147}$=1.69, *P*=.17; T3: $F_{1,147}$=2.94, *P*=.08), self-compassion (T2: $F_{1,147}$=2.97, *P*=.09; T3: $F_{1,147}$=1.46, *P*=.23), and self-esteem (T2: $F_{1,147}$=0.15, *P*=.70; T3: $F_{1,147}$=1.36, *P*=.25) did not differ significantly between both study arms at both assessment points.

Subgroup Analysis

More than three-fourths of the participants (77.3%; IG: 58/75; WCG: 58/75) showed symptoms above the cut-off for clinically relevant symptoms of depression at baseline. Between-group effect sizes for depression in this subgroup were moderate to large, both for T2 (*d*=0.67, 95% CI 0.34-1.00) and T3 (*d*=0.73, 95% CI 0.40-1.06). Treatment response was achieved by 36 (62%; T2) and 33 of 58 participants (57%, T3) in the IG compared with 14 of 58 participants (24%; T2 and T3) in the WCG, resulting in an NNT to achieve one additional treatment response in the IG compared with the WCG of 2.64 (95% CI 1.83-4.70) for T2 (χ^2_1=17.0, *P*<.001) and 3.05 for T3 (95% CI 2.02-6.28, χ^2_1=12.9, *P*<.001).

Table 2. Means and SDs of the intervention group (intervention) and waitlist control group (control) for the intention-to-treat-sample at baseline, posttest (7 weeks), and 3-month follow-up.

Outcome and assessment point	Intervention (N=75)	Control (N=75)
	Mean (SD)	Mean (SD)
Primary outcome		
Perceived stress (low to high 0-16)		
Baseline	11.13 (1.93)	11.03 (1.87)
7 weeks	7.43 (2.93)	9.49 (3.06)
3 months	6.96 (2.73)	8.66 (3.26)
Mental health		
Depression (0-45)		
Baseline	24.31 (9.06)	23.97 (8.63)
7 weeks	15.88 (8.85)	21.47 (8.96)
3 months	16.79 (8.72)	21.92 (9.53)
Anxiety (6-24)		
Baseline	16.05 (3.37)	15.77 (4.22)
7 weeks	13.37 (3.51)	16.03 (3.48)
3 months	13.33 (3.59)	15.50 (4.10)
Well-being (0-30)[a]		
Baseline	8.01 (4.34)	8.81 (3.69)
7 weeks	11.93 (5.03)	9.36 (4.35)
3 months	12.62 (5.34)	10.57 (4.81)
Emotional exhaustion (5-30)		
Baseline	21.63 (4.49)	22.27 (4.31)
7 weeks	18.43 (5.64)	22.36 (3.77)
3 months	20.04 (5.08)	22.30 (4.45)
Risk and protective factors		
Dysfunctional perfectionism (8-56)		
Baseline	44.29 (7.90)	43.89 (7.50)
7 weeks	43.02 (7.22)	43.45 (7.34)
3 months	41.05 (5.94)	44.33 (6.67)
Resilience (0-8)[a]		
Baseline	4.80 (1.72)	4.79 (1.87)
7 weeks	5.38 (1.85)	5.05 (1.97)
3 months	5.56 (1.36)	5.17 (1.61)
Self-compassion (12-60)[a]		
Baseline	33.95 (3.47)	34.54 (3.23)
7 weeks	34.95 (5.67)	34.16 (3.72)
3 months	35.25 (3.26)	34.78 (3.96)
Self-esteem (10-40)[a]		
Baseline	29.25 (2.58)	29.20 (2.78)
7 weeks	29.14 (3.61)	28.93 (2.61)
3 months	30.10 (3.03)	29.54 (2.68)

Outcome and assessment point	Intervention (N=75)	Control (N=75)
	Mean (SD)	Mean (SD)
College-related outcomes		
Academic work impairment (10-50)		
Baseline	28.29 (5.34)	27.88 (5.39)
7 weeks	25.54 (5.83)	27.55 (6.13)
3 months	24.74 (5.06)	27.44 (6.22)
Academic productivity (percent)[a]		
Baseline	52.79 (27.04)	54.30 (23.03)
7 weeks	60.36 (24.12)	52.36 (24.16)
3 months	67.76 (17.27)	58.21 (23.62)
Academic self-efficacy (7-28)[a]		
Baseline	17.04 (4.46)	16.34 (4.04)
7 weeks	18.35 (4.03)	16.43 (4.12)
3 months	18.60 (3.86)	16.37 (4.12)
Academic worrying (0-40)		
Baseline	22.35 (6.63)	22.01 (6.01)
7 weeks	18.29 (6.16)	21.71 (5.94)
3 months	17.82 (6.97)	21.14 (6.18)
Treatment expectancies (0-100)		
Baseline	62.34 (13.51)	62.44 (16.03)

[a]Higher scores indicate better outcomes.

Completer Analysis

The results of the completer analyses were similar to the ITT analyses, with moderate to large between-group effect sizes for the primary outcome at T2 (IG: mean=6.72, SD 2.86; WCG: mean=9.32, SD 3.16; $F_{1,88}$=18.60, P<.001; d=0.85, 95% CI 0.44-1.27) and T3 (IG: mean=6.41, SD 2.84; WCG: mean=8.65, SD 3.43; $F_{1,88}$=13.41, P<.001; d=0.69, 95% CI 0.29-1.10). In contrast to the main analysis, however, resilience had increased significantly in the IG compared with the WCG at T2 ($F_{1,88}$=8.56, P=.004; d=0.46, 95% CI 0.06-0.86).

Process Evaluation

Adherence to the Intervention

On average, participants in the IG completed 5.05 modules (SD 2.78), which equals 72.1% of the intervention. Participants completed optional add-on modules in the majority (82.1%) of sessions in which they were available. Most participants completed rumination & worrying (59%, 44/75), whereas only 8 of the 75 participants completed social support (11%). In all, 46 of the 75 participants in the IG (61%) downloaded and logged into the diary app at least once. Activation of the automated SMS messages was requested by 4 of 75 participants in the IG (5%) during the study.

Client Satisfaction

Overall client satisfaction with the intervention was high (see Table 4).

Perceived Usefulness, Difficulty, and Duration of Sessions

Most participants described the 8 treatment modules as useful and not overly complex (see Multimedia Appendix 4). Reported session duration was high, with participants having spent the most time on module 6 (Self-compassion; 28% spending more than 1 hour 30 min, 10/36).

Table 3. Results for the intention-to-treat sample for analyses of covariance for between-group effects, effect sizes (Cohen's *d*) for primary and secondary outcomes at posttest (7 weeks; T2) and 3-month follow-up (T3).

Outcome and assessment point	Effect size		ANCOVA[a]	
	Cohen's *d*	95% CI	F_{1147}	*P* value
Primary outcome				
Perceived stress				
7 weeks	0.69	0.36 to 1.02	19.70	<.001
3 months	0.57	0.24 to 0.89	15.10	<.001
Mental health				
Depression				
7 weeks	0.63	0.30 to 0.96	22.31	<.001
3 months	0.56	0.24 to 0.89	16.62	<.001
Anxiety				
7 weeks	0.76	0.43 to 1.09	28.20	<.001
3 months	0.56	0.24 to 0.89	14.68	<.001
Well-being				
7 weeks	0.55	0.22 to 0.87	21.06	<.001
3 months	0.40	0.08 to 0.73	12.14	.001
Emotional exhaustion				
7 weeks	0.82	0.49 to 1.15	30.67	<.001
3 months	0.59	0.26 to 0.92	8.93	.003
Risk and protective factors				
Dysfunctional perfectionism				
7 weeks	0.06	−0.26 to 0.38	0.38	.54
3 months	0.52	0.19 to 0.84	15.79	<.001
Resilience				
7 weeks	0.17	−0.15 to 0.49	1.69	.17
3 months	0.26	−0.06 to 0.58	2.94	.08
Self-compassion				
7 weeks	0.17	−0.16 to 0.49	2.97	.09
3 months	0.13	−0.19 to 0.45	1.46	.23
Self-esteem				
7 weeks	0.07	−0.25 to 0.39	0.15	.70
3 months	0.19	−0.13 to 0.51	1.36	.25
College-related outcomes				
Academic work impairment				
7 weeks	0.34	0.01 to 0.66	6.57	.01
3 months	0.48	0.15 to 0.80	10.57	.001
Academic productivity				
7 weeks	0.33	0.01 to 0.65	4.29	.04
3 months	0.46	0.14 to 0.79	9.68	.002
Academic self-efficacy				
7 weeks	0.49	0.16 to 0.81	12.74	<.001
3 months	0.56	0.23 to 0.88	17.98	<.001

Outcome and assessment point	Effect size		ANCOVA[a]	
	Cohen's *d*	95% CI	F_{1147}	*P* value
Academic worrying				
7 weeks	0.56	0.24 to 0.89	27.41	<.001
3 months	0.50	0.18 to 0.83	16.04	<.001

[a]ANCOVA: analysis of covariance.

Table 4. Clients' satisfaction with the intervention (T2; Intervention Group only).

Ratings	n (%)
Quality of the training rated as excellent or good	59 (92)
Indication that the training was the kind of intervention they wanted to receive (generally or definitely)	51 (80)
Indication that the own needs were almost all or mostly met	47 (73)
Inclination to recommend the training to a friend in need of similar help	58 (91)
Satisfaction with the amount of help received (mostly or very satisfied)	51 (80)
Indication that the training has helped (a great deal) to deal more effectively with problems	53 (83)
Satisfaction with the training in a general, overall sense (mostly or very satisfied)	55 (86)
Inclination to use the training again if in need for help	49 (77)

Discussion

Principal Findings

Results of this study indicate moderate to large intergroup effects for the reduction of perceived stress and other relevant health- and college-related outcomes, as well as substantial effects in individuals with clinically relevant symptoms of depression, which were highly prevalent in our recruited sample. No significant effects were found for self-compassion, perfectionism (T2), resilience, and self-esteem.

The benefits of this intervention were larger than those found in previous trials evaluating internet-based stress interventions in college students [19,20], albeit somewhat smaller than the reported overall effect of technology-delivered skill training interventions on perceived stress [18], and comparable to internet-based stress interventions in general, as reported in a recent meta-analysis, with a pooled standardized mean difference of d=0.64 (95% CI 0.50-0.79; perceived stress) in guided internet- and mobile-based interventions [66]. The study further contributes to current literature by showing that targeting perceived stress in students does not only result in better mental health–related outcomes and well-being but can also have a substantial beneficial impact on college-related outcomes which, to the best of our knowledge, have not been investigated so far.

Participants' adherence to the intervention was satisfying, and the intervention was well accepted among the large majority of students. Participant feedback on the length of specific modules, however, suggests that participants may have spent more time than anticipated on some of the modules. This may indicate that some modules could be shortened to attempt to further improve adherence. Whether shortening might, in fact, result in higher adherence to the intervention, however, is not fully clear. Earlier research has reported higher adherence rates for shorter interventions, albeit focusing on the number of modules used [67]. Whether this effect also holds true for the length of specific modules remains unknown. Shortening some of the modules might potentially optimize adherence but may also compromise the intervention's overall efficacy due to less potentially beneficial information or techniques being conveyed and trained. It has been argued that various ways in which participants prioritize provided content may lead to positive outcomes, and the ability to progress through interventions at one's own pace might represent a key asset of internet-delivered treatment [68]. Qualitative interviews conducted with participants of this intervention suggest that the elective mini-modules for various student-relevant topics were very well accepted [69]. Providing larger amounts of content in a flexible way, allowing participants to tailor the intervention to their specific needs, could, therefore, be a promising approach to optimize intervention usage patterns [70]. However, research is warranted to test whether this might further increase adherence.

Limitations

This study has some limitations. First, women were overrepresented in the study sample, as frequently seen in preventive interventions. Second, study dropout in the IG at 3-month follow-up was relatively high and larger in the IG. Albeit being a common limitation in clinical trial research [71], and with differential dropout rates having been reported before for internet-based stress intervention trials in tertiary education students [72,73], this restricts the generalizability of our findings on long-term effects. Attrition analysis, however, did not result in any significant baseline differences between dropout and nondropout cases, which may be an indicator that results were not overly biased due to unequal dropout [74]. Third, because of ethical reasons, participants had full access to treatment-as-usual. Thus, we cannot rule out potential cointervention effects due to utilization of health services.

Finally, because of feasibility and ethical reasons, such as not denying one half of the sample access to the intervention they sought after, participants in the IG of this study were compared with a WCG to assess effects of the training. The influence of treatment and change expectancies have been discussed as an artifact in clinical evaluation trials using WCGs because they potentially discourage participants with delayed access to treatment to initiate health-related behavior changes, and thus lead to accentuate effects [75].

Most college students with depression do not seek treatment through conventional health care channels [76], and attitudinal barriers, such as fear of stigmatization, have been shown to have a large impact on treatment utilization [77]. As our findings indicate that (1) a large number of students in this sample did not use conventional treatment options before, (2) the majority of students who were willing to use this intervention reported clinically significant symptoms of depression, and (3) among this group, the treatment response was favorable; future studies should explore whether internet-delivered stress interventions, labeled as providing "support for coping with academic stress," might potentially attract students with symptoms of depression who would not use formal mental health treatment and whether they can reduce the incidence of depressive disorders [8]. Future studies should thus investigate the utility of internet and mobile-based interventions in affected students, that is, with symptoms of major depression, as an indicated preventative or early intervention approach to narrow the treatment gap and improve academic functioning [4].

Conclusions

In conclusion, internet- and mobile-based interventions could be an acceptable, effective, and potentially cost-effective approach to reduce the negative consequences associated with college-related stress.

Acknowledgments

The study was partly funded by BARMER, a major health care insurance company in Germany.

Conflicts of Interest

DDE reports to have received consultancy fees or served in the scientific advisory board from several companies such as Minddistrict, Lantern, Schön Kliniken, and German health insurance companies (BARMER, Techniker Krankenkasse). DDE and MB are also stakeholders of the Institute for health trainings online (GET.ON), which aims to implement scientific findings related to digital health interventions into routine care. HB reports to have received consultancy fees and fees for lectures or workshops from chambers of psychotherapists and training institutes for psychotherapists. In the past 3 years, RCK received support for his epidemiological studies from Sanofi Aventis, was a consultant for Johnson & Johnson Wellness and Prevention, Sage Pharmaceuticals, Shire, Takeda, and served on an advisory board for the Johnson & Johnson Services Inc, and Lake Nona Life Project. RCK is a co-owner of DataStat, Inc, a market research firm that carries out health care research.

Multimedia Appendix 1

Trial registration.

[PDF File (Adobe PDF File), 75KB - jmir_v20i4e136_app1.pdf]

Multimedia Appendix 2

Intervention modules.

[PDF File (Adobe PDF File), 59KB - jmir_v20i4e136_app2.pdf]

Multimedia Appendix 3

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 930KB - jmir_v20i4e136_app3.pdf]

Multimedia Appendix 4

Participants' perceived usefulness, complexity and duration for each intervention module.

[PDF File (Adobe PDF File), 62KB - jmir_v20i4e136_app4.pdf]

References

1. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]

2. Blanco C, Okuda M, Wright C, Hasin DS, Grant BF, Liu S, et al. Mental health of college students and their non-college-attending peers: results from the National Epidemiologic Study on Alcohol and Related Conditions. Arch Gen Psychiatry 2008 Dec;65(12):1429-1437 [FREE Full text] [doi: 10.1001/archpsyc.65.12.1429] [Medline: 19047530]
3. World Health Organization. The Global Burden of Disease: 2004 update. Geneva: WHO; 2008.
4. Bruffaerts R, Mortier P, Kiekens G, Auerbach RP, Cuijpers P, Demyttenaere K, et al. Mental health problems in college freshmen: prevalence and academic functioning. J Affect Disord 2018 Jan 01;225:97-103. [doi: 10.1016/j.jad.2017.07.044] [Medline: 28802728]
5. Breslau J, Lane M, Sampson N, Kessler RC. Mental disorders and subsequent educational attainment in a US national sample. J Psychiatr Res 2008 Jul;42(9):708-716 [FREE Full text] [doi: 10.1016/j.jpsychires.2008.01.016] [Medline: 18331741]
6. Ebert DD, Zarski AC, 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]
7. Eisenberg D, Downs MF, Golberstein E, Zivin K. Stigma and help seeking for mental health among college students. Med Care Res Rev 2009 Oct;66(5):522-541. [doi: 10.1177/1077558709335173] [Medline: 19454625]
8. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using Internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
9. Ebert DD, Gollwitzer M, Riper H, Cuijpers P, Baumeister H, Berking M. For whom does it work? Moderators of outcome on the effect of a transdiagnostic internet-based maintenance treatment after inpatient psychotherapy: randomized controlled trial. J Med Internet Res 2013 Oct 10;15(10):e191 [FREE Full text] [doi: 10.2196/jmir.2511] [Medline: 24113764]
10. Buntrock C, Berking M, Smit F, Lehr D, Nobis S, Riper H, et al. Preventing depression in adults with subthreshold depression: health-economic evaluation alongside a pragmatic randomized controlled trial of a web-based intervention. J Med Internet Res 2017 Jan 04;19(1):e5 [FREE Full text] [doi: 10.2196/jmir.6587] [Medline: 28052841]
11. Chan JK, Farrer LM, Gulliver A, Bennett K, Griffiths KM. University students' views on the perceived benefits and drawbacks of seeking help for mental health problems on the Internet: a qualitative study. JMIR Hum Factors 2016;3(1):e3 [FREE Full text] [doi: 10.2196/humanfactors.4765] [Medline: 27026140]
12. Batterham PJ, Calear AL. Preferences for Internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]
13. Königbauer J, Letsch J, Doebler P, Ebert D, Baumeister H. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
14. Sander L, Rausch L, Baumeister H. Effectiveness of Internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
15. Välimäki M, Anttila K, Anttila M, Lahti M. Web-based interventions supporting adolescents and young people with depressive symptoms: systematic review and meta-analysis. JMIR Mhealth Uhealth 2017 Dec 08;5(12):e180 [FREE Full text] [doi: 10.2196/mhealth.8624] [Medline: 29222079]
16. Seo DC, Niu J. Evaluation of Internet-based interventions on waist circumference reduction: a meta-analysis. J Med Internet Res 2015 Jul 21;17(7):e181 [FREE Full text] [doi: 10.2196/jmir.3921] [Medline: 26199208]
17. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
18. Conley CS, Durlak JA, Shapiro JB, Kirsch AC, Zahniser E. A meta-analysis of the impact of universal and indicated preventive technology-delivered interventions for higher education students. Prev Sci 2016 Dec;17(6):659-678. [doi: 10.1007/s11121-016-0662-3] [Medline: 27225631]
19. Hintz S, Frazier PA, Meredith L. Evaluating an online stress management intervention for college students. J Couns Psychol 2015 Apr;62(2):137-147. [doi: 10.1037/cou0000014] [Medline: 24635586]
20. Chiauzzi E, Brevard J, Thum C, Decembrele S, Lord S. MyStudentBody-Stress: an online stress management intervention for college students. J Health Commun 2008 Sep;13(6):555-572. [doi: 10.1080/10810730802281668] [Medline: 18726812]
21. Ryan ML, Shochet IM, Stallman HM. Universal online interventions might engage psychologically distressed university students who are unlikely to seek formal help. Adv Ment Health 2010;9(1):73-83. [doi: 10.5172/jamh.9.1.73]
22. Apolinário-Hagen J, Vehreschild V, Alkoudmani RM. Current views and perspectives on e-mental health: an exploratory survey study for understanding public attitudes toward Internet-based psychotherapy in Germany. JMIR Ment Health 2017 Feb 23;4(1):e8 [FREE Full text] [doi: 10.2196/mental.6375] [Medline: 28232298]
23. Auerbach RP, Benjet C, Bruffaerts R, Cuijpers P, Ebert DD, Kessler RC. Harvard Medical School Department of Health Care Policy. 2017. The WHO World Mental Health International College Student Project URL: https://www.hcp.med.harvard.edu/wmh/college_student_survey.php [WebCite Cache ID 6w0NBawLL]

24. Deutsches Register Klinischer Studien [German Clinical Trials Register]. DRKS. 2016 Dec 05. Online based self-help stress management program for students with feedback on demand URL: http://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00010212 [accessed 2018-03-05] [WebCite Cache ID 6w55Ewhjd]
25. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
26. Ebert DD, Lehr D, Smit F, Zarski AC, Riper H, Heber E, et al. Efficacy and cost-effectiveness of minimal guided and unguided internet-based mobile supported stress-management in employees with occupational stress: a three-armed randomised controlled trial. BMC Public Health 2014;14:807 [FREE Full text] [doi: 10.1186/1471-2458-14-807] [Medline: 25099533]
27. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York: Springer; 1984.
28. Berking M, Ebert D, Cuijpers P, Hofmann SG. Emotion regulation skills training enhances the efficacy of inpatient cognitive behavioral therapy for major depressive disorder: a randomized controlled trial. Psychother Psychosom 2013;82(4):234-245. [doi: 10.1159/000348448] [Medline: 23712210]
29. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
30. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin Psychol Sci Pract 2016 Nov 17;24(1):27-45. [doi: 10.1111/cpsp.12173]
31. Zarski AC, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to Internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
32. Nielsen MG, Ørnbøl E, Vestergaard M, Bech P, Larsen FB, Lasgaard M, et al. The construct validity of the Perceived Stress Scale. J Psychosom Res 2016 May;84:22-30. [doi: 10.1016/j.jpsychores.2016.03.009] [Medline: 27095155]
33. Denovan A, Dagnall N, Dhingra K, Grogan S. Evaluating the Perceived Stress Scale among UK university students: implications for stress measurement and management. Stud High Educ 2017 Jun 22:1-14 [FREE Full text] [doi: 10.1080/03075079.2017.1340445]
34. Reis D, Lehr D, Heber E, Ebert DD. The German version of the Perceived Stress Scale (PSS-10): evaluation of dimensionality, validity, and measurement invariance with exploratory and confirmatory bifactor modeling. Assessment 2017 Jun 01:1073191117715731. [doi: 10.1177/1073191117715731] [Medline: 28627220]
35. Lee E. Review of the psychometric evidence of the Perceived Stress Scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
36. Hewitt PL, Flett GL, Mosher SW. The Perceived Stress Scale: factor structure and relation to depression symptoms in a psychiatric sample. J Psychopathol Behav Assess 1992 Sep;14(3):247-257 [FREE Full text] [doi: 10.1007/BF00962631]
37. Pbert L, Doerfler LA, DeCosimo D. An evaluation of the perceived stress scale in two clinical populations. J Psychopathol Behav Assess 1992 Dec;14(4):363-375 [FREE Full text] [doi: 10.1007/BF00960780]
38. Klein EM, Brähler E, Dreier M, Reinecke L, Müller KW, Schmutzer G, et al. The German version of the Perceived Stress Scale - psychometric characteristics in a representative German community sample. BMC Psychiatry 2016 Dec 23;16:159 [FREE Full text] [doi: 10.1186/s12888-016-0875-9] [Medline: 27216151]
39. Warttig SL, Forshaw MJ, South J, White AK. New, normative, English-sample data for the Short Form Perceived Stress Scale (PSS-4). J Health Psychol 2013 Dec;18(12):1617-1628. [doi: 10.1177/1359105313508346] [Medline: 24155195]
40. Mitchell AM, Crane PA, Kim Y. Perceived stress in survivors of suicide: psychometric properties of the Perceived Stress Scale. Res Nurs Health 2008 Dec;31(6):576-585. [doi: 10.1002/nur.20284] [Medline: 18449942]
41. Hautzinger M, Bailer M, Hofmeister D, Keller F. Center for Epidemiological Studies Depression Scale (CES-D; Radloff, L. S., 1977) - German adaption. Psychiatr Prax 2012;39(6):302-304 [FREE Full text] [doi: 10.1055/s-0032-1326702]
42. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep;31 (Pt 3):301-306. [Medline: 1393159]
43. Andrews B, Wilding JM. The relation of depression and anxiety to life-stress and achievement in students. Br J Psychol 2004 Nov;95(Pt 4):509-521. [doi: 10.1348/0007126042369802] [Medline: 15527535]
44. Brähler E, Mühlan H, Albani C, Schmidt S. Teststatistische Prüfung und Normierung der deutschen Versionen des EUROHIS-QOL Lebensqualität-Index und des WHO-5 Wohlbefindens-Index [Test statistic testing and standardization of the German versions of the EUROHIS-QOL quality of life index and the WHO-5 well-being index]. Diagn 2007 Apr;53(2):83-96 [FREE Full text] [doi: 10.1026/0012-1924.53.2.83]
45. Gumz A, Erices R, Brähler E, Zenger M. [Factorial structure and psychometric criteria of the German translation of the Maslach Burnout Inventory--Student Version by Schaufeli et al. (MBI-SS)]. Psychother Psychosom Med Psychol 2013 Feb;63(2):77-84. [doi: 10.1055/s-0032-1323695] [Medline: 23408302]
46. Smith MM, Saklofske DH, Yan G, Sherry SB. Does perfectionism predict depression, anxiety, stress, and life satisfaction after controlling for neuroticism? J Individ Differ 2017 Apr;38(2):63-70 [FREE Full text] [doi: 10.1027/1614-0001/a000223]
47. Slaney RB, Rice KG, Mobley M, Trippi J, Ashby JS. The revised almost perfect scale. Meas Eval Couns Dev 2001;34(3):130-145.

48. Galatzer-Levy IR, Burton CL, Bonanno GA. Coping flexibility, potentially traumatic life events, and resilience: a prospective study of college student adjustment. J Soc Clin Psychol 2012 Jun;31(6):542-567 [FREE Full text] [doi: 10.1521/jscp.2012.31.6.542]
49. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
50. Neff KD, Rude SS, Kirkpatrick KL. An examination of self-compassion in relation to positive psychological functioning and personality traits. J Res Pers 2007 Aug;41(4):908-916 [FREE Full text] [doi: 10.1016/j.jrp.2006.08.002]
51. Trompetter HR, de Kleine E, Bohlmeijer ET. Why does positive mental health buffer against psychopathology? An exploratory study on self-compassion as a resilience mechanism and adaptive emotion regulation strategy. Cognit Ther Res 2017;41(3):459-468 [FREE Full text] [doi: 10.1007/s10608-016-9774-0] [Medline: 28515539]
52. Hupfeld J, Ruffieux N. Validierung einer deutschen Version der Self-Compassion Scale (SCS-D) [Validation of a German version of the Self-Compassion Scale (SCS-D)]. Z Klin Psychol Psychother 2011 Apr;40(2):115-123 [FREE Full text] [doi: 10.1026/1616-3443/a000088]
53. Thoits PA. Self, identity, stress, and mental health. In: Aneshensel CS, Phelan JC, Bierman A, editors. Handbook of the Sociology of Mental Health. Dordrecht: Springer; 2013:357-377.
54. Ferring D, Filipp SH. Messung des selbstwertgefühls: befunde zu reliabilität, validität und stabilität der Rosenberg-Skala [Measurement of self-esteem: findings on reliability, validity, and stability of the Rosenberg Scale]. Diagn 1996;42(3):284-292.
55. Matsushita M, Adachi H, Arakida M, Namura I, Takahashi Y, Miyata M, et al. Presenteeism in college students: reliability and validity of the Presenteeism Scale for students. Qual Life Res 2011 Apr;20(3):439-446. [doi: 10.1007/s11136-010-9763-9] [Medline: 20945160]
56. Jerusalem M, Schwarzer R. Selbstwirksamkeit. Skalen zur Befindlichkeit und Persönlichkeit [Self-Efficacy. Scales for Mental State and Personality]. Berlin: Institute for Psychology, Free University Berlin; 1986.
57. Wolitzky-Taylor K. University of Texas at Austin. 2009. Randomized Clinical Trial Investigating the Efficacy of Self-Administered Interventions for Reducing Pathological Academic Worry URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/7513/wolitzky-taylork81714.pdf;sequence=2 [WebCite Cache ID 6w4xzl8lt]
58. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
59. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [Medline: 11132119]
60. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001 Apr 17;134(8):663-694 [FREE Full text] [doi: 10.7326/0003-4819-134-8-200104170-00012]
61. Leonhart R. [Estimating effect sizes in clinical trials]. Rehabilitation (Stuttg) 2004 Aug;43(4):241-246. [doi: 10.1055/s-2004-828293] [Medline: 15318292]
62. Rosnow RL, Rosenthal R. Assessing the effect size of outcome research. In: Nezu AM, Nezu CM, editors. Evidence-Based Outcome Research: A Practical Guide to Conducting Randomized Controlled Trials for Psychosocial Interventions. New York, NY, US: Oxford University Press; 2008.
63. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
64. 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. [Medline: 2002127]
65. Lehr D, Hillert A, Schmitz E, Sosnowsky N. Screening depressiver Störungen mittels Allgemeiner Depressions-Skala (ADS-K) und State-Trait Depressions Scales (STDS-T) [Screening of depressive disorders using the General Depression Scale (ADS-K) and State-Trait Depression Scales (STDS-T)]. Diagn 2008 Apr;54(2):61-70 [FREE Full text] [doi: 10.1026/0012-1924.54.2.61]
66. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
67. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
68. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31 [FREE Full text] [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
69. Fleischmann RJ, Harrer M, Zarski AC, Baumeister H, Lehr D, Ebert DD. Patients' experiences in a guided Internet- and app-based stress intervention for college students: a qualitative study. Internet Interv 2018:A (forthcoming) [FREE Full text] [doi: 10.1016/j.invent.2017.12.001]

70. Farrer L, Gulliver A, Chan JK, Bennett K, Griffiths KM. A virtual mental health clinic for university students: a qualitative study of end-user service needs and priorities. JMIR Ment Health 2015;2(1):e2 [FREE Full text] [doi: 10.2196/mental.3890] [Medline: 26543908]
71. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
72. Cavanagh K, Strauss C, Cicconi F, Griffiths N, Wyper A, Jones F. A randomised controlled trial of a brief online mindfulness-based intervention. Behav Res Ther 2013 Sep;51(9):573-578. [doi: 10.1016/j.brat.2013.06.003] [Medline: 23872699]
73. Frazier P, Meredith L, Greer C, Paulsen JA, Howard K, Dietz LR, et al. Randomized controlled trial evaluating the effectiveness of a web-based stress management program among community college students. Anxiety Stress Coping 2015;28(5):576-586. [doi: 10.1080/10615806.2014.987666] [Medline: 25420030]
74. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. Br Med J 2013 Jan 21;346:e8668 [FREE Full text] [Medline: 23338004]
75. Mohr DC, Ho J, Hart TL, Baron KG, Berendsen M, Beckner V, et al. Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. Transl Behav Med 2014 Dec;4(4):407-423 [FREE Full text] [doi: 10.1007/s13142-014-0262-3] [Medline: 25584090]
76. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]
77. 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]

Abbreviations

ANCOVA: analysis of covariance
CES-D: Center for Epidemiological Studies' Depression Scale
IG: intervention group
ITT: intention-to-treat
NNT: numbers needed to treat
PSS-4: Perceived Stress Scale, 4-item version
SMD: standardized mean difference
SMS: short message service
T1: baseline assessment
T2: postassessment (7 weeks)
T3: 3-month follow-up assessment
TAU: treatment as usual
WCG: waitlist control group

Edited by G Eysenbach; submitted 30.10.17; peer-reviewed by P Frazier, K Baron, T Furukawa; comments to author 14.12.17; revised version received 07.02.18; accepted 12.02.18; published 23.04.18.

Please cite as:
Harrer M, Adam SH, Fleischmann RJ, Baumeister H, Auerbach R, Bruffaerts R, Cuijpers P, Kessler RC, Berking M, Lehr D, Ebert DD
Effectiveness of an Internet- and App-Based Intervention for College Students With Elevated Stress: Randomized Controlled Trial
J Med Internet Res 2018;20(4):e136
URL: http://www.jmir.org/2018/4/e136/
doi:10.2196/jmir.9293
PMID:29685870

Original Paper

Effects of a 12-Week Digital Care Program for Chronic Knee Pain on Pain, Mobility, and Surgery Risk: Randomized Controlled Trial

Gabriel Mecklenburg[1*], MPhil; Peter Smittenaar[1*], PhD; Jennifer C Erhart-Hledik[2], PhD; Daniel A Perez[1], BSc; Simon Hunter[1], PhD

[1]Hinge Health, Inc, San Francisco, CA, United States

[2]Department of Orthopaedic Surgery, Stanford University Medical Center, Stanford University, Stanford, CA, United States

[*]these authors contributed equally

Corresponding Author:
Gabriel Mecklenburg, MPhil
Hinge Health, Inc
818 Mission St, Ste 200
San Francisco, CA,
United States
Phone: 1 415 349 7757
Email: gabriel@hingehealth.com

Abstract

Background: Chronic knee pain, most commonly caused by knee osteoarthritis, is a prevalent condition which in most cases can be effectively treated through conservative, non-surgical care involving exercise therapy, education, psychosocial support, and weight loss. However, most people living with chronic knee pain do not receive adequate care, leading to unnecessary use of opiates and surgical procedures.

Objective: Assess the efficacy of a remotely delivered digital care program for chronic knee pain.

Methods: We enrolled 162 participants into a randomized controlled trial between January and March 2017. Participants were recruited from participating employers using questionnaires for self-assessment of their knee pain, and randomized into treatment (n=101) and control (n=61) groups. Participants in the treatment group were enrolled in the Hinge Health digital care program for chronic knee pain. This is a remotely delivered, home-based 12-week intervention that includes sensor-guided exercise therapy, education, cognitive behavioral therapy, weight loss, and psychosocial support through a personal coach and team-based interactions. The control group received three education pieces regarding self-care for chronic knee pain. Both groups had access to treatment-as-usual. The primary outcome was the Knee Injury and Osteoarthritis Outcome Score (KOOS) Pain subscale and KOOS Physical Function Shortform (KOOS-PS). Secondary outcomes were visual analog scales (VAS) for pain and stiffness respectively, surgery intent, and self-reported understanding of the condition and treatment options. Outcome measures were analyzed by intention to treat (excluding 7 control participants who received the digital care program due to administrative error) and per protocol.

Results: In an intent-to-treat analysis the digital care program group had a significantly greater reduction in KOOS Pain compared to the control group at the end of the program (greater reduction of 7.7, 95% CI 3.0 to 12.3, *P*=.002), as well as a significantly greater improvement in physical function (7.2, 95% CI 3.0 to 11.5, *P*=.001). This was also reflected in the secondary outcomes VAS pain (12.3, 95% CI 5.4 to 19.1, *P*<.001) and VAS stiffness (13.4, 95% CI 5.6 to 21.1, *P*=.001). Participants' self-reported likelihood (from 0% to 100%) of having surgery also reduced more strongly in the digital care program group compared to the control group over the next 1 year (–9.4 percentage points, pp, 95% CI –16.6 to –2.2, *P*=.01), 2 years (–11.3 pp, 95% CI –20.1 to –2.5, *P*=.01), and 5 years (–14.6 pp, 95% CI –23.6 to –5.5, *P*=.002). Interest in surgery (from 0 to 10) also reduced more so in the digital care program compared to control group (–1.0, 95% CI –1.7 to –0.2, *P*=.01). Participants' understanding of the condition and treatment options (on a scale from 0 to 4) increased more substantially for participants in the digital care program than those in the control group (0.9, 95% CI 0.6 to 1.3, *P*<.001). In an analysis on participants that completed the intervention (per protocol analysis) all primary and secondary outcomes remained significant at greater effect magnitudes compared to intention to treat, with those completing the program showing a 61% (95% CI 48 to 74) reduction in VAS pain compared to 21% (95% CI 5 to 38) in the control group (*P*<.001). Accounting for the cost of administering the program, we estimate net cost savings on surgery alone of US $4340 over 1 year and $7900 over 5 years for those participants completing the digital care program compared

to those in the control group receiving treatment-as-usual. In an exploratory subgroup analysis including only participants exhibiting clinical symptoms of osteoarthritis the program proved equally effective.

Conclusions: This trial provides strong evidence that a comprehensive 12-week digital care program for chronic knee pain, including osteoarthritis, yields significantly improved outcomes for pain, physical function, stiffness, surgery risk, and understanding of the condition, compared to a control group.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 13307390; http://www.isrctn.com/ISRCTN13307390 (Archived by WebCite at http://www.webcitation.org/6ycwjGL73)

(J Med Internet Res 2018;20(4):e156) doi:10.2196/jmir.9667

KEYWORDS

osteoarthritis, knee; chronic pain; exercise; education; cognitive behavioral therapy; computers, handheld; coaching; non-invasive; digital health; digital therapy; digital care program

Introduction

Background

Chronic knee pain (CKP), often caused by knee osteoarthritis, affects 1 in 4 individuals over the age of 55 [1], and is a major health condition [2] that is becoming increasingly prevalent [3]. The effects of CKP are far-reaching and not limited only to the knee joint. Rather, chronic pain can result in negative effects on general health status including social functioning, energy and vitality, general health perception, limitations due to emotional and physical problems [4], negative effects on quality of life [5], productivity [6], emotional well-being [7], and health care costs [8].

Current recommendations for management of chronic pain suggest that treatments addressing multiple aspects of pain, including physical, psychological, and social, are most effective as compared to a single therapy [9,10]. Recommended components of effective non-pharmacological care for chronic musculoskeletal pain include physical activity, patient education, weight reduction, and self-management and coping strategies [9,11–13]. Thus, an effective treatment algorithm for CKP is a comprehensive program consisting of the main components of recommended conservative care.

Such comprehensive programs for chronic pain - including knee osteoarthritis (OA), one of the most common diagnosis for CKP [11] - have been shown to improve pain and function [15–23] and reduce utilization of total knee arthroplasty (TKA) [12]. However, despite research into comprehensive programs for CKP, utilization of such programs outside of the research arena is rare. For example, it is estimated that 80% of individuals with CKP due to knee OA are not adequately treated with conservative care [13]. This, in turn, leads many patients to undergo costly knee surgeries that could have been otherwise avoided [14]. Thus, there is a significant need to improve access and increase use of a comprehensive treatment program for the large population of individuals affected by CKP.

Digital technology has the potential to effectively provide comprehensive CKP programs. A digital care program (DCP) incorporating multiple components of recommended care could allow for more efficient, effective, and economical treatment by overcoming barriers to behavior change often observed in traditional in-person care, such as travel time, missed work, cost of care, and limited access to healthcare. Furthermore, a DCP incorporating remote sensing would allow for monitoring of patient adherence, a critical barrier limiting long-term effectiveness of treatment programs [15,16]. Only a few studies, however, have examined the use of digital technologies for CKP, investigating web-based platforms for physical activity and exercise [17,18], pain coping training [19], and more comprehensive programs incorporating education and exercise [20-22]. In particular, there are limited studies using a more rigorous randomized controlled design [18,19,22], and the use of digital health in musculoskeletal conditions is regarded as early stage [23].

We have developed a 12-week program for CKP called the Hinge Health DCP [24]. It consists of recommended components of non-pharmacological care for chronic musculoskeletal pain and includes sensor-guided exercise therapy promoting local muscle strengthening and stretching, education, cognitive behavioral therapy, psychosocial support through teams and personal health coaches, weight loss, and activity tracking. We have previously shown that the Hinge Health 12-week DCP improves clinical outcomes of pain, function, and stiffness over a period of 6 months after initiation of the program in a single-arm study of individuals with CKP [24]. The purpose of this study was to assess the short-term effectiveness (12 weeks after initiation) of the Hinge Health DCP in improving knee pain and disability in subjects with CKP, as compared to a control group receiving treatment as usual and knee care education only. We employed a randomized controlled trial with the hypothesis that the DCP would cause a greater improvement in outcome than the control treatment.

Methods

Study Design

This study was a two-armed, randomized, controlled, unblinded trial of participants with chronic knee pain. Online applications were invited from employees and their dependents at participating employers spread out over 12 office locations. Participants were recruited through emails and posters distributed through the participants' employers between January and March 2017. The trial was approved by the Western Institutional Review Board, and participants completed the intervention at home. The trial was performed in compliance with the Helsinki Declaration for research involving human subjects and in line with ICH-GCP guidelines. The trial was

preregistered at International Standard Randomized Controlled Trial Number (ISRCTN) 13307390. We followed CONSORT guidelines for reporting this trial.

Study Population

We assessed the eligibility of all applicants that completed the baseline questionnaire for CKP through their web browser. Participants provided informed consent as part of this questionnaire using a checkbox and digital information sheet. The inclusion criteria were: (1) age over 18, (2) knee pain for at least 1 month in the last 12 months, (3) participating in the collaborating employers' health plans, and (4) provision of informed consent. The exclusion criteria were (1) a prior diagnosis of rheumatoid arthritis, (2) surgery on the knee less than 3 months ago, and (3) an injury to the knee less than 3 months ago. We did not include knee OA as an inclusion criterion, though we did assess the presence of OA through 6 self-reported clinical criteria, whereby 3 or more positive criteria suggested OA: age over 50 years, stiffness for <30 minutes in the morning, crepitus, bony tenderness, bony enlargement, and no palpable warmth [11]. As there were a limited number of places available on the program, eligible applicants were prioritized for enrollment, with those exhibiting greater pain, disability, and surgery intent prioritized over those showing less. Applicants that were not prioritized were placed on the waitlist (n=73). Participants were not paid for their time, other than an incentive offered to complete the outcome questionnaire for those that did not complete it within 4 days of first invitation.

Randomization

Eligible applicants were randomized into the trial twice weekly during the signup period. Batches of selected participants were then randomized into treatment and control using a 60:40 (treatment: control) ratio (n=115) or using an 80:20 ratio (n=47). The 80:20 ratio represents a deviation from the study protocol due to an administrative error and was only used for a restricted time. The effective allocation ratio was therefore 62:38 (treatment: control). When a batch of applicants was randomized, an algorithm shuffled the batch and selected the first 60% (or 80%) to enter the treatment, and the remaining 40% (or 20%) to enter control. As such, the person(s) reviewing the applicants had no way of knowing whether any given applicant would end up in the treatment or control group (concealed allocation). After randomization, participants in the treatment group received an email inviting them to complete their profile and receive their kit to participate in the DCP, whereas those in the control group received an email with three education articles to help them care for their knee. Due to the nature of the study, neither the study staff nor the participants were blinded to group allocation.

Study Intervention

The treatment group received the Hinge Health 12-week DCP for CKP. The contents of this program have been described previously [24]. In short, participants received a tablet computer with the Hinge Health application installed, and two custom Bluetooth sensors with straps to be used on the upper and lower leg during the in-app exercise therapy. Participants were assigned a personal coach that provided support and accountability throughout the program and were placed in a team to provide peer support through a discussion feed within the app. Participation was completed entirely remotely through the app, at times and places chosen by the participant. Reminders were provided by text message and email if the participant was not engaging at the recommended intensity with the program. On a weekly basis, participants in the DCP were set the goal of completing 3 sessions of sensor-guided exercise therapy, reading one to two education articles, logging their symptoms at least twice, performing cognitive behavioral therapy (CBT; subset of weeks only), working at weight loss (if overweight), and tracking at least three 30-minute sessions of aerobic activities. Details of each of these components of the DCP are described elsewhere [24]. Each participant also maintained access to treatment as usual (TAU).

The control group received three pieces of education, presented digitally, that is also part of the Hinge Health DCP. These articles discussed the importance of self-care, how to deal with setbacks in knee pain, and how to manage communication and relationships when living with CKP. The control group maintained access to TAU and were informed that they would be reconsidered for the program when new places became available following the 12-week study.

The application was developed, owned, and sponsored by Hinge Health, Inc. The 12-week program received extensive testing over a 2-year period prior to starting the trial. All participants received the same version of the program, and there were no major application updates during the course of the trial.

Study Outcomes

Primary Outcomes

The preregistered primary outcomes were the Knee Injury and Osteoarthritis Outcome Score (KOOS) Pain subscale [25], and the KOOS Physical Function Shortform (KOOS-PS, referred to as "KOOS short version" in preregistration) scale [26]. Both scales span from 0 (no pain or impaired function, respectively) to 100 (extreme pain or impaired function, respectively), and were assessed at baseline and at the end of the 12-week DCP in the intervention and control groups. Both surveys are composite measures which may confound multiple conditions, however the digital nature of the program precluded individual clinical evaluation. Additionally, those in the treatment group were asked to complete both questionnaires at weeks 4 and 8 as part of the DCP. To conclude a positive effect of treatment we required a significant effect on *both* primary outcomes, though we note this was not specified in the preregistration.

Secondary Outcomes

We describe multiple preregistered secondary outcomes. Firstly, a visual analogue scale (VAS) for the question "Over the past 24 hours, how bad was your knee pain?" from 0 ("none") to 100 ("worst imaginable"). Secondly, a VAS for the question "Over the past 24 hours, how bad was your knee stiffness?" from 0 ("none") to 100 ("worst imaginable"). Thirdly, we assessed surgery intent using multiple questions: "What do you think are the chances you'll have knee surgery in the next year, in %?" as well as the same question for 2 and 5 years into the future. We also asked "On a scale of 0 to 10 how interested are

you in knee surgery?" with labels "not at all" at 0, and "definitely going to get surgery" at 10. Lastly, we asked "Thinking about your symptoms, how well do you feel you understand your condition and your treatment options?" with answers "Not at all", "Slightly", "Moderately", "Very well", "Completely", coded from 0 to 4. All data were assessed at baseline and at the end of the 12-week DCP in both the intervention and control groups. Additionally, those in the treatment group were asked to complete these questions at various points during the DCP: the VAS twice each week, and the questions related to surgery and understanding of their condition at week 6.

Sample Size

The minimal clinically important difference for KOOS is considered to be approximately 10 points on the 100-point scale, and a standard deviation of 15 is recommended for power calculations [27]. Although we did not use the full KOOS scale, we assumed its derivate questionnaires would have similar properties and used a standard deviation of 15 to perform power calculations. The number of participants needed in each group to detect a 10-point difference given a Type I error rate of 0.05 and power of 0.8 is 36. Given our unequal allocation ratio, this would need to be at least 54 in the treatment group and 36 in the control group for a total of at least 90 participants in the trial. As we had two primary endpoints albeit not independent of one another, we significantly over-recruited participants into the trial. We opted for an unequal allocation ratio to ensure we would be able to enter a certain minimum number of people into the treatment arm, a criterion mandated by the commercial nature of the deployments.

Statistical Analysis

Our primary analysis was conducted using a modified intent-to-treat approach. This analysis included all participants that were randomized, including those in the treatment group that never started the DCP. However, we excluded 7 participants in the control group that were enrolled in the DCP due to an administrative error (including these participants does not materially affect the statistical significance of the results). We describe baseline characteristics for the treatment and control groups using frequencies, means, and standard deviation. For those participants in the treatment group that performed at least one session of exercise therapy and those that completed the week 12 outcome questionnaire respectively, we also provide descriptive statistics of their engagement with the DCP. The analysis of primary and secondary outcomes was performed using a linear mixed model using the Linear Mixed-Effects Model ("lme4") package in R [28] with within-subject factor "time point" (baseline or outcome) and between-subject factor "group" (treatment or control) and their interaction. We modeled a separate baseline for each participant, effectively examining the change scores only (in lme4 this was performed as "score ~ timepoint*group + (1|participant)", where (1|participant) models an intercept for each participant separately). We assessed normality of the residuals based on quantile-quantile (QQ)-plots. If we did not have outcome data for a participant, we used last observation carried forward (LOCF). For those in the control group, this meant their baseline was carried forward; for those in the treatment group this meant either their baseline was carried forward, or data collected during the course of the DCP. We also analyzed all primary and secondary outcomes with baseline carried forward (BOCF) also for the treatment group (rather than LOCF). We also omitted LOCF and instead allowed the mixed-effects model to account for the missing data, which yielded an identical pattern of results as using LOCF and BOCF. We also report the primary and secondary outcomes following a per-protocol analysis to assess the effect of the program on those that complete it. Lastly, we performed an exploratory subgroup analysis using the same primary and secondary outcomes on participants that met the criteria for knee OA as defined by having at least 3 out of 6 clinical criteria: age >50 years, stiffness in the morning <30 min, crepitus, bony tenderness, bony enlargement, and no palpable warmth [11].

Surgery Cost Savings

We report the expected savings on surgery costs based on participants' self-assessment of their likelihood to have surgery. The calculation estimates the cost of knee surgery at US $40,000 [29]. For example, a 10-percentage point reduction in self-reported 1-year surgery likelihood would translate into a cost saving of US $4000 in the first year, minus the costs of the program per participant. The net cost saving is not considered a primary or secondary outcome of the clinical trial and is only calculated for those participants completing the trial (per protocol).

Results

Study Population

A total of 309 people completed the screening for CKP in January or February 2017, of which 162 entered the trial and were randomized (Figure 1). Of those 162 individuals, 62% entered the treatment arm (101/162) and 38% entered the control arm (61/162). Seven participants in the control arm were given the DCP due to administrative error and have all been excluded from all following results. The errors afflicted this set of participants completely at random (ie, not as a function of symptoms, demographics, or otherwise) and their exclusion, therefore, does not bias the findings; in contrast, including these participants would have led to an underestimation of the true effect of treatment. The baseline demographics were comparable between groups (Table 1), with the average participant 46 years of age and overweight. At baseline, all but 1 participant believed the DCP would help them delay surgery, and 87% (135/155) believed the DCP could help them avoid surgery altogether. A substantial minority (41%) had already undergone knee surgery in the past, though none were actively rehabilitating. The only difference in demographics between both groups was the gender balance; there were more women in the treatment compared to control group (43% versus 26% respectively).

Table 1. Demographics of the control and treatment groups. The term SD refers to standard deviation.

Characteristic	Treatment	Control	All
Number of participants	101	54	155
Age in years, mean (SD)	46 (12)	47 (12)	46 (12)
Body Mass Index (kg/m2), mean (SD)	27 (5)	28 (4)	27 (5)
Female, n (%)	43 (43)	14 (26)	57 (37)
Physical Training-like exercise at screening[a], n (%)	33 (33)	17 (31)	50 (32)
Fear avoidance[b], n (%)	6 (6)	6 (11)	12 (8)
Godin activity score[c], mean (SD)	34 (23)	39 (25)	36 (24)
Hours sedentary per day, mean (SD)	6 (3)	6 (3)	6 (3)
Think Digital Care Program can delay surgery, n (%)	100 (99)	54 (100)	154 (99)
Think Digital Care Program can avoid surgery, n (%)	87 (86)	48 (89)	135 (87)
Taking antidepressants, n (%)	2 (2)	0 (0)	2 (1)
Taking opioids, n (%)	5 (5)	3 (6)	8 (5)
Self-efficacy[d], mean (SD)	11 (3)	10 (3)	11 (3)
Surgery on the knee in the past, n (%)	45 (45)	19 (35)	64 (41)
Knee osteoarthritis[e], n (%)	75 (74)	44 (81)	119 (77)

[a]Positive answer to the question “Do you currently do any physical therapy-style exercises?”

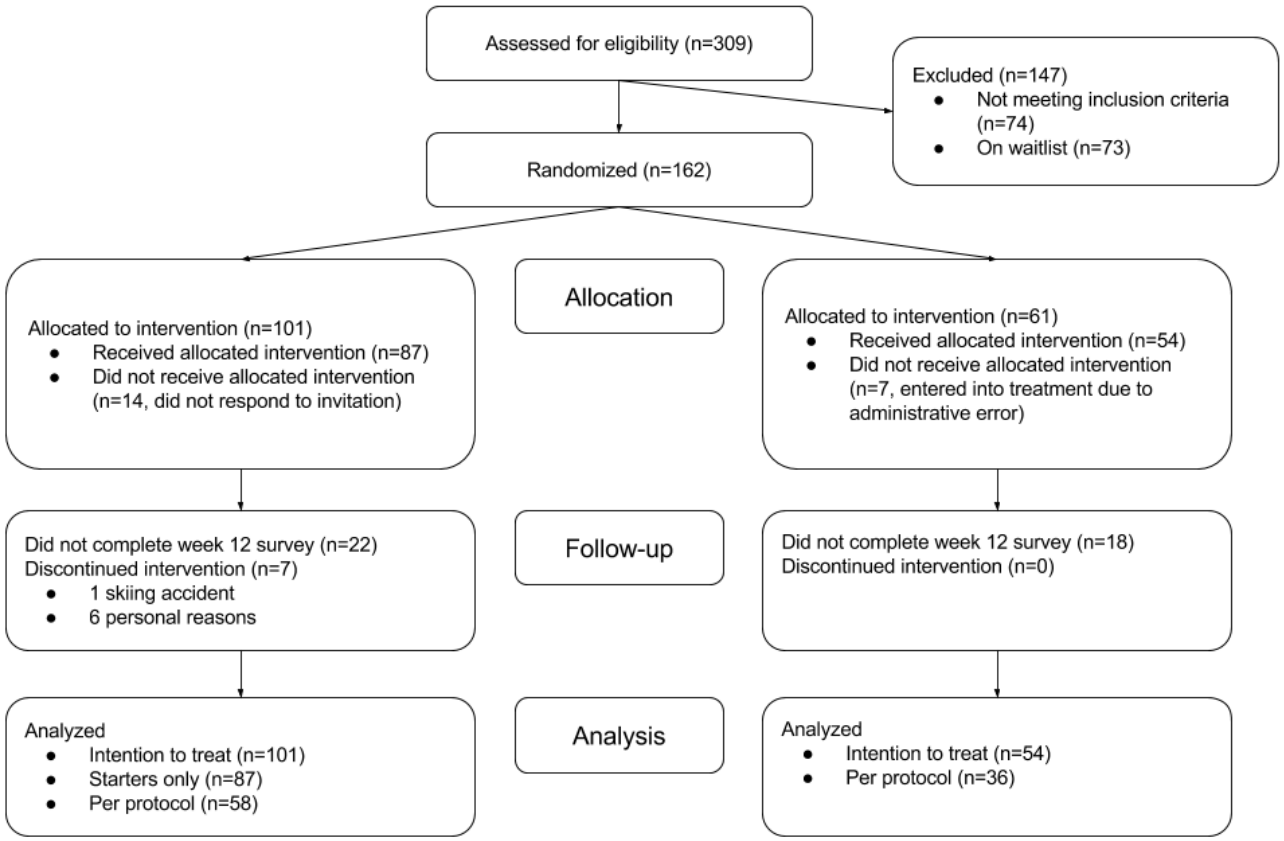
[b]Positive answer to the question “It's really not safe for a person with a condition like mine to be physically active.”

[c]Composite score; 24 indicates “active”, 14-23 indicates “Moderately active”, and <14 indicates “Insufficiently active/sedentary” [30].

[d]Health self-efficacy assessment, scores from 0 (no self-efficacy) to 15 (high self-efficacy) [31].

[e]Defined as satisfying at least 3 out of 6 clinical criteria for osteoarthritis [11].

Figure 1. Participant flow diagram.



Participant Flow

Participants in the treatment group were lost at three stages. First, of the 101 participants randomized to treatment, 14 did not respond to our invitation to take part in the DCP or subsequent follow-up communications. Second, 7 participants actively withdrew during the course of the DCP, due to injury or for personal reasons (eg, time constraints or stress at work, Figure 1). Third, 22 participants did not complete the week 12 online outcomes survey.

In the control group, 7 participants were placed in the DCP due to an administrative error. A further 18 participants did not complete the week 12 outcomes survey.

Engagement

For participants in the treatment group, we tracked completion of each component of the DCP. Participants that started the DCP (n=87), defined as performing at least one sensor-guided workout, performed an average of 33 in-app workouts, or an average of 2.5 workouts per week from week 0 (introduction to the program) to 12. Users that completed the outcome questionnaires at 12 weeks (n=59) performed 43 sensor-guided workouts (3.3 workouts per week), compared to the 3 times per week that is recommended in the DCP. Average weekly engagement with the DCP was 76% for those that started the program, and 95% for those that completed it. Participants that completed the 12-week follow-up read approximately 10 education articles, completed 2 Cognitive Behavioral Therapy (CBT) sessions, posted on the feed 8 times, and contacted their coach over text message or in-app message about 7 times.

Outcomes

Primary Outcomes

Both primary outcomes improved significantly more in the treatment group compared to the control group (Table 3). We observed a statistically significant mean group difference for KOOS Pain whereby the treatment group improved by 7.7 (95% CI 3.0 to 12.3) points more than the control group. Similarly, the treatment group improved by 7.2 (95% CI 3.0 to 11.5) points more than the control group on the KOOS-PS scale.

Secondary Outcomes

Each of the secondary outcomes also showed the DCP to be superior to control (Table 3). The VAS pain and stiffness scores, although improved in the control as well as the treatment group, improved by 12 (95% CI 5.4 to 19.1) points more for those in the DCP than in control. The self-reported likelihood of surgery in the next 1, 2, and 5 years, as well as interest in having surgery, all decreased more for those in the DCP than those in control over the 12-week period. Lastly, participants on the DCP improved their understanding of their condition and treatment options more than those in the control group. All primary and secondary outcomes remained statistically significant in an analysis using baseline carried forward for the treatment group rather than last observation carried forward.

Per Protocol Analysis

We also provide results for only those participants that received their allocated intervention and completed the outcome questionnaires (n=58 for treatment, n=36 for control) in Table 4. The per protocol results are fully consistent with the intent-to-treat analysis.

Based on the reductions in surgery likelihood we also calculated the net savings per participant of the program after accounting for the costs of delivering the program. The 1-year net saving of the digital care program is US $4340 (13.1% * US $40,000, corrected for the cost of the digital care program), the 2-year savings are US $4660, and the 5-year savings are US $7900.

Subgroup Analysis: Knee Osteoarthritis

An intent-to-treat analysis on participants with knee osteoarthritis (75/101 in the treatment group; 44/54 in the control group) showed results consistent with the original intent-to-treat analysis presented in Table 3 in terms of the magnitude of the group difference and statistical significance.

Table 2. Engagement indicators for each of the aspects of the Digital Care Program. “Starters” indicates participants performed at least one sensor-guided workout. “Finishers” indicates participants that completed the outcomes questionnaires at week 12 follow-up.

Indicator	All Starters (n=87)	All Finishers (n=59)
Number of workouts, mean (SD)	33.1 (24)	42.9 (17.3)
Users engaging with the program per week, n (%)	66.3 (76.2)	55.9 (94.8)
Users active with sensor-guided exercise in weeks 1-4, n (%)	78 (89.7)	58 (98.3)
Users active with sensor-guided exercise in weeks 5-8, n (%)	69 (79.3)	56 (94.9)
Users active with sensor-guided exercise in weeks 9-12, n (%)	60 (69)	55 (93.2)
Offline activities logged in hours, mean (SD)	9.6 (9.1)	13.2 (8.8)
Education articles read, mean (SD)	7.3 (4.5)	9.6 (3.1)
Cognitive Behavioral Therapy session completed, mean (SD)	1.4 (1.2)	1.8 (1.1)
Team posts and comments, mean (SD)	6.1 (7.2)	8.4 (7.6)
In-app messages sent to coach, mean (SD)	5.9 (5.6)	6.6 (5.7)

Table 3. Primary and secondary outcomes for the intent-to-treat analysis of treatment and control groups. The mean group difference as well as the *P* value for the interaction are derived from the linear mixed effects model. Each of the primary and secondary outcomes favors the treatment over the control group. *P* values uncorrected for multiple tests. KOOS: Knee injury and Osteoarthritis Outcome Score; PS: Physical Function Shortform; VAS: visual analogue scale.

Outcome	Treatment at baseline, mean (SD)	Treatment at outcome, mean (SD)	Control at baseline, mean (SD)	Control at outcome, mean (SD)	Group difference, mean (95% CI)	Interaction, *P* value
Primary Outcomes						
KOOS Pain (0-100)	41.0 (14.1)	30.3 (17.1)	41.4 (16.5)	38.4 (17.2)	–7.7 (–12.3 to –3)	.002
KOOS-PS (0-100)	53.8 (12.3)	44.6 (16.7)	54.5 (15.7)	52.5 (16.2)	–7.2 (–11.5 to –3)	.001
Secondary Outcomes						
VAS Pain score (0-100)	45.2 (21.4)	26.6 (22)	44.7 (20.3)	38.3 (22.2)	–12.3 (–19.1 to –5.4)	.001
VAS Stiffness score (0-100)	42.6 (23.4)	25.1 (22.3)	47.4 (21.9)	43.2 (21.6)	–13.4 (–21.1to –5.6)	.001
Surgery chance next year, %	24.5 (26.9)	14.7 (25)	24.3 (26.2)	23.9 (29.1)	–9.4 (–16.6 to –2.2)	.01
Surgery chance next two years, %	32.1 (31)	19.1 (26.9)	31.7 (27.9)	30 (28.9)	–11.3 (–20.1 to –2.5)	.01
Surgery chance next five years, %	47.8 (35)	27.5 (32.9)	49.8 (32.7)	44.1 (33.6)	–14.6 (–23.6 to –5.5)	.002
Surgery interest (0-10)	3.03 (3.41)	1.92 (2.93)	3.02 (3.32)	2.89 (3.21)	–1.0 (–1.7 to –0.2)	.01
Understanding of condition and treatment options (0-4)	1.92 (1.01)	2.68 (0.937)	1.94 (1.04)	1.76 (1.03)	0.9 (0.6 to 1.3)	<.001

Table 4. Per protocol results. All participants that completed their assigned treatment and completed the week 12 outcome questionnaire are included. KOOS: Knee Injury and Osteoarthritis Outcome Score; PS: Physical Function Shortform; VAS: visual analogue scale.

Outcome	Treatment at baseline, mean (SD)	Treatment at outcome, mean (SD)	Control at baseline, mean (SD)	Control at outcome, mean (SD)	Group difference, mean (95% CI)	Interaction, *P* value
Primary outcomes						
KOOS Pain (0-100)	39.6 (14.5)	21.8 (13.4)	39.2 (14.7)	34 (12.9)	–12.6 (–18.7 to –6.5)	<.001
KOOS-PS (0-100)	52.9 (12.6)	37.4 (16.1)	51.8 (16.4)	48.4 (15.9)	–12.1 (–17.7 to –6.6)	<.001
Secondary outcomes						
VAS Pain score (0-100)	44.1 (21.5)	17.2 (16.2)	45.5 (19.6)	35.8 (21.8)	–17.3 (–26.3 to –8.3)	<.001
VAS Stiffness score (0-100)	42.4 (24.3)	15.9 (17.3)	47.4 (22.1)	40.8 (20.9)	–19.9 (-30.4 to –9.4)	<.001
Surgery chance next year, %	21.6 (24.9)	7.59 (18.5)	20.8 (21.9)	20 (26.3)	–13.1 (–24.1 to –2.2)	.02
Surgery chance next two years, %	28.1 (29.1)	12.1 (21.5)	27.4 (25)	25.3 (26.9)	–13.9 (–26.6 to –1.3)	.03
Surgery chance next five years, %	48.7 (33.9)	18.2 (26.5)	48.6 (29.9)	40.1 (30.9)	–22 (–35 to –9.1)	.001
Surgery interest (0-10)	2.93 (3.28)	1.31 (2.39)	3 (3.3)	2.78 (3.14)	–1.4 (–2.5 to –0.3)	.01
Understanding of condition and treatment options (0-4)	1.88 (1.03)	3.09 (0.657)	1.83 (1.06)	1.56 (0.998)	1.5 (1.1 to 1.9)	<.001

Discussion

Principal Findings

While CKP is a prevalent cause of disability worldwide [1,2], comprehensive conservative programs for the condition are lacking. The Hinge Health DCP has been designed to address this lack of chronic pain programs and incorporates components of best-practice conservative care for CKP in a digital format that provides flexibility to the user. The results of this randomized controlled trial demonstrated large improvements in knee pain, physical function, and stiffness in individuals with CKP on the Hinge Health DCP that were significantly greater than a control group receiving knee care education and treatment as usual over a period of 12 weeks after program initiation. Participant surgery interest also significantly decreased and understanding of their condition increased in the treatment group as compared to the control group. The positive results of this study demonstrate the potential of the Hinge Health DCP as a treatment for a large number of individuals affected by CKP that otherwise would be at risk for surgery.

Analysis of primary study outcomes demonstrated large improvements in both KOOS pain and function scales in the treatment group. Similarly, significantly greater improvements in physical function (KOOS-PS) scores were observed in the treatment group as compared to the control group. When considering individuals who started on and completed the study

as per protocol, the improvements observed in KOOS Pain and function scores were 45% in the treatment group as compared to 13% in control, and 29% in treatment as compared to 7% in control, respectively. The improvements observed in the treatment group exceeded recommended minimal clinically important changes for KOOS [27], while the small improvements in the control group did not. The group difference in KOOS did not quite reach the minimal clinically important difference due to the relatively large drop-off which 'diluted' the improvements of those that completed the program. Nonetheless, the clinically significant improvements demonstrated at the end of the 12-week program for not only the primary outcome measures but also secondary metrics provide strong evidence for the benefits of the Hinge Health DCP for individuals affected by CKP.

Similar large improvements that were significantly greater in the treatment group than in control were noted in secondary outcomes of VAS pain and stiffness scales. At the end of the 12-week program, per protocol participants in the treatment group had 61% and 63% improvements in VAS pain and stiffness, as compared to 21% and 14%, respectively, in the control group. Subjects' perception of surgery interest and surgery requirements also changed favorably at the end of the 12-week Hinge Health DCP, with a 63% reduction in the belief that they would require surgery within the next 5 years in the per protocol treatment group as compared to a reduction of 17% in the control group. It is also important to note that the Hinge Health DCP was safe for participants, as there were no reported adverse events during the 12-week program.

The results of this study demonstrated comparable or greater improvements in pain, physical function, and stiffness as compared to other treatment programs for CKP. Bossen et al [18] demonstrated improvements of 15% in physical function and 35% in pain at 3 months after initiation of a 9-week web-based behavior graded physical activity intervention in patients with knee and/or hip OA. Hughes et al [32] found improvements in Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain, stiffness, and function scores of 23%, 17%, and 23%, respectively, at the end of an 8-week exercise and behavior-change program for OA. Similarly, Hurley et al [33] demonstrated per protocol improvements in WOMAC pain and function of 31% and 26%, respectively, at the end of a 6-week rehabilitation program combining self-management and exercise for CKP. While the current study did not investigative the longer-term effect of the Hinge Health DCP past the end of the 12-week program, data from prior studies of treatment programs of similar duration (6-12 weeks) [18,22,32-34] showed improvement in outcomes can be maintained as long as 30 months after program completion [33]. Thus, it is likely that the results of the Hinge Health DCP would be maintained after the 12-week program.

The improvements observed as a result of the Hinge Health DCP have the potential to translate into substantial economic savings, however, in lieu of long-term data are based on participant self-report. Based on results for participants' self-reported likelihood of having surgery, the potential savings per completing participant due to surgery avoidance alone equate to US $4340 net cost savings on surgery over the first year in individuals using the Hinge Health DCP as compared to treatment as usual. Other integrated rehabilitation programs for CKP have also demonstrated lower healthcare costs as compared to usual care [33]. Thus, while the long-term effect of the Hinge Health DCP may in part be dependent on continued adherence to the program [35], it is anticipated that the behavioral, educational, and psychosocial components of the program have the potential for long-term clinical and economic effects [33].

When interpreting the results of this study, its strengths and limitations should be considered. Strengths of this study include the randomized controlled study design, and that the study was designed, conducted, and analyzed according to a pre-specified protocol. Further, the digital format of the program provides flexibility to individuals to participate at times and places convenient to them. While the results of this study demonstrate significantly greater improvements in primary and secondary outcomes with the Hinge Health DCP as compared to control, this study did not investigate the long-term effect of the Hinge Health DCP past the end of the 12-week program. Thus, further work is needed to evaluate the long-term impact of the Hinge Health DCP as compared to control. Our prior work suggests the potential for long-term effects as it demonstrated improved patient-reported outcomes in a single-arm study 6 months after program initiation [24]. We are in the process of collecting multi-year data and these results will be published in due course. A second point to note is that preliminary analyses not shown here suggest that BMI, gender, and surgery risk all affect the risk of dropping out. We plan to investigate risk factors for failure to adhere to the DCP in an upcoming study. Finally, around 20% of participants had less than 3 months of pain over the past 12 months, which does not strictly meet a definition of chronic pain. In a larger cohort, the efficacy of the program on long-term versus intermittent knee pain should be examined.

The study enrolled a representative population with CKP problems. While CKP is a hallmark symptom of knee OA [11], a diagnosis of knee OA was not required for inclusion in this study. Analysis of participant baseline data demonstrated that 74% of individuals in the treatment arm and 80% of individuals in the control arm had knee OA as defined by clinical diagnosis for knee OA derived from the American College of Rheumatology criteria for OA of the knee [11]. A sub-group analysis of these participants confirmed the successful outcome of the Hinge Health DCP in the primary and secondary outcomes over the 12-week period as compared to control, demonstrating the applicability of the program to highly prevalent knee OA.

The comprehensive nature of the Hinge Health DCP addresses multiple components of recommended management for CKP [9,10]. However, it is therefore unknown if all components of the program (sensor-guided exercise therapy, education, cognitive behavioral therapy, weight loss, and psychosocial support) are necessary to attain the reported results. Similar to other studies investigating interventions for CKP and knee OA, [18,22,32,34,36] due to the nature of this study, participants could not be blinded as to the intervention, and thus we cannot rule out the possibility of an attention effect. The attrition rate for week 12 patient-reported outcomes was in line with other studies for CKP [18]. Further, as the rate was similar in both

control and treatment it is not anticipated to have impacted the findings of the study.

Conclusion

Individuals with CKP who used the Hinge Health DCP for 12 weeks experienced significantly greater improvement in self-reported clinical outcome measures of pain, physical function, stiffness, as well as surgery intent and understanding of their condition, as compared to a control group receiving knee education articles and treatment as usual. Given the observed benefits, the Hinge Health DCP may be an effective comprehensive treatment program for individuals with CKP.

Acknowledgments

We would like to thank all participants for their efforts, and the participating companies for their contributions to the success of the deployments.

Authors' Contributions

GM, PS, and DP conceived of the study. All authors helped design the study. PS was responsible for data collection and analysis. All authors contributed to the manuscript.

Conflicts of Interest

All authors except JH work at Hinge Health, Inc. Author JH is a paid domain expert consultant.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 648KB - jmir_v20i4e156_app1.pdf]

References

1. Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. Ann Rheum Dis 2001 Feb;60(2):91-97 [FREE Full text] [Medline: 11156538]
2. Global B. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]
3. Nguyen UDT, Zhang Y, Zhu Y, Niu J, Zhang B, Felson DT. Increasing prevalence of knee pain and symptomatic knee osteoarthritis: survey and cohort data. Ann Intern Med 2011 Dec 06;155(11):725-732 [FREE Full text] [doi: 10.7326/0003-4819-155-11-201112060-00004] [Medline: 22147711]
4. Dawson J, Linsell L, Zondervan K, Rose P, Carr A, Randall T, et al. Impact of persistent hip or knee pain on overall health status in elderly people: a longitudinal population study. Arthritis Rheum 2005 Jun 15;53(3):368-374 [FREE Full text] [doi: 10.1002/art.21180] [Medline: 15934104]
5. 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]
6. Agaliotis M, Mackey MG, Jan S, Fransen M. Burden of reduced work productivity among people with chronic knee pain: a systematic review. Occup Environ Med 2014 Sep;71(9):651-659. [doi: 10.1136/oemed-2013-101997] [Medline: 24872332]
7. Gureje O, Von KM, Simon GE, Gater R. Persistent pain and well-being: a World Health Organization Study in Primary Care. JAMA 1998 Jul 8;280(2):147-151. [Medline: 9669787]
8. 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]
9. Booth J, Moseley G, Schiltenwolf M, Cashin A, Davies M, Hübscher M. Exercise for chronic musculoskeletal pain: A biopsychosocial approach. Musculoskeletal Care 2017 Dec;15(4):413-421. [doi: 10.1002/msc.1191] [Medline: 28371175]
10. 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]
11. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986 Aug;29(8):1039-1049. [Medline: 3741515]
12. Deyle GD, Henderson NE, Matekel RL, Ryder MG, Garber MB, Allison SC. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. A randomized, controlled trial. Ann Intern Med 2000 Feb 01;132(3):173-181. [Medline: 10651597]
13. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]

14. Riddle D, Jiranek W, Hayes C. Use of a validated algorithm to judge the appropriateness of total knee arthroplasty in the United States: a multicenter longitudinal cohort study. Arthritis Rheumatol 2014 Aug;66(8):2134-2143 [FREE Full text] [doi: 10.1002/art.38685] [Medline: 24974958]
15. Aitken D, Buchbinder R, Jones G, Winzenberg T. Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. Aust Fam Physician 2015;44(1-2):39-42 [FREE Full text] [Medline: 25688958]
16. Sabaté E. World Health Organization, editors. Adherence to long-term therapiesvidence for action. Geneva: World Health Organization; 2003:e.
17. Brooks MA, Beaulieu JE, Severson HH, Wille CM, Cooper D, Gau JM, et al. Web-based therapeutic exercise resource center as a treatment for knee osteoarthritis: a prospective cohort pilot study. BMC Musculoskelet Disord 2014 May 17;15:158 [FREE Full text] [doi: 10.1186/1471-2474-15-158] [Medline: 24884547]
18. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a Web-Based Physical Activity Intervention in Patients With Knee and/or Hip Osteoarthritis: Randomized Controlled Trial. J Med Internet Res 2013 Nov 22;15(11):e257. [doi: 10.2196/jmir.2662]
19. Rini C, Porter LS, Somers TJ, McKee DC, DeVellis RF, Smith M, et al. Automated Internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015 May;156(5):837-848 [FREE Full text] [doi: 10.1097/j.pain.0000000000000121] [Medline: 25734997]
20. Dahlberg LE, Grahn D, Dahlberg JE, Thorstensson CA. A Web-Based Platform for Patients With Osteoarthritis of the Hip and Knee: A Pilot Study. JMIR Res Protoc 2016 Jun 03;5(2):e115 [FREE Full text] [doi: 10.2196/resprot.5665] [Medline: 27261271]
21. Dobson F, Hinman RS, French S, Rini C, Keefe F, Nelligan R, et al. Internet-mediated physiotherapy and pain coping skills training for people with persistent knee pain (IMPACT - knee pain): a randomised controlled trial protocol. BMC Musculoskelet Disord 2014 Aug 13;15:279 [FREE Full text] [doi: 10.1186/1471-2474-15-279] [Medline: 25125068]
22. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
23. Li LC, Cott C, Jones CA, Badley EM, Davis AM, PEOPLE Meeting Participants. Improving Primary Health Care in Chronic Musculoskeletal Conditions through Digital Media: The PEOPLE Meeting. JMIR Res Protoc 2013 Mar 08;2(1):e13 [FREE Full text] [doi: 10.2196/resprot.2267] [Medline: 23612113]
24. Smittenaar P, Erhart-Hledik J, Kinsella R, Hunter S, Mecklenburg G, Perez D. Translating Comprehensive Conservative Care for Chronic Knee Pain Into a Digital Care Pathway: 12-Week and 6-Month Outcomes for the Hinge Health Program. JMIR Rehabil Assist Technol 2017 Apr 05;4(1):e4 [FREE Full text] [doi: 10.2196/rehab.7258] [Medline: 28582253]
25. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88-96. [doi: 10.2519/jospt.1998.28.2.88] [Medline: 9699158]
26. Perruccio AV, Stefan LL, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008 May;16(5):542-550 [FREE Full text] [doi: 10.1016/j.joca.2007.12.014] [Medline: 18294869]
27. Roos E, Lohmander L. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003 Nov 03;1:64 [FREE Full text] [doi: 10.1186/1477-7525-1-64] [Medline: 14613558]
28. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen and S4. R Package Version ?. 2014. others URL: https://rdrr.io/cran/lme4/ [WebCite Cache ID 6ykM0M2jq]
29. Blue CBS. Blue Health Intelligence. 2015 Jan. A Study of Cost Variations for Knee and Hip Replacement Surgeries in the U.S. Internet URL: https://www.bcbs.com/the-health-of-america/reports/study-of-cost-variations-knee-and-hip-replacement-surgeries-the-us [WebCite Cache ID 6ykMCIDzm]
30. Godin G, Shephard R. Godin leisure-time exercise questionnaire. Med Sci Sports Exerc 1997;29(6):38.
31. Schwarzer R, Renner B. Social-cognitive predictors of health behavior: action self-efficacy and coping self-efficacy. Health Psychol 2000 Sep;19(5):487-495. [Medline: 11007157]
32. Hughes SL, Seymour RB, Campbell RT, Huber G, Pollak N, Sharma L, et al. Long-term impact of Fit and Strong! on older adults with osteoarthritis. Gerontologist 2006 Dec;46(6):801-814. [Medline: 17169935]
33. Hurley MV, Walsh NE, Mitchell H, Nicholas J, Patel A. Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: a pragmatic, cluster randomized, controlled trial. Arthritis Care Res (Hoboken) 2012 Feb;64(2):238-247 [FREE Full text] [doi: 10.1002/acr.20642] [Medline: 21954131]
34. Skou ST, Rasmussen S, Laursen MB, Rathleff MS, Arendt-Nielsen L, Simonsen O, et al. The efficacy of 12 weeks non-surgical treatment for patients not eligible for total knee replacement: a randomized controlled trial with 1-year follow-up. Osteoarthritis Cartilage 2015 Sep;23(9):1465-1475 [FREE Full text] [doi: 10.1016/j.joca.2015.04.021] [Medline: 25937024]
35. Roddy E, Zhang W, Doherty M, Arden NK, Barlow J, Birrell F, et al. Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee--the MOVE consensus. Rheumatology (Oxford) 2005 Jan;44(1):67-73. [doi: 10.1093/rheumatology/keh399] [Medline: 15353613]

36. Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Patel A, Williamson E, et al. Clinical effectiveness of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain: a cluster randomized trial. Arthritis Rheum 2007 Oct 15;57(7):1211-1219 [FREE Full text] [doi: 10.1002/art.22995] [Medline: 17907147]

Abbreviations

BMI: body-mass index
CBT: cognitive behavioral therapy
CKP: chronic knee pain
CONSORT: Consolidated Standards of Reporting Trials
DCP: digital care program
KOOS: Knee injury and Osteoarthritis Outcome Score
OA: osteoarthritis
PS: Physical Function Shortform
VAS: visual analogue scale

Edited by G Eysenbach; submitted 15.12.17; peer-reviewed by J Dahlberg, L Dahlberg, S Rasmussen; comments to author 23.01.18; revised version received 20.03.18; accepted 07.04.18; published 25.04.18.

Please cite as:
Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S
Effects of a 12-Week Digital Care Program for Chronic Knee Pain on Pain, Mobility, and Surgery Risk: Randomized Controlled Trial
J Med Internet Res 2018;20(4):e156
URL: http://www.jmir.org/2018/4/e156/
doi:10.2196/jmir.9667
PMID:29695370

Original Paper

Guided Web-Based Cognitive Behavior Therapy for Perfectionism: Results From Two Different Randomized Controlled Trials

Alexander Rozental[1,2], PhD; Roz Shafran[2], PhD; Tracey D Wade[3], PhD; Radha Kothari[2], PhD; Sarah J Egan[4], PhD; Linda Ekberg[5], MSc; Maria Wiss[5], MSc; Per Carlbring[1], PhD; Gerhard Andersson[5,6], PhD

[1]Department of Psychology, Stockholm University, Stockholm, Sweden

[2]Institute of Child Health, University College London, London, United Kingdom

[3]School of Psychology, Flinders University, Adelaide, Australia

[4]School of Psychology, Curtin University, Perth, Australia

[5]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[6]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Roz Shafran, PhD
Institute of Child Health
University College London
30 Guilford Street
London, WC1N 1EH
United Kingdom
Phone: 44 0207 905 2232
Email: r.shafran@ucl.ac.uk

Abstract

Background: Perfectionism can become a debilitating condition that may negatively affect functioning in multiple areas, including mental health. Prior research has indicated that internet-based cognitive behavioral therapy can be beneficial, but few studies have included follow-up data.

Objective: The objective of this study was to explore the outcomes at follow-up of internet-based cognitive behavioral therapy with guided self-help, delivered as 2 separate randomized controlled trials conducted in Sweden and the United Kingdom.

Methods: In total, 120 participants randomly assigned to internet-based cognitive behavioral therapy were included in both intention-to-treat and completer analyses: 78 in the Swedish trial and 62 in the UK trial. The primary outcome measure was the Frost Multidimensional Perfectionism Scale, Concern over Mistakes subscale (FMPS CM). Secondary outcome measures varied between the trials and consisted of the Clinical Perfectionism Questionnaire (CPQ; both trials), the 9-item Patient Health Questionnaire (PHQ-9; Swedish trial), the 7-item Generalized Anxiety Disorder scale (GAD-7; Swedish trial), and the 21-item Depression Anxiety Stress Scale (DASS-21; UK trial). Follow-up occurred after 6 months for the UK trial and after 12 months for the Swedish trial.

Results: Analysis of covariance revealed a significant difference between pretreatment and follow-up in both studies. Intention-to-treat within-group Cohen *d* effect sizes were 1.21 (Swedish trial; 95% CI 0.86-1.54) and 1.24 (UK trial; 95% CI 0.85-1.62) for the FMPS CM. Furthermore, 29 (59%; Swedish trial) and 15 (43%; UK trial) of the participants met the criteria for recovery on the FMPS CM. Improvements were also significant for the CPQ, with effect sizes of 1.32 (Swedish trial; 95% CI 0.97-1.66) and 1.49 (UK trial; 95% CI 1.09-1.88); the PHQ-9, effect size 0.60 (95% CI 0.28-0.92); the GAD-7, effect size 0.67 (95% CI 0.34-0.99); and the DASS-21, effect size 0.50 (95% CI 0.13-0.85).

Conclusions: The results are promising for the use of internet-based cognitive behavioral therapy as a way of targeting perfectionism, but the findings need to be replicated and include a comparison condition.

(J Med Internet Res 2018;20(4):e154) doi:10.2196/jmir.9823

KEYWORDS

cognitive behavior therapy; internet; perfectionism; follow-up studies; cognitive therapy

Introduction

Perfectionism has many positive features, such as striving for excellence, but it can also have a negative impact in many areas, including mental health [1]. Characterized by perfectionistic strivings and perfectionistic concerns, perfectionism, instead of helping the individual fulfill their goals, is associated with avoidance, worry, procrastination, and self-criticism [2]. Perfectionism has also been found to be related to anxiety disorders, depression, and eating disorders, in part by increasing the odds of their occurrence, but also by interfering with treatment progress, particularly by making behavior change more difficult. In particular, certain populations tend to have elevated perfectionism, such as people with eating disorders, depression, social anxiety disorder, panic disorder, and obsessive-compulsive disorder, compared with healthy controls [3]. It has therefore been suggested that the use of psychological treatments to address perfectionism could improve outcomes for other conditions as well through transdiagnostic processes. Moreover, the two higher-order dimensions of perfectionism have also been found to be related to psychiatric disorders in different ways. Perfectionistic strivings seem to be particularly related to eating disorders, while perfectionism concerns are primarily associated with depression and anxiety disorders [2]. This could imply that psychological treatments targeting perfectionism might have to be adapted depending on the psychiatric disorder [3]. Furthermore, elevated levels of perfectionism can also be found among athletes and specific sociodemographics [2], suggesting that there may be populations that are particularly vulnerable to developing problems due to perfectionism, and for whom psychological intervention could prevent the development of further psychopathology.

A systematic review and meta-analysis by Lloyd et al [4] provided evidence for the efficacy of cognitive behavioral therapy (CBT) for managing perfectionism and for targeting symptoms of depression and anxiety. Data gathered from 6 clinical trials that used the Frost Multidimensional Perfectionism Scale, Concern over Mistakes subscale (FMPS CM) [5], a widely distributed self-report measure of perfectionism, showed that the average within-group Hedges *g* effect size between pre- and posttreatment was 1.32 (95% CI 1.02-1.64). Using the Clinical Perfectionism Questionnaire (CPQ), a self-report measure to assess clinical perfectionism [6], 3 clinical trials obtained comparable outcomes, with Hedges *g* effect sizes ranging from 0.90 to 1.24. However, even though the findings are promising, the meta-analysis did not provide any estimates for follow-up data, making it unclear whether the outcomes were maintained. A clinical trial by Egan et al [7], which recruited participants through self-referral and included an assessment at 6 months following treatment, indicated that CBT, administered face-to-face or via the internet without any guidance, had within-group Cohen *d* effect sizes of 2.11 (face-to-face; 95% CI 1.26-2.88) and 0.43 (unguided self-help; 95% CI –0.28 to 1.12) on the FMPS CM. For the CPQ, Cohen *d* effect sizes were 1.61 (face-to-face; 95% CI 0.83-2.32) and 0.73 (unguided self-help; 95% CI –0.01 to 1.42). In terms of recovery, based on a statistical cutoff and exceeding the Reliable Change Index (RCI), 67% (face-to-face) and 40% (unguided self-help) met the criteria for recovery at follow-up. This implies that the benefits can be maintained and possibly even improved from posttreatment. However, given the small sample size and high attrition rate, these findings need to be interpreted cautiously.

To extend the understanding of the treatment of perfectionism and to evaluate the efficacy of internet-based CBT (ICBT), 2 clinical trials of treatment with guided self-help were conducted in Sweden and the United Kingdom. The participants were self-referred and assigned to self-help with guidance from a therapist or to a waitlist control group. The results from pre- to posttreatment were encouraging, obtaining within-group Cohen *d* effect sizes of 1.03 (Swedish trial; 95% CI 0.69-1.36) and 1.47 (UK trial; 95% CI 1.06-1.86) for the FMPS CM. For the CPQ, Cohen *d* effect sizes were 1.44 (Swedish trial; 95% 1.08-1.78) and 1.67 (UK trial; 95% CI 1.25-2.07) [8,9].

This study aimed to assess the long-term benefits of treating perfectionism by determining the outcomes at follow-up for those who received ICBT with guided self-help in the Swedish and UK studies. However, given the differences between the 2 clinical trials on several key characteristics, for example their inclusion and exclusion criteria, we present the results separately rather than combined.

Methods

Participants and Procedure

Participants in both clinical trials consisted of self-referrals recruited through advertisements in the media, social media, and on campus grounds. Those interested in participating entered a website to complete a screening process and provide electronic informed consent. The websites were connected to a secure Web-based interface where self-report measures were presented, asynchronous communication with the study supervisors was carried out, and treatment content was delivered [10]. To log on, the participants had to use an autogenerated identification code (eg, 1234abcd), a strong personal password, and a one-time personal identification number sent to their mobile phone, ensuring safety and anonymity. Inclusion and exclusion criteria differed somewhat between the clinical trials, as did a few other key characteristics (see Table 1) [11,12].

Individuals fulfilling the inclusion criteria and being deemed eligible to participate were randomly assigned to ICBT with guided self-help or to a waitlist control group. In total, 156 were included in the Swedish trial (guided self-help n=78), compared with 120 in the UK trial (guided self-help n=62). Those assigned to the waitlist control later received the same treatment content, either in a second wave of treatment (Swedish trial) or by being given a self-help book (UK trial); we did not consider these participants in this study. Further details regarding recruitment and eligibility can be obtained elsewhere [8,9], as well as in Table 1 and the flowchart in Figure 1. Table 2 shows the baseline characteristics of the participants at pretreatment for each clinical trial; the 2 trials differed on several baseline characteristics. Overall, participants in the UK trial were more likely to be single, be younger, report prior mental health problems, have previous experience of undergoing psychological

treatment and using psychotropic medication, and have greater symptom severity on the FMPS CM, but not on the CPQ.

Measures

The main outcome measure in both trials was the FMPS CM [5], comprising 9 items related to worries of making mistakes that are scored on a 5-point Likert scale (range 1-5). The FMPS CM has a Cronbach alpha of .88 [5]; in the 2 trials it was .85 (Swedish trial) and .74 (UK trial). Both trials also administered the CPQ [6], comprising 12 items associated with a specific construct of clinical perfectionism that are scored on a 4-point Likert scale (range 0-3). The CPQ has a Cronbach alpha of .73 [6]; in the 2 trials it was .66 (Swedish trial) and .74 (UK trial).

The secondary outcome measures differed between the 2 clinical trials but are nonetheless reported to examine the benefits on conditions other than perfectionism. In the Swedish trial, these were the 9-item Patient Health Questionnaire (PHQ-9) [13] and the 7-item Generalized Anxiety Disorder scale (GAD-7) [14]. The PHQ-9 has 9 items related to depression that are scored on a 4-point Likert scale (range 0-3), with a range in scores from 0 to 27. The PHQ-9 has a Cronbach alpha of .89 [13]; in Swedish trial it was .84. The GAD-7 has 7 items measuring worry and anxiety that are scored on a 4-point Likert scale (range 0-3), with a range in scores from 0 to 21. The GAD-7 has a Cronbach alpha of .92 [14]; in the Swedish study it was .87. The UK trial used the 21-item Depression Anxiety Stress Scale (DASS-21) [15]. The DASS-21 has 21 items measuring depression, anxiety, and stress that are scored on 4-point Likert scale (range 0-3), with a range in scores from 0 to 63. The DASS-21 has a Cronbach alpha of .88 [16]; in the UK study it was .91.

Table 1. Key characteristics of the clinical trials. FMPS CM: Frost Multidimensional Perfectionism Scale, Concern over Mistakes subscale; SMS: short message service text messaging.

Characteristic	Swedish trial	UK trial
Recruitment process	Self-referrals	Self-referrals
Cutoff criteria for maladaptive perfectionism	No	Yes (≥29 on the FMPS CM)
Telephone interview assessment	Yes (Mini-International Neuropsychiatric Interview)	No
Inclusion criteria		
Minimum age	Yes (>18 years)	Yes (>18 years)
Concurrent psychological treatment	Not allowed	Allowed
Change in pharmacological treatment[a]	Not allowed	Allowed
Other more severe conditions	Not allowed (eg, anorexia nervosa)	Allowed (except suicidality)
Randomization	Yes (once)	Yes (continuous)
Confirmation to commence treatment	Yes	No
Starting date	Simultaneous for all participants	Individual starting dates
Guidance from a therapist	Yes (on 2 predetermined weekdays)	Yes (weekly but with no specification)
Therapist level	Master's degree	Mixed (undergraduate to PhD)
Therapist supervision	Yes (weekly)	Yes (monthly)
Feedback checked by supervisor	No	Yes
Treatment protocol	Egan et al [11][b]	Egan et al [11]
Treatment period	8 weeks	12 weeks
Follow-up	12 months	6 months
Follow-up reminders	Telephone, email, and SMS	Email
Monetary compensation	No	Yes (£10)[c]
Ethics approval	Yes (Dnr 2015/419-31)	Yes (project identifier 6222:001)
Study protocol	No	Yes [12]
Registered at ClinicalTrials.gov	No	Yes (registration no. NCT02756871)
Informed consent obtained	Yes	Yes

[a]Any change 3 months prior to the screening process.

[b]Minor change in order of modules and greater emphasis on behavioral interventions.

[c]If participant completed posttreatment assessment.

Figure 1. Flow of participants through the study. ICBT: Internet-based cognitive behavioral therapy; MINI: Mini-International Neuropsychiatric Interview.

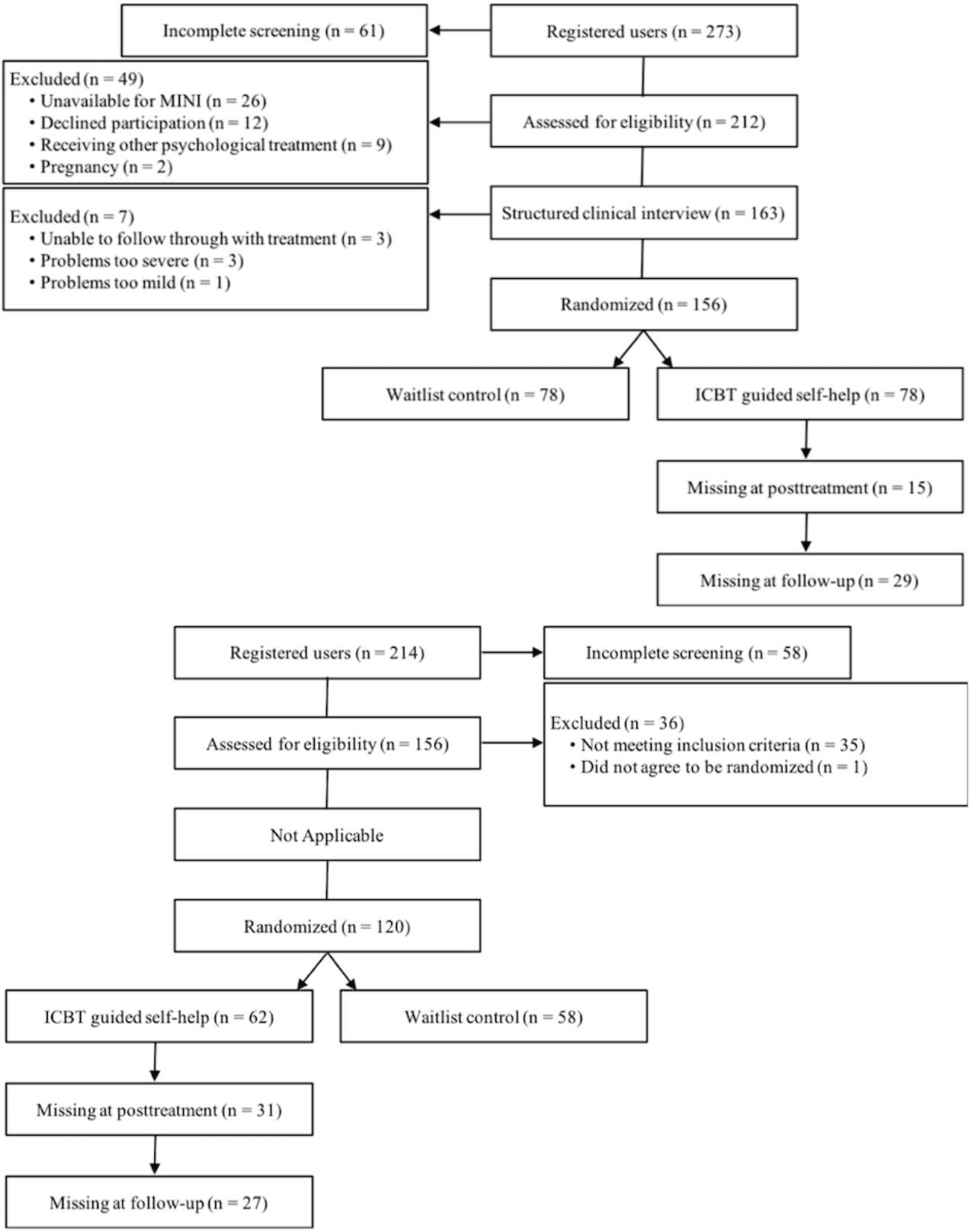


Table 2. Baseline characteristics of the participants at pretreatment assessments. FMPS CM: Frost Multidimensional Perfectionism Scale, Concern over Mistakes subscale; N/A: not applicable.

Baseline characteristic	UK trial (n=62)	Swedish trial (n=78)	χ^2	*df*	95% CI
Female sex, n (%)	49 (79.0)	64 (82.1)	0.2	1	N/A
Age (years), mean (SD)	28.6 (8.3)	34.22 (9.9)[a]	N/A	N/A	2.6 to -8.7
Marital status, n (%)			32.3	2	N/A
Single	48 (77.4)[a]	23 (29.5)			
Married/partner	14 (22.6)	52 (66.7)[a]			
Divorced/widowed	0 (0.0)	3 (3.8)			
Prior mental health problem (yes), n (%)	20 (27.4)	4 (5.1)	17.9	1	N/A
Psychological treatment (yes), n (%)	16 (25.8)[a]	0 (0)	22.7	1	N/A
Psychotropic medication (yes), n (%)	11 (17.7)[a]	5 (6.4)	36.5	1	N/A
Primary outcome measures, mean (SD)					
FMPS CM	36.7 (4.4)[a]	33.4 (6.4)	N/A	N/A	1.5 to 5.1
Clinical Perfectionism Questionnaire	35.7 (4.7)	38.3 (4.6)[a]	N/A	N/A	–4.1 to –1.0

[a]The clinical trial with significantly higher values on a specific baseline characteristic.

Table 3. Modules and order of presentation in the clinical trials.

Module	Swedish trial	UK trial
1	Understanding your perfectionism	Understanding your perfectionism
2	Your own model, values, and motivation	Your own model, values, and motivation
3	Surveys and experiments	Surveys and experiments
4	Dealing with perfectionistic behaviors	New ways of thinking
5	New ways of thinking	Dealing with perfectionistic behaviors
6	Self-criticism and self-compassion	Self-criticism and self-compassion
7	Self-worth	Self-worth
8	Maintain and continue positive change	Maintain and continue positive change

The outcome measures were completed at pre- and posttreatment and follow-up (12 months for the Swedish trial and 6 months for the UK trial). The CPQ was also distributed weekly during the treatment period for the UK trial.

Treatment and Therapists

In both clinical trials, the treatment content was derived from Egan et al [11], administered as an ICBT with guided self-help. It consisted of 8 modules, with 1 module delivered weekly, including psychoeducation, exercises, and homework assignments related to perfectionism. Each module was approximately 12 pages, totaling 121 pages. However, there were some minor differences between the 2 clinical trials in the order the modules were distributed (see Table 3). In addition, the Swedish trial put a greater emphasis on behavioral interventions than did the UK trial. Also, the treatment period was 8 weeks in the Swedish trial and 12 weeks in the UK trial. Therapists in the Swedish trial were Master's degree students having completed 1.5 years of clinical training, compared with a mixed set of therapists in the UK trial (undergraduates, master's degree students, doctoral candidates in clinical psychology, and PhDs). In both cases, the therapists provided the participants with weekly feedback, but while this was done on predetermined weekdays in the Swedish trial, this was not the case in the UK trial. The amount of therapist supervision also differed between the 2 clinical trials: weekly for the Swedish trial and monthly for the UK trial. Table 1 shows an overview of the differences. A study protocol for the UK trial is also available [12].

Statistical Analysis

Given the significant baseline differences between the 2 trials, as well as the different lengths of follow-up, we did not combine the data for any analyses. We calculated a priori power for the 2 clinical trials to detect significant differences compared with the waitlist control at posttreatment, but not for a between-study difference at follow-up. We explored differences between the Swedish and the UK trial at baseline, as well as baseline predictors of completion of the self-report measures at follow-up (ie, dropouts, regardless of the number of modules they had completed during the treatment period), by using 2-sided

independent *t* tests for continuous variables and chi-square tests for nominal variables.

We performed intention-to-treat analyses for the main and secondary outcome measures using multiple imputation to account for missing values and we conducted completer analyses using available data. The imputation models used all available self-report measures to create 10 imputed datasets. We then used analysis of covariance (ANCOVA) to investigate the change scores between pretreatment and follow-up, with scores at baseline as covariates, averaging the parameter estimates for the 10 analyses. All statistical analyses were done with IBM SPSS version 24.0.0.1 (IBM Corporation).

Using the results from the ANCOVAs, we calculated within-group effect sizes for the means of 2 assessments, divided by the pooled standard deviations. All results are presented with 95% CIs, where applicable, including effect sizes, and discussed in relation to similar findings in other clinical trials [17]. For the FMPS CM, we defined recovery as having a score at follow-up within 1 SD of that of the general population (<29 [18]; ie, a clinically significant change), with a change score also exceeding the RCI (ie, 1.96 times the standard error of the instrument [19]). This is assumed to reflect those participants moving from a dysfunctional to a functional distribution in terms of perfectionism. Furthermore, we determined improvement using only the RCI as a cutoff. Also, nonresponse corresponded to those participants not exceeding the RCI in any direction, while we calculated reliable deterioration using the RCI but in a negative direction [20].

All presented results concern the 6- and 12-month follow-ups. For a review of the outcomes at posttreatment, see Rozental et al [8] and Shafran et al [9].

Results

Attrition and Adherence

There was no statistical difference between the 2 clinical trials in attrition, defined as the number of participants who were randomly assigned but did not complete the assessment at follow-up. In the Swedish trial, 49 (63%) completed the follow-up self-report measures, compared with 35 (57%) in the UK trial (χ^2_1=0.6, *P*=.45). None of the baseline characteristics were related to attrition in either of the 2 clinical trials. With regard to adherence, defined as the number of completed modules during the treatment period, the 2 clinical trials differed significantly, with a mean difference of 3.14 modules (95% CI 2.31-3.97), demonstrating that the participants in the Swedish trial completed more modules than those in the UK trial, on average.

Treatment Results

Table 4 shows the descriptive statistics for the primary and secondary outcome measures. The ANCOVAs for the intention-to-treat analyses revealed significant differences between pretreatment and follow-up for all of the primary and secondary outcome measures in both clinical trials. Table 5 shows within-group Cohen *d* effect sizes and their respective 95% CIs.

The completer analyses revealed significant differences for both clinical trials between pretreatment and follow-up, revealing a mean difference of 8.98 points (Swedish trial; 95% CI 7.07-10.89) and 10.35 points (UK trial; 95% CI 7.25-13.44) for the FMPS CM. Results were similar for the CPQ: 8.69 points (Swedish trial; 95% CI 6.61-10.77) and 11.10 (UK trial; 95% CI 9.14-13.07). For the secondary outcome measures, the mean differences in the Swedish trial were 3.57 points for the PHQ-9 (95% CI 2.28-4.86) and 3.22 points for the GAD-7 (95% CI 2.33-4.12). In the UK trial, the mean difference for the DASS was 8.78 points (95% CI 4.16-13.40).

Table 4. Observed and estimated scores for each primary outcome measure, by clinical trial, intention-to-treat analysis, and completer analysis. FMPS: Frost Multidimensional Perfectionism Scale, Concern over Mistakes subscale.

Measure and condition	Intention-to-treat analysis				Completer analysis			
	Pretreatment		Follow-up		Pretreatment		Follow-up	
	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n
Swedish trial[a]								
FMPS CM	33.42 (6.44)	78	25.14 (7.23)	78	33.42 (6.44)	78	23.61 (7.60)	49
Clinical Perfectionism Questionnaire	38.26 (4.63)	78	29.63 (8.00)	78	38.26 (4.63)	78	29.51 (6.70)	49
9-item Patient Health Questionnaire	9.59 (5.63)	78	6.45 (4.73)	78	9.59 (5.63)	78	5.47 (4.77)	49
7-item Generalized Anxiety Disorder scale	7.83 (4.85)	78	4.95 (3.72)	78	7.83 (4.85)	78	3.84 (3.41)	49
UK trial[b]								
FMPS CM	36.71 (4.42)	62	28.83 (7.80)	62	36.71 (4.42)	62	25.52 (8.07)	29
Clinical Perfectionism Questionnaire	35.69 (4.73)	62	27.25 (6.44)	62	35.69 (4.73)	62	24.55 (5.25)	29
21-item Depression Anxiety Stress Scale	26.31 (12.82)	62	19.89 (13.11)	62	26.31 (12.82)	62	15.93 (12.52)	27

[a]12-month follow-up.

[b]6-month follow-up.

Table 5. Within-group effect sizes, Cohen *d* (95% CI). FMPS: Frost Multidimensional Perfectionism Scale, Concern over Mistakes subscale.

Measure and condition	Intention-to-treat analysis	Completer analysis
	Cohen *d* (95% CI)	Cohen *d* (95% CI)
Swedish trial		
FMPS CM	1.21 (0.86-1.54)	1.42 (1.01-1.81)
Clinical Perfectionism Questionnaire	1.32 (0.97-1.66)	1.59 (1.17-1.98)
9-item Patient Health Questionnaire	0.60 (0.28-0.92)	0.77 (0.40-1.14)
7-item Generalized Anxiety Disorder scale	0.67 (0.34-0.99)	0.92 (0.54-1.29)
UK trial		
FMPS CM	1.24 (0.85-1.62)	1.92 (1.38-2.43)
Clinical Perfectionism Questionnaire	1.49 (1.09-1.88)	2.27 (1.70-2.80)
21-item Depression Anxiety Stress Scale	0.50 (0.13-0.85)	0.82 (0.34-1.28)

Improvement and Deterioration

Recovery on the FMPS CM was defined as those participants having a score at follow-up within 1 SD of the general population (<29; ie, clinically significant change) and exceeding the RCI. According to this definition, 29/49 (59%) participants in the Swedish trial and 15/35 (43%) in the UK trial met the criteria for recovery at follow-up. Improvement—that is, having a change score beyond the RCI at follow-up—was achieved by 31/49 (63%) in the Swedish trial and 18/35 (51%) in the UK trial. Meanwhile, 17/49 (35%) in the Swedish trial and 11/35 (31%) in the UK trial did not respond. In none of these cases was there a significant difference between the 2 clinical trials: χ^2_2 range 3.09 to 3.99, *P* value range .14 to .21. Only 1/84 (1%) participant deteriorated.

Discussion

Principal Findings

This study evaluated the long-term benefits of ICBT with guided self-help for perfectionism, indicating that the results at follow-up were similar to and possibly even somewhat improved from posttreatment. This suggests that ICBT with guidance from a therapist could help individuals manage and overcome their perfectionism in the long term. Compared with the findings of Lloyd et al [4], in this study the within-group effect sizes for perfectionism are similar, although that systematic review and meta-analysis included data only from pre- and posttreatment and not from follow-up. This is also true for depression and anxiety, with average Hedges *g* effect sizes being 0.64 (95% CI 0.35-0.92) and 0.52 (95% CI 0.23-0.81), respectively. In other words, the findings from our study are in line with previous research, suggesting that some long-term benefits can be achieved for anxiety, depression, and stress as well. Given that we could not combine the outcomes from the 2 clinical trials due to the differences in several key characteristics, most notably not implementing the same follow-up period, the result s should be interpreted with caution. Nevertheless, the results from this study are comparable with those of Egan et al [7] at 6-month follow-up, although, in that study, unguided self-help yielded much lower effects than face-to-face CBT. Therefore, the use of guided self-help when providing ICBT for perfectionism might be assumed to be better than unguided self-help for outcome. This idea was supported by a study on the differences between various levels of guidance by a therapist in ICBT for patients with depression [21] and, similarly, by a systematic review and meta-analysis of self-guided interventions for obsessive-compulsive disorder [22], indicating that more support yields larger effects overall. However, Titov et al [23] did not find such a difference in outcome, arguing that carefully controlled ICBT without any support can be just as beneficial. More research on the importance of guidance in relation to the treatment of perfectionism with ICBT is thus warranted, including an investigation of its influence on adherence.

The rates of recovery at follow-up, that is, those participants meeting the criteria of moving from a dysfunctional to a functional distribution in terms of perfectionism (clinically significant change and exceeding the RCI) were 59% (FMPS CM) in the Swedish trial and 43% (FMPS CM) in the UK trial. This can be compared with the results of Egan et al [7] of 67% for face-to-face CBT and 40% for unguided self-help. However, given that, to our knowledge, no other clinical trial has investigated recovery at follow-up, these numbers should be interpreted cautiously and warrant replication. Furthermore, given the high rate of attrition in this study, the estimates of recovery might be unreliable because we derived them from those participants completing the follow-up assessment, possibly inflating the actual rates. In addition, it is unclear whether the cutoff for determining clinically significant change (<29 on the FMPS CM [18]) is sensitive enough to accurately differentiate those belonging to a clinical population from those belonging to a nonclinical population. In this sense, recovery in this study should primarily be regarded as reaching a statistical criterion, in line with the recommendations by Jacobson and Truax [24], rather than true recovery in terms of no longer fulfilling the criteria for a psychiatric disorder. Whether participants are in fact recovered from perfectionism is, however, an issue that requires both more empirical data and a better conceptual idea of where the dysfunctional and functional distributions meet. Nevertheless, the recovery rates indicate that ICBT works well and does have an impact on perfectionism that should be clinically relevant.

Study Limitations

This study makes an important contribution to the research on the treatment of perfectionism, as it is one of few studies that included follow-up data and the only one to date, to our knowledge, with follow-up at 12 months. There are, however, also limitations that need to be addressed. First, issues related to the design limit the conclusions that can be made. The 2 clinical trials differed on some key characteristics, including different length of the treatment period (8 vs 12 weeks), therapist experience, therapist supervision, and, especially, the timing of the follow-up assessment. This was primarily due to different conventions and logistical issues among the researchers involved, meaning that the samples could not be combined and thus limit power. Nevertheless, given the many similarities between them in terms of treatment content and delivery, the findings are arguably relevant to present together. Additionally, there was no comparison group. However, replicating this study is warranted, possibly by recruiting a larger and more heterogeneous sample in the context of a randomized controlled trial.

Second, attrition was high in both clinical trials, with only 63% (Swedish trial) and 57% (UK trial) completing the self-report measures at 6- and 12-month follow-up, respectively, potentially affecting the conclusions that can be drawn. This can be compared with a study on ICBT for procrastination [25], which had 32% attrition at 12-month follow-up, suggesting that our clinical trials both had higher rates of dropouts. In addition, an individual patient data meta-analysis of 10 studies on ICBT for depression indicated that 40% dropped out before completing one-fourth of the modules [26]. Multiple imputation was used to account for missing values, given the indication that data were missing at random. In our study, none of the baseline characteristics were associated with attrition, but this does not preclude other variables that we did not explore from being related to completing the self-report measures at follow-up. Preventing attrition is thus important, and future research should try to implement ways of improving these rates, perhaps by adding more support or the use of tailored modules [27]. Additional issues related to the different attrition rates in the 2 clinical trials have also been addressed by Shafran et al [9], such as not explicitly asking the participants to confirm their participation and the absence of a telephone interview assessment in the UK trial, aspects that may have to be addressed to a greater extent in future studies.

Third, perfectionism is not a psychiatric disorder in itself [1]. This complicates the issue of assessing eligibility to participate in a clinical trial, but also of determining whether it is a condition that actually requires a stand-alone treatment like the one provided. However, given the close connection with eating disorders, depression, and anxiety disorders [3], it is reasonable to assume that elevated perfectionism is a common denominator for many psychiatric disorders. Hence, interventions targeting its mechanisms should be beneficial for many individuals by providing a more transdiagnostic approach. This is supported by the finding that benefits can be observed on many different outcomes in clinical trials of perfectionism, such as depression [7-9]. Still, more research needs to be done to investigate whether a transdiagnostic approach adds something to a disorder-specific treatment, for instance comparing their efficacy in a head-to-head comparison for a particular psychiatric disorder.

Fourth, the participants in this study were all self-referred, and although they scored high on perfectionism at pretreatment, they may not be representative of individuals with this problem in general. Arnberg et al [28] stressed that most individuals receiving help via ICBT have a high educational level and are more likely to be women and of a particular socioeconomic group, possibly limiting the generalizability of the results. However, Titov et al [29] compared the baseline characteristics of individuals receiving ICBT with both those receiving CBT at an outpatient clinic and those taking an epidemiological survey, noting that the differences were small and not necessarily important. In comparison with the general population, those receiving treatment, regardless of format, had higher severity levels overall, but ICBT and CBT did not differ from each other. Also, with the only exceptions of age, sex, and marital status, such aspects as educational level and employment status were not different between ICBT and CBT treatment groups, suggesting that there may not be anything specific to receiving treatment via the internet with regard to those who seek help in general. Further research should nonetheless be performed on recruitment and diversity, preferably by limiting the number of inclusion and exclusion criteria and by reaching out to a more heterogeneous sample.

Conclusion

We examined the long-term benefits of ICBT with guided self-help for perfectionism, depression, and anxiety. The results at follow-up were comparable with posttreatment assessment, obtaining medium to large within-group effect sizes. The results from these 2 different cases of ICBT with guided self-help are thus promising but warrant replication using a larger and more heterogeneous sample.

Acknowledgments

The authors would like to acknowledge Linköping University and a Professor's grant to the Principal Investigator (GA), which made the completion of the study possible. The study was also supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. Further, we thank George Vlaescu for his outstanding webmaster service, and Chloe Shu for providing the basis for the original website for the study.

Conflicts of Interest

None declared.

References

1. Hewitt P, Flett G, Mikail S. Perfectionism: A Relational Approach to Conceptualization, Assessment, and Treatment. New York, NY: Guilford Publications; 2017.
2. Limburg K, Watson HJ, Hagger MS, Egan SJ. The relationship between perfectionism and psychopathology: a meta-analysis. J Clin Psychol 2017 Oct;73(10):1301-1326. [doi: 10.1002/jclp.22435] [Medline: 28026869]
3. Egan SJ, Wade TD, Shafran R. Perfectionism as a transdiagnostic process: a clinical review. Clin Psychol Rev 2011 Mar;31(2):203-212. [doi: 10.1016/j.cpr.2010.04.009] [Medline: 20488598]
4. Lloyd S, Schmidt U, Khondoker M, Tchanturia K. Can psychological interventions reduce perfectionism? A systematic review and meta-analysis. Behav Cogn Psychother 2015 Nov;43(6):705-731. [doi: 10.1017/S1352465814000162] [Medline: 26393777]
5. Frost RO, Marten P, Lahart C, Rosenblate R. The dimensions of perfectionism. Cogn Ther Res 1990 Oct;14(5):449-468. [doi: 10.1007/Bf01172967]
6. Fairburn C, Cooper Z, Shafran R. Clinical Perfectionism Questionnaire. Oxford, UK: Department of Psychiatry, University of Oxford; 2003.
7. Egan SJ, van Noort E, Chee A, Kane RT, Hoiles KJ, Shafran R, et al. A randomised controlled trial of face to face versus pure online self-help cognitive behavioural treatment for perfectionism. Behav Res Ther 2014 Dec;63:107-113. [doi: 10.1016/j.brat.2014.09.009] [Medline: 25461785]
8. Rozental A, Shafran R, Wade T, Egan S, Nordgren LB, Carlbring P, et al. A randomized controlled trial of internet-based cognitive behavior therapy for perfectionism including an investigation of outcome predictors. Behav Res Ther 2017 Aug;95:79-86. [doi: 10.1016/j.brat.2017.05.015] [Medline: 28586713]
9. Shafran R, Wade TD, Egan SJ, Kothari R, Allcott-Watson H, Carlbring P, et al. Is the devil in the detail? A randomised controlled trial of guided internet-based CBT for perfectionism. Behav Res Ther 2017 Aug;95:99-106. [doi: 10.1016/j.brat.2017.05.014] [Medline: 28618299]
10. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Intervent 2016 Nov;6:107-114. [doi: 10.1016/j.invent.2016.09.006]
11. Egan S, Wade T, Shafran R, Antony M. Cognitive-Behavioral Treatment of Perfectionism. New York, NY: Guilford Publications; 2014.
12. Kothari R, Egan S, Wade T, Andersson G, Shafran R. Overcoming perfectionism: protocol of a randomized controlled trial of an internet-based guided self-help cognitive behavioral therapy intervention. JMIR Res Protoc 2016 Nov 11;5(4):e215 [FREE Full text] [doi: 10.2196/resprot.6378] [Medline: 27836815]
13. 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 Jul;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8] [Medline: 15183601]
14. Dear BF, Titov N, Sunderland M, McMillan D, Anderson T, Lorian C, et al. Psychometric comparison of the generalized anxiety disorder scale-7 and the Penn State Worry Questionnaire for measuring response during treatment of generalised anxiety disorder. Cogn Behav Ther 2011;40(3):216-227. [doi: 10.1080/16506073.2011.582138] [Medline: 21770844]
15. Lovibond S, Lovibond P. Manual for the Depression Anxiety Stress Scales. Sydney, Australia: Psychology Foundation of Australia; 1996.
16. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
17. Cumming G. The new statistics: why and how. Psychol Sci 2014 Jan;25(1):7-29. [doi: 10.1177/0956797613504966] [Medline: 24220629]
18. Suddarth B, Slaney R. An investigation of the dimensions of perfectionism in college students. Meas Eval Couns Dev 2001;34(3):157-165.
19. Evans C, Margison F, Barkham M. The contribution of reliable and clinically significant change methods to evidence-based mental health. Evid Based Ment Health 1998 Aug 01;1(3):70-72. [doi: 10.1136/ebmh.1.3.70]
20. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of Internet interventions. Internet Intervent 2014 Mar;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]
21. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
22. Pearcy CP, Anderson RA, Egan SJ, Rees CS. A systematic review and meta-analysis of self-help therapeutic interventions for obsessive-compulsive disorder: is therapeutic contact key to overall improvement? J Behav Ther Exp Psychiatry 2016 Jun;51:74-83. [doi: 10.1016/j.jbtep.2015.12.007] [Medline: 26794856]

23. Titov N, Fogliati VJ, Staples LG, Gandy M, Johnston L, Wootton B, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided. self-guided internet-delivered cognitive-behavioural therapy. BJPsych Open 2016 Jan;2(1):50-58 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002139] [Medline: 27703754]
24. 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. [Medline: 2002127]
25. Rozental A, Forsell E, Svensson A, Andersson G, Carlbring P. Overcoming procrastination: one-year follow-up and predictors of change in a randomized controlled trial of Internet-based cognitive behavior therapy. Cogn Behav Ther 2017 Apr;46(3):177-195. [doi: 10.1080/16506073.2016.1236287] [Medline: 27682415]
26. 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]
27. Nordgren LB, Hedman E, Etienne J, Bodin J, Kadowaki A, Eriksson S, et al. Effectiveness and cost-effectiveness of individually tailored Internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: a randomized controlled trial. Behav Res Ther 2014 Aug;59:1-11 [FREE Full text] [doi: 10.1016/j.brat.2014.05.007] [Medline: 24933451]
28. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
29. Titov N, Andrews G, Kemp A, Robinson E. Characteristics of adults with anxiety or depression treated at an internet clinic: comparison with a national survey and an outpatient clinic. PLoS One 2010;5(5):e10885 [FREE Full text] [doi: 10.1371/journal.pone.0010885] [Medline: 20526371]

Abbreviations

ANCOVA: analysis of covariance
CBT: cognitive behavioral therapy
CPQ: Clinical Perfectionism Questionnaire
DASS-21: 21-item Depression Anxiety Stress Scale
FMPS CM: Frost Multidimensional Perfectionism Scale, Concern over Mistakes subscale
ICBT: internet-based cognitive behavioral therapy
PHQ-9: 9-item Patient Health Questionnaire
RCI: Reliable Change Index

Edited by G Eysenbach; submitted 17.01.18; peer-reviewed by J Apolinário-Hagen, A Basu; comments to author 22.02.18; revised version received 06.03.18; accepted 07.03.18; published 26.04.18.

Please cite as:
Rozental A, Shafran R, Wade TD, Kothari R, Egan SJ, Ekberg L, Wiss M, Carlbring P, Andersson G
Guided Web-Based Cognitive Behavior Therapy for Perfectionism: Results From Two Different Randomized Controlled Trials
J Med Internet Res 2018;20(4):e154
URL: http://www.jmir.org/2018/4/e154/
doi:10.2196/jmir.9823
PMID:29699968

Original Paper

A Single-Session, Web-Based Parenting Intervention to Prevent Adolescent Depression and Anxiety Disorders: Randomized Controlled Trial

Mairead C Cardamone-Breen[1], BPsySc (Hons); Anthony F Jorm[2], DSc, PhD; Katherine A Lawrence[1], DPsych (Clin); Ronald M Rapee[3], PhD; Andrew J Mackinnon[4], PhD; Marie Bee Hui Yap[1,2], MPsych (Clin), PhD

[1]Monash Institute of Cognitive and Clinical Neurosciences, School of Psychological Sciences, Monash University, Melbourne, Australia

[2]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

[3]Centre for Emotional Health, Macquarie University, Sydney, Australia

[4]Black Dog Institute, University of New South Wales, Sydney, Australia

Corresponding Author:
Marie Bee Hui Yap, MPsych (Clin), PhD
Monash Institute of Cognitive and Clinical Neurosciences
School of Psychological Sciences
Monash University
18 Innovation Walk
Clayton
Melbourne, 3800
Australia
Phone: 61 399050723
Email: marie.yap@monash.edu

Abstract

Background: Depression and anxiety disorders are significant contributors to burden of disease in young people, highlighting the need to focus preventive efforts early in life. Despite substantial evidence for the role of parents in the prevention of adolescent depression and anxiety disorders, there remains a need for translation of this evidence into preventive parenting interventions. To address this gap, we developed a single-session, Web-based, tailored psychoeducation intervention that aims to improve parenting practices known to influence the development of adolescent depression and anxiety disorders.

Objective: The aim of this study was to evaluate the short-term effects of the intervention on parenting risk and protective factors and symptoms of depression and anxiety in adolescent participants.

Methods: We conducted a single-blind, parallel group, superiority randomized controlled trial comparing the intervention with a 3-month waitlist control. The intervention is fully automated and consists of two components: (1) completion of an online self-assessment of current parenting practices against evidence-based parenting recommendations for the prevention of adolescent depression and anxiety disorders and (2) an individually tailored feedback report highlighting each parent's strengths and areas for improvement based on responses to the self-assessment. A community sample of 349 parents, together with 327 adolescents (aged 12-15 years), were randomized to either the intervention or waitlist control condition. Parents and adolescents completed online self-reported assessments of parenting and adolescent symptoms of depression and anxiety at baseline, 1-month (parent-report of parenting only), and 3-month follow-up.

Results: Compared with controls, intervention group parents showed significantly greater improvement in parenting risk and protective factors from baseline to 1-month and 3-month follow-up ($F_{2,331.22}$=16.36, P<.001), with a small to medium effect size at 3-month follow-up (d=0.33). There were no significant effects of the intervention on adolescent-report of parenting or symptoms of depression or anxiety in the adolescents (all P>.05).

Conclusions: Findings suggest that a single-session, individually tailored, Web-based parenting intervention can improve parenting factors that are known to influence the development of depression and anxiety in adolescents. However, our results do not support the effectiveness of the intervention in improving adolescent depression or anxiety symptoms in the short-term. Long-term studies are required to adequately assess the relationship between improving parenting factors and adolescent depression

and anxiety outcomes. Nonetheless, this is a promising avenue for the translation of research into a low-cost, sustainable, universal prevention approach.

Trial Registration: Australian New Zealand Clinical Trials Registry: ACTRN12615000247572; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12615000247572 (Archived by WebCite at http://www.webcitation.org/6v1ha19XG)

(J Med Internet Res 2018;20(4):e148) doi:10.2196/jmir.9499

KEYWORDS

adolescent; mental health; depression, anxiety, parenting; family; preventive health services; Internet

Introduction

Background

Depression and anxiety disorders are among the leading contributors to global burden of disease [1]. Importantly, these disorders have peak onset early in life, accounting for the greatest proportion of disability in young people (aged 12-24 years [2-4]). Early onset of these disorders is associated with deleterious sequelae across the lifespan, leading to substantial cost at individual, social, and economic levels [4,5]. Preventive efforts targeting these disorders early in life are therefore a global priority [4]. Given that the onset of these disorders peaks during adolescence, this is an ideal period to target such preventive efforts [4].

Due to their frequent comorbidity, overlap of symptoms, and shared aetiological factors, depression and anxiety disorders are often grouped under an *internalizing* cluster [6-8]. Transdiagnostic approaches to prevention may therefore enhance the efficacy and cost-effectiveness of preventive programs [6,8,9]. In particular, a growing body of evidence highlights the important role of families in the prevention of internalizing problems in children and adolescents. Many risk factors for depression and anxiety involve the family environment (eg, interparental conflict and family connectedness [9-11]) or parenting (eg, parental warmth, aversiveness, and over-involvement [9,12-14]). Other factors can be detected and responded to by parents (eg, negative affectivity, coping style, behavioral inhibition, and excessive reassurance seeking [8,15-17]), or are directly influenced or modeled by parents (eg, parental response to child emotions and modeling or reinforcement of anxious behaviours [18-21]). As these factors are potentially within parents' control, they are amenable to preventive intervention. Furthermore, parents are a strategic target for preventive approaches for several reasons: most adolescents live with their parents, increasing proximity and exposure to preventive strategies; parents are intrinsically motivated to take actions to promote their child's health [21]; and parents may have the foresight to appreciate the value of prevention [21]. Given these factors, it is unsurprising that the role of family and parenting in the promotion of youth mental health has been recognized as a key research translation priority [4,14,22,23].

Promisingly, existing research demonstrates that depression and anxiety disorders in children and adolescents can be prevented. A number of recent systematic reviews and meta-analyses support the efficacy of existing preventive programs (eg, [23,24-28]). Most of these programs are delivered directly to the child (eg, through schools), often utilizing a cognitive behavioral approach and delivered face-to-face by trained professionals (eg, [25,26-28]). In programs that do include a parent component, this often involves teaching parents how to understand and implement the content that is delivered to the child, rather than targeting specific parenting factors. Some exceptions to this include programs that target family conflict, parental over-involvement, or parent-child relationship as part of broader cognitive behavioral or psychoeducational prevention programs for parents of children at risk of anxiety or depression [29-34]. One program that aims to improve parent emotion socialization practices has also shown improvements in family conflict and internalizing symptoms in adolescents [35,36].

Additionally, a recent meta-analysis of preventive parenting interventions (ie, where the parent received >50% of the intervention) for internalizing symptoms in children and adolescents (aged 0-18 years) found lasting preventive effects from 6 months to 11 years postintervention [37]. Notably, of the 51 studies included in the review, only 3 targeted parents of adolescents. Thus, in comparison with programs for parents of younger children, there is a lack of preventive parenting interventions for internalizing disorders in adolescents. Although the amount of variance explained by individual parenting or family factors is small [9,38-40], targeting multiple factors in one intervention may result in larger effect sizes. Furthermore, preventive parenting programs could be used alongside existing preventive interventions (eg, school-based programs delivered to the child), thereby increasing the number of risk and protective factors targeted for each child or family.

Existing parenting interventions face many challenges in engaging parents because of barriers such as time and scheduling constraints, geographical distance, childcare provision, and financial cost [41-43]. In addition, mental health interventions are often associated with added concerns about privacy, the perception of being a "bad" parent, and stigma [44-46]. Online delivery is one potential strategy to overcome some of these barriers. Web-based programs reduce geographic and time constraint barriers, can be accessed privately and anonymously, and can be disseminated widely at low cost. Furthermore, research shows that parents are already seeking information about both parenting and mental health online [47-50]; therefore, this is an ideal space for delivery of preventive parenting programs. Moreover, there is evidence for the efficacy of online interventions for the treatment (eg, [51,52]) and prevention (eg,

[53,54]) of depression and anxiety, as well as the delivery of parenting programs (eg, [31,55-57]).

The Partners in Parenting Intervention

To address some of the gaps discussed above, our team recently developed the Partners in Parenting (PiP) intervention—a Web-based, multi-level public health approach that aims to support parents in prevention and early intervention for adolescent depression and anxiety disorders (see [58] for further details). The PiP intervention comprises three components: (1) a parenting self-assessment that assesses current parenting practices against evidence-based parenting recommendations for the prevention of adolescent depression and anxiety [59]; (2) an individually tailored feedback report based on each parent's responses to the self-assessment, highlighting parenting strengths and areas for improvement; and (3) a set of interactive modules to support parents in applying the parenting recommendations. The development of PiP was guided by the persuasive systems design (PSD) model that aims to use technology to promote behavior change through principles such as tailoring, personalization, and suggestion [60]. The content of all components of PiP is based on the evidence-based parenting guidelines *How to Prevent Depression and Clinical Anxiety in Your Teenager: Strategies for Parents* (referred to henceforth as *the Guidelines* [61]). The Guidelines represent the synthesis of high-quality research evidence and international expert consensus. They were developed via a systematic review and meta-analysis of parenting risk and protective factors for adolescent depression and anxiety, published by researchers internationally [9], followed by a Delphi international expert consensus study of parenting strategies important for the prevention of depression and anxiety in adolescents [21]. The core content of the Guidelines (and subsequently PiP) should therefore be applicable to an international audience. However, as we initially developed PiP to trial with an Australian population, some aspects of the language and examples provided may need to be adapted for use in other English-speaking countries (eg, specific terms, types of health professionals, and links to additional local resources).

Preventive interventions can be universal (ie, delivered to all individuals regardless of level of risk), selective (ie, targeting individuals with known risk factors), or indicated (ie, targeting individuals showing signs or symptoms [62]). The PiP multi-level intervention is designed to support parents across the prevention continuum [58]. The first two components (parenting self-assessment and personalized feedback report) can serve as a stand-alone single-session intervention, which is most appropriate as a universal prevention approach [58]. This is likely to be an acceptable level of intervention for parents who are motivated, educated, and whose child does not have existing mental health problems [58]. The brevity of the single-session intervention may be particularly appealing to parents who are not willing or able to commit to the more intensive program (ie, series of interactive modules). As well as the Web programming ensuring intervention fidelity, the single-session intervention increases the likelihood of users achieving 100% adherence compared with interventions that require participants to return for multiple sessions. This is a notable advantage over many existing online interventions where adherence is often low [63,64]. Within PiP, parents can "step up" to the next level of intervention (interactive modules) based on need (ie, selective or indicated prevention) or parent preference (ie, parents who do not find a single-session intervention sufficient). The single-session intervention is also designed to serve as a prompt for parents to seek further assistance if required (eg, mental health support for themselves or their child or the PiP interactive modules).

Evaluation of the full PiP intervention is currently underway. Three-month postintervention results demonstrate that PiP can improve parenting risk and protective factors compared with an active control condition [65]. This paper evaluates the effectiveness of the single-session component of PiP as a stand-alone intervention. This is important given the proposed multi-level model (ie, some parents will only complete the single-session component), as well as the high attrition rates typically observed in online intervention trials (eg, [63,66,67]). As this is a preliminary trial of the newly developed intervention, and to align with a concurrent randomized controlled trial (RCT) of the full PiP intervention, this paper only examines the short-term (3-month) effects of the intervention. To our knowledge, the intervention evaluated in this paper is the first single-session, Web-based, tailored psychoeducation intervention targeting the wide range of parenting risk and protective factors for the development of depression and anxiety in adolescents.

Study Aims and Hypotheses

The primary aim of this study was to evaluate the short-term effect of the intervention on evidence-based parenting risk and protective factors for adolescent depression and anxiety disorders, as assessed by parent-reported concordance with the Guidelines. The secondary aims of the study were to examine the effects of the intervention on parent and child reported symptoms of depression and anxiety in adolescent participants and adolescent report of parenting. To do this, we conducted an RCT comparing the intervention with a waitlist control condition. We hypothesized that compared with waitlist controls, intervention group parents would show a greater increase in concordance with the Guidelines from baseline to postintervention and follow-up. We also predicted a greater reduction in symptoms of adolescent depression and anxiety in the intervention compared with the control group.

Methods

Study Design and Participants

The trial was a single-blind, parallel-group, superiority RCT with two conditions: intervention and 3-month waitlist control. Assessments took place at baseline (pre-intervention), 1-month postintervention, and 3-month follow-up, with data collection from April 2015 to November 2016. Participants were parents or primary caregivers of at least one adolescent aged 12 to 15 years, who resided in Australia, had regular Internet access, and had an email account. Computer and Internet literacy was implicit in the eligibility criteria and registration process. Only one parent and one adolescent from each family was allowed to participate. No other exclusion criteria were specified. Parents were able to participate if their adolescent declined participation;

therefore, not all parent participants had an associated adolescent participant. Recruitment was primarily via secondary schools across Australia, as well as online networks and social media. Schools were requested to distribute recruitment materials (flyers and participant explanatory statements) via their usual methods of communication with parents. This predominantly involved school newsletters (electronic and hard copy), online parent portals, and email communication. Hard copy flyers were also made available (eg, at parent information evenings). Other recruitment methods included an email to the mailing list of Mental Health First Aid Australia, a social media (Facebook and twitter) post by *beyondblue* (the Australian national depression and anxiety initiative), and an advertisement in an Australian parenting magazine (*Exploring Teens*, electronic and hard copy versions). A power analysis (conducted using Stata-based software [68]) indicated a required sample size of 294 parents and adolescents to detect a small effect size (Cohen *d*=0.20) with power of .80 and alpha of .05 for the primary outcome. To allow for approximately 15% attrition, we aimed to recruit 340 parents and adolescents. Our final sample comprised 349 parents together with 327 adolescents at randomization. The trial was approved by the Monash University Human Research Ethics Committee (approval number CF14/3886-2014002023) and prospectively registered with the Australian and New Zealand Clinical Trials Registry (registration number ACTRN12615000247572).

Intervention

As discussed, the single-session intervention evaluated in this paper is the first component of the multi-level PiP program [58]. The intervention provides individually tailored psychoeducation to each parent. Parents first complete an online parenting scale that assesses their current parenting practices and beliefs against the recommendations in the Guidelines (the Parenting to Reduce Adolescent Depression and Anxiety Scale, PRADAS [59]). On the basis of their responses to the PRADAS, each parent receives an individually tailored feedback report highlighting parenting strengths and areas for improvement. The feedback report covers the nine domains of parenting in the Guidelines (see Table 1) and recommends specific actions parents could take to improve their concordance with the Guidelines (see Multimedia Appendix 1 for screenshots of the intervention). Feedback messages are designed to be brief, and links are provided for parents to seek further information if they wish. The intervention is fully automated and designed to tailor the content of the Guidelines for each parent, according to principles of the PSD model [56]. For example, feedback messages suggest specific actionable strategies that parents could implement (suggestion and tuneling principles), the content of the Guidelines is reduced to a shorter feedback report covering the areas deemed most relevant to each parent (reduction principle), feedback messages are tailored to each parent to increase relevance of the recommendations (tailoring principle), each feedback section includes praise for areas of strength (praise principle), and both the PRADAS and feedback report are personalized with the adolescent's name and gender (personalization principle).

The development of the intervention included consultation with a reference group of parents of adolescents to ensure acceptability to target end users. This involved conducting three 2-hour workshops in which 22 parents of adolescents aged 11 to 18 years (19 mothers, 3 fathers; n=7-8 per workshop) were shown a draft of the PRADAS and feedback messages. Parents completed the PRADAS for themselves and read sample feedback reports before participating in facilitated discussion regarding ways to improve the intervention. Feedback from parents was incorporated into the final version of the intervention. This included changes to the language used, content of the feedback report, additional practical strategies or ideas on how to implement these, and the addition of a section at the beginning of the feedback report suggesting how parents may wish to work through the information provided.

Measures

Primary Outcome Measure: The Parenting to Reduce Adolescent Depression and Anxiety Scale

The PRADAS is a self-report, criterion-referenced measure of parental concordance with the Guidelines [59]. The scale comprises a total of 73 items across eight subscales (parent-child relationship, involvement, family rules, home environment, health habits, dealing with problems, coping with anxiety, and professional help-seeking). One of the original nine subscales (relationships with others) was removed from the final version of the PRADAS because of its unsatisfactory psychometric properties. Most items assess specific recommendations in the Guidelines scored on a Likert-type frequency scale (never, rarely, sometimes, and often). Being a criterion-referenced measure, each item has a cut-off score for mastery, with items scored as either concordant (1) or nonconcordant (0) with the Guidelines. Items can be summed to form eight subscale scores and a total score, ranging from 0 to 73 (higher scores indicate greater concordance with the Guidelines). The total score has demonstrated high reliability, as measured by the agreement coefficient (.97), and acceptable one-month test-retest reliability (.78). It has also shown convergent validity with two other parenting measures and a small association with adolescent depression and anxiety symptoms [59]. We utilized the total score as it has the strongest psychometric properties. Reliability was high in our sample (agreement coefficient .97).

Secondary Outcome Measures

Parenting to Reduce Adolescent Depression and Anxiety Scale-Adolescent Report

The PRADAS-Adolescent report (PRADAS-A) was developed by our team to assess the adolescent's perspective on the parenting factors assessed by the PRADAS. The original PRADAS items were reworded to reflect the adolescent's perspective at a developmentally appropriate level. Items not applicable to the adolescent (eg, parental knowledge, attitudes, beliefs, or hypothetical questions) were not included. The PRADAS-A comprises 43 items across the same eight subscales as the PRADAS. As with the PRADAS, one original subscale (relationships with others) was removed from the final version of the scale. Response options and scoring are similar to the PRADAS; most items are on a Likert-type frequency scale (never, rarely, sometimes, and often) and are scored as either concordant (1) or nonconcordant (0) with the Guidelines. Higher scores indicate greater concordance with the Guidelines.

Reliability has been shown to be high for the total score (agreement coefficient=.81) in a community sample of 670 adolescents aged 12 to 15 years [69].

Table 1. Parenting domains covered in the intervention, corresponding Guidelines topics, parenting risk or protective factors addressed, and example parenting recommendations.

Intervention domain	Guidelines topic	Risk or protective factors covered	Example recommended parenting strategy
Your relationship with [Child][a]	Establish and maintain a good relationship with your teenager	Parental warmth, aversiveness, affection, emotional availability	Making time each day to ask [Child] about [his/her] day and what [he/she] has been doing, regardless of [his/her] response.
Your involvement in [Child]'s life	Be involved and support increasing autonomy	Parental over-involvement, autonomy granting, monitoring	Gradually increasing [Child]'s responsibilities and independence over time to allow [him/her] to mature.
[Child]'s relationships with others	Encourage supportive relationships	Parental encouragement of sociability	Take some time to talk through any social problems [Child] may have.
Your family rules	Establish family rules and consequences	Consistency of discipline	Noticing when [Child] behaves well, and rewarding [him/her] with positive consequences (eg, praise or privileges).
Your home environment	Minimize conflict in the home	Interparental conflict, parent-child conflict management, criticism, parental modeling of conflict management	Try not to argue with your partner if [Child] can hear. Frequent and intense conflict between parents increases a teenager's risk of depression and clinical anxiety.
Health habits	Encourage good health habits	Diet, physical activity, sleep hygiene (7 items); responding to alcohol or drug use (5 items)	Set an example for [Child] by having good health habits (ie, healthy diet, regular exercise, and responsible use of alcohol) yourself.
Dealing with problems in [Child]'s life	Help your teenager to deal with problems	Problem solving, emotion regulation, stress management, modeling of problem solving approaches	When talking with [Child] about problems that [he/she] has dealt with, recognize and praise [his/her] problem-solving efforts (ie, what [he/she] did well when trying to solve the problem) rather than focusing on the outcome [he/she] achieved.
Coping with anxiety	Help your teenager to deal with anxiety	Anxiety management (avoidance, exposure), modeling of anxiety, management strategies	Try not to step in to help [Child] at the first sign of any stress or anxiety, as the way you respond to [Child]'s anxiety may unintentionally increase [his/her] anxiety. Instead, let [him/her] try to manage the situation [himself/herself] and provide help if [he/she] asks you to or if the anxiety persists.
Getting help with needed	Encourage professional help-seeking when needed	Professional help-seeking knowledge and behaviors (parent and child)	If you do notice a persistent change in [Child]'s mood or behavior: try to determine whether the change in mood or behavior is caused by a temporary situation or a more ongoing problem.

[a]Square brackets denote personalization with the adolescent's name and gender.

The scale also has good 3-month test-retest reliability (.81) and moderate correlations with child-report symptom measures of depression and anxiety (r=.45 and .31, respectively [69]). The correlation between the PRADAS and PRADAS-A in the community sample was significant but small (r=.25; [69]). In the current sample, reliability of the total score was high (agreement coefficient=.81), and test-retest reliability (waitlist control group, baseline to 3 months) was good (.81). The correlation between PRADAS and PRADAS-A at baseline was .25 (P<.001).

Spence Children's Anxiety Scale

The Spence Children's Anxiety Scale (SCAS) is a widely used child- and parent-report measure of child anxiety [70,71]. The SCAS-Child report (SCAS-C) comprises 45 items, including six nonscored filler items, whereas the SCAS-Parent report (SCAS-P) has a total of 39 items. Items examine the degree to which the child experiences specific anxiety symptoms on a 4-point frequency scale (never, sometimes, often, and always). Items are scored from 0 (never) to 3 (always) and can be summed to form six subscale scores and a total anxiety score (ranging from 0-114, with higher scores indicating greater anxiety symptomology). We utilized the total anxiety score, which had high reliability in our sample, as measured by coefficient omega (a less biased reliability index than Cronbach alpha [72,73]; SCAS-C: omega=95; SCAS-P: omega=.93). The correlation between baseline SCAS-C and SCAS-P total scores in the current sample was .44 (P<.001).

Short Mood and Feelings Questionnaire

The Short Mood and Feelings Questionnaire (SMFQ) is a Child-report (SMFQ-C) and Parent-report (SMFQ-P) measure

of depressive symptoms in children and adolescents [74]. The 13 items assess frequency of depressive symptoms in the previous 2 weeks on a 3-point scale of not true (0), sometimes true (1), or true (2). Items are summed to form a total score (ranging from 0-26, with higher scores indicating higher symptom levels). Reliability was high in our sample (omega=.92 for SMFQ-C and .91 for SMFQ-P). The correlation between baseline SMFQ-C and SMFQ-P scores was .47 (P<.001).

Process Evaluation: Parent Use and Satisfaction With the Intervention

Parents in both groups were asked five questions immediately postintervention (ie, after receiving their feedback at baseline or 3-month follow-up for the intervention and control groups, respectively). The questions were scored on Likert-type scales (see Multimedia Appendix 2) and assessed: (1) how much of the feedback was read, (2) satisfaction with the feedback, (3) perceived usefulness of the feedback, (4) intention to change based on the feedback provided, and (5) confidence in ability to implement the recommended changes. Additionally, at 1-month and 3-month follow-up, parents were asked whether they had attempted to make changes to their parenting since the previous assessment and their perceived success in making these changes. Parents were also asked if they had accessed any additional parenting resources or sought professional help for their own or their child's mental health since the last assessment.

Procedure

Registration and Consent

Parents self-selected by responding to advertisements and registering themselves and their adolescent via the publicly accessible trial website. Parents created an account with their email address and self-selected password and were required to verify their account via an account activation link sent to their email. Parents provided online consent and contact details for the adolescent; however, they were informed that they could still participate if their adolescent declined. Adolescents were then contacted by phone to explain the study requirements and obtain verbal assent. Adolescents were informed that their decision to participate or not would not affect their parent's participation. Multimedia Appendices 3-5 present the participant informed consent information.

Baseline Assessments

All assessments were completed online via a dedicated trial website, and participants were required to log in with their username and password. Adolescents were guided through completion of their baseline assessment over the phone, with assistance provided by a member of the research team as necessary. The adolescent baseline assessment included the PRADAS-A, the SCAS-C, and the SMFQ-C. Adolescents were reimbursed with a $10 AUD e-voucher for completion of the assessment. Submission of the adolescent baseline assessment triggered an automated email sent to parent participants containing the link to their baseline assessment, which included the PRADAS, the SCAS-P, and the SMFQ-P. In cases where the adolescent did not participate, the research team cancelled the adolescent assessment, which triggered the automated email to be sent to the parent participant.

Randomization and Blinding

Upon completion of the parent baseline assessment, parents were randomly allocated to the intervention or control group via a computer-generated unblocked, unstratified randomization procedure on a 1:1 ratio. Parents were not blinded to their allocation. As assessments were conducted entirely online, assessor was not relevant. Researchers who phoned adolescent participants were blinded to allocation.

Intervention

Immediately following completion of the baseline assessment, intervention group parents were shown their individually tailored feedback on screen. They were also emailed the feedback report and a copy of the Guidelines in PDF format. Parents in the waitlist control group were informed via a website message that they would receive the intervention in 3 months' time.

Follow-Up Assessments

Follow-up assessments were conducted with parents at 1-month postintervention and with both parents and adolescents at 3-months postintervention. Parents were sent an email containing the link to their online assessment, and adolescents were contacted via phone to guide them through completion of the assessment. For all assessments, parents who had not completed their assessment received reminder emails 7 and 14 days following the initial invitation and a phone call or text message 21 days after initial invitation.

One-Month (Postintervention) Follow-Up

Thirty days after baseline, parents were sent the invitation to complete their postintervention assessment that included the PRADAS to assess for changes in parenting. One month was chosen to allow sufficient time to read the feedback and attempt to implement the recommended changes. Mean duration to follow-up was 41.47 days (SD 10.76).

Three-Month Follow-Up

At 3-month follow-up, both parents and adolescents completed their respective versions of the PRADAS, SCAS, and SMFQ. Parents in the waitlist control group received their individually tailored feedback report and a copy of the Guidelines immediately following submission of their 3-month follow-up assessment. Parents and adolescents were reimbursed with a $10 AUD e-voucher each on completion of the assessment. Mean duration to follow-up was 102.74 days (SD 16.51) for parents and 96.14 days (SD 13.81) for adolescents.

Symptom Elevation Procedure

At baseline and 3-month follow-up (ie, when the SCAS and SMFQ were administered), participants who reported elevated adolescent symptoms on the SCAS and/or SMFQ were contacted. For the SCAS, elevated symptom status was defined as ≥1.5 SDs above the mean based on published Australian community sample norms [75]. As there are no published norms or consistent cut-off scores for the SMFQ, we considered a score of ≥8 on the SMFQ-C or SMFQ-P as elevated (based on the original SMFQ-C cut-off suggested by the developing authors [74]). In cases where both the child and parent reported elevation on either or both scales, parents were notified via email and encouraged to seek professional assistance for the adolescent

as appropriate (n=38 at baseline, no significant differences between groups). All participants were provided with a list of websites and potential referral sources at baseline, and participants who reported elevated symptoms were encouraged to consult this. In cases where the child reported extreme elevation on the SMFQ (score > 20), a graduate clinical psychology student also phoned the adolescent to conduct a risk assessment and provide referral information as required (n=29 at baseline, no significant differences between groups).

Statistical Analysis

Missing data rates were low (<1%) at both the item and participant level for all measures. Item-level missing data were replaced with subscale mean imputation for cases with less than 20% missing on a given measure. This is considered an acceptable approach for this amount of missing data [76]. Cases with greater than 20% missing on a measure were considered missing entirely from the analyses.

All analyses were conducted with an a priori type I error rate of .05. Independent samples *t* tests (for continuous variables) and chi-square tests (for categorical variables) were conducted to examine differences between completers and non-completers on outcome measures and participant characteristics. We also assessed for differences between complete parent-adolescent dyads and parents who participated alone.

Primary and secondary outcomes were analyzed with Mixed effect Model Repeated Measures (MMRM) analyses using the MIXED procedure of SPSS version 23 (IBM Corp) with an unstructured covariance matrix. MMRM is consistent with intention-to-treat principles, using all available data from each participant, including those who did not complete follow-up assessments [77]. This is a preferred analytic method for repeated measures designs as it yields unbiased results when data are missing at random and accounts for correlations between repeated measurements of the same participant [78]. For the SCAS and SMFQ, the distribution of model residuals violated the assumption of normality. We therefore applied log transformations to these variables, which improved the distribution of residuals. We ran the MMRM analyses on both the raw and transformed data. As the results did not change substantially and conclusions remained the same, we have reported results based on the raw data.

Finally, as the majority of parents were female, and there were more fathers in the control compared with intervention group (21 vs 7), we conducted sensitivity analyses by running all primary and secondary outcome analyses with data from mothers only. Similarly, as there were differences in participant characteristics between complete parent-adolescent dyads and parents who participated alone, we also ran all analyses using data from complete dyads only.

Results

Participant Characteristics

A total of 349 parents and 327 adolescents registered and completed baseline assessments. Parents had a mean age of 45.11 years (SD 6.11), and the majority were female (320/349, 91.7%). Adolescents had a mean age of 13.60 years (SD 1.03), with 49.2% (161/327) female. Table 2 presents baseline participant characteristics and Table 3 presents participant mental health characteristics. Mean scores on the SCAS-C and SCAS-P were within one SD of published Australian community sample norms for the age range of our sample [75]. Mean SMFQ scores were within one SD of other nonclinical samples (eg, [74,79,80]) and below clinical cut-off scores suggested by the developing authors (SMFQ-C score of 8 [74]) and other authors (eg, SMFQ-C score of 11 [79] and SMFQ-P score of 9 [81]). Mean SFMQ-C scores in our sample were also similar to those reported in large community sample of Australian and American adolescents [79]. Symptom levels in our sample are therefore considered to be within the normal range for this age.

Comparison of Complete Parent-Adolescent Dyads and Parent-Only Participants

We compared participant characteristics and baseline parent-report measures between complete dyads (ie, both parent and adolescent participated at baseline; 93.7%, 327/349) and parents who participated without their adolescent. Parent-only participants had slightly older children (mean=14.1 vs 13.6 years, t_{347}=2.38, *P*=.02), higher baseline SCAS-P and SMFQ-P scores (SCAS-P: mean=25.59 vs 16.17, $t_{21.71}$=2.08, *P*=.049; SMFQ-P: mean=7.59 vs 4.06, $t_{22.01}$=2.27, *P*=.03), were more likely to speak a language other than English at home (22.7% vs 7.6%; χ^2_1[N=349]=5.97, *P*=.02), and were more likely to report that their child had a history of or current anxiety disorder or concern regarding a previous undiagnosed mental health problem (all *P*<.05). This suggests that adolescents may have declined to participate because of their mental health or English proficiency. There were no differences for any other demographic variables, and complete dyads were balanced between groups (intervention group: 93.9%, 154/164; control group: 93.5%, 173/185; χ^2_1[N=349]=0.22, *P*=.88). As we intentionally allowed parents to participate without their adolescent to attain a more diverse sample, we kept data from parent-only participants in our main analyses. See sensitivity analyses below for comparison of results with complete dyads only.

Attrition

Figure 1 presents the participant flow diagram. Overall, the attrition rate was low at 6.0% (n=21) for parents (intervention group: n=13, 7.9%; control group: n=8, 4.3%) and 6.7% (n=22) for adolescents (intervention group: n=8, 5.2%; control group: n=14, 8.1%) at 3-month follow-up. Chi-square tests indicated no significant differences in attrition between groups at any time point (all *P*>.05). Chi-square tests and *t* tests were conducted to assess for differences in demographic characteristics and outcome measures between participants who completed and those who withdrew. There were no significant differences for any demographic variables (all *P*>.05). There were also no differences in scores on the PRADAS-A, SCAS-C, SCAS-P, or SMFQ-C. Significant differences were found for baseline PRADAS and SMFQ-P scores. Specifically, intervention group parents who withdrew from the 1-month follow-up assessment scored significantly lower on the baseline PRADAS than those who completed 1-month follow-up. This

was not the case at 3-month follow-up or for control group parents. Similarly, intervention group parents who did not complete the 3-month assessment reported higher baseline SMFQ-P scores than those who did complete 3-month follow-up.

Table 2. Participant characteristics at baseline.

Parent or child characteristic	n (%)
Parent relationship to child	
Mother	315 (90.3)
Father	28 (8.0)
Other[a]	6 (1.7)
Parent relationship status	
Married or defacto	271 (77.7)
Separated or divorced	58 (16.6)
Single	16 (4.6)
Widowed	4 (1.1)
Family situation	
Intact family, child living with both parents	240 (68.8)
Separated parents, shared care	36 (10.3)
Child living with one parent (participant)	61 (17.5)
Child living with one parent (not participant)	6 (1.7)
Other	6 (1.7)
Parent employment status	
Working full time	159 (45.6)
Working part time	155 (44.4)
Unemployed	35 (10.0)
Parent study status	
Studying full time	57 (16.3)
Studying part time	11 (3.2)
Not studying	281 (80.5)
Parent education level	
Secondary school year 7-12	27 (7.7)
Trade or apprenticeship	1 (0.3)
Technical and further education certificate or other technical qualification	40 (11.5)
Diploma	61 (17.5)
Bachelor degree	105 (30.1)
Postgraduate	115 (33.0)
Language other than English spoken at home	30 (8.6)
Parent Indigenous status	
Yes	4 (1.1)
No	340 (97.4)
Prefer not to say	5 (1.4)
State of residence	
New South Wales	113 (32.4)
Victoria	74 (21.2)
Queensland	61 (17.5)
Tasmania	49 (14.0)
Australian Capital Territory	22 (6.3)
South Australia	16 (4.6)

Parent or child characteristic	n (%)
Western Australia	13 (3.7)
Northern Territory	1 (0.3)

[a]Other parent-child relationship category includes step-mother, step-father, grandmother, and legal guardian.

Intervention Use

As the intervention was a once-off personalized feedback report, all parents allocated to the intervention group received the intervention. Following presentation of their feedback report, parents were asked how much of the feedback they had read. Of the 128 (78.0%) intervention group parents who answered this question, 102 (79.7%) reported reading all of it, 24 (18.8%) reporting reading about half of it, and 2 (1.6%) stated that they would read it later.

Primary Outcome Measure: Parenting to Reduce Adolescent Depression and Anxiety Scale

Observed scores for the PRADAS at each occasion are presented in Multimedia Appendix 6 (Table 1). Figure 2 presents the estimated marginal means of PRADAS scores at baseline, 1-month, and 3-month follow-up, estimated under the group-by-measurement-occasion mixed model. The overall group-by-measurement-occasion interaction effect was significant ($F_{2,331.22}$=16.36, P<.001), indicating a different pattern of change over time between groups. Planned contrasts between groups at each occasion indicated a significantly greater increase in PRADAS scores in the intervention compared with control group at both 1-month and 3-month follow-up (1-month follow-up, $t_{1,332.13}$= 5.27, P<.001; 3-month follow-up, $t_{1,339.10}$=4.87, P<.001). Effect sizes were small to medium (1-month follow-up, d=0.30 [95% CI 0.06-0.50]; 3-month follow-up, d=0.33[95% CI 0.05-0.49]). Figure 3 presents estimated marginal means of PRADAS-A scores from baseline to 3-month follow-up, estimated under the group-by-measurement-occasion mixed model.

At 1-month and 3-month follow-up, parents in both groups were also asked whether they had tried to make any changes to their parenting since the previous assessment time point and their perceived success in making changes. Chi-square analyses of the postintervention (1-month) assessment revealed a significant difference in attempts to change parenting (χ^2_3[N=307]=19.65, P<.001), with significantly more intervention group parents reporting making some changes to their parenting. This finding aligns with the parent-report of specific parenting behaviors on the PRADAS. There was, however, no significant difference between groups for parent-reported success in making changes to their parenting, χ^2_3(N=307)=6.26, P=.10. At 3-month follow-up, there were no significant group differences in reported attempts to change parenting (χ^2_3[N=323]=6.03, P=.11) or perceived success in making changes (χ^2_3[N=313]=6.03, P=.11). The proportion of parents who reported accessing additional parenting resources, or seeking professional help for their own or their child's mental health, did not differ between groups at either 1-month or 3-month follow-up (all P>.05).

Secondary Outcome Measures

Observed values for all secondary outcome measures are presented in Multimedia Appendix 6 (Table 2). Table 4 presents the estimated marginal means and MMRM group-by-measurement-occasion interaction results for these measures.

Parenting to Reduce Adolescent Depression and Anxiety Scale-Adolescent Report

The group-by-measurement-occasion interaction effect for PRADAS-A was nonsignificant ($F_{1,303.82}$=0.02, P=.88), demonstrating no difference in the pattern of change over time between groups. A significant main effect for time was observed ($F_{1,303.82}$=8.49, P=.004), with planned contrasts showing a significant reduction in PRADAS-A scores from baseline to 3-month follow-up for both groups (intervention, P=.04; control, P=.046).

Adolescent Symptoms

As shown in Table 4, the group-by-measurement-occasion interaction effects were not significant for any of the symptom measure analyses (all P>.05), suggesting no significant difference between the groups over time. Significant main effects for time were found for the SMFQ-P ($F_{1,324.76}$=12.19, P<.001), SCAS-P ($F_{1,325.55}$=70.54, P<.001), and SCAS-C ($F_{1,302.74}$=5.55, P=.02), with planned contrasts demonstrating a reduction in symptoms for both groups from baseline to 3 months. There were no significant main effects for the SMFQ-C (all P>.05). Figure 4 presents the estimated marginal means for all symptom measures from baseline to 3-month follow-up.

Table 3. Parent and adolescent mental health characteristics, Spence Children's Anxiety Scale (SCAS), and Short Mood and Feelings Questionnaire (SMFQ) scores at baseline. P and C indicate parent and child report, respectively.

Parent or child characteristic	Statistics
Parental concern about child's risk of developing depression, n (%)	
Not at all	77 (22.1)
A little	175 (50.1)
Yes	68 (19.5)
Very much so	27 (7.7)
Missing (declined to answer)	2 (0.6)
Parental concern about child's risk of developing an anxiety disorder, n (%)	
Not at all	79 (22.6)
A little	165 (47.3)
Yes	74 (21.2)
Very much so	29 (8.3)
Missing (declined to answer)	2 (0.6)
Parental history or current mental health problem (as reported by parent), n (%)	
None	145 (41.5)
Yes, past history	138 (39.5)
Yes, current	33 (9.5)
Yes, past and current	31 (8.9)
Missing (declined to answer)	2 (0.6)
Child history of mental health problem or behavioral disorder diagnosis (as reported by parent), n (%)	
None	231 (66.2)
Depression	2 (0.6)
Any anxiety disorder	10 (2.9)
Autism spectrum disorder (including Asperger syndrome)	6 (1.7)
Other mental health or behavioral disorder[a]	9 (2.6)
Multiple diagnoses	11 (3.2)
No formal diagnosis; however, I believe my child has experienced some emotional or behavioral problems	76 (21.8)
Missing (declined to answer)	4 (1.1)
Child current mental health or behavioral problems (as reported by parent), n (%)	
None	243 (69.6)
Depression	1 (0.3)
Any anxiety disorder	18 (5.2)
Autism spectrum disorder (including Asperger syndrome)	7 (2.0)
Other mental health or behavioral disorder[a]	7 (2.0)
Multiple diagnoses	16 (4.6)
No formal diagnosis; however, I believe my child is currently experiencing some emotional or behavioral problems	51 (14.6)
Missing (declined to answer)	6 (1.7)
Baseline symptom measures, mean (SD)	
SCAS-P score (n=349)	16.76 (11.66)
SCAS-C score (n=326)	29.78 (17.87)
SMFQ-P score (n=349)	4.28 (4.58)
SMFQ-C score (n=325)	6.18 (5.66)

[a]This category includes attention-deficit/hyperactivity disorder, oppositional defiant disorder, conduct disorder, learning difficulties, or any other disorder specified by parents.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) participant flow diagram. ITT: intention-to-treat.

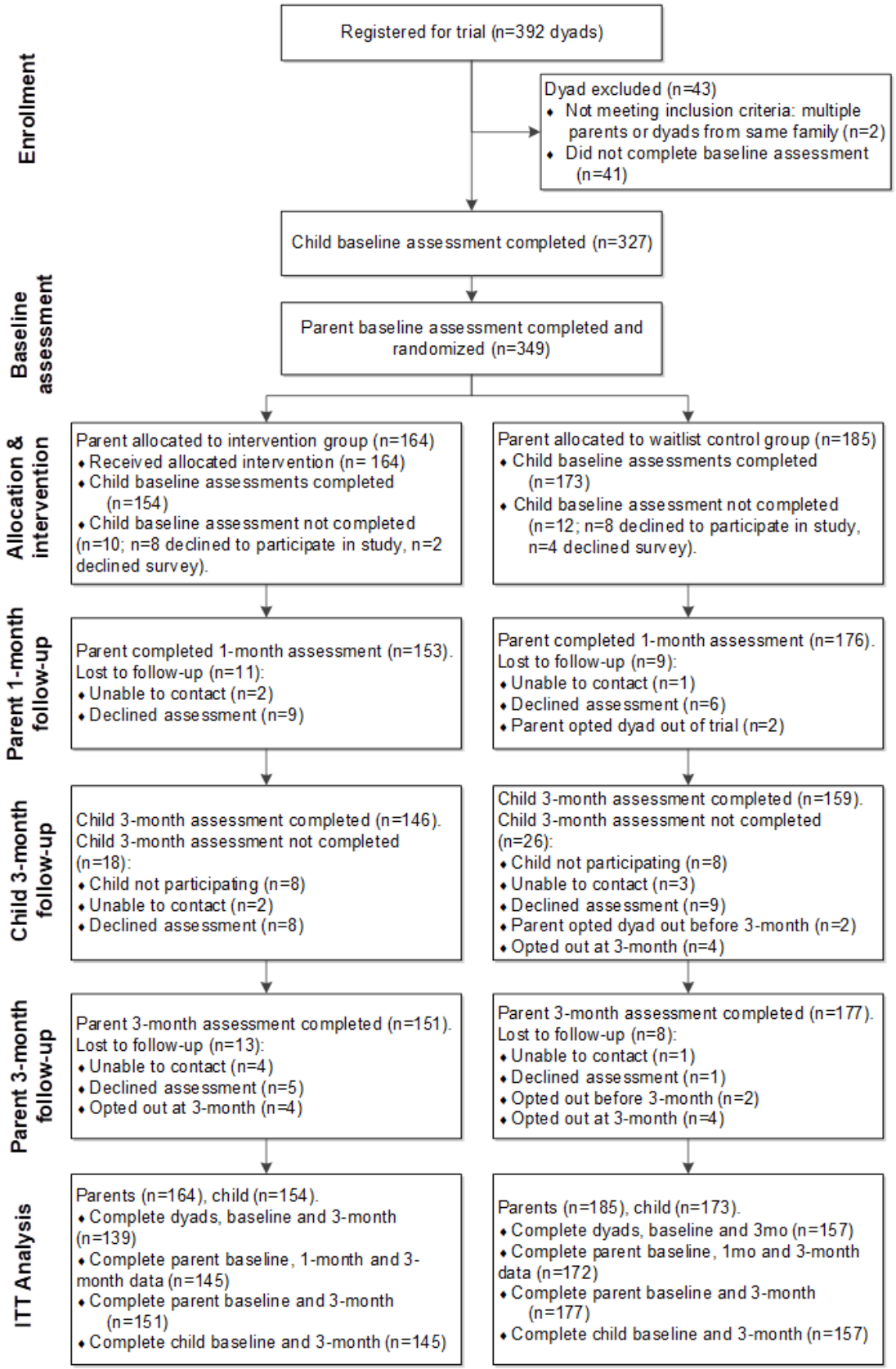


Figure 2. Estimated marginal means for Parenting to Reduce Adolescent Depression and Anxiety Scale (PRADAS) scores at baseline (n=349), 1-month (n=329), and 3-month (n=328) follow-up, estimated under group-by-measurement-occasion mixed model. Error bars represent standard errors. *Planned contrast significant at P<.001 level.

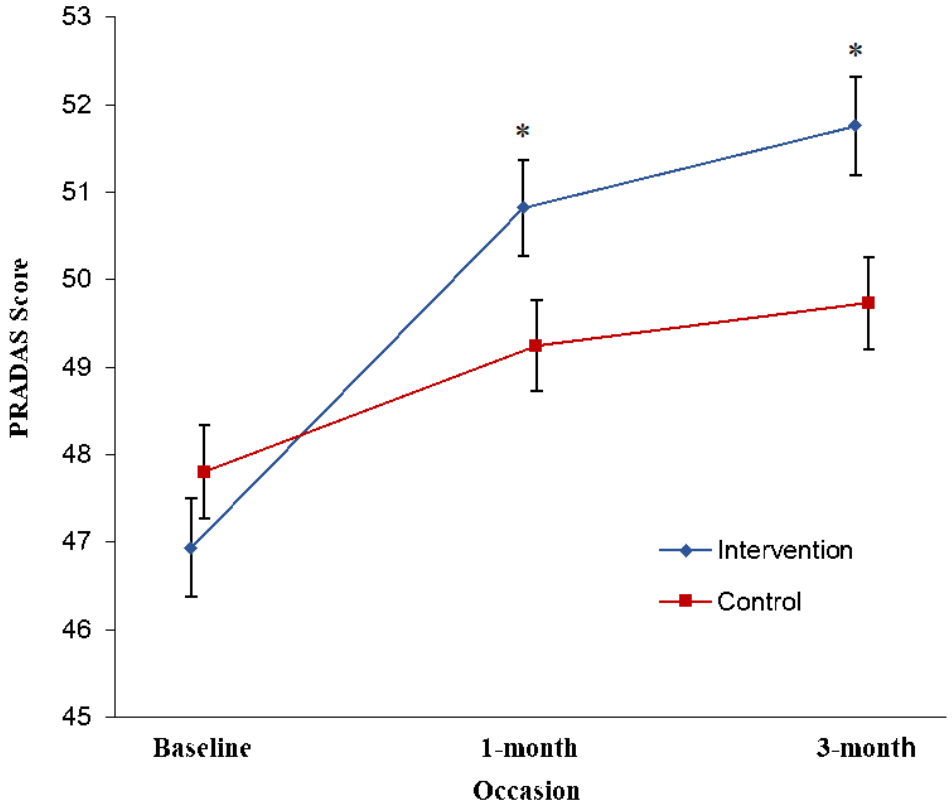


Figure 3. Estimated marginal means of Parenting to Reduce Adolescent Depression and Anxiety Scale-Adolescent report (PRADAS-A) scores from baseline to 3-month follow-up, estimated under group-by-measurement-occasion mixed model. Error bars represent standard errors.

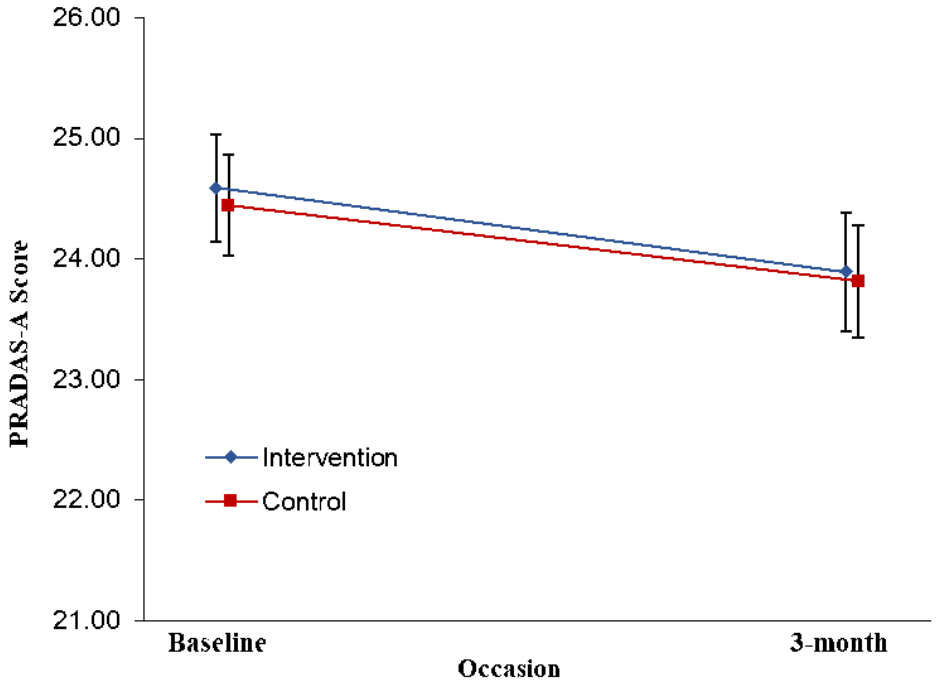
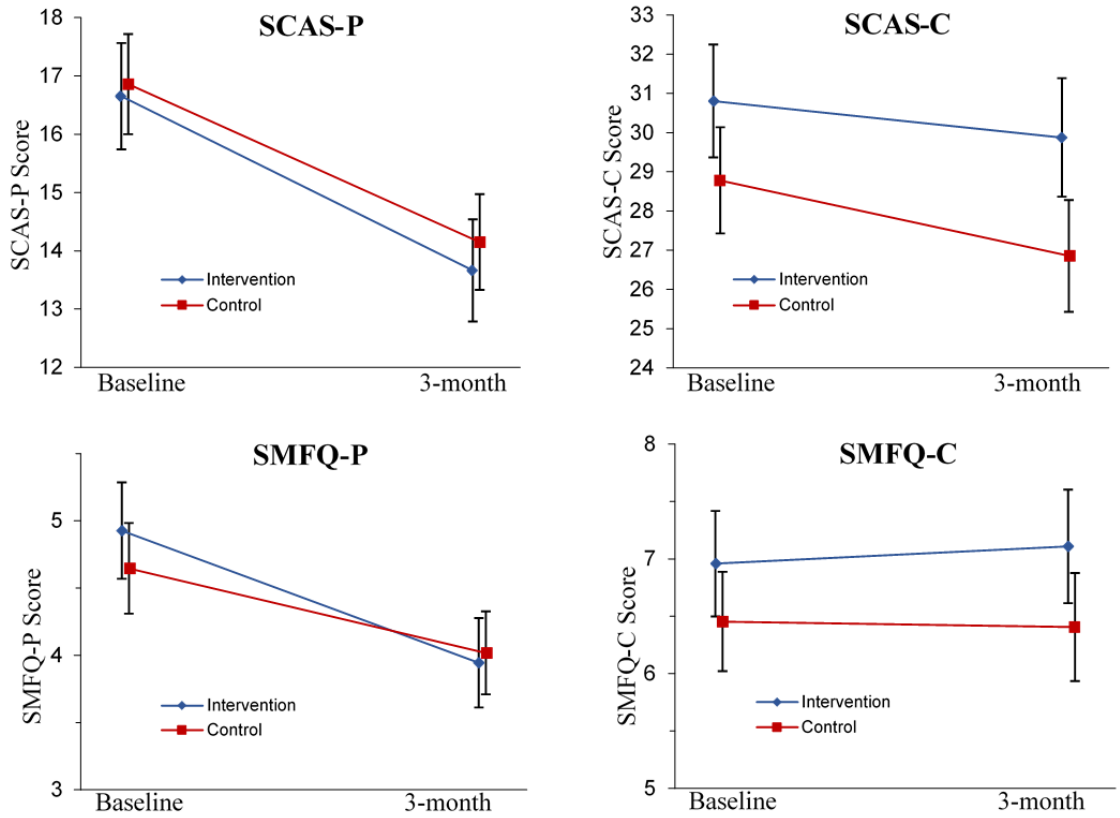


Table 4. Estimated marginal means, standard errors, and Mixed effect Model Repeated Measures (MMRM) test of group-by-measurement-occasion interaction for all secondary outcome measures. PRADAS-A: Parenting to Reduce Adolescent Depression and Anxiety Scale-Adolescent report; SCAS-C: Spence Children's Anxiety Scale-Child-report; SCAS-P: Spence Children's Anxiety Scale-Parent-report; SMFQ-C: Short Mood and Feelings Questionnaire-Child-report; SMFQ-P: Short Mood and Feelings Questionnaire-Parent-report.

Measure and occasion	Estimated marginal means (SE), n		MMRM group-by-measurement-occasion interaction effect			Cohen *d*[a] (95% CI)
	Intervention	Control	*F*	Degrees of freedom	*P* value	
PRADAS-A			0.02	1,303.82	.88	
Baseline	24.59 (0.44), 154	24.45 (0.42), 173				
3-month	23.89 (0.49), 146	23.82 (0.47), 159				0.00 (−0.22 to 0.23)
SCAS-P			0.17	1,325.55	.68	
Baseline	16.65 (0.91), 164	16.86 (0.86), 185				
3-month	13.66 (0.88), 151	14.15 (0.82), 177				−0.06 (−0.29 to 0.15)
SCAS-C			0.67	1,302.74	.41	
Baseline	30.81 (1.44), 153	28.78 (1.35), 173				
3-month	29.88 (1.51), 145	26.86 (1.43), 159				0.18 (−0.05 to 0.40)
SMFQ-P			0.59	1,324.76	.44	
Baseline	4.43 (0.36), 164	4.15 (0.34), 185				
3-month	3.44 (0.33), 151	3.52 (0.31), 177				−0.04 (−0.26 to 0.17)
SMFQ-C			0.18	1,301.25	.68	
Baseline	6.46 (0.46), 152	5.96 (0.43), 173				
3-month	6.61 (0.50), 146	5.91 (0.47), 159				0.12 (−0.10 to 0.35)

[a]Cohen *d* effect size calculated based on observed means at end point (3-month follow-up) and SD of the control group.

Figure 4. Estimated marginal means for Spence Children's Anxiety Scale-Parent report (SCAS-P), Spence Children's Anxiety Scale-Child report (SCAS-C), Short Mood and Feelings Questionnaire-Parent report (SMFQ-P), and Short Mood and Feelings Questionnaire-Child report (SMFQ-C) from baseline to 3-month follow-up, estimated under group-by-measurement-occasion mixed model. Error bars represent standard errors.



Parent Satisfaction and Acceptability of the Intervention

Frequencies for parental responses to the process questions asked immediately postintervention (ie, after receiving their feedback at baseline or 3-month follow-up, for the intervention and waitlist control groups, respectively) are presented in Multimedia Appendix 2. Of the parents who answered these questions, the majority (93.6%) reported that they were somewhat or very satisfied with the feedback received, and 95.1% reported that they found the feedback either somewhat, very, or extremely useful. Most parents (90.2%) also reported that they were somewhat or very likely to change their parenting based on the feedback provided, although fewer parents were confident to do so (84.1% either very or moderately confident).

Sensitivity Analyses

All results from primary and secondary outcome analyses using data from mothers only, as well as data from complete parent-adolescent dyads only, were consistent with results reported above.

Discussion

Principal Findings

This study aimed to evaluate the short-term effects of a single-session, Web-based, tailored psychoeducational parenting intervention on parenting factors and symptoms of depression and anxiety in adolescents. Results provide preliminary support for the effectiveness of the intervention for improving evidence-based parenting risk and protective factors for adolescent depression and anxiety disorders. In support of our first hypothesis, parents in the intervention group demonstrated a greater increase in self-reported concordance with the Guidelines' recommendations than control group parents. This effect was evident at 1-month postintervention and 3-month follow-up, with small to medium effect sizes. Additionally, most parents were satisfied with the intervention and found it useful. The intervention did not, however, have a significant effect on any of the secondary outcome measures; namely adolescent-report of parenting, or symptoms of depression or anxiety in the adolescents. Our findings therefore do not support the effectiveness of the intervention for reducing adolescent symptoms in the short-term.

Effect of the Intervention on Parenting

The current findings add to a growing body of literature supporting the efficacy of preventive parenting interventions for the improvement of modifiable parenting factors (eg, [33,35,82-84]). This study contributes a number of novel findings to the literature. To our knowledge, the intervention is the first of its kind to target a wide range of parenting factors known to influence the development of depression and anxiety disorders in adolescents. Additionally, the intervention is considerably briefer than existing parenting interventions and was conducted entirely online with no therapist support. It is therefore promising to find an improvement in self-reported parenting, particularly given that many existing interventions are labor-intensive (eg, delivered by trained professionals) and expensive to disseminate. Together, these factors support the potential of a single-session, online intervention as a strategy to translate research regarding parenting risk and protective factors for adolescent depression and anxiety into an accessible, low-cost preventive approach. Importantly, this study also adds to the paucity of research investigating preventive interventions for parents of adolescents.

Although we found a significant effect of the intervention on self-reported parental concordance with the Guidelines, there was no significant effect of the intervention on adolescent-report of parenting. In fact, adolescents in both groups reported a slight reduction in their parent's concordance with the Guidelines' recommendations. This discrepancy is in line with the frequently reported discordance between parent- and child-report of parenting (eg, [85-87]). It is also reflected by the modest correlation between the PRADAS and PRADAS-A (r=.25), which is consistent with associations between parent and child report in the literature (eg, correlations of .20 to .40 are typical [85,86,88]).

There are several possible explanations for the discrepancy between parent- and adolescent-report found in this study. First, the perspectives captured by parent- and child-report may reflect the different focus or importance that parents and adolescents place on various parenting practices. For example, parents may report on subtle behaviors, attitudes, and beliefs that may not be noticed by the adolescent. Similarly, adolescents' reports may be influenced by their experiences over a longer period of time, rather than specific recent parenting behaviors. Second, the PRADAS-A covers less content and is less specific than the PRADAS, with fewer items (43 compared with 73 items of the PRADAS). It may therefore be less sensitive to detect subtle or gradual changes in parenting, which may not be easily detected by adolescents, particularly in the short-term. As this is the first study to utilize the PRADAS-A, we do not have prior evidence to support its ability to detect change. It is also possible that adolescents in our sample, who had a mean age of 13.64 years, were too young to accurately report change in the parenting practices covered in the PRADAS-A. Finally, it could be that our results only demonstrate parents' perceived improvement in parenting, as measured by the self-report PRADAS. Without objective measures of behavior and with the discrepancy between respondents, we cannot be certain that the intervention did result in improved parenting practices. Unfortunately, this is a limitation inherent in many parenting intervention studies, which often use single-informant parent-report measures of family and parenting factors.

Effect of the Intervention on Adolescent Symptoms

The lack of effect of the intervention on adolescent symptoms contrasts with a growing body of literature suggesting that preventive parenting interventions can reduce internalizing problems in children and adolescents (eg, [31,32,35,37,89]). There are a number of potential reasons for this finding. First, we only collected data up to 3-month follow-up. It is likely that a longer time period is required for changes in parenting to influence adolescent symptoms. Much of the evidence base for the parenting factors targeted in the intervention stems from longitudinal research [9]; hence, the parenting factors are theorized to have a long-term impact on adolescent outcomes. For example, factors such as consistent discipline or supporting

age-appropriate autonomy would likely take time to have an effect on adolescent symptoms. In line with this hypothesis, several studies that have found beneficial effects of parenting interventions on child mental health outcomes have had longer follow-up periods (eg, 6-12 months or more; see [37,84] for reviews) or have found that effects increase over time (eg, [29,30,90]). It is possible that long-term effects of preventive parenting interventions occur because of "sleeper effects." That is, the effects of the intervention on adolescent outcomes may increase over time, potentially because of the bidirectional nature of parenting behaviors and child temperament or behavior on subsequent child internalizing outcomes (eg, [14,37,91,92]). To adequately assess the relationship between improving parenting factors and adolescent depression and anxiety outcomes, longer-term studies, ideally with follow-up into late adolescence, are required.

It should also be noted that although there is strong evidence regarding the role of parenting in the development of child and adolescent depression and anxiety, the amount of variance explained by individual parenting factors is small (eg, [9,38-40]). The development of these disorders is multifaceted and influenced by many factors other than parenting (eg, gender, genetic predisposition, parent psychopathology, early-life events, and socioeconomic factors [10,93-95]), as well as interactions among these factors [93,95]. In line with this, the correlations between the PRADAS and adolescent symptoms at baseline were small (SCAS-P: r= −.13, P=.02; SCAS-C: r=−.09, P=.10; SMFQ-P: r=−.22, P<.001; SMFQ-C: r=−.11, P=.06). It is therefore possible that the change in parenting found in this study was not large enough to result in reduced adolescent symptoms in the short-term. Additionally, to ensure independence of observations, we only allowed one parent per family to participate. This differs to face-to-face parenting interventions, which typically invite both parents to participate (although participation of mothers is unequivocally higher [96]), and emphasize the importance of parenting consistency. Thus, while changing the parenting of one parent was not effective in reducing adolescent symptoms in this study, it is possible that a consistent change across both parents could have greater influence on adolescent outcomes.

Implications for Universal Prevention

Evidence to date is conflicting regarding the comparative efficacy of universal, selective, and indicated prevention approaches for depression and anxiety in young people. Some reviews have found selective or indicated programs to be more effective (eg, [28,97-99]), whereas others have found no significant differences between the approaches (eg, [23,37,100]). Although effect sizes may be larger in selective and indicated programs, universal approaches have the benefit of reaching a greater proportion of the population. Thus, even with small effect sizes, the population mean can be shifted, resulting in significant impact at a population level [101]. The small effect on parent-reported behaviors found in this study could therefore lead to population-level changes in parenting if the intervention is delivered universally.

In practice, the optimal plan for dissemination of preventive parenting programs for adolescent depression and anxiety would likely involve a combination of universal, selective, and indicated approaches [93]. In this way, universal programs are able to reach a greater proportion of the population, whereas selective and indicated programs, which are typically more intensive and costly to deliver, are available for those at elevated risk. As discussed, this is the aim of the multilevel PiP program [58]. The single-session intervention evaluated in this paper can serve as a first step, which can be accessed by all parents regardless of their adolescent's risk or current symptoms. This may have the added benefit of minimizing stigma, if promoted as a resource for all parents of adolescents (eg, at the transition to secondary school). Given the difficulty engaging and retaining parents in online interventions, a model such as this may allow more parents to receive at least some intervention (ie, the single-session component), even if they do not continue to complete the PiP modules. Parents then have the option of "stepping up" to the next level of PiP based on personal preference or need. Results from this study support the value of including the single-session component as a stand-alone intervention within the PiP model. In comparison to a recent RCT comparing the full PiP intervention with an active control condition, the single-session intervention achieved a similar effect at 3-month follow-up (although the overall effect size was larger for the full intervention [65]).

Strengths, Limitations, and Future Directions

This study has several strengths. We recruited a large community sample via methods similar to how the intervention could be disseminated (ie, primarily via schools and online networks). The demographics of our sample are therefore likely to be representative of the expected end users of the program. Similarly, there were no exclusion criteria, and participants were not precluded from accessing other resources or services. These factors suggest good external validity of the findings. Furthermore, parents were generally satisfied with the intervention and found the feedback to be useful. This may partially account for the low attrition rate, which is notable given the high attrition often reported in online interventions (eg, [63,66,67]).

Study limitations include the overrepresentation of highly educated mothers from intact families, which may limit the generalizability of findings to fathers, parents with lower educational attainment, and different family situations. Although not representative of the general population, our sample characteristics are similar to other studies of online parenting interventions (eg, [56,57,102,103]), suggesting that these types of programs may be most appealing to this demographic. This limitation is particularly relevant given the aim of promoting the single-session intervention (and PiP more broadly) as a universal prevention approach. The impact of the intervention at a population level would be limited if it is only accessed by a certain demographic, particularly given that parent educational attainment is associated with higher positive parenting practices [59,104,105] and lower rates of child mental health problems [106,107]. To adequately assess the potential of the intervention as a universal approach, research with more diverse samples is required. Our team is currently planning a trial of PiP with parents of lower socioeconomic status. It is possible that parents with more risk factors (ie, lower concordance with the

Guidelines' recommendations) may show greater improvement with intervention, as they have more room for improvement. Conversely, the single-session approach may not be a sufficient level of intervention for such parents, although it may be a useful way to initially engage or identify parents who could benefit from a more intensive parenting support program.

We assessed parenting with newly developed measures, designed specifically to assess concordance with the Guidelines. This was necessary to assess the wide range of parenting factors covered in the intervention with minimal burden to participants. However, it was in place of existing validated measures; therefore, future research is required to assess whether there are effects of the intervention beyond what is captured by the PRADAS. Although not feasible in this study, the addition of objective measures of parenting, such as behavioral observation of parent-adolescent interactions, would also be of value in future research. Additionally, we only collected data up to 3-month follow-up, and the ethical need to limit the duration of waitlist precludes the possibility of between-group comparison of long-term follow-up. We also used symptom measures, rather than measures of diagnostic status. Future research would benefit from including diagnostic measures, as well as broader measures of quality of life and general functioning. Examining the effectiveness of the program when both parents participate would also be of value.

Conclusions

This RCT provides preliminary support for the effectiveness of a single-session, Web-based, tailored psychoeducation parenting intervention for improving evidence-based parenting risk and protective factors for adolescent depression and anxiety disorders. At this stage, there is no evidence that the program reduces symptoms of depression and anxiety in adolescents in the short-term. However, given the empirical and theoretical basis for the parenting factors targeted in the intervention, it is possible that altering these parenting behaviors in early adolescence could result in reduced risk in the long-term. Given the brevity and ease of dissemination of the program, this is a promising avenue for the translation of research into a sustainable, low-cost intervention that can be disseminated widely as a public health prevention strategy.

Acknowledgments

The authors acknowledge funding from the National Health and Medical Research Council (NHMRC) for the Web development of the intervention and the partnership of *beyondblue, the national depression and anxiety initiative* in the development of the parenting guidelines. The authors received salary support from the NHMRC for a Career Development Fellowship (MBHY, APP1061744) and a Senior Principal Research Fellowship (AFJ, APP1059785), an Australian Research Council Laureate Fellowship (RMR, FL150100096), an Australian Government Research Training Program Scholarship (MCB), and a Windermere Foundation Doctoral Scholarship in Health (MCB). The RCT was supported by Monash University's Faculty of Medicine, Nursing, and Health Sciences Faculty Strategic Grant Scheme funding (SGS15-0149) and an Advancing Women's Research Success Grant. None of the funding sources had any role in the conduct or publication of this study. The authors also acknowledge the RCT project managers Mrs Jennifer Hanson-Peterson, Dr Shireen Mahtani, Ms Claire Nicolas, and Ms Jacqueline Green; the team of research assistants who assisted with data collection; Mental Health First Aid Australia and the schools who assisted with recruitment; and the reference group of parents who contributed to the development of the intervention.

Conflicts of Interest

Authors MCB, MBHY, AFJ and KL codeveloped the intervention. None of the authors derive any financial benefit from the intervention.

Multimedia Appendix 1

Screenshots of the intervention.

[PDF File (Adobe PDF File), 594KB - jmir_v20i4e148_app1.pdf]

Multimedia Appendix 2

Parental responses to process evaluation questions.

[PDF File (Adobe PDF File), 30KB - jmir_v20i4e148_app2.pdf]

Multimedia Appendix 3

Parent explanatory statement.

[PDF File (Adobe PDF File), 408KB - jmir_v20i4e148_app3.pdf]

Multimedia Appendix 4

Child explanatory statement.

[PDF File (Adobe PDF File), 316KB - jmir_v20i4e148_app4.pdf]

Multimedia Appendix 5

Online registration and consent form.

[PDF File (Adobe PDF File), 392KB - jmir_v20i4e148_app5.pdf]

Multimedia Appendix 6

Observed scores for primary and secondary outcome measures at each measurement occasion.

[PDF File (Adobe PDF File), 35KB - jmir_v20i4e148_app6.pdf]

Multimedia Appendix 7

CONSORT-EHEALTH Checklist (V 1.6.1).

[PDF File (Adobe PDF File), 9MB - jmir_v20i4e148_app7.pdf]

References

1. World Health Organization. Geneva: World Health Organization; 2017. Depression and other common mental disorders: global health estimates URL: http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf [WebCite Cache ID 6vCqDVUUD]
2. Kessler RC, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
3. Mathews RR, Hall WD, Vos T, Patton GC, Degenhardt L. What are the major drivers of prevalent disability burden in young Australians? Med J Aust 2011 Mar 07;194(5):232-235. [Medline: 21381994]
4. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. Lancet 2007 Apr 14;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7] [Medline: 17434406]
5. World Health Organization. Geneva: Word Health Organization; 2011. Global burden of mental disorders and the need for a comprehensive, coordinated response from health and societal sectors at the country level URL: http://apps.who.int/gb/ebwha/pdf_files/EB130/B130_R8-en.pdf [WebCite Cache ID 6vCqTwGp1]
6. Schleider JL, Weisz JR. Family process and youth internalizing problems: a triadic model of etiology and intervention. Dev Psychopathol 2017 Feb;29(1):273-301. [doi: 10.1017/S095457941600016X] [Medline: 27048767]
7. Reardon JM, Williams NL. The specificity of cognitive vulnerabilities to emotional disorders: anxiety sensitivity, looming vulnerability and explanatory style. J Anxiety Disord 2007;21(5):625-643. [doi: 10.1016/j.janxdis.2006.09.013] [Medline: 17070666]
8. Dozois DJ, Seeds PM, Collins KA. Transdiagnostic approaches to the prevention of depression and anxiety. J Cogn Psychother 2009 Feb 01;23(1):44-59. [doi: 10.1891/0889-8391.23.1.44]
9. Yap MB, Pilkington PD, Ryan SM, Jorm AF. Parental factors associated with depression and anxiety in young people: a systematic review and meta-analysis. J Affect Disord 2014 Mar;156:8-23. [doi: 10.1016/j.jad.2013.11.007] [Medline: 24308895]
10. Van Voorhees BW, Paunesku D, Kuwabara SA, Basu A, Gollan J, Hankin BL, et al. Protective and vulnerability factors predicting new-onset depressive episode in a representative of U.S. adolescents. J Adolesc Health 2008 Jun;42(6):605-616. [doi: 10.1016/j.jadohealth.2007.11.135] [Medline: 18486870]
11. Sheeber L, Hops H, Alpert A, Davis B, Andrews J. Family support and conflict: prospective relations to adolescent depression. J Abnorm Child Psychol 1997 Aug;25(4):333-344. [Medline: 9304449]
12. Wood JJ, McLeod BD, Sigman M, Hwang W, Chu BC. Parenting and childhood anxiety: theory, empirical findings, and future directions. J Child Psychol Psychiatry 2003 Jan;44(1):134-151. [Medline: 12553416]
13. Rapee RM. Potential role of childrearing practices in the development of anxiety and depression. Clin Psychol Rev 1997;17(1):47-67. [Medline: 9125367]
14. Restifo K, Bögels S. Family processes in the development of youth depression: translating the evidence to treatment. Clin Psychol Rev 2009 Jun;29(4):294-316. [doi: 10.1016/j.cpr.2009.02.005] [Medline: 19356833]
15. Beesdo K, Knappe S, Pine DS. Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V. Psychiatr Clin North Am 2009 Sep;32(3):483-524 [FREE Full text] [doi: 10.1016/j.psc.2009.06.002] [Medline: 19716988]
16. Biederman J, Hirshfeld-Becker DR, Rosenbaum JF, Hérot C, Friedman D, Snidman N, et al. Further evidence of association between behavioral inhibition and social anxiety in children. Am J Psychiatry 2001 Oct;158(10):1673-1679. [doi: 10.1176/appi.ajp.158.10.1673] [Medline: 11579001]

17. Hankin BL. Adolescent depression: description, causes, and interventions. Epilepsy Behav 2006 Feb;8(1):102-114. [doi: 10.1016/j.yebeh.2005.10.012] [Medline: 16356779]
18. Schwartz OS, Dudgeon P, Sheeber LB, Yap MBH, Simmons JG, Allen NB. Parental behaviors during family interactions predict changes in depression and anxiety symptoms during adolescence. J Abnorm Child Psychol 2012 Jan;40(1):59-71. [doi: 10.1007/s10802-011-9542-2] [Medline: 21789522]
19. Fisak Jr B, Grills-Taquechel AE. Parental modeling, reinforcement, and information transfer: risk factors in the development of child anxiety? Clin Child Fam Psychol Rev 2007 Sep;10(3):213-231. [doi: 10.1007/s10567-007-0020-x] [Medline: 17487582]
20. Rapee RM. Family factors in the development and management of anxiety disorders. Clin Child Fam Psychol Rev 2012 Mar;15(1):69-80. [doi: 10.1007/s10567-011-0106-3] [Medline: 22116624]
21. Yap MBH, Pilkington PD, Ryan SM, Kelly CM, Jorm AF. Parenting strategies for reducing the risk of adolescent depression and anxiety disorders: a Delphi consensus study. J Affect Disord 2014 Mar;156:67-75. [doi: 10.1016/j.jad.2013.11.017] [Medline: 24359862]
22. Avenevoli S, Merikangas KR. Implications of high-risk family studies for prevention of depression. Am J Prev Med 2006 Dec;31(6 Suppl 1):S126-S135. [doi: 10.1016/j.amepre.2006.07.003] [Medline: 17175407]
23. Fisak BJ, Richard D, Mann A. The prevention of child and adolescent anxiety: a meta-analytic review. Prev Sci 2011 Sep;12(3):255-268. [doi: 10.1007/s11121-011-0210-0] [Medline: 21437675]
24. Stockings EA, Degenhardt L, Dobbins T, Lee YY, Erskine HE, Whiteford HA, et al. Preventing depression and anxiety in young people: a review of the joint efficacy of universal, selective and indicated prevention. Psychol Med 2016 Jan;46(1):11-26. [doi: 10.1017/S0033291715001725] [Medline: 26315536]
25. Christensen H, Pallister E, Smale S, Hickie IB, Calear AL. Community-based prevention programs for anxiety and depression in youth: a systematic review. J Prim Prev 2010 Jun;31(3):139-170. [doi: 10.1007/s10935-010-0214-8] [Medline: 20437102]
26. Werner-Seidler A, Perry Y, Calear AL, Newby JM, Christensen H. School-based depression and anxiety prevention programs for young people: a systematic review and meta-analysis. Clin Psychol Rev 2017 Feb;51:30-47 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.005] [Medline: 27821267]
27. Hetrick SE, Cox GR, Witt KG, Bir JJ, Merry SN. Cognitive behavioural therapy (CBT), third-wave CBT and interpersonal therapy (IPT) based interventions for preventing depression in children and adolescents. Cochrane Database Syst Rev 2016 Aug 09(8):CD003380. [doi: 10.1002/14651858.CD003380.pub4] [Medline: 27501438]
28. Merry SN, Hetrick SE, Cox GR, Brudevold-Iversen T, Bir JJ, McDowell H. Psychological and educational interventions for preventing depression in children and adolescents. Cochrane Database Syst Rev 2011 Dec 07(12):CD003380. [doi: 10.1002/14651858.CD003380.pub3] [Medline: 22161377]
29. Beardslee WR, Gladstone TR, Wright EJ, Cooper AB. A family-based approach to the prevention of depressive symptoms in children at risk: evidence of parental and child change. Pediatrics 2003 Aug;112(2):e119-e131. [Medline: 12897317]
30. Ginsburg GS. The Child Anxiety Prevention Study: intervention model and primary outcomes. J Consult Clin Psychol 2009 Jun;77(3):580-587 [FREE Full text] [doi: 10.1037/a0014486] [Medline: 19485597]
31. Morgan AJ, Rapee RM, Salim A, Goharpey N, Tamir E, McLellan LF, et al. Internet-delivered parenting program for prevention and early intervention of anxiety problems in young children: randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2017 May;56(5):417-25.e1. [doi: 10.1016/j.jaac.2017.02.010] [Medline: 28433091]
32. Rapee RM, Kennedy S, Ingram M, Edwards S, Sweeney L. Prevention and early intervention of anxiety disorders in inhibited preschool children. J Consult Clin Psychol 2005 Jun;73(3):488-497. [doi: 10.1037/0022-006X.73.3.488] [Medline: 15982146]
33. Sandler I, Gunn H, Mazza G, Tein J, Wolchik S, Berkel C, et al. Effects of a program to promote high quality parenting by divorced and separated fathers. Prev Sci 2017 Sep 15:1-11. [doi: 10.1007/s11121-017-0841-x] [Medline: 28913663]
34. Wolchik SA, West SG, Sandler IN, Tein JY, Coatsworth D, Lengua L, et al. An experimental evaluation of theory-based mother and mother-child programs for children of divorce. J Consult Clin Psychol 2000 Oct;68(5):843-856. [Medline: 11068970]
35. Kehoe CE, Havighurst SS, Harley AE. Tuning in to teens: improving parent emotion socialization to reduce youth internalizing difficulties. Social Development 2013 Oct 15;23(2):413-431. [doi: 10.1111/sode.12060]
36. Havighurst SS, Kehoe CE, Harley AE. Tuning in to teens: improving parental responses to anger and reducing youth externalizing behavior problems. J Adolesc 2015 Jul;42:148-158. [doi: 10.1016/j.adolescence.2015.04.005] [Medline: 26005933]
37. Yap MB, Morgan AJ, Cairns K, Jorm AF, Hetrick SE, Merry S. Parents in prevention: a meta-analysis of randomized controlled trials of parenting interventions to prevent internalizing problems in children from birth to age 18. Clin Psychol Rev 2016 Oct 21;50:138-158. [doi: 10.1016/j.cpr.2016.10.003] [Medline: 27969003]
38. McLeod BD, Weisz JR, Wood JJ. Examining the association between parenting and childhood depression: a meta-analysis. Clin Psychol Rev 2007 Dec;27(8):986-1003. [doi: 10.1016/j.cpr.2007.03.001] [Medline: 17449154]
39. McLeod BD, Wood JJ, Weisz JR. Examining the association between parenting and childhood anxiety: a meta-analysis. Clin Psychol Rev 2007 Mar;27(2):155-172. [doi: 10.1016/j.cpr.2006.09.002] [Medline: 17112647]

40. Yap MB, Jorm AF. Parental factors associated with childhood anxiety, depression, and internalizing problems: a systematic review and meta-analysis. J Affect Disord 2015 Apr 01;175:424-440. [doi: 10.1016/j.jad.2015.01.050] [Medline: 25679197]
41. Axford N, Lehtonen M, Kaoukji D, Tobin K, Berry V. Engaging parents in parenting programs: lessons from research and practice. Child Youth Serv Rev 2012 Oct;34(10):2061-2071 [FREE Full text] [doi: 10.1016/j.childyouth.2012.06.011]
42. McCurdy K, Daro D. Parent involvement in family support programs: an Integrated Theory. Fam Relat 2001 Apr;50(2):113-121 [FREE Full text] [doi: 10.1111/j.1741-3729.2001.00113.x]
43. Spoth R, Redmond C. Research on family engagement in preventive interventions: toward improved use of scientific findings in primary prevention practice. J Prim Prev 2000;21(2):267-284. [doi: 10.1023/A:1007039421026]
44. Cuijpers P, van Straten A, Warmerdam L, van Rooy MJ. Recruiting participants for interventions to prevent the onset of depressive disorders: possible ways to increase participation rates. BMC Health Serv Res 2010;10:181 [FREE Full text] [doi: 10.1186/1472-6963-10-181] [Medline: 20579332]
45. Heinrichs N, Bertram H, Kuschel A, Hahlweg K. Parent recruitment and retention in a universal prevention program for child behavior and emotional problems: barriers to research and program participation. Prev Sci 2005 Dec;6(4):275-286. [doi: 10.1007/s11121-005-0006-1] [Medline: 16075192]
46. Mytton J, Ingram J, Manns S, Thomas J. Facilitators and barriers to engagement in parenting programs: a qualitative systematic review. Health Educ Behav 2014 Apr;41(2):127-137. [doi: 10.1177/1090198113485755] [Medline: 23640123]
47. Reavley NJ, Cvetkovski S, Jorm AF. Sources of information about mental health and links to help seeking: findings from the 2007 Australian National Survey of Mental Health and Wellbeing. Soc Psychiatry Psychiatr Epidemiol 2011 Dec;46(12):1267-1274. [doi: 10.1007/s00127-010-0301-4] [Medline: 20978883]
48. Powell J, Clarke A. Internet information-seeking in mental health: population survey. Br J Psychiatry Journal of Psychiatry 2006 Sep;189:273-277. [doi: 10.1192/bjp.bp.105.017319] [Medline: 16946364]
49. 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;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
50. Yap MB, Martin PD, Jorm AF. Online parenting guidelines to prevent adolescent depression and anxiety: evaluating user characteristics and usefulness. Early Interv Psychiatry 2017 Oct 27:1-8. [doi: 10.1111/eip.12503] [Medline: 29076246]
51. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
52. 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]
53. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
54. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
55. Breitenstein SM, Gross D, Christophersen R. Digital delivery methods of parenting training interventions: a systematic review. Worldviews Evid Based Nurs 2014 Jun;11(3):168-176. [doi: 10.1111/wvn.12040] [Medline: 24842341]
56. Sanders MR, Baker S, Turner KM. A randomized controlled trial evaluating the efficacy of Triple P Online with parents of children with early-onset conduct problems. Behav Res Ther 2012 Nov;50(11):675-684. [doi: 10.1016/j.brat.2012.07.004] [Medline: 22982082]
57. Enebrink P, Högström J, Forster M, Ghaderi A. Internet-based parent management training: a randomized controlled study. Behav Res Ther 2012 Apr;50(4):240-249. [doi: 10.1016/j.brat.2012.01.006] [Medline: 22398153]
58. Yap MB, Lawrence KA, Rapee RM, Cardamone-Breen MC, Green J, Jorm AF. Partners in parenting: a multi-level web-based approach to support parents in prevention and early intervention for adolescent depression and anxiety. JMIR Ment Health 2017 Dec 19;4(4):e59 [FREE Full text] [doi: 10.2196/mental.8492] [Medline: 29258974]
59. Cardamone-Breen MC, Jorm AF, Lawrence KA, Mackinnon AJ, Yap MB. The parenting to reduce adolescent depression and anxiety scale: assessing parental concordance with parenting guidelines for the prevention of adolescent depression and anxiety disorders. PeerJ 2017;5:e3825 [FREE Full text] [doi: 10.7717/peerj.3825] [Medline: 28951815]
60. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Communications of the Association for Information Systems 2009;24:485-500 [FREE Full text]
61. Parenting Strategies Program. Melbourne; 2013. How to prevent depression and clinical anxiety in your teenager: strategies for parents URL: https://www.parentingstrategies.net/depression/media/pdf/Guidelines.pdf [accessed 2017-11-24] [WebCite Cache ID 6vCqbkNHX]
62. National Research Council. In: O'Connell ME, Boat T, Warner KE, editors. Preventing mental, emotional, and behavioral disorders among young people: progress and possibilities. Washington, DC: National Academies Press; 2009.
63. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]

64. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
65. Yap MB, Mahtani S, Rapee RM, Nicolas C, Lawrence KA, Mackinnon A, et al. A tailored web-based intervention to improve parenting risk and protective factors for adolescent depression and anxiety problems: postintervention findings from a randomized controlled trial. J Med Internet Res 2018 Jan 19;20(1):e17 [FREE Full text] [doi: 10.2196/jmir.9139] [Medline: 29351895]
66. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
67. Nicholas J, Proudfoot J, Parker G, Gillis I, Burckhardt R, Manicavasagar V, et al. The ins and outs of an online bipolar education program: a study of program attrition. J Med Internet Res 2010;12(5):e57 [FREE Full text] [doi: 10.2196/jmir.1450] [Medline: 21169169]
68. Royston P, Babiker A. A menu-driven facility for complex sample size calculation in randomized controlled trials with a survival or a binary outcome. Stata J 2002;2(2):151-163.
69. Cardamone-Breen MC, Jorm AF, Lawrence KA, Mackinnon AJ, Yap MBH. Development and validation of the parenting to reduce adolescent depression and anxiety scale - adolescent-report (PRADAS-A). Manuscript in preparation 2018 (forthcoming).
70. Spence SH. Structure of anxiety symptoms among children: a confirmatory factor-analytic study. J Abnorm Psychol 1997 May;106(2):280-297. [Medline: 9131848]
71. Nauta MH, Scholing A, Rapee RM, Abbott M, Spence SH, Waters A. A parent-report measure of children's anxiety: psychometric properties and comparison with child-report in a clinic and normal sample. Behav Res Ther 2004 Jul;42(7):813-839. [doi: 10.1016/S0005-7967(03)00200-6] [Medline: 15149901]
72. McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods 2017 May 29:e1-22. [doi: 10.1037/met0000144] [Medline: 28557467]
73. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 2014 Aug;105(3):399-412. [doi: 10.1111/bjop.12046] [Medline: 24844115]
74. Angold A, Costello E, Messer S, Pickles A, Winder F, Silver D. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. Int J Methods Psychiatr Res 1995;5:237-249.
75. Spence SH. Scaswebsite. Normative data for total SCAS Scores URL: http://scaswebsite.com/docs/normstotalscas.pdf [accessed 2018-01-17] [WebCite Cache ID 6wWj7fKyP]
76. 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]
77. Mallinckrodt CH, Sanger TM, Dubé S, DeBrota DJ, Molenberghs G, Carroll RJ, et al. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. Biol Psychiatry 2003 Apr 15;53(8):754-760. [Medline: 12706959]
78. 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]
79. McKenzie DP, Toumbourou JW, Forbes AB, Mackinnon AJ, McMorris BJ, Catalano RF, et al. Predicting future depression in adolescents using the Short Mood and Feelings Questionnaire: a two-nation study. J Affect Disord 2011 Nov;134(1-3):151-159 [FREE Full text] [doi: 10.1016/j.jad.2011.05.022] [Medline: 21669461]
80. Rhew IC, Simpson K, Tracy M, Lymp J, McCauley E, Tsuang D, et al. Criterion validity of the Short Mood and Feelings Questionnaire and one- and two-item depression screens in young adolescents. Child Adolesc Psychiatry Ment Health 2010 Feb 09;4(1):8 [FREE Full text] [doi: 10.1186/1753-2000-4-8] [Medline: 20181135]
81. Thapar A, McGuffin P. Validity of the shortened Mood and Feelings Questionnaire in a community sample of children and adolescents: a preliminary research note. Psychiatry Res 1998 Nov 16;81(2):259-268. [Medline: 9858042]
82. Sandler IN, Schoenfelder EN, Wolchik SA, MacKinnon DP. Long-term impact of prevention programs to promote effective parenting: lasting effects but uncertain processes. Annu Rev Psychol 2011;62:299-329 [FREE Full text] [doi: 10.1146/annurev.psych.121208.131619] [Medline: 20822438]
83. Boyd RC, Gillham JE. Review of interventions for parental depression from toddlerhood to adolescence. Curr Psychiatry Rev 2009 Nov 01;5(4):226-235 [FREE Full text] [Medline: 20824114]
84. Sandler I, Ingram A, Wolchik S, Tein J, Winslow E. Long-Term effects of parenting-focused preventive interventions to promote resilience of children and adolescents. Child Dev Perspect 2015 Jul 15;9(3):164-171. [doi: 10.1111/cdep.12126] [Medline: 24843434]
85. Human LJ, Dirks MA, DeLongis A, Chen E. Congruence and incongruence in adolescents' and parents' perceptions of the family: using response surface analysis to examine links with adolescents' psychological adjustment. J Youth Adolesc 2016 Oct;45(10):2022-2035. [doi: 10.1007/s10964-016-0517-z] [Medline: 27287397]
86. De Los Reyes A, Ohannessian CM. Introduction to the special issue: discrepancies in adolescent-parent perceptions of the family and adolescent adjustment. J Youth Adolesc 2016 Oct;45(10):1957-1972. [doi: 10.1007/s10964-016-0533-z] [Medline: 27384957]

87. Bögels SM, Melick MV. The relationship between child-report, parent self-report, and partner report of perceived parental rearing behaviors and anxiety in children and parents. Pers Individ Dif 2004 Dec;37(8):1583-1596. [doi: 10.1016/j.paid.2004.02.014]
88. Ohannessian CM, Lerner JV, Lerner RM, von Eye A. Adolescent-parent discrepancies in perceptions of family functioning and early adolescent self-competence. Int J Behav Dev 2016 Jun 30;24(3):362-372. [doi: 10.1080/016502500501183 58]
89. Siegenthaler E, Munder T, Egger M. Effect of preventive interventions in mentally ill parents on the mental health of the offspring: systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry 2012 Jan;51(1):8-17.e8. [doi: 10.1016/j.jaac.2011.10.018] [Medline: 22176935]
90. Compas BE, Forehand R, Keller G, Champion JE, Rakow A, Reeslund KL, et al. Randomized controlled trial of a family cognitive-behavioral preventive intervention for children of depressed parents. J Consult Clin Psychol 2009 Dec;77(6):1007-1020 [FREE Full text] [doi: 10.1037/a0016930] [Medline: 19968378]
91. Yap MB, Allen NB, Sheeber L. Using an emotion regulation framework to understand the role of temperament and family processes in risk for adolescent depressive disorders. Clin Child Fam Psychol Rev 2007 Jun;10(2):180-196. [doi: 10.1007/s10567-006-0014-0] [Medline: 17265137]
92. Lengua LJ, Kovacs EA. Bidirectional associations between temperament and parenting and the prediction of adjustment problems in middle childhood. J Appl Dev Psychol 2005 Jan;26(1):21-38. [doi: 10.1016/j.appdev.2004.10.001]
93. Garber J. Depression in children and adolescents: linking risk research and prevention. Am J Prev Med 2006 Dec;31(6 Suppl 1):S104-S125. [doi: 10.1016/j.amepre.2006.07.007] [Medline: 17175406]
94. Mathew AR, Pettit JW, Lewinsohn PM, Seeley JR, Roberts RE. Co-morbidity between major depressive disorder and anxiety disorders: shared etiology or direct causation? Psychol Med 2011 Oct;41(10):2023-2034 [FREE Full text] [doi: 10.1017/S0033291711000407] [Medline: 21439108]
95. Rapee RM, Schniering CA, Hudson JL. Anxiety disorders during childhood and adolescence: origins and treatment. Annu Rev Clin Psychol 2009;5:311-341. [doi: 10.1146/annurev.clinpsy.032408.153628] [Medline: 19152496]
96. Panter-Brick C, Burgess A, Eggerman M, McAllister F, Pruett K, Leckman JF. Practitioner review: engaging fathers--recommendations for a game change in parenting interventions based on a systematic review of the global evidence. J Child Psychol Psychiatry 2014 Nov;55(11):1187-1212 [FREE Full text] [doi: 10.1111/jcpp.12280] [Medline: 24980187]
97. Teubert D, Pinquart M. A meta-analytic review on the prevention of symptoms of anxiety in children and adolescents. J Anxiety Disord 2011 Dec;25(8):1046-1059. [doi: 10.1016/j.janxdis.2011.07.001]
98. Horowitz JL, Garber J. The prevention of depressive symptoms in children and adolescents: a meta-analytic review. J Consult Clin Psychol 2006 Jun;74(3):401-415. [doi: 10.1037/0022-006X.74.3.401] [Medline: 16822098]
99. Gladstone TR, Beardslee WR. The prevention of depression in children and adolescents: a review. Can J Psychiatry 2009 Apr;54(4):212-221. [doi: 10.1177/070674370905400402] [Medline: 19321027]
100. van Zoonen K, Buntrock C, Ebert DD, Smit F, Reynolds CF, Beekman AT, et al. Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. Int J Epidemiol 2014 Apr;43(2):318-329 [FREE Full text] [doi: 10.1093/ije/dyt175] [Medline: 24760873]
101. Rose G. Mental disorder and the strategies of prevention. Psychol Med 1993 Aug;23(3):553-555. [Medline: 8234564]
102. Dittman CK, Farruggia SP, Palmer ML, Sanders MR, Keown LJ. Predicting success in an online parenting intervention: the role of child, parent, and family factors. J Fam Psychol 2014 Apr;28(2):236-243. [doi: 10.1037/a0035991] [Medline: 24611694]
103. 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 [FREE Full text] [doi: 10.1093/jpepsy/jsn099] [Medline: 18794187]
104. Bornstein MH, editor. Socioeconomic status and parenting. In: Handbook of parenting Volume 2: Biology and ecology of parenting. Mahwah, New Jersey: Laurence Erllbaum Associates; 2002:231-252.
105. Conger RD, Conger KJ, Martin MJ. Socioeconomic status, family processes, and individual development. J Marriage Fam 2010 Jun;72(3):685-704 [FREE Full text] [doi: 10.1111/j.1741-3737.2010.00725.x] [Medline: 20676350]
106. Lawrence D, Johnson S, Hafekost J, de HK, Sawyer M, Ainley J. Health.gov. Canberra: Australian Government Department of Health; 2015. The mental health of children and adolescents: report on the second Australian Child and adolescent survey of mental health and wellbeing URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/9DA8CA21306FE6EDCA257E2700016945/$File/child2.pdf [WebCite Cache ID 6whiEgRII]
107. Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]

Abbreviations

RCT: randomized controlled trial
PRADAS: Parenting to Reduce Adolescent Depression and Anxiety Scale (-A: Adolescent report)
SCAS: Spence Children's Anxiety Scale (-C: Child report; -P: Parent report)

SMFQ: Short Mood and Feelings Questionnaire (-C: Child report; -P: Parent report)
MMRM: Mixed effect Model Repeated Measures
PiP: Partners in Parenting

Edited by G Eysenbach; submitted 29.11.17; peer-reviewed by C Gilligan, Y Perry, A Calear, M Deady, C Buntrock, J Dudeney; comments to author 16.12.17; revised version received 19.02.18; accepted 08.03.18; published 26.04.18.

Please cite as:
Cardamone-Breen MC, Jorm AF, Lawrence KA, Rapee RM, Mackinnon AJ, Yap MBH
A Single-Session, Web-Based Parenting Intervention to Prevent Adolescent Depression and Anxiety Disorders: Randomized Controlled Trial
J Med Internet Res 2018;20(4):e148
URL: http://www.jmir.org/2018/4/e148/
doi:10.2196/jmir.9499
PMID:29699964

Original Paper

Efficacy of a Community-Based Technology-Enabled Physical Activity Counseling Program for People With Knee Osteoarthritis: Proof-of-Concept Study

Linda C Li[1,2], PT, PhD; Eric C Sayre[2], PhD; Hui Xie[2,3], PhD; Ryan S Falck[1], MSc; John R Best[1], PhD; Teresa Liu-Ambrose[1], PT, PhD; Navi Grewal[2]; Alison M Hoens[1], BSc (PT), MSc; Greg Noonan[4], BSc (PT); Lynne M Feehan[1,2], PT, PhD

[1]Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada

[2]Arthritis Research Canada, Richmond, BC, Canada

[3]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[4]Mary Pack Arthritis Program, Vancouver General Hospital, Vancouver, BC, Canada

Corresponding Author:
Linda C Li, PT, PhD
Arthritis Research Canada
5591 No. 3 Road
Richmond, BC, V6X 2C7
Canada
Phone: 1 604 207 4020
Email: lli@arthritisresearch.ca

Abstract

Background: Current practice guidelines emphasize the use of physical activity as the first-line treatment of knee osteoarthritis; however, up to 90% of people with osteoarthritis are inactive.

Objective: We aimed to assess the efficacy of a technology-enabled counseling intervention for improving physical activity in people with either a physician-confirmed diagnosis of knee osteoarthritis or having passed two validated criteria for early osteoarthritis.

Methods: We conducted a proof-of-concept randomized controlled trial. The immediate group received a brief education session by a physical therapist, a Fitbit Flex, and four biweekly phone calls for activity counseling. The delayed group received the same intervention 2 months later. Participants were assessed at baseline (T0) and at the end of 2 months (T1), 4 months (T2), and 6 months (T3). Outcomes included (1) mean time on moderate-to-vigorous physical activity (MVPA ≥3 metabolic equivalents [METs], primary outcome), (2) mean time on MVPA ≥4 METs, (3) mean daily steps, (4) mean time on sedentary activities, (5) Knee Injury and Osteoarthritis Outcome Score (KOOS), and (6) Partners in Health scale. Mixed-effects repeated measures analysis of variance was used to assess five planned contrasts of changes in outcome measures over measurement periods. The five contrasts were (1) immediate T1-T0 vs delayed T1-T0, (2) delayed T2-T1 vs delayed T1-T0, (3) mean of contrast 1 and contrast 2, (4) immediate T1-T0 vs delayed T2-T1, and (5) mean of immediate T2-T1 and delayed T3-T2. The first three contrasts estimate the between-group effects. The latter two contrasts estimate the effect of the 2-month intervention delay on outcomes.

Results: We recruited 61 participants (immediate: n=30; delayed: n=31). Both groups were similar in age (immediate: mean 61.3, SD 9.4 years; delayed: mean 62.1, SD 8.5 years) and body mass index (immediate: mean 29.2, SD 5.5 kg/m^2; delayed: mean 29.2, SD 4.8 kg/m^2). Contrast analyses revealed significant between-group effects in MVPA ≥3 METs (contrast 1 coefficient: 26.6, 95% CI 4.0-49.1, *P*=.02; contrast 3 coefficient: 26.0, 95% CI 3.1-49.0, *P*=.03), daily steps (contrast 1 coefficient: 1699.2, 95% CI 349.0-3049.4, *P*=.02; contrast 2 coefficient: 1601.8, 95% CI 38.7-3164.9, *P*=.045; contrast 3 coefficient: 1650.5, 95% CI 332.3-2968.7; *P*=.02), KOOS activity of daily living subscale (contrast 1 coefficient: 6.9, 95% CI 0.1-13.7, *P*=.047; contrast 3 coefficient: 7.2, 95% CI 0.8-13.6, *P*=.03), and KOOS quality of life subscale (contrast 1 coefficient: 7.4, 95% CI 0.0-14.7, *P*=.049; contrast 3 coefficient: 7.3, 95% CI 0.1-14.6, *P*=.048). We found no significant effect in any outcome measures due to the 2-month delay of the intervention.

Conclusions: Our counseling program improved MVPA ≥3 METs, daily steps, activity of daily living, and quality of life in people with knee osteoarthritis. These findings are important because an active lifestyle is an important component of successful self-management.

Trial Registration: ClinicalTrials.gov NCT02315664; https://clinicaltrials.gov/ct2/show/NCT02315664 (Archived by WebCite at http://www.webcitation.org/6ynSgUyUC)

(J Med Internet Res 2018;20(4):e159) doi:10.2196/jmir.8514

KEYWORDS

osteoarthritis; physical activity; wearables; goal setting; physiotherapy; eHealth

Introduction

Arthritis is the most common cause of severe chronic pain and disability worldwide. Analysis by the Arthritis Alliance of Canada estimates one new diagnosis of osteoarthritis (OA) every 60 seconds, resulting in nearly 30% of the employed labor force having difficulties working due to OA [1]. Current evidence supports the use of physical activity to manage OA due to its beneficial effects on pain, mobility, and quality of life [2,3]. It has been shown that moderate weight-bearing activities improve joint health by preserving glycosaminoglycan content in cartilage [4,5]. Furthermore, specific training that involves functional activities improves balance and proprioception, which in turn can contribute to improving pain and mobility [6,7]. The OA Research Society International recommends the use of physical activity and therapeutic exercise as a first-line treatment of knee OA [8]. Public health guidelines recommend more than 150 minutes a week of moderate-to-vigorous physical activity (MVPA) performed in bouts of 10 minutes or more [9]; however, a 2013 systematic review reported that only 13% of people with OA met this recommendation [10]. This concurs with another study using accelerometers that more than 90% of people with knee OA did not meet the physical activity guidelines [11]. The findings are particularly alarming because the evidence on the first-line treatment for OA has been consistent over a decade [12]. This represents a major knowledge-to-action gap.

Several modifiable risk factors are associated with low physical activity participation in people with arthritis. These include lack of motivation [13], doubts about the effectiveness of exercise [14], and lack of health professional advice [15]. Once patients start being active, they need feedback on their progress. A Cochrane review reported that "graded exercise activity," which initially focuses on simple activities and then gradually increases to more challenging ones, is effective for improving adherence in people with chronic musculoskeletal conditions [16]. Progression of activities and goals can be guided by a physical therapist (PT) [16].

We recently demonstrated feasibility of a physical activity counseling program with the use of a Fitbit wrist band in 34 people with knee OA [17]. Compared to controls, those who received the program showed a trend of increased MVPA and perceived self-management capacity after 1 month [17]. The findings supported further research on this program. The purpose of the current study was to assess the efficacy of the program for improving physical activity participation, disease status, and perceived self-management capacity in people with knee OA.

Methods

Study Design and Participant Eligibility

Monitor-OA was a proof-of-concept study that used a randomized, delayed-control design, whereby the randomization determined the timing of when the intervention was provided (ie, immediately vs a 2-month delay). As such, efficacy was assessed within a conventional randomized controlled trial (RCT), with an intervention group and a control group, at 2 months, whereas all participants received the intervention beyond this time. This study design is best suited for interventions that include components that are likely beneficial and low risk to participants, such as physical activity counseling.

Eligible individuals were those who had a physician-confirmed diagnosis of knee OA or passed two criteria for early OA: (1) aged 50 years and (2) had experienced pain or discomfort in or around the knee during the previous year lasting more than 28 separate or consecutive days. In a community-based study [18], 191 of 195 (97%) urban-dwelling participants who met these criteria also met the American College of Rheumatology clinical criteria for knee OA [19].

We excluded individuals who:

1. had a diagnosis of inflammatory arthritis, connective tissue diseases, fibromyalgia, or gout;
2. had used disease-modifying antirheumatic drugs or gout medications;
3. had knee arthroplasty;
4. were on a waitlist to receive knee or hip arthroplasty;
5. had any surgery in the back, hip, knee, foot, or ankle joint in the past 12 months;
6. had acute knee injury in the past 6 months;
7. had received a steroid injection or hyaluronate injection in a knee in the last 6 months;
8. had a body mass index (BMI) of 40 kg/m^2 or higher;
9. did not have an email address or daily access to a personal computer with Internet access;
10. were unable to attend the required education session in person;
11. were using medications that impaired activity tolerance (eg, beta-blockers); and
12. had an inappropriate level of risk for increasing their unsupervised physical activity.

Potential participants completed the Physical Activity Readiness Questionnaire (PAR-Q [20]; 2014 version). If a potential risk was identified by the PAR-Q, physician confirmation in writing

was required to ensure that the person was able to be physically active without supervision of a health professional.

Participants were recruited from the Mary Pack Arthritis Program in Vancouver, BC, Canada. Study information was also posted on social media (Facebook, Twitter, Kijiji, and Craigslist) and the Arthritis Research Canada website. In addition, emails about the study were sent by the Arthritis Consumer Experts, a nonprofit patient education organization, to their members. After completing the baseline assessment, eligible participants were randomly assigned to the immediate group or the delayed group (ie, control) in 1:1 allocation ratio. The delayed group received the same intervention as the immediate group after a 2-month wait. We performed randomization using computer-generated random numbers in variable block sizes.

Intervention

The intervention involved participants attending a 1.5-hour session, where they received (1) 15-minute standardized education about physical activity, (2) a Fitbit Flex, and (3) individual counseling with a study PT who was trained in motivational interviewing [21]. The choice of a face-to-face session over the use of videoconferencing technology was to maximize the opportunity for participants and PTs to established rapport [22,23]. The education portion, delivered in groups of two to four participants, addressed the benefits of an active lifestyle, the detrimental effect of sedentary behavior, and ways to be active without aggravating OA symptoms. The individual counseling portion followed the Brief Action Planning approach [24], whereby PTs guided participants to identify activity goals, develop an action plan, and identify barriers and solutions. The PTs used the SMART (specific, measurable, attainable, relevant, time-bound) principle during goal setting (eg, 30 minutes of brisk walking in the neighborhood in the evening three times a week). Participants were then asked to rate their confidence in executing the plan on a zero to 10 scale, with 10 meaning very confident. The process was repeated until the confidence rating reached 7 or higher out of 10. For sedentary behaviors, the PTs began by asking participants to estimate their sitting time in a normal day and identify ways to break up the sitting time. They then repeated the goal setting and confidence assessment.

Participants then received a Fitbit Flex to be worn at the wrist of the nondominant side 24 hours a day except during water-based activity or when charging. The physical activity data were wirelessly synchronized with Fitbit's online Dashboard that could be viewed only by the participants and their study PTs. During the intervention period, the PT reviewed the participant's physical activity on the Dashboard and progressively modified their SMART goals during four biweekly 20-minute phone calls. Participants could also contact the PT via email in-between the scheduled calls. At the end of the intervention, participants could keep the Fitbit.

Four PTs with a primary caseload consisting of patients with arthritis were trained to deliver the education and counseling components. Three of them were from the public sector (Vancouver Coastal Health Authority, Vancouver, BC, Canada), and one worked in a mix of public and private practices. One PT also participated in our previous feasibility study [17] and provided feedback to refine the intervention for the current project. The PTs attended a 2-day introductory motivational interview course offered by the University of British Columbia Extended Learning program. Before data collection, we held two orientation sessions (2 hours each) for the PTs to review the study protocol and practice the counseling component.

Outcome Measures

Participants were assessed at baseline (T0) and the end of 2 months (T1), 4 months (T2), and 6 months (T3). Our primary outcome measure was mean daily time performing MVPA at ≥3 metabolic equivalents, or METs (MVPA ≥3 METs; performed in both ambulatory and nonambulatory activities throughout the day) measured with SenseWear Mini, a multisensor monitor that was worn on the upper arm over the triceps. SenseWear integrates triaxial accelerometer data, physiological sensor data, and personal demographic information to provide estimates of steps and energy expenditure. Tierney et al [25] showed that SenseWear was a valid tool for estimating energy expenditure during activities of daily living in people with arthritis (intraclass correlation coefficient [ICC]=0.72). A strong relationship was also found between SenseWear and indirect calorimetry measures of energy expenditure for activities of daily living (Pearson r=.85) [25]. SenseWear can be worn 24 hours a day; hence, it can capture a full picture of physical activity and the off-body time throughout the day [26,27]. Participants wore a SenseWear for 7 days at each assessment. Almeida et al [28] determined that a minimum of 4 days of wear was required to reliably assess energy expenditure from different levels of physical activity in people with arthritis (ICC >0.80).

We performed additional analysis with a MVPA cut-off at ≥4 METs, which reflected an activity level of brisk walking and higher (ie, purposeful ambulatory activities) [29]. Other secondary outcomes included the daily mean time spent in sedentary activities [30], the Knee Injury and OA Outcome Score (KOOS) [31,32], and the Partners in Health scale [33]. An important feature of SenseWear is its ability to differentiate between sedentary and light activities [34], making it an ideal instrument to assess sedentary activities. For sedentary activities, we calculated the mean daily time spent with an energy expenditure of ≤1.5 METs, occurring in bouts of 20 minutes or more during waking hours [35-38].

The KOOS consists of five subscales: knee pain, stiffness, activity of daily living, sports/recreation, and quality of life. It was originally developed for people recovering from anterior cruciate ligament and meniscus injury and has been validated in people with OA [31,32]. The Partners in Health scale is a 12-item measure designed to assess perceived self-management capacity via subjective knowledge of the health condition and treatment, and perceived self-management behaviors (eg, adopting a healthy lifestyle; Cronbach alpha=.82) [33]. We also tracked self-reported adverse events (falls, cardiovascular and musculoskeletal events) [39] using a monthly log.

Sample Size Justification and Data Analysis

Our recruitment strategy enabled the study to enroll at least 60 eligible participants over 12 months. For a proof-of-concept

study, it is reasonable to expect a moderately large difference between groups after the intervention. Based on our feasibility study, we estimated the standard deviation of the change in MVPA ≥3 METs (primary outcome measure) from T0-T1 to be 40 minutes. This resulted in 81.5% power to detect a 30-minute difference between groups in the T1-T0 change via a two-sided test at alpha level of .05.

Descriptive analysis was used to summarize participant characteristics, comorbid conditions, and adverse events. We generated plots that included means and standard errors for the outcome measures at each time point for both groups for all outcome measures.

An intention-to-treat analysis was performed by a biostatistician who was blinded to the group assignment. Quantile-quantile plots were used to assess normality of the outcome variables. Mixed-effects repeated measures analysis of variance was used to assess five planned contrasts of changes in outcome measures over measurement periods. They were:

1. Contrast 1: immediate group T1-T0 vs delayed group T1-T0;
2. Contrast 2: delayed group T2-T1 vs delayed group T1-T0;
3. Contrast 3: mean of contrast 1 and contrast 2;
4. Contrast 4: immediate group T1-T0 vs delayed group T2-T1; and
5. Contrast 5: mean of immediate group T2-T1 and delayed group T3-T2.

The first three contrasts estimate the effect of the intervention versus control. The latter two contrasts estimate the effect of the delay on the intervention. Contrast 1 is the between-group contrast estimate for the 2-month effect of the program, similar to analysis from a parallel design. Contrast 2 is the within-group contrast estimate for the 2-month effect of the program. One benefit of the delayed-control design is that experimental units can serve as their own control to improve efficiency and precision of treatment effect estimation. To further improve the treatment effect estimation efficiency, we combined contrast 1 and 2 by taking their mean (contrast 3). Contrast 4 examines the 2-month effect of the program when it was delayed for 2 months. Contrast 5 combines data from both groups to assess the lastingness of the effect, when the program ended 2 months before assessment. No adjustment was made for multiple comparisons because type II error is a greater concern than type I error in proof-of-concept studies [40,41].

Ethics Approval

The research protocol was approved by the University of British Columbia Behavioral Research Ethics Board (application number: H14-01762) and was published in ClinicalTrials.gov (NCT02315664).

Results

In 2015-2016, 278 people indicated an interest to participate and 64 met the eligibility criteria (Figure 1). Of those, we recruited 61 participants (immediate group: n=30, 73%, 22/30 women; delayed group: n=31, 90%, 28/31 women). Both groups were similar in age (immediate group: mean 61.3, SD 9.4 years; delayed group: mean 62.1, SD 8.5 years) and BMI (immediate group: mean 29.2, SD 5.5 kg/m^2; delayed group: mean 29.2, SD 4.8 kg/m^2; Table 1).

Multimedia Appendix 1 and Figures 2-11 present the results of outcome measures from four time points. Prespecified contrast analyses revealed significant effects as follows: mean time on MVPA ≥3 METs (contrast 1 coefficient: 26.6, 95% CI 4.0-49.1, *P*=.02; contrast 3 coefficient: 26.0, 95% CI 3.1-49.0, *P*=.03), mean daily steps (contrast 1 coefficient: 1699.2, 95% CI 349.0-3049.4, *P*=.02; contrast 2 coefficient: 1601.8, 95% CI 38.7-3164.9, *P*=.045; contrast 3 coefficient: 1650.5, 95% CI 332.3-2968.7, *P*=.02), KOOS activity of daily living subscale (contrast 1 coefficient: 6.9, 95% CI 0.1-13.7, *P*=.047; contrast 3 coefficient: 7.2, 95% CI 0.8-13.6, *P*=.03), KOOS quality of life subscale (contrast 1 coefficient: 7.4, 95% CI 0.0-14.7, *P*=.049; contrast 3 coefficient: 7.3, 95% CI 0.1-14.6, *P*=.048). We found no significant effect in any outcome measures in the other contrast analyses. No adverse event associated with the intervention (eg, falls, cardiovascular and musculoskeletal events) was reported by participants during the study.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) flowchart.

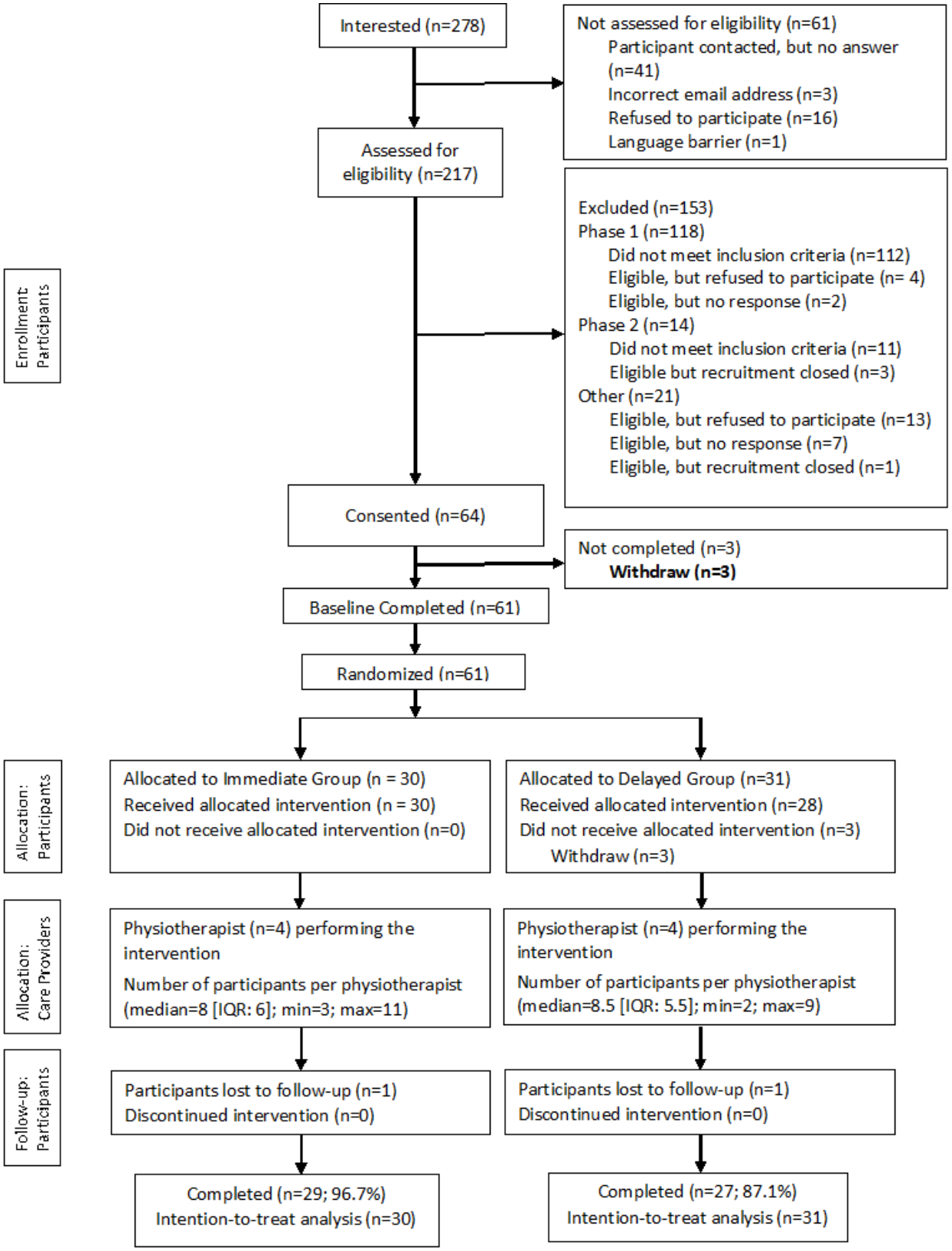


Table 1. Baseline characteristics of immediate group and delayed group participants.

Characteristics	All (N=61)	Immediate group (n=30)	Delayed group (n=31)	*P* value[a]
Gender (female), n (%)	50 (82)	22 (73)	28 (90)	.11
Age (years), mean (SD)	61.7 (8.9)	61.3 (9.4)	62.1 (8.5)	.72
Marital status, n (%)				.27
Married/Common law	33 (54)	19 (63)	14 (45)	
Separated/Divorced	15 (25)	7 (23)	8 (26)	
Widowed / Never married / Other	13 (21)	4 (13)	9 (29)	
University degree, n (%)	30 (49)	14 (47)	16 (52)	.80
Gross annual household income (CAN$), n (%)				.62
≤12,000	0	0	0	
12,001-24,000	3 (5)	1 (3)	2 (7)	
24,001-40,000	6 (10)	4 (13)	2 (7)	
40,001-60,000	11 (18)	5 (17)	6 (19)	
60,001-80,000	13 (21)	9 (30)	4 (13)	
80,001-100,000	5 (8)	2 (7)	3 (10)	
>100,000	11 (18)	5 (17)	6 (19)	
No answer	12 (20)	4 (13)	8 (26)	
Diagnosed with OA, n (%)				>.99
Yes	52 (85)	26 (87)	26 (84)	
No, but met the “likely OA” criteria	9 (15)	4 (13)	5 (16)	
In general, would you say your health is... n (%)				.68
Excellent	1 (2)	1 (3)	0	
Very good	29 (48)	15 (50)	14 (45)	
Good	24 (39)	10 (33)	14 (45)	
Fair	7 (12)	4 (13)	3 (10)	
Poor	0	0	0	
Compared to one year ago, how would you rate your health in general now? n (%)				.78
Much better	4 (7)	2 (7)	2 (6)	
Somewhat better	9 (15)	6 (20)	3 (10)	
About the same	27 (44)	12 (40)	15 (48)	
Somewhat worse	21 (34)	10 (33)	11 (36)	
Much worse	0	0	0	
Number of comorbid conditions, median (IQR)	3.0 (2.0-4.0)	2.0 (1.0-4.0)	4.0 (2.0-5.0)	.06
Body mass index (kg/m^2), mean (SD)	29.2 (5.1)	29.2 (5.5)	29.2 (4.8)	.95

[a]*P* values based on exact chi-square tests for categorical variables (nonmissing data) and two-sample *t* tests for continuous variables.

Figure 2. Bouted moderate to vigorous physical activity (≥3 metabolic equivalent tasks [METs]).

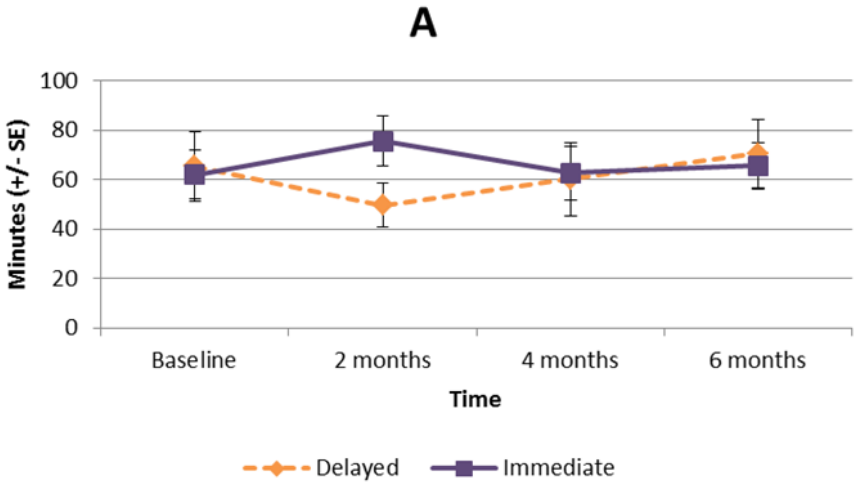


Figure 3. Bouted moderate to vigorous physical activity (≥4 METs).

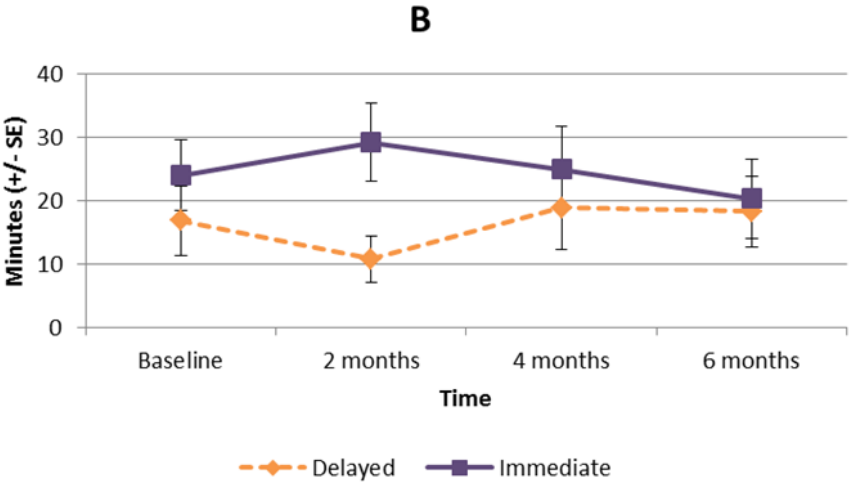


Figure 4. Bouted sedentary time.

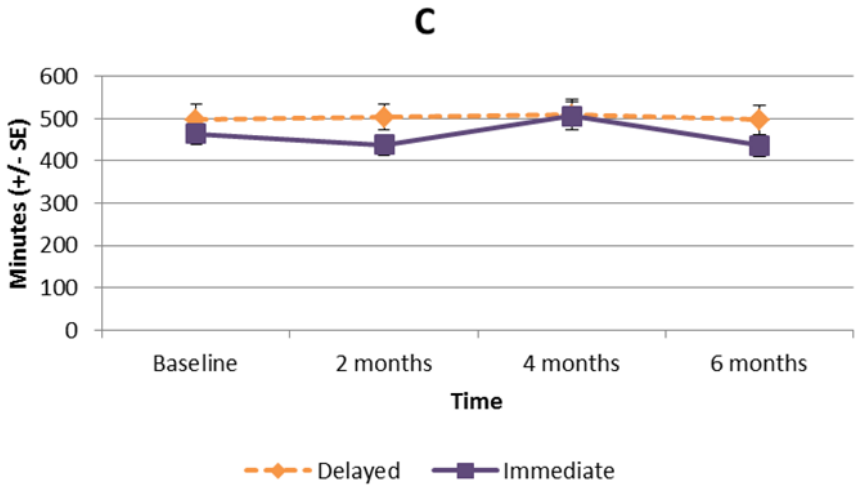


Figure 5. Mean daily step count.

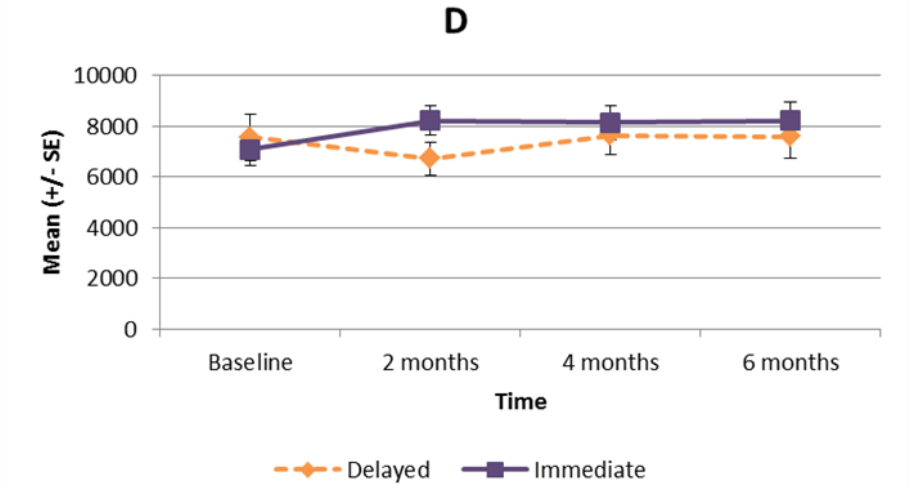


Figure 6. Knee Injury and Osteoarthritis Outcome Score (KOOS) symptoms subscale.

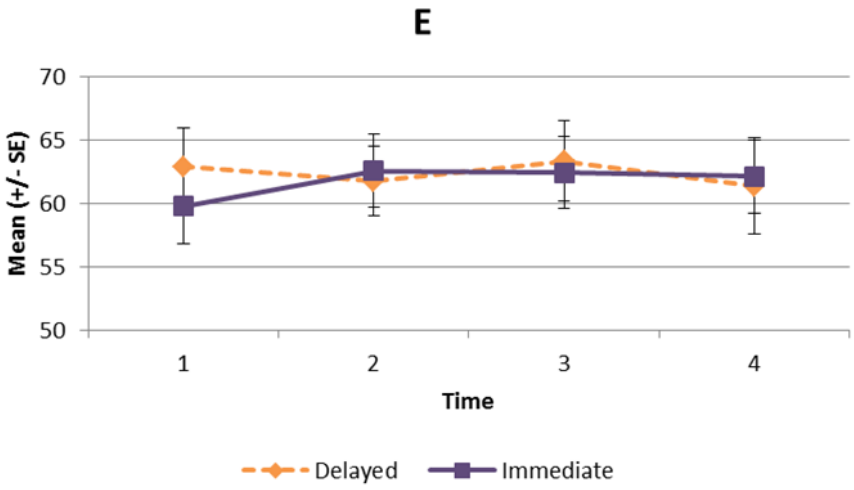


Figure 7. KOOS pain subscale.

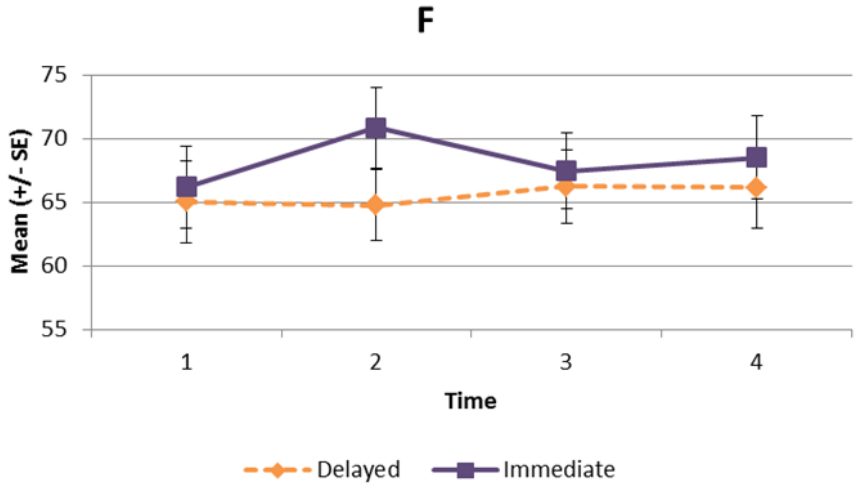


Figure 8. KOOS sports and recreation subscale.

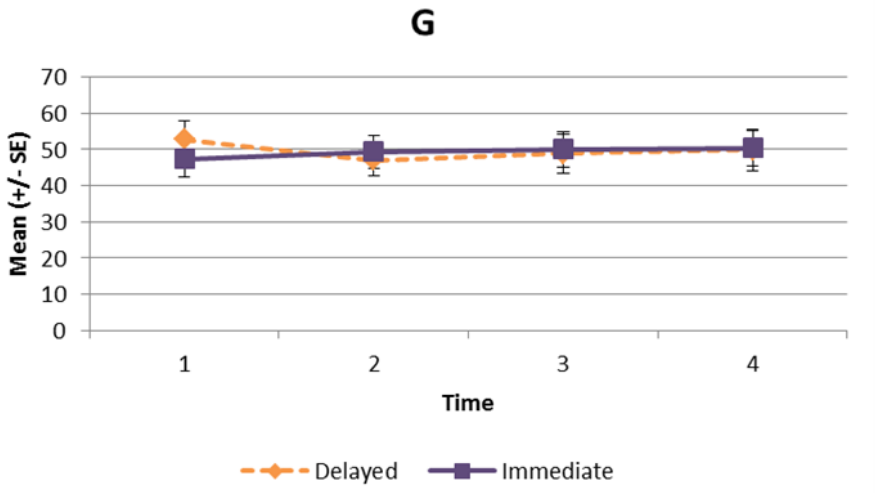


Figure 9. Partners in Health scale.

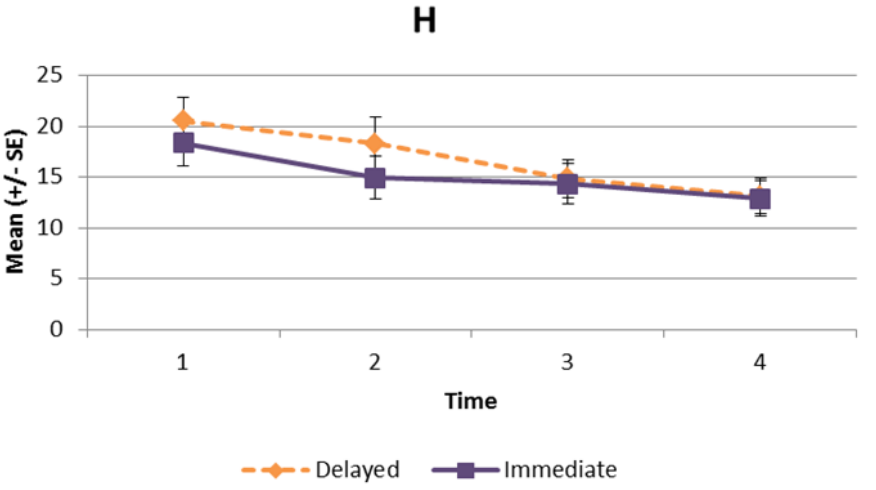


Figure 10. KOOS activities of daily living subscale.

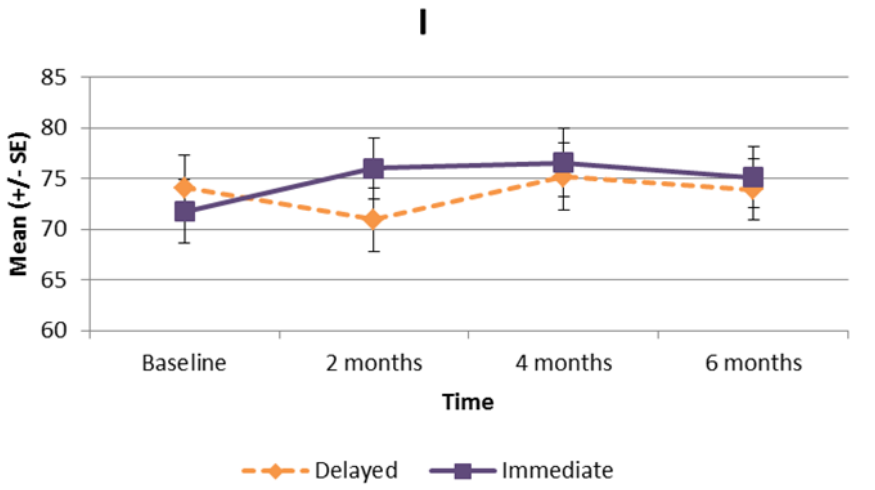
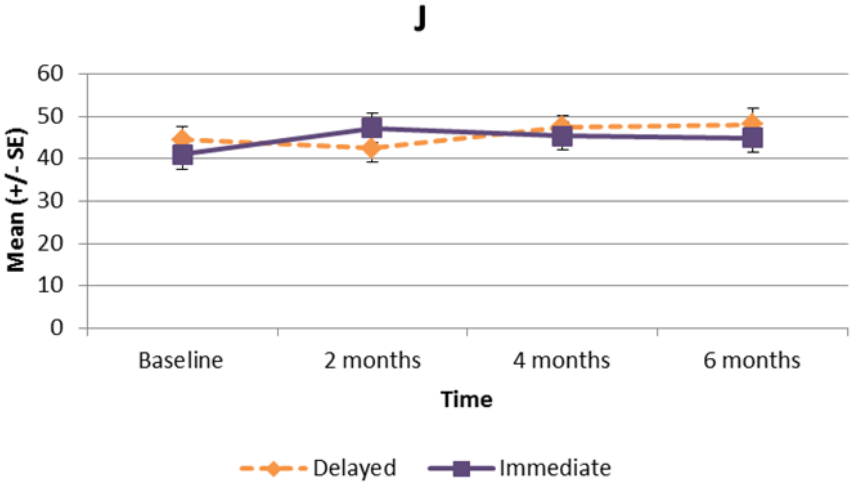


Figure 11. KOOS quality of life subscale.



Discussion

This proof-of-concept study showed that our intervention had a significant effect on time spent on physical activity (MVPA ≥3 METs) by people with knee OA compared to controls. The program also had a positive effect on step count, as well as activity of daily living and quality of life measured by the KOOS. Furthermore, the 2-month delay did not affect how participants responded to the program. This is noteworthy for the design of future RCTs in which the use of a "delay control" is considered. Our findings match those of previous studies that found individualized programs and self-management techniques could enhance physical activity adherence among people with chronic musculoskeletal conditions [16,17]. However, the findings should be viewed in the context that compared to the general population with knee OA, our participants were more active (baseline MVPA ≥3 METs for more than 60 minutes per day). A 2011 study in the United States using accelerometers found that more than 90% of people with knee OA failed to meet the physical activity guidelines of 150 minutes of MVPA per week [11].

Our positive results may be explained by two reasons. First, we employed a variety of behavior change techniques to promote physical activity. In a review of 13 consumer wearables, Lyons et al [42] found that Fitbit used behavior change techniques such as goal setting, feedback, self-monitoring, social support, social comparison, as well as providing instructions and rewards. Although many of these techniques are in line with recommendations from social cognitive theory [43], techniques such as action planning, problem solving, and behavioral practice/rehearsal are absent with the use of wearables alone [42]. In a 2015 systematic review, Lewis et al [44] reported that wearable-only interventions tend to produce only modest effects on improving physical activity behavior. Therefore, we included PT counselors with expertise in arthritis care and motivational interview skills to prompt participants to plan and practice their activities, and to identify potential barriers and solutions during the follow-up calls. Our findings also echo those of a recent RCT on a 12-week intervention involving an Internet-based physical activity program with the use of an accelerometer and remote coaching. Broekhuizen et al [45] reported that the intervention improved the emotional and mental health among community-dwelling older adults (mean age 65 years).

The second reason was related to the experience and training of our study PTs. Previous studies in health counseling have stressed the importance of the experience and skill of counselors in behavioral interventions [46,47]. In this study, all PTs had a clinical caseload primarily in arthritis, received the essential training on motivational interviewing, and were familiar with the counseling protocol. Taken together, these reasons might have contributed to the positive effects of the intervention.

However, we did not observe a significant effect in the mean time spent in sedentary activities. From the feasibility study, we learned that increasing physical activity and reducing sedentary behavior (eg, prolonged sitting) required distinct counseling approaches, and that practice would be required by the PTs to deliver the intervention [17]. Hence, all study PTs were provided opportunities to practice the counseling procedure with the investigators before data collection. We did not make major modification to the counseling protocol, and it was possible that some challenges persisted for participants to identify ways to break up their sitting time. It should be noted that sedentary behavior is linked to individuals' habitual routines [48-51]. Although the general recommendation of MVPA can be accomplished in short daily episodes (eg, 30 minutes of brisk walking), reducing sedentary time requires people adjusting their habits throughout the day (eg, computer/mobile device use, television viewing). Therefore, different counseling strategies are likely required for the two behaviors.

In light of the findings, we suggest two modifications for the physical activity counseling program. First, the counseling conversation should begin by examining the individual's habitual routines during a typical workday and a non-work/weekend day, and then focus on periods when prolonged sitting occurs. This would allow the conversation to center on identifying opportunities, challenges, and solutions to break up sitting time when it is feasible for the person. Second, a more flexible system would be needed for setting personal goals to reduce sedentary time. The current Fitbit-manufactured apps, both the Web-based and mobile versions, reward users if they take 250 steps or more

(approximately 2-3 minutes of walking) in a given hour. Although the parameter meets the general recommendation of standing up and moving after 30 minutes of uninterrupted sitting, it does not allow the flexibility to individualize goals to break up sitting. We suggest that future Fitbit-compatible apps should include functions for users to set personalized goals on the frequency and duration to break up their sitting time during the day, and to receive feedback on their goal attainment. This would provide users with positive reinforcement to adopt the current recommendation of reducing extended periods of sitting.

This study has a few limitations. With the use of a delayed-control design in which the participants in the control arm received treatment after a 2-month delay, efficacy of the physical activity counseling intervention could only be assessed at 2 months. Hence, the long-term effect of the intervention remained unclear. Furthermore, our sample was relatively active; hence, the results may not be generalizable to people with knee OA who are more sedentary. The results also may not be generalizable to men because 82% of the participants were women. Finally, we have identified several shortcomings in the counseling program that may have limited its potential to affect the behavioral and health-related outcomes. Despite these limitations, this proof-of-concept study has demonstrated a significant effect of a multifaceted counseling intervention on improving physical activity participation. We have since applied these learnings to improve the next iteration of the program. Specifically, it now includes a new sedentary counseling strategy and a Fitbit-compatible Web app with enhanced functionality for setting goals and rewarding behaviors that break up prolonged sitting [52]. The modified program is currently being tested in a RCT involving people with rheumatoid arthritis and systematic lupus erythematosus (ClinicalTrial.gov identifier: NCT02554474) [53].

Acknowledgments

The authors are grateful for the collaboration of patient partner Cheryl Koehn (Arthritis Consumer Experts) and physical therapists Karen Tsui, Valerie Gray, Carla Wiebe, and Wendy Watson. This study was supported by a Vancouver Coastal Health Research Institute Innovation & Translational Research Award.

LCL and TL-A are supported by the Canada Research Chair Program and the Michael Smith Foundation for Health Research (MSFHR) Scholar Award. LCL is also supported by the Harold Robinson/Arthritis Society Chair in Arthritic Diseases. HX is supported by the Maureen and Milan Ilich / Merck Chair in Statistics for Arthritis and Musculoskeletal Diseases. RSF is funded by the University of British Columbia Rehabilitation Sciences Scholarship and the Omer H Patrick II Memorial Award.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Participant outcomes and results of contrast analyses.

[PDF File (Adobe PDF File), 54KB - jmir_v20i4e159_app1.pdf]

References

1. Bombardier C, Hawker G, Mosher D. The Impact of Arthritis in Canada: Today and Over the Next 30 Years. Toronto, ON: Arthritis Alliance of Canada; 2011.
2. Brosseau L, MacLeay L, Robinson V, Wells G, Tugwell P. Intensity of exercise for the treatment of osteoarthritis. Cochrane Database Syst Rev 2003(2):CD004259. [doi: 10.1002/14651858.CD004259] [Medline: 12804510]
3. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007 Sep;15(9):981-1000 [FREE Full text] [doi: 10.1016/j.joca.2007.06.014] [Medline: 17719803]
4. Roos E, Dahlberg L. Positive effects of moderate exercise on glycosaminoglycan content in knee cartilage: a four-month, randomized, controlled trial in patients at risk of osteoarthritis. Arthritis Rheum 2005 Nov;52(11):3507-3514 [FREE Full text] [doi: 10.1002/art.21415] [Medline: 16258919]
5. Munukka M, Waller B, Häkkinen A, Nieminen M, Lammentausta E, Kujala U, et al. Physical activity Is related with cartilage quality in women with knee osteoarthritis. Med Sci Sports Exerc 2017 Jul;49(7):1323-1330. [doi: 10.1249/MSS.0000000000001238] [Medline: 28240703]
6. Skou S. Roos EM: Good Life with osteoArthritis in Denmark (GLA:D)vidence-based education and supervised neuromuscular exercise delivered by certified physiotherapists nationwide. BMC Musculoskelet Disord 2017;18(1):72. [doi: 10.1186/s12891-017-1439-y] [Medline: 5297181]
7. Davis AM, Kennedy D, Wong R, Robarts S, Skou S, McGlasson R, et al. Cross-cultural adaptation and implementation of Good Life with osteoarthritis in Denmark (GLA:D™): group education and exercise for hip and knee osteoarthritis is feasible in Canada. Osteoarthritis Cartilage 2018 Feb;26(2):211-219. [doi: 10.1016/j.joca.2017.11.005] [Medline: 29146385]

8. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008 Feb;16(2):137-162 [FREE Full text] [doi: 10.1016/j.joca.2007.12.013] [Medline: 18279766]
9. Haskell W, Lee I, Pate R, Powell K, Blair S, Franklin B, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
10. Wallis JA, Webster KE, Levinger P, Taylor NF. What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis. Osteoarthritis Cartilage 2013 Nov;21(11):1648-1659 [FREE Full text] [doi: 10.1016/j.joca.2013.08.003] [Medline: 23948979]
11. Dunlop DD, Song J, Semanik PA, Chang RW, Sharma L, Bathon JM, et al. Objective physical activity measurement in the osteoarthritis initiative: are guidelines being met? Arthritis Rheum 2011 Nov;63(11):3372-3382 [FREE Full text] [doi: 10.1002/art.30562] [Medline: 21792835]
12. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative. Semin Arthritis Rheum 2014 Jun;43(6):701-712. [doi: 10.1016/j.semarthrit.2013.11.012] [Medline: 24387819]
13. Gyurcsik N, Brawley L, Spink K, Brittain D, Fuller D, Chad K. Physical activity in women with arthritis: examining perceived barriers and self-regulatory efficacy to cope. Arthritis Rheum 2009 Aug 15;61(8):1087-1094 [FREE Full text] [doi: 10.1002/art.24697] [Medline: 19644901]
14. Der Ananian C, Wilcox S, Saunders R, Watkins K, Evans A. Factors that influence exercise among adults with arthritis in three activity levels. Prev Chronic Dis 2006 Jul;3(3):A81 [FREE Full text] [Medline: 16776882]
15. Henchoz Y, Zufferey P, So A. Stages of change, barriers, benefits, and preferences for exercise in RA patients: a cross-sectional study. Scand J Rheumatol 2013;42(2):136-145. [doi: 10.3109/03009742.2012.724707] [Medline: 23244196]
16. Jordan J, Holden M, Mason EJ, Foster NE. Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. Cochrane Database Syst Rev 2010 Jan 20(1):CD005956. [doi: 10.1002/14651858.CD005956.pub2] [Medline: 20091582]
17. Li L, Sayre E, Xie H, Clayton C, Feehan LM. A community-based physical activity counselling program for people with knee osteoarthritis: feasibility and preliminary efficacy of the track-OA study. JMIR Mhealth Uhealth 2017 Jun 26;5(6):e86 [FREE Full text] [doi: 10.2196/mhealth.7863] [Medline: 28652228]
18. Marra C, Cibere J, Tsuyuki R, Soon J, Esdaile J, Gastonguay L, et al. Improving osteoarthritis detection in the community: pharmacist identification of new, diagnostically confirmed osteoarthritis. Arthritis Rheum 2007 Oct 15;57(7):1238-1244 [FREE Full text] [doi: 10.1002/art.23019] [Medline: 17907209]
19. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986 Aug;29(8):1039-1049. [Medline: 3741515]
20. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
21. Rollnick S, Miller W, Butler CC. Motivational Interviewing in Health Care: Helping Patients Change Behavior. New York: Guilford Press; Jul 02, 2009.
22. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. 3rd ed. New York: Guilford Press; 2012.
23. Richards J, Hillsdon M, Thorogood M, Foster C. Face-to-face interventions for promoting physical activity. Cochrane Database Syst Rev 2013 Sep 30(9):CD010392. [doi: 10.1002/14651858.CD010392.pub2] [Medline: 24085592]
24. Gutnick D, Reims K, Davis C, Gainforth H, Jay M, Cole S. Brief action planning to facilitate behavior change and support patient self-management. J Clin Outcomes Manag 2014;21:17-29.
25. Tierney M, Fraser A, Purtill H, Kennedy N. Study to determine the criterion validity of the SenseWear Armband as a measure of physical activity in people with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2013 Jun;65(6):888-895 [FREE Full text] [doi: 10.1002/acr.21914] [Medline: 23213019]
26. Holsgaard-Larsen A, Roos EM. Objectively measured physical activity in patients with end stage knee or hip osteoarthritis. Eur J Phys Rehabil Med 2012 Dec;48(4):577-585 [FREE Full text] [Medline: 22641252]
27. Semanik P, Lee J, Manheim L, Dipietro L, Dunlop D, Chang RW. Relationship between accelerometer-based measures of physical activity and the Yale Physical Activity Survey in adults with arthritis. Arthritis Care Res (Hoboken) 2011 Dec;63(12):1766-1772 [FREE Full text] [doi: 10.1002/acr.20644] [Medline: 21954166]
28. Almeida G, Wasko M, Jeong K, Moore C, Piva SR. Physical activity measured by the SenseWear Armband in women with rheumatoid arthritis. Phys Ther 2011 Sep;91(9):1367-1376 [FREE Full text] [doi: 10.2522/ptj.20100291] [Medline: 21719635]
29. Ainsworth B, Haskell EW. Compendium of Physical Activities Tracking Guide. URL: https://sites.google.com/site/compendiumofphysicalactivities/ [accessed 2018-04-06] [WebCite Cache ID 6yU2kq5Q1]
30. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Dec 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]

31. Roos E, Roos P, Lohmander L, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88-96. [doi: 10.2519/jospt.1998.28.2.88] [Medline: 9699158]
32. Roos E, Roos P, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version. Scand J Med Sci Sports 1998 Dec;8(6):439-448. [Medline: 9863983]
33. 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]
34. Feehan L, Goldsmith CH, Leung AF, Li LC. SenseWearMini and Actigraph GT3X accelerometer classification of observed sedentary and light-intensity physical activities in a laboratory setting. Physiother Can 2016 May;68(2):116-123. [doi: 10.3138/PTC.2015-12]
35. Owen N. Sedentary behavior: understanding and influencing adults' prolonged sitting time. Prev Med 2012 Dec;55(6):535-539. [doi: 10.1016/j.ypmed.2012.08.024] [Medline: 22968124]
36. Dunstan D, Kingwell B, Larsen R, Healy G, Cerin E, Hamilton M, et al. Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 2012 May;35(5):976-983 [FREE Full text] [doi: 10.2337/dc11-1931] [Medline: 22374636]
37. Latouche C, Jowett J, Carey A, Bertovic D, Owen N, Dunstan D, et al. Effects of breaking up prolonged sitting on skeletal muscle gene expression. J Appl Physiol (1985) 2013 Feb 15;114(4):453-460 [FREE Full text] [doi: 10.1152/japplphysiol.00978.2012] [Medline: 23271697]
38. Howard B, Fraser S, Sethi P, Cerin E, Hamilton M, Owen N, et al. Impact on hemostatic parameters of interrupting sitting with intermittent activity. Med Sci Sports Exerc 2013 Jul;45(7):1285-1291. [doi: 10.1249/MSS.0b013e318285f57e] [Medline: 23439415]
39. Ory M, Resnick B, Jordan P, Coday M, Riebe D, Ewing GC, et al. Screening, safety, and adverse events in physical activity interventions: collaborative experiences from the behavior change consortium. Ann Behav Med 2005 Apr;29 Suppl:20-28. [doi: 10.1207/s15324796abm2902s_5] [Medline: 15921486]
40. Schoenfeld D. Statistical considerations for pilot studies. Int J Radiat Oncol Biol Phys 1980 Mar;6(3):371-374. [Medline: 7390914]
41. Carter R, Woolson RF. Statistical design considerations for pilot studies transitioning therapies from the bench to the bedside. J Transl Med 2004 Oct 28;2(1):37 [FREE Full text] [doi: 10.1186/1479-5876-2-37] [Medline: 15511289]
42. Lyons J, Lewis H, Mayrsohn G, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
43. Task Force on Community Preventive Services. Recommendations to increase physical activity in communities. Am J Prev Med 2002 May;22(4 Suppl):67-72. [Medline: 11985935]
44. Lewis Z, Lyons E, Jarvis J, Baillargeon J. Using an electronic activity monitor system as an intervention modality: A systematic review. BMC Public Health 2015 Jun 24;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]
45. Broekhuizen K, de Gelder J, Wijsman C, Wijsman L, Westendorp R, Verhagen E, et al. An Internet-based physical activity intervention to improve quality of life of inactive older adults: a randomized controlled trial. J Med Internet Res 2016 Apr 27;18(4):e74 [FREE Full text] [doi: 10.2196/jmir.4335] [Medline: 27122359]
46. Tulloch H, Fortier M, Hogg W. Physical activity counseling in primary care: who has and who should be counseling? Patient Educ Couns 2006 Dec;64(1-3):6-20. [doi: 10.1016/j.pec.2005.10.010] [Medline: 16472959]
47. Hutchison A, Breckon J, Johnston LH. Physical activity behavior change interventions based on the transtheoretical model: a systematic review. Health Educ Behav 2009 Oct;36(5):829-845. [doi: 10.1177/1090198108318491] [Medline: 18607007]
48. O'Donoghue G, Perchoux C, Mensah K, Lakerveld J, van der Ploeg H, Bernaards C, DEDIPAC Consortium. A systematic review of correlates of sedentary behaviour in adults aged 18-65 years: a socio-ecological approach. BMC Public Health 2016 Feb 17;16:163 [FREE Full text] [doi: 10.1186/s12889-016-2841-3] [Medline: 26887323]
49. Rutten G, Savelberg H, Biddle S, Kremers SP. Interrupting long periods of sitting: good STUFF. Int J Behav Nutr Phys Act 2013 Jan 02;10:1 [FREE Full text] [doi: 10.1186/1479-5868-10-1] [Medline: 23281722]
50. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2016;10(1):89-112 [FREE Full text] [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
51. Gardner B, Lally P, Wardle J. Making health habitual: the psychology of 'habit-formation' and general practice. Br J Gen Pract 2012 Dec;62(605):664-666 [FREE Full text] [doi: 10.3399/bjgp12X659466] [Medline: 23211256]
52. Gupta A, Tong X, Shaw C, Li L, Feehan L. In: Stephanidis C, editor. HCI International 19th International Conference, Proceedings, Part II. Cham, Switzerland: Springer; 2017:232-240.
53. Li LC, Feehan LM, Shaw C, Xie H, Sayre EC, Aviña-Zubeita A, et al. A technology-enabled counselling program versus a delayed treatment control to support physical activity participation in people with inflammatory arthritis: study protocol for the OPAM-IA randomized controlled trial. BMC Rheumatol 2017 Nov 28;1(1):1. [doi: 10.1186/s41927-017-0005-4]

Abbreviations

BMI: body mass index
ICC: intraclass correlation coefficient
KOOS: Knee Injury and Osteoarthritis Outcome Score
MET: metabolic equivalent
MVPA: moderate-to-vigorous physical activity
OA: osteoarthritis
PAR-Q: Physical Activity Readiness Questionnaire
PT: physical therapist
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 22.01.18; peer-reviewed by H Nero, A Schäfer; comments to author 21.02.18; revised version received 19.03.18; accepted 20.03.18; published 30.04.18.

Please cite as:
Li LC, Sayre EC, Xie H, Falck RS, Best JR, Liu-Ambrose T, Grewal N, Hoens AM, Noonan G, Feehan LM
Efficacy of a Community-Based Technology-Enabled Physical Activity Counseling Program for People With Knee Osteoarthritis: Proof-of-Concept Study
J Med Internet Res 2018;20(4):e159
URL: http://www.jmir.org/2018/4/e159/
doi:10.2196/jmir.8514
PMID:29712630

Original Paper

Identification of Users for a Smoking Cessation Mobile App: Quantitative Study

SK Leon Chevalking[1*], MSc; Somaya Ben Allouch[1*], PhD; Marjolein Brusse-Keizer[2*], PhD; Marloes G Postel[3], PhD; Marcel E Pieterse[4], PhD

[1]Research Group Technology, Health & Care, Saxion University of Applied Sciences, Enschede, Netherlands

[2]Medical School Twente, Medisch Spectrum Twente, Enschede, Netherlands

[3]Tactus Addiction Treatment, University of Twente, Enschede, Netherlands

[4]Centre for eHealth and Well-Being Research, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

[*]these authors contributed equally

Corresponding Author:
Somaya Ben Allouch, PhD
Research Group Technology, Health & Care
Saxion University of Applied Sciences
MH Tromplaan 28
7513 AB
Enschede, 70.000 PO Box
Netherlands
Phone: 31 612292901
Email: s.benallouch@saxion.nl

Abstract

Background: The number of mobile apps that support smoking cessation is growing, indicating the potential of the mobile phone as a means to support cessation. Knowledge about the potential end users for cessation apps results in suggestions to target potential user groups in a dissemination strategy, leading to a possible increase in the satisfaction and adherence of cessation apps.

Objective: This study aimed to characterize potential end users for a specific mobile health (mHealth) smoking cessation app.

Methods: A quantitative study was conducted among 955 Dutch smokers and ex-smokers. The respondents were primarily recruited from addiction care facilities and hospitals through Web-based media via websites and forums. The respondents were surveyed on their demographics, smoking behavior, and personal innovativeness. The intention to use and the attitude toward a cessation app were determined on a 5-point Likert scale. To study the association between the characteristics and intention to use and attitude, univariate and multivariate ordinal logistic regression analyses were performed.

Results: The multivariate ordinal logistic regression showed that the number of previous quit attempts (odds ratio [OR] 4.1, 95% CI 2.4-7.0, and OR 3.5, 95% CI 2.0-5.9) and the score on the Fagerstrom Test of Nicotine Dependence (OR 0.8, 95% CI 0.8-0.9, and OR 0.8, 95% CI 0.8-0.9) positively correlates with the intention to use a cessation app and the attitude toward cessation apps, respectively. Personal innovativeness also positively correlates with the intention to use (OR 0.3, 95% CI 0.2-0.4) and the attitude towards (OR 0.2, 95% CI 0.1-0.4) a cessation app. No associations between demographics and the intention to use or the attitude toward using a cessation app were observed.

Conclusions: This study is among the first to show that demographic characteristics such as age and level of education are not associated with the intention to use and the attitude toward using a cessation app when characteristics related specifically to the app, such as nicotine dependency and the number of quit attempts, are present in a multivariate regression model. This study shows that the use of mHealth apps depends on characteristics related to the content of the app rather than general user characteristics.

(J Med Internet Res 2018;20(4):e118) doi:10.2196/jmir.7606

KEYWORDS

mobile applications; telemedicine; mHealth; eHealth; tobacco; smoking cessation; health informatics

Introduction

Background

Smoking is a serious health problem. The World Health Organization estimates that over 1 billion people worldwide smoke tobacco, and the deaths of 6 million people per year can be attributed to causes linked to smoking [1]. The global prevalence of tobacco use is estimated to be 19%: 1 of 5 people older than 15 years of age uses tobacco, and in developed countries, this statistic is even higher, that is, 1 of 4 people older than 15 years uses tobacco [2]. The prevalence of tobacco use in the Netherlands was estimated to be 26.1% in 2015 among Dutch citizens older than 17 years of age [2].

Attempts to quit smoking are not monitored worldwide by a single organization, but statistics are available for specific nations. In the United States, among current smokers and former smokers, in 2012, an estimated 53% had made a quit attempt longer than 1 day [3]. Estimations from surveys of Dutch smokers [4] show that in 2014, approximately 1 in 3 smokers, accounting for approximately 500,000 smokers, attempted to quit smoking.

Independent quit attempts are unregistered, and because the uptake of interventions (digital and face-to-face) are generally unpredictable [5,6], there is no scientific literature available that relates quit attempts to categories of interventions to show which interventions are popular among smokers. Although there is a worldwide growing uptake of eHealth interventions, objective scientific data about who uses these kinds of interventions and the success of these interventions are almost nonexisting, despite international efforts to gather these data [7-9].

Mobile Cessation Apps

Recent studies identify a growing body of literature concerning various types of health care apps, resulting in several classifications and taxonomies [10-12]. One of the more recent types of eHealth interventions used for smoking cessation is cessation apps. As Eysenbach [13] noted in 2001, eHealth applies to the dynamic environment in which more is encompassed than just the internet and medicine, which is certainly true for the upcoming mHealth apps that are now part of the broader eHealth environment. There are several mobile phone apps for cessation available to a broad audience; however, results show [8] that most of them are not based on scientific, established guidelines for smoking cessation and do not have a tailored approach meeting individual needs. Due to this gap in the existing literature, the objective of our study was to develop a cessation app based on behavioral strategies and a user-centered approach. In this paper, we report on the characteristics of the potential end users for our mHealth cessation app.

Mobile apps on mobile devices (such as mobile phones and tablets) are fairly common. Research on mobile cessation apps has shown that the number of cessation apps available to consumers and their use are rising steadily [8,9,14,15]. An explanation for this growth can be found in the market potential for cessation apps, as smokers are likely to own a mobile phone, and smokers who intend to quit use their mobile phone more frequently than smokers who have no intention to quit [16].

Studies on cessation apps focus on different outcomes such as app usage [17-19], changes in psychological state related to cessation behavior [20-22], and satisfaction with the treatment with an app [18,20]. To date, effectiveness outcomes have been reported in 3 studies with self-reported abstinence [17,18,20,23], and 1 study protocol with biochemically validated abstinence outcome was found [22]. Thus, although an initial body of research on the use of cessation apps exists, little is known about the factors that influence the use of cessation apps and the characteristics of the users of these mHealth apps.

Studies [18] show that 50% of users found a cessation app useful as a means of support to quit smoking. In addition, participants have been given the choice to use a mobile cessation app to support them, and the result was that when given the choice, 60% of the participants actually used the mobile apps [17]. Both studies [17,18] note that their study populations might not be completely representative of the general smoking population because the study populations included respondents with an affinity for digital smoking support.

Potential Users

There is little evidence on what population of smokers uses cessation apps and if they share characteristics that can predict mobile cessation app use or explain why such a population is motivated to use cessation apps [7]. The only information available to identify the characteristics of potential users of a smoking cessation app is evidence from the general literature on the use of mobile apps. Research has shown that demographics and personality characteristics are associated with the adoption and use of mobile apps [24-30]. Income and level of education are positively correlated with mobile phone appropriation and adoption, whereas age is negatively correlated with mobile phone appropriation and adoption. However, gender is not correlated with appropriation or adoption [24,27]. We do not know if these findings apply for potential users of mobile cessation apps, which makes it difficult to predict the usefulness of specific mobile app elements, such as a reward system, color schemes, and gamification elements.

Recent findings show that the Big 5 personality traits [31] are related to the adoption of certain categories of mobile apps [26] or the use of a specific app such as Facebook [30,32]. The purpose of this study is to characterize the potential end users for a specific mHealth smoking cessation app, building on the factors known from previous studies on general mobile apps. However, because these are somewhat generic characteristics, for this study, factors related to smoking behavior were added.

These specific characteristics include the level of nicotine dependence, the number of previous quit attempts, and the modes of cessation support used in the past. Such variables have been shown to predict quit attempts in smokers and the likelihood of sustaining abstinence and might therefore also determine the decision to adopt and use a mobile cessation app. In the context of the Integrated Change model (or I-Change Model), these generic and smoking behavior-related characteristics are predisposing factors, which are considered

to be distal determinants affecting indirectly behavior through more proximal cognitive determinants such as intention, attitude, and self-efficacy expectancy. Although these proximal determinants are generally considered more powerful predictors of behavior on which the content of an app should be based, distal determinants are particularly useful for defining the segments of a population that should be targeted.

Research Questions

This study hypothesizes that the adoption of a mobile smoking cessation app is associated with the personal characteristics of smokers and ex-smokers. On the basis of evidence from past research, we will explore the relationship between the intention to use the app and several predisposing factors that are, in nature, demographic (age, gender, educational level, residential area), technology-related (personal innovativeness), or smoking behavior-related (nicotine dependency, number of previous quit attempts). We will also explore the relationship between the attitude toward a mobile smoking cessation app and the aforementioned predisposing factors.

Methods

Recruitment and Data Collection

This study used a quantitative approach by using a Web-based questionnaire among smokers and ex-smokers from December 2015 to March 2016. The link to the questionnaire was distributed during the entire period via different social media (Twitter, LinkedIn, WhatsApp, etc), recruiting participants through the digital snowballing method.

Before starting the questionnaire, the respondents read an introductory text on the project and the context wherein the questionnaire was taken. The participants were informed that when they entered their name and email address at the end of the questionnaire, they could be randomly selected to claim a gift card.

The inclusion criteria for this study were that the participants were able to access the Web-based questionnaire and that they were smokers or ex-smokers. Smokers are defined as respondents who smoked regularly at the time they completed the questionnaire. Ex-smokers are defined as respondents who smoked regularly but who have quit. There were no restrictions placed on the quit date. Smokers who are currently attempting to quit are registered as ex-smokers if their quit attempt lasted longer than 24 hours at the time they completed the questionnaire.

Measures

The questionnaire consisted of 8 constructs, which can be found in Table 1. When available, we used validated scales used in previous research on the adoption of mobile apps or cessation behavior. The complete questionnaire can be found in Multimedia Appendix 1.

Factor Analysis and Internal Consistency

The dependent variables were used on the scalar level. Factor analysis was used to ensure that the measured scales are unidimensional. The use of factor analysis is applicable, because the sample size is larger than 500 [36]. Additionally, Cronbach alpha was used to determine the internal consistency of the items composing the scales for behavioral intention, attitude, personal innovativeness, and the Fagerstrom Test of Nicotine Dependence.

Univariate and Multivariate Ordinal Logistic Regression

Due to the ordinal nature of the dependent variables, ordinal logistic regression was used to measure the relationship between the dependent and independent variables [37].The analyses were performed with IBM SPSS statistics for Windows, version 23.0, as the statistical package.

The associations between categorical independent and dependent variables were tested with chi-square tests, followed by a univariate ordinal regression model. Continuous variables were also tested using a univariate ordinal regression model.

Variables associated with $P<.05$ were all entered in a multivariate ordinal regression model. Subsequently, the variables with the highest P value were deleted from the model until the model fit decreased significantly.To prevent univariate and multivariate models from becoming unstable or unreliable because of empty cells, categories of categorical variables were omitted or grouped together. In the Results section, these steps are described in detail.

Table 1. Summary of study variables. N/A: not available.

Study variables	Item number	Reference
Independent variables		
Current smoking status	1-2	N/A
Nicotine dependence	3-8	[33]
Number of quit attempts	9	N/A
Previous use of digital cessation support	10	N/A
Personal innovativeness of IT	11-14	[34]
Demographics	23-26	N/A
Dependent variables		
Behavioral intention to use the cessation app	15-18	[35]
Attitude toward using the cessation app	19-22	[35]

Results

Sample

A total of 955 respondents who met the inclusion criteria started the Web-based survey, and 79.6% (760/955) respondents filled out the complete survey. The questionnaire was constructed such that research variables such as smoking characteristics, the intention to use a mobile cessation app, and attitude toward a mobile cessation app were at the beginning of the questionnaire, whereas demographic information was at the end of the questionnaire, allowing the respondents to be included based on the inclusion criteria, regardless of the completion of their questionnaire. Of the 955 respondents who started the questionnaire, 73.6% (703/955) still smoked, and 26.4% (252/955) respondents had stopped smoking. Analyses with missing values were conducted on a list-wise deletion per analysis. Tables 2 and 3 show the number of respondents per variable.

A total of 898 respondents provided information about their cessation attempt(s): 61.0% (583/955) claimed to have made more than 1 attempt to quit smoking, 24.0% (229/955) made only 1 attempt to quit, and 9.0% (86/955) respondents had not yet made a quit attempt. Crosstabs of cessation attempts with the current smoking status surprisingly revealed that of the 225 respondents who had stopped smoking, 13 respondents claimed to have not made a quit attempt. This could be due to the explanation of the question, which stated that a quit attempt had to be done consciously and should have lasted longer than 24 hours. Because the questionnaire was distributed in a time period that included New Year's Day, the respondents might have been in the process of their first quit attempt but not longer than 24 hours. Therefore, these respondents are treated as smokers, without a quit attempt. The alternative would be to combine this group of 13 respondents with the 229 respondents who made 1 quit attempt, which would lead to a shift of 1% in the ratios.

When asked about their previous experiences with digital cessation support, 802 respondents provided information. A total of 66.4% (634/955) indicated no previous experience with digital cessation support. The most commonly used digital cessation support was a Web-based self-management program or cessation app, for 13.0% (124/955) respondents, followed by looking up Web-based information about cessation, for 6.0% (57/955) respondents, having digitally contacted a care professional, 1.1% (10/955) respondents, and looking for Web-based peer support (eg, through a Web-based forum), for 0.9% (9 /955) respondents.

Factor Analysis

Factory analysis was conducted to determine whether the multi-item independent and dependent variables measured with a Likert-scale were unidimensional. Although factor analysis on ordinal variables has risks of over-dimensionalization [38], considering the limited number of items (3-4) and the large number of respondents, it is applied in this study.

All 4 items regarding behavioral intention (n=730) correlated to at least .5, suggesting factorability [39,40]. The Kaiser-Meyer-Olkin measure of sampling adequacy was .76, which is above the suggested value of .6. Bartlett test of sphericity was significant (P<.001). Communalities were all above .5. Given these findings, factor analysis was deemed suitable for these 4 items. Principal component analysis (PCA) shows that the largest factor explains 72% of the variance, being the only factor with an initial eigenvalue total higher than 1. All items load onto a single component, with a minimum value of .77. This leads to the conclusion that behavioral intention is a unidimensional scale.

The items regarding attitude (n=730) correlate to at least .7, except item 2 (min. .17). The Kaiser-Meyer-Olkin measure of sampling adequacy was .76, which is above the suggested value of .6. Bartlett test of sphericity was significant (P<.001).

Table 2. Descriptive characteristics of the study population (N=955).

Variable	Category	Number of participants (%)
Age	Scale	719 (75.3)
Gender	Categorical	719 (75.3)
Educational level	Ordinal	719 (75.3)
Residential area	Categorical	719 (75.3)
Personal innovativeness	Ordinal	857 (89.7)
Nicotine dependence (daytime smoking)	Ordinal	680 (71.2)
Current smoking status	Categorical	898 (94.0)
Number of quit attempts	Ordinal	898 (94.0)
Previous use of digital smoking cessation support	Categorical	802 (84.0)
Behavioral intention to use a mobile smoking cessation app	Ordinal	730 (76.4)
Attitude toward using a mobile smoking cessation app	Ordinal	730 (76.4)

Table 3. Descriptive demographic characteristics of the study population (N=955). SD: standard deviation; PhD: Doctor of Philosophy.

Variable	Total
Age, mean in years (SD)	38.0 (13.6)
Gender, n (%)[a]	
Female	470 (49.2)
Male	245 (25.6)
Invalid responses	4 (0.4)
Highest completed educational level, n (%)[a]	
Elementary school or secondary education	250 (26.2)
Vocational degree	220 (23.0)
Polytechnic education or university of applied science	201 (21.0)
Scientific degree (Master's and PhD)	48 (5.0)

[a]Data missing of N=236.

Communalities were all above .8, except for item 2 (.1). These findings suggest that factor analysis is deemed suitable, with the indication that item 2 does not load onto the same factor as the others. PCA confirms this point, showing that the first factor explains 65% of the total variance, being the only factor with an initial eigenvalue total higher than 1. All items load onto a single component, with a minimum value of .90, except item 2 (.34). Item 2 is therefore eliminated from the scale measuring attitude to preserve the unidimensional scale.

The personal innovativeness scale (n=857) consists of 4 items that all correlate to at least.5. The Kaiser-Meyer-Olkin measure of sampling adequacy was .80, which is above the suggested value of.6. Bartlett test of sphericity was significant (P<.001). Communalities were all above.5. Factor analysis was deemed suitable based on these findings. PCA shows that the first factor explains 65% of the total variance, being the only factor with an initial eigenvalue total higher than 1. All items load onto a single component, with a minimum value of.76. This leads to the conclusion that personal innovativeness is a unidimensional scale.

Research using confirmatory factor analysis on the Fagerstrom Test of Nicotine Dependence indicates that the scale is best modeled as 2 correlated factors with a cross-loading [39]. These 2 factors are a morning smoking factor and a daytime smoking factor.

We performed exploratory factor analysis and found that the Fagerstrom Test of Nicotine Dependence scale (n=680) consisted of 6 items that loaded onto 2 different factors, with initial eigenvalues of 2.4 and 1.1. The first factor consisted of the following items (the rotated factor loadings are noted in parentheses): *How soon after you wake up do you smoke your first cigarette?* (.37), *Which cigarette would you hate most to give up?* (.60), and *Do you smoke more frequently during the first hours after waking than during the rest of the day?* (.45), resulting in a *morning smoking* factor. The second factor consisted of the following items: *How soon after you wake up do you smoke your first cigarette?* (.72), *On average, how many cigarettes are you currently smoking each day?* (.73), *Do you find it difficult to keep from smoking in places where it is not allowed?* (.39), and *Do you smoke if you are so ill that you are in bed most of the day?* (.50), resulting in a "daytime smoking factor" on which we will conduct our further analysis.

Internal Consistency

Both dependent variables show good internal consistency, as shown in Table 4. The measure of attitude consists of 3 items, after item 2 was deleted. Both variables are taken as scales in the ordinal logistic regression analyses as follows: the scores for each item are added together, dividing the total score by the number of items and rounding the result to the nearest integer, for example, 1.5 becomes 2, creating a scale with 5 categories.

Table 4. Cronbach alpha values for the scale variables.

Variable	Number of respondents	Number of items	Cronbach alpha
Dependent variables			
Behavioral intention to use	730	4	.87
Attitude toward using	730	3	.91
Independent variables			
Personal innovativeness of Information Technology	857	4	.82
Morning smoking factor	680	3	.51
Daytime smoking factor	680	4	.68

The internal consistency of the independent variables of personal innovativeness of Information Technology (IT), morning smoking, and daytime smoking is shown in Table 4. The internal consistency of the personal innovativeness scale is good, whereas the internal consistency of morning smoking is questionable and that of daytime smoking is acceptable. The remainder of the analysis is therefore conducted based on the daytime smoking factor, as this factor best captures nicotine dependency [41,42], and morning smoking has a questionable Cronbach alpha value of .51.

Univariate Ordinal Logistic Regression

Crosstabs between the dependent and independent variables showed empty cells for the variables of *educational level*, *residential area*, and *previous use of digital cessation support* for both the dependent variables and the personal innovativeness items 1, 3, and 4 on attitude . Empty cells in ordinal logistic regression may cause goodness-of-fit predictors to become unreliable. Univariate analysis of single personal innovativeness items on attitude is therefore omitted. The variable *educational level* was reduced to 4 categories to eliminate all empty cells: (1) elementary school and secondary education, (2) vocational degree, (3) polytechnic or university of applied science, and (4) scientific degree or higher.

The variable *residential area* had only 1 empty cell, and the parameter estimates showed no significant results at the level of $P<.05$, making it unnecessary to reduce the number of categories for either the dependent or independent variable. The categories for the variable *previous use of digital cessation support* are grouped into 2 groups: respondents who have used digital cessation support (n=168) and respondents who have not (n=634). Additionally, the categories of the dependent variables were reduced from 5 to 3 to eliminate the remaining empty cells: (1) agree, (2) neutral, and (3) disagree.

The results in Tables 5 and 6 should be interpreted as corresponding to the coding of the questionnaire where favorable outcomes (agree) were coded with low values and undesirable outcomes (disagree) with high values. This means that an odds ratio, OR <1 implies a positive correlation.

Table 5 shows that there is no significant effect for the odds that males have a higher intention to use the cessation app than females. For educational level, there is a significant effect that smokers and ex-smokers with a lower educational level have a higher intention to use the cessation app. For residential area, no significant effects that smokers and ex-smokers living outside the city limits have a higher intention to use the app than smokers and ex-smokers living in a village or in a city were found. The respondents who smoke but have the intention to quit have a higher intention to use the cessation app than ex-smokers. There is no evidence that not having the intention to quit smoking is associated with the intention to use a mobile smoking cessation app.

Table 6 shows that the associations found between the independent variables and attitude are the same as the associations between the independent variables and behavioral intention.

Scale Ordinal Logistic Regression

As multiple items on the personal innovativeness scale and the Fagerstrom Test of Nicotine Dependence showed significant relationships with the dependent variables, ordinal logistic regression is performed with the scale variables to determine whether the results change.

Crosstabs for the personal innovativeness scale and both behavioral intention and attitude show empty cells at the end categories for the personal innovativeness scale. The personal innovativeness scale is reduced to 3 categories: (1) low personal innovativeness (ranging from 1 to 2.5), (2) average personal innovativeness (ranging from 2.5 to 3.5), and (3) high personal innovativeness (ranging from 3.5 to 5).

Multivariate Ordinal Logistic Regression

Multivariate ordinal logistic regression is performed with the independent variables *educational level*, *nicotine dependency (daytime smoking)*, *current smoking status*, *and number of quit attempts* and *personal innovativeness* for *behavioral intention* as the dependent variable. *Previous use of digital cessation support*, although significant, is omitted from the multivariate analysis because the outcome measures were adjusted to reduce the number of empty cells. Analysis on the dependent variable of *attitude* includes the same independent variables, excluding *educational level* because of an unacceptable goodness-of-fit.

Multicollinearity tests for both behavioral intention and attitude as the dependent variable show that *current smoking status* has a tolerance of 0.012 (*behavioral intention*) and 0.011 (*attitude*) and variance inflation factor scores >10. The variable is therefore omitted from the multivariate analysis on both dependent variables because its predictability is largely accounted for by the other variables.

Table 5. Univariate ordinal regression analyses on behavioral intention. An odds ratio lower than 1 positively correlates with the independent variable.

Variable	Odds ratio (95% CI)
Age (n=715)	1.00 (0.99-1.01)
Gender (n=715)	
Male	0.94 (0.70-1.26)
Female (base level)	1 (0)
Educational level[a] (n=711)	
Elementary + secondary education	0.48 (0.30-0.76)
Vocational degree	0.51 (0.33-0.81)
Polytechnic + university applied sciences	0.52 (0.33-0.83)
Scientific education (base level)	1 (0)
Residential area (n=717)	
City	1.73 (0.78-3.80)
Village	1.66 (0.73-3.76)
Outside city/village limits (base level)	1 (0)
Personal innovativeness[a] (n=730)	
High	0.31 (0.21-0.49)
Moderate	0.47 (0.33-0.67)
Low (base level)	1 (0)
Nicotine dependency[a], daytime smoking; 0=low dependency (n=554)	0.89 (0.77-0.89)
Current smoking status[a] (n=730)	
Smokes but intention to quit	0.21 (0.15-0.29)
Smokes and no intention to quit[b]	1.29 (0.82-2.03)
Quit smoking (base level)	1 (0)
Number of quit attempts[a] (n=730)	
0	3.38 (2.12-5.40)
1	1.38 (1.00-1.90)
More than 1 (base level)	1 (0)
Previous use of digital cessation support[a] (n=657)	
Has experience with digital cessation support; 0=has experience	0.48 (0.33-0.71)
No experience (base level); 1=no experience	1 (0)

[a]*P* value <.05.

[b]No significant results due to goodness of fit.

Table 6. Univariate ordinal regression analyses on attitude. An odds ratio lower than 1 positively correlates with the independent variable.

Variable	Odds ratio (95% CI)
Age (n=715)	1.00 (0.99-1.01)
Gender (n=715)	
Male	0.92 (0.68-1.24)
Female (base level)	1 (0)
Educational level[a] (n=711)	
Elementary + secondary education	0.61 (0.38-0.98)
Vocational degree	0.60 (0.38-0.96)
Polytechnic + university applied sciences	0.58 (0.36-0.93)
Scientific education (base level)	1 (0)
Residential area (n=717)	
City	1.24 (0.56-2.75)
Village	1.33 (0.58-3.06)
Outside city/village limits (base level)	1 (0)
Personal innovativeness[b] (n=730)	
High	0.24 (0.16-0.36)
Moderate	0.44 (0.31-0.63)
Low (base level)	1 (0)
Nicotine dependency[b], daytime smoking; 0=low dependency (n=554)	0.82 (0.76-0.89)
Current smoking status[b] (n=730)	
Smokes but intention to quit	0.23 (0.16-0.32)
Smokes and no intention to quit[b]	1.28 (0.82-2.00)
Quit smoking (base level)	1 (0)
Number of quit attempts[b] (n=730)	
0	3.21 (2.02-5.12)
1	1.67 (1.20-2.32)
More than 1 (base level)	1 (0)
Previous use of digital cessation support[b] (n=657)	
Has experience with digital cessation support	0.31 (0.18-0.51)
No experience (base level) 1	1 (0)

[a]No significant results due to goodness of fit.

[b]*P* value <.05.

Tables 7 and 8 should be interpreted in the same manner as Tables 5 and 6, with OR<1 meaning a positive correlation with the dependent variable. Table 7 shows the outcome of the ordinal logistic regression of the variables on *behavioral intention*, which included 540 respondents. *Educational level* is the only nonsignificant variable (*P*=.67 for polytechnic or university of applied science, *P*=.262 for vocational degree, and *P*=.157 for elementary school and secondary education, measured against scientific education as the base level).

Table 8 shows the outcome of the regression with *attitude* as the dependent variable (n=554). All independent variables remain as unique predictors of the respondent's attitude toward the use of a cessation app. The respondents who score high on *personal innovativeness* are the most positive toward the use of a cessation app compared with respondents who score low on *personal innovativeness*. The respondents who have not made a quit attempt are more negative toward the use of a cessation app compared with respondents who have attempted to quit more than once.

Table 7. Results of the multivariate ordinal logistic regression on behavioral intention (n=540). An odds ratio lower than 1 positively correlates with the independent variable.

Variable	Odds ratio (95% CI)
Educational level	
Elementary + secondary education	0.64 (0.35-1.19)
Vocational degree	0.71 (0.39-1.29)
Polytechnic + university applied sciences	0.57 (0.31-1.04)
Scientific education (base level)	1 (0)
Personal innovativeness[a]	
High	0.26 (0.16-0.40)
Moderate	0.40 (0.26-0.61)
Low (base level)	1 (0)
Nicotine dependency[a], daytime smoking; 0=low dependency	0.83 (0.77-0.90)
Number of quit attempts[a]	
0	4.07 (2.37-7.01)
1	1.67 (1.13-2.48)
More than 1 (base level)	1 (0)

[a]*P* value <.05.

Table 8. Results of the multivariate ordinal logistic regression on attitude (n=554). An odds ratio lower than 1 positively correlates with the independent variable.

Variable	Odds ratio (95% CI)
Personal innovativeness[a]	
High	0.23 (0.14-0.36)
Moderate	0.39 (0.26-0.61)
Low (base level)	1 (0)
Nicotine dependency[a], daytime smoking; 0=low dependency	0.84 (0.77-0.91)
Number of quit attempts[a]	
0	3.48 (2.04-5.94)
1	1.75 (1.18-2.60)
More than 1 (base level)	1 (0)

[a]*P* value <.05.

Discussion

Potential Users of mHealth Cessation Apps

This study presents a conceptualization of the characteristics of potential end users for a specific mHealth cessation app. By providing a comprehensive overview of the characteristics that influence mHealth app acceptance and behavioral intention from a wide variety of disciplines relevant for health technology acceptance behavior, we tried to provide more insight into the specific variables that can play a role in the complex process of peoples' technology use for health-related reasons.

Our data were based on a large sample of Dutch smokers and ex-smokers (n=955) in which several categories of variables yielded interesting results. First, correlations were found between the *number of quit attempts*, *nicotine dependency*, *previous use of digital cessation support*, and *personal innovativeness* on attitude toward using a cessation app and the intention to use a cessation app. This shows that smokers, with high dependency and experience with failed attempts, are aware of the need to seek support in attempting to quit. First-time quitters lack this awareness and may need to be targeted with more persuasive messages from either human (professional/private) or digital channels to help them become aware of using an mHealth app for smoking cessation.

This study showed that no significant relationship between demographic characteristics and attitude toward or intention to use a cessation app exists. These findings are supported by a similar study [7] that found similar results based on real

cessation app data in Australia, the United States, and the United Kingdom.

Second, examining smoking-related characteristics, we found that the total number of quit attempts and the score on the Fagerstrom test for daytime smoking positively correlated with the attitude toward and intention to use a cessation app. Therefore, it is believed that the findings of this study can provide a basis for the further development of a model that may deepen our understanding of the influence of several factors on people's use of mHealth apps.

Study Limitations

Despite the results indicating that the number of quit attempts is positively associated with the intention to use and the attitude toward a cessation app, the definition of "quit attempt" in this study, though not uncommon [43,44], is chosen somewhat arbitrarily. First, the respondents were instructed to count a quit attempt when they did not smoke for over 24 hours. Second, the respondents were counted as ex-smokers when they had not smoked within the past 24 hours. Third, 13 respondents out of 225 provided inconsistent data on their current smoking status and number of quit attempts, stating that although they quit smoking, they did not make a quit attempt.

Moreover, due to some limitations of the dataset caused by empty cells in the ordinal regression analysis, certain correlations could not be analyzed. For example, although personal innovativeness positively correlates with intention to use and attitude toward using, whether there is a correlation between personal innovativeness and the previous use of digital support, which could explain the positive correlation with intention to use and attitude toward using, could not be established.

Another study limitation is that the respondents in this study were mostly recruited through Web-based media, such as email and social media. This might have created a sample biased toward the use of digital means, such as a mobile cessation app. The generalizability of the results is therefore limited.

Conclusions

This study showed that personal factors influence the adoption of a mobile smoking cessation app. Demographics did not seem to correlate with the intention to use or the attitude toward using a mobile smoking cessation app. Personal characteristics related to smoking or quitting behavior as well as personal innovativeness correlated with the adoption of a mobile smoking cessation app. This study provides useful insights into the concept of an mHealth app for smoking cessation from a user perspective. Future research could employ a study to investigate the potential influence that personal characteristics have on the adoption of mHealth apps to increase the satisfaction and effectiveness of mHealth interventions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire.

[PDF File (Adobe PDF File), 70KB - jmir_v20i4e118_app1.pdf]

References

1. Bilano V, Gilmour S, Moffiet T, d'Espaignet ET, Stevens GA, Commar A, et al. Global trends and projections for tobacco use, 1990-2025: an analysis of smoking indicators from the WHO Comprehensive Information Systems for Tobacco Control. Lancet 2015 Mar 14;385(9972):966-976. [doi: 10.1016/S0140-6736(15)60264-1] [Medline: 25784347]
2. Volksgezondheidenzorg (Public Healthcare). Trend in smoking adults URL: https://www.volksgezondheidenzorg.info/onderwerp/roken/cijfers-context/trends#!node-trend-roken-volwassenen [accessed 2018-02-25] [WebCite Cache ID 6xUqR1sJs]
3. Agaku IT, King BA, Dube SR. CDC. 2014. Current Cigarette Smoking Among Adults - United States, 2005-2012 URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6302a2.htm [accessed 2018-02-25] [WebCite Cache ID 6xUqjFln5]
4. CBS Statline. 2015. Lifestyle and (preventive) health research; personal characteristics URL: http://statline.cbs.nl/Statweb/publication/?VW=T&DM=SLNL&PA=83021NED&D1=0-16&D2=0-2,5-13,37-41&D3=0&D4=l&HD=150430-1352&HDR=T&STB=G1,G2,G3 [accessed 2015-08-26]
5. McClure JB, Greene SM, Wiese C, Johnson KE, Alexander G, Strecher V. Interest in an online smoking cessation program and effective recruitment strategies: results from Project Quit. J Med Internet Res 2006 Aug 22;8(3):e14 [FREE Full text] [doi: 10.2196/jmir.8.3.e14] [Medline: 16954124]
6. Balmford J, Borland R, Benda P, Howard S. Factors associated with use of automated smoking cessation interventions: findings from the eQuit study. Health Educ Res 2013;28(2):288-299 [FREE Full text] [doi: 10.1093/her/cys104]
7. BinDhim NF, McGeechan K, Trevena L. Who uses smoking cessation apps? A feasibility study across three countries via smartphones. JMIR Mhealth Uhealth 2014 Feb 6;2(1):e4 [FREE Full text] [doi: 10.2196/mhealth.2841] [Medline: 4114511]
8. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]

9. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GW, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016;18(5):1025-1031. [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
10. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
11. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
12. Plachkinova M, Andres S, Chatterjee S. A Taxonomy of mHealth Apps- Security and Privacy Concerns. Hawaii Presented at: 2015 48th Hawaii International Conference on System Sciences; 5-8 January; Kauai, HI, USA p. 3187-3196. [doi: 10.1109/HICSS.2015.385]
13. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
14. Jacobs MA, Cobb CO, Abroms L, Graham AL. Facebook apps for smoking cessation: a review of content and adherence to evidence-based guidelines. J Med Internet Res 2014;16(9):e205 [FREE Full text] [doi: 10.2196/jmir.3491] [Medline: 25205129]
15. Das S, Tonelli M, Ziedonis D. Update on smoking cessation: e-cigarettes, emerging tobacco products trends, and new technology-based interventions. Curr Psychiatry Rep 2016 May;18(5):51. [doi: 10.1007/s11920-016-0681-6] [Medline: 27040275]
16. Borrelli B, Bartlett YK, Tooley E, Armitage CJ, Wearden A. Prevalence and frequency of mHealth and eHealth use among US and UK smokers and differences by motivation to quit. J Med Internet Res 2015;17(7):e164 [FREE Full text] [doi: 10.2196/jmir.4420] [Medline: 26149323]
17. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. Telemed J E Health 2014 Mar;20(3):206-214 [FREE Full text] [doi: 10.1089/tmj.2013.0169] [Medline: 24350804]
18. 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. [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
19. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. Am J Drug Alcohol Abuse 2015 Jan;41(1):68-73 [FREE Full text] [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]
20. Baskerville NB, Struik LL, Hammond D, Guindon GE, Norman CD, Whittaker R, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial study protocol. JMIR Res Protoc 2015;4(1):e10 [FREE Full text] [doi: 10.2196/resprot.3823] [Medline: 25599695]
21. Ploderer B, Smith W, Pearce J, Borland R. A mobile app offering distractions and tips to cope with cigarette craving: a qualitative study. JMIR Mhealth Uhealth 2014;2(2):e23 [FREE Full text] [doi: 10.2196/mhealth.3209] [Medline: 25099632]
22. Hassandra M, Lintunen T, Kettunen T, Vanhala M, Toivonen HM, Kinnunen K, et al. Effectiveness of a mobile phone app for adults that uses physical activity as a tool to manage cigarette craving after smoking cessation: a study protocol for a randomized controlled trial. JMIR Res Protoc 2015;4(4):e125 [FREE Full text] [doi: 10.2196/resprot.4600] [Medline: 26494256]
23. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
24. de Reuver M, Nikou S, Bouwman H. Domestication of smartphones and mobile applications: a quantitative mixed-method study. Mob Media Commun 2016;4(3):347-370 [FREE Full text] [doi: 10.1177/2050157916649989]
25. Xu R, Frey RM, Ilic A. Individual Differences and Mobile Service Adoption: An Empirical Analysis. Presented at: 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService); 23 May, 2016; Oxford, United Kingdom. [doi: 10.1109/BigDataService.2016.15]
26. Xu R, Frey RM, Fleisch E, Ilic A. Understanding the impact of personality traits on mobile app adoption – insights from a large-scale field study. Comput Human Behav 2016;62:244-256. [doi: 10.1016/j.chb.2016.04.011]
27. Rice RE, Pearce KE. Divide and diffuse: comparing digital divide and diffusion of innovations perspectives on mobile phone adoption. Mob Media Commun 2015;3(3):401-424 [FREE Full text] [doi: 10.1177/2050157915590469]
28. Seneviratne S, Seneviratne A, Mohapatra P, Mahanti A. Predicting user traits from a snapshot of apps installed on a smartphone. Mob Comput Commun Rev 2014;18(2):1-8. [doi: 10.1145/2636242.2636244]
29. Seneviratne S, Seneviratne A, Mohapatra P, Mahanti A. Your installed apps reveal your gender and more!. Mob Comput Commun Rev 2015 Jan 13;18(3):55-61. [doi: 10.1145/2721896.2721908]
30. 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. [doi: 10.1016/j.chb.2014.11.067]
31. McCrae RR, Costa PT. Validation of the five-factor model of personality across instruments and observers. J Pers Soc Psychol 1987;52(1):81-90. [Medline: 3820081]
32. Ryan T, Xenos S. Who uses Facebook? An investigation into the relationship between the big five, shyness, narcissism, loneliness, and Facebook usage. Comput Human Behav 2011;27(5):1658-1664. [doi: 10.1016/j.chb.2011.02.004]

33. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerstrom test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. Br J Addict 1991;86(9):1119-1127. [Medline: 1932883]
34. Agarwal R, Prasad A. A conceptual and operational definition of personal innovativeness in the domain of information technology. Inf Syst Res 1998;9:204-215 [FREE Full text] [doi: 10.1287/isre.9.2.204]
35. Park Y, Chen J. Acceptance and adoption of the innovative use of smartphone. IMDS 2007;107(9):1349-1365 [FREE Full text] [doi: 10.1108/02635570710834009]
36. Comrey AL, Lee HB. A First Course in Factor Analysis (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.
37. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression, 3rd edition. London: Wiley; 2013.
38. van der Eijk C, Rose J. Risky business: factor analysis of survey data - assessing the probability of incorrect dimensionalisation. PLoS One 2015;10(3):e0118900 [FREE Full text] [doi: 10.1371/journal.pone.0118900]
39. Field A. Discovering statistics using IBM SPSS statistics. London: Sage; 2013.
40. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull 1988;103(2):265-275. [doi: 10.1037/0033-2909.103.2.265] [Medline: 3363047]
41. Richardson CG, Ratner PA. A confirmatory factor analysis of the Fagerstrom Test for Nicotine Dependence. Addict Behav 2005 May;30(4):697-709. [doi: 10.1016/j.addbeh.2004.08.015] [Medline: 15833575]
42. Haddock CK, Lando H, Klesges RC, Talcott GW, Renaud EA. A study of the psychometric and predictive properties of the Fagerström Test for Nicotine Dependence in a population of young smokers. Nicotine Tob Res 1999 Mar;1(1):59-66. [Medline: 11072389]
43. Skov-Ettrup LS, Egan KK, Dalum P, Tolstrup JS. Stress-related expectations about smoking cessation and future quit attempts and abstinence - a prospective study in daily smokers who wish to quit. Prev Med Rep 2017 Jun;6:187-190. [doi: 10.1016/j.pmedr.2017.03.004]
44. Zhuang Y, Cummins SE, Sun J, Zu SH. Long-term e-cigarette use and smoking cessation: a longitudinal study with US population. Tob Control 2016;25:i90-i95. [doi: 10.1136/bmj.i4645]

Abbreviations

I-Change Model: Integrated Change model
OR: odds ratio
PCA: principal component analysis

Edited by M Focsa; submitted 28.02.17; peer-reviewed by H Tange, KL Cheung, A Jahan; comments to author 29.03.17; revised version received 31.10.17; accepted 15.11.17; published 09.04.18.

Please cite as:
Chevalking SKL, Ben Allouch S, Brusse-Keizer M, Postel MG, Pieterse ME
Identification of Users for a Smoking Cessation Mobile App: Quantitative Study
J Med Internet Res 2018;20(4):e118
URL: http://www.jmir.org/2018/4/e118/
doi:10.2196/jmir.7606
PMID:29631988

Review

Use of Smartphones for Early Detection of Melanoma: Systematic Review

Cédric Rat[1,2], MD, PhD; Sandrine Hild[1], MD; Julie Rault Sérandour[1], MD; Aurélie Gaultier[3]; Gaelle Quereux[2,4], MD, PhD; Brigitte Dreno[2,4], MD, PhD; Jean-Michel Nguyen[2,3], MD, PhD

[1]Department of General Practice, Faculty of Medicine, University of Nantes, Nantes, France

[2]Unit 1232 – Team 2, Centre de Recherche en Cancérologie, French National Institute of Health and Medical Research, Nantes, France

[3]Department of Epidemiology and Biostatistics, Nantes University Hospital, CHU Nantes, Nantes, France

[4]Oncodermatology Department, Nantes University Hospital, CHU Nantes, Nantes, France

Corresponding Author:
Cédric Rat, MD, PhD
Department of General Practice
Faculty of Medicine
University of Nantes
1 rue Gaston Veil
Nantes, 44000 Nantes
France
Phone: 33 6 61 86 79 53
Email: cedric.rat@univ-nantes.fr

Abstract

Background: The early diagnosis of melanoma is associated with decreased mortality. The smartphone, with its apps and the possibility of sending photographs to a dermatologist, could improve the early diagnosis of melanoma.

Objective: The aim of our review was to report the evidence on (1) the diagnostic performance of automated smartphone apps and store-and-forward teledermatology via a smartphone in the early detection of melanoma, (2) the impact on the patient's medical-care course, and (3) the feasibility criteria (focusing on the modalities of picture taking, transfer of data, and time to get a reply).

Methods: We conducted a systematic search of PubMed for the period from January 1, 2007 (launch of the first smartphone) to November 1, 2017.

Results: The results of the 25 studies included 13 concentrated on store-and-forward teledermatology, and 12 analyzed automated smartphone apps. Store-and-forward teledermatology opens several new perspectives, such as it accelerates the care course (less than 10 days vs 80 days), and the related procedures were assessed in primary care populations. However, the concordance between the conclusion of a teledermatologist and the conclusion of a dermatologist who conducts a face-to-face examination depended on the study (the kappa coefficient range was .20 to .84, median κ=.60). The use of a dermoscope may improve the concordance (the kappa coefficient range was .29 to .87, median κ=.74). Regarding automated smartphone apps, the major concerns are the lack of assessment in clinical practice conditions, the lack of assessment in primary care populations, and their low sensitivity, ranging from 7% to 87% (median 69%). In this literature review, up to 20% of the photographs transmitted were of insufficient quality. The modalities of picture taking and encryption of the data were only partially reported.

Conclusions: The use of store-and-forward teledermatology could improve access to a dermatology consultation by optimizing the care course. Our review confirmed the absence of evidence of the safety and efficacy of automated smartphone medical apps. Further research is required to determine quality criteria, as there was major variability among the studies.

(J Med Internet Res 2018;20(4):e135) doi:10.2196/jmir.9392

KEYWORDS

smartphone; melanoma; screening; teledermatology; telemedicine; mobile app

Introduction

Background

The incidence of melanoma has increased in all Western countries over the last 30 years and has increased 3 to 5 times depending on the country [1,2], presently affecting 13.8 people in North America, 14.6 in northern Europe and 35.1 out of the 100,000 people in Australia [1,2]. Melanoma 5-year survival depends on the stage at the time of diagnosis, decreasing by 84% at a localized stage to 13% at the metastatic stage [2]. It is therefore essential that clinicians and policy makers concentrate their efforts to ensure early detection of the disease [3]. Numerous factors associated with a delayed diagnosis are patient-related [4-12]. Other factors are related to the opportunity to consult a dermatologist rather than a general practitioner [4,13]. However, various authors have reported difficulties in obtaining an appointment with a dermatologist [12,14-17].

Many countries have tested the use of telemedicine in dermatology as a way to increase access to health care services when distance is a critical factor [18-26]. Telemedicine in dermatology can be based either on videoconferences or on store-and-forward teledermatology procedures. Videoconferences, which allow a patient and a dermatologist to be connected for a consultation, are time-consuming for the dermatologist and may require an expensive setup [20,27]. Store-and-forward teledermatology procedures are based on sending information and photographs to a dermatologist for a deferred medical opinion [20].

Although smartphones have now revolutionized the daily life of physicians in all Western countries [19,24,28,29], one issue is to determine whether smartphone use has been assessed for store-and-forward teledermatology procedures. As various apps exist that provide scores, decision aids, and management advice, an alternative both to videoconferences and to store-and-forward teledermatology might be automated smartphone apps with no need of a dermatologist opinion. An issue to be addressed is whether such apps can help in the early diagnosis of melanoma.

Objective

We initiated a review focusing on the use of a smartphone in sustaining melanoma early detection (either store-and-forward teledermatology or automated apps). The aim was to report evidence on (1) the diagnostic performance of the procedures, (2) the impact on the patient's medical-care course and delays before the dermatological consultation, and (3) the limitations of either store-and-forward teledermatology or automated apps.

Methods

This review was conducted according to the key steps required for systematic reviews [30]. Considering that evidence might be sparse, the literature review was based on a broad scope and was not restricted to randomized controlled trials.

Study Identification and Selection

We conducted a systematic search of PubMed for the period from January 1, 2007 (launch of the first smartphone) to November 1, 2017. The keywords were as follows: [smartphone OR cell phone OR remote OR telemedicine] AND [dermatology OR skin disease OR melanoma OR skin neoplasm OR skin abnormalities]. We also searched the reference lists of reviews and studies identified during the initial search by hand. Abstracts and full texts were reviewed independently by 2 reviewers (SH and JR) for inclusion. Any disagreements on inclusion or exclusion were resolved by consensus, and a third reviewer (CR) was consulted to resolve any remaining disagreements.

Inclusion and Exclusion Criteria

In this manuscript, the term "smartphones" refers to mobile phones that have an internet data communication system and a digital camera (compared to mobile phones that would not have these two specific devices). The inclusion criteria for the studies included in this review were as follows: (1) photographs concerning pigmented suspicious lesions, (2) photographs taken or analyzed using a smartphone, (3) patients older than 18 years, (4) studies written in French or English, and (5) abstracts available.

The exclusion criteria were (1) no use of a smartphone, (2) the research area was not related to melanoma early detection, (3) dermatology teleconsultation in the form of a videoconference, (4) study based on histology, (5) studies consisting of assessing patients or caregivers' preferences through interviews or surveys, and (6) an editorial or a letter to the editor.

Considering that evidence might be sparse, the literature review was based on a broad scope and the inclusion criteria were not restricted to a "Patients Intervention Comparison Outcome" presentation [30].

Data Extraction and Synthesis

Studies were critically appraised by 2 reviewers (SH and JR), and discrepancies were resolved by consensus. The studies were first classified by the type of procedure assessed (app or store-and-forward teledermatology). Then the following data were extracted, such as design, population of the sample, whether the study had been conducted in the context of primary care, whether it was a descriptive or comparative study, and main outcome measures (Tables 1 and 2).

The main outcomes in studies focusing on store-and-forward teledermatology were either (1) the diagnostic concordance between the teledermatology procedure and the reference (the kappa coefficient of concordance is a measure of agreement between 2 raters, based on the following formula [(observed probability–expected probability)/(1–expected probability)]) or (2) the impact on the patient's medical-care course and delays before dermatological consultation. The proportion of uninterpretable photographs was also reported.

The main outcomes in studies focusing on apps were sensitivity (defined as the number of true positive assessments/number of all positive assessments), specificity (defined as the number of true negative assessments/number of all negative assessments), and accuracy (defined as the number of correct assessments/number of all assessments). The proportion of uninterpretable photographs was also reported. When necessary, the authors were contacted to obtain information not reported in the studies.

Analysis of Bias

We assessed the risk of bias using the quality assessment of diagnostic accuracy 2nd edition (QUADAS-2) [55]. The reporting of risk of bias focused on patient selection, index test, reference standard, and flow and timing. For each item, signaling questions helped to estimate whether the risk of bias was low or high. Unclear was used if no information was available. The applicability of the study intervention was also assessed using QUADAS-2, focusing on patient selection, index test, and reference standard.

Table 1. Studies based on store-and-forward teledermatology procedures (design, patients, comparison, and outcome). N/A: not applicable. RCT: randomized controlled trial.

Authors	Design	Patients, n (%)	Population of sample	Comparison	Main outcomes
Store-and-forward teledermatology without teledermoscopy					
Boyce et al, 2011 [31]	Prospective study	N/A	Patients at an elevated risk of melanoma	Face to face	Concordance
Lamel et al, 2011 [32]	Prospective study	1 (0.7)	Patients from a melanoma screening campaign	Face to face	Concordance
Janda et al, 2014 [33]	RCT	1 (1)	Patients at an elevated risk of melanoma	Face to face	Concordance
Store-and-forward teledermatology that included teledermoscopy					
Ford et al, 2015 [34]	Quasi-experiment	22 (11.3)	Patients recruited in primary care centers	Face to face	Secondary care referral
Börve et al, 2013 [29]	Prospective study	12 (17)	Patients referred for an excision	Face to face[a]	Concordance
Börve et al, 2015 [35]	Quasi-experiment	55 (3.52)	Patients recruited in primary care centers	Teledermatology versus paper referral	Delays
Hue et al, 2016 [36]	Descriptive study	1 (0.3)	Patients from a melanoma screening campaign	No comparison	Feasibility and delays
Janda et al, 2013 [37]	Descriptive study	N/A	Patients at an elevated risk of melanoma	No comparison	Feasibility
Kroemer at al, 2011 [38]	Prospective study	6 (5)	Patients referred to the dermatologist	Face to face[a]	Concordance
Manahan et al, 2015 [39]	RCT	0 (0.0)	Patients with a dermatological follow-up	Face to face	Concordance
Markun et al, 2017 [40]	Prospective study	1 (0.05)	Patients from a melanoma screening campaign	Face to face[a]	Concordance
Massone et al, 2007 [41]	Prospective study	2 (11)	Patients with a dermatological follow-up	Face to face	Concordance
Wu et al, 2015 [42]	Prospective study	N/A	Patients with a dermatological follow-up	Face to face	Concordance

[a]For these studies, suspicious lesions were referred for excision and histopathology results were analyzed as a secondary outcome in the study.

Table 2. Studies based on automated smartphone apps (design, photographs and patients, comparison, and outcomes). N/A: not applicable.

Authors	Design	n (%)[a]	Patients, N (characteristics)	Comparison	Main outcomes
Photographs issued from a database					
Do et al, 2014 [43]	Case-control study	29 (36)	N/A	Histopathology	Accuracy[b]
Doukas et al, 2012 [44]	Case-control study	800 (26.67)	N/A	Clinical evaluation	Accuracy[b]
Ferrero et al, 2013 [45]	Descriptive study	93 (100)	N/A	Histopathology	Sensitivity, Specificity
Ramlakhan et al, 2011 [46]	Case-control study	46 (55)	N/A	Unclear	Sensitivity, Specificity
Wadhawan et al, 2011a [47]	Case-control study	388 (29.85)	N/A	Histopathology	Sensitivity, Specificity
Wadhawan et al, 2011b [48]	Case-control study	110 (31.7)	N/A	Histopathology	Sensitivity, Specificity
Wolf et al, 2013 [49]	Case-control study	60 (31.9)	N/A	Histopathology	Sensitivity, Specificity
Photographs taken of patients in real condition					
Dorairaj et al, 2017 [50]	Prospective study	9 (28)	N/A (referred for an excision)	Teledermatologist[c]	Sensitivity, Specificity
Maier et al, 2015 [51]	Prospective study	26 (18.1)	N/A (with a dermatological follow-up)	Face to face[c]	Sensitivity, Specificity
Ngoo et al, 2017 [52]	Prospective study	1 (2)	30 (with a dermatological follow-up)	Teledermatologist	Sensitivity, Specificity
Robson et al, 2012 [53]	Prospective study	2 (6)	31 (referred to the dermatologist)	Face to face[c]	Sensitivity, Specificity
Thissen et al, 2017 [54]	Prospective study	6 (1.8)	256 (referred to the dermatologist)	Face to face[c]	Sensitivity, Specificity

[a]Included photographs, proportion with melanoma.

[b]Accuracy=(True Negatives+True Positives)/(True Negatives+True Positives+False Negatives+False Positive).

[c]For these studies, suspicious lesions were referred for excision and histopathology results were analyzed as a secondary outcome in the study.

Results

Overview

In total, 1450 titles and abstracts were screened for eligibility, utilizing the inclusion and exclusion criteria. A previous review was identified [2] and the related studies from the references list were included in this review. A total of 25 studies [29,31-54,56] were included in the review (Figure 1). Of these, 15 studies had been published as original papers [29,31,32,34-36,38-42,49,51,52,54], 5 were conference papers [43,44,46-48] and 5 were research letters [33,37,45,50,53]. In these, 12 studies had been conducted in European countries, that is, Great Britain [34,53], Austria [38,41], Sweden [29,35], Ireland [50], Germany [51], Switzerland [40], Greece [44], the Netherlands [54], and France [36]; 7 in United States of America [32,42,45-49]; 5 in Australia [31,33,37,39,52]; and one in Singapore [43].

Store and Forward Teledermatology

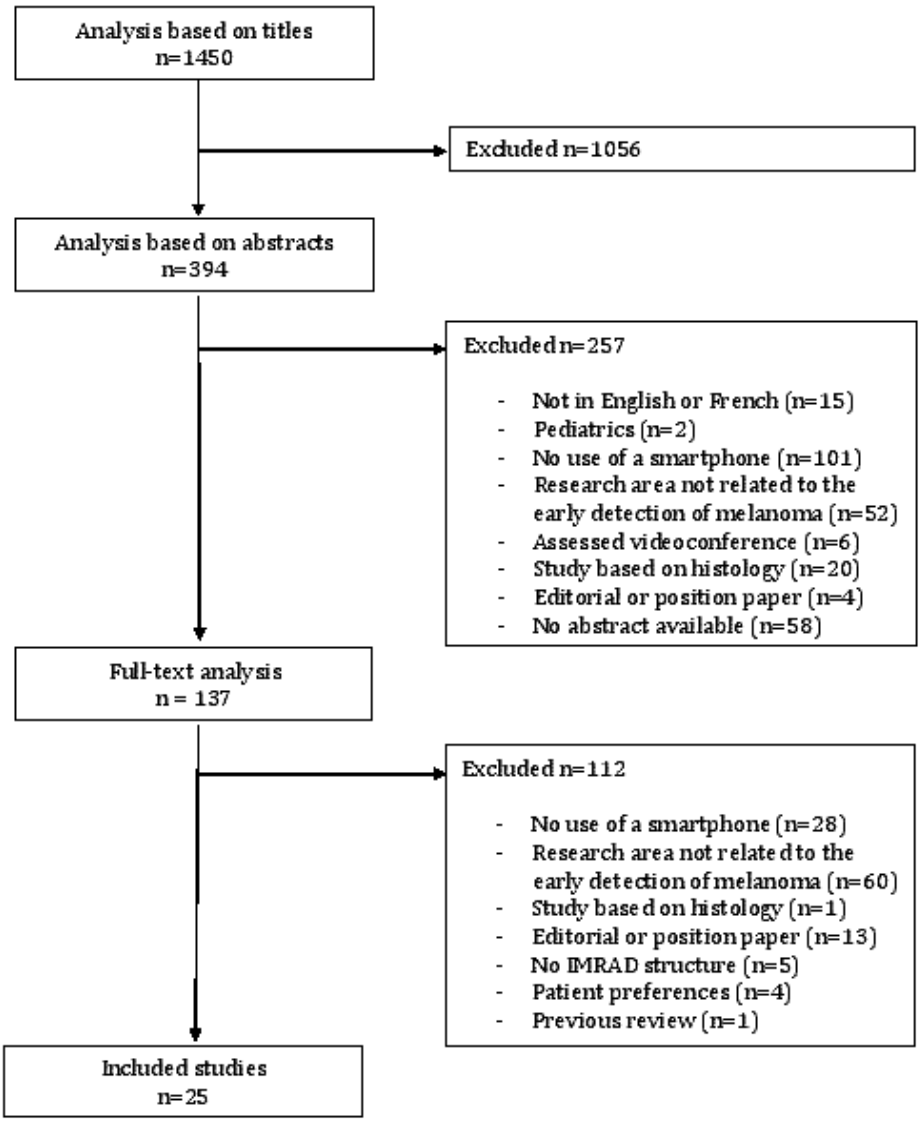
A total of 13 studies assessed store-and-forward teledermatology [29,31-42] (Table 1). There were 12 studies that specified the smartphone model used, that is, 9 tested iPhones [29,33-37,39,40,42] and 3 tested other brands of telephones [32,38,41].

The study population were patients recruited in primary care in 9 studies, that is, either in the context of a screening campaign [32,36,40] or during targeted screening focusing on patients at an elevated risk of melanoma [31,33,37] or during opportunistic screening conducted in general practice [34,35,38]. For the other 4 studies, the patients had already consulted a dermatologist [29,39,41,42]. The prevalence of melanomas in the related populations varied greatly, ranging from 0% [39] to 17.3% [29].

Ten studies compared the conclusion of the teledermatologist with the conclusion of a dermatologist who conducted a face-to-face examination (Table 1). There were 2 studies that focused on feasibility without providing any comparison [36,37]. From the 13 studies conducted, 10 studies used a mobile teledermoscope [29,34-42], whereas 3 studies only transferred the pictures taken without the teledermoscope [31-33].

There were 7 studies that provided information on diagnostic concordance for store-and-forward teledermatology based solely on clinical photographs. The diagnostic concordance between the conclusions of the teledermatologist and the dermatologist (face-to-face) ranged from 62% [32] to 89% [41]. This concordance was analyzed further using the kappa coefficient [29,31,32,38-40,42], which ranged from .20 [40] to .84 [38]. Börve reported 58% concordance between the conclusions of 2 independent teledermatologists [29]. Focusing on whether the patients could take pictures of their lesions themselves, Boyce et al reported 69% concordance between the conclusion of a dermatologist (face-to-face) and the conclusion of a teledermatologist (κ=.23) [31].

Figure 1. Flowchart of studies identified in this systematic review focusing on the use of smartphones in the early detection of melanoma. IMRAD: Introduction, Methods, Results, And Discussion.



For teledermatology based on pictures taken with a teledermoscope, the diagnostic concordance between the teledermatologist's conclusions and the conclusion of a dermatologist (face-to-face) ranged from 51% [29] to 97% [42]. The kappa coefficient varied from .29 [40] to .87 [42]. Massone reported 94% concordance between the conclusions of 2 independent teledermatologists who analyzed photographs taken using a teledermoscope [41]. Focusing on whether the patients could take pictures of their lesions themselves, Manahan and Wu reported concordance ranging from 90% [39] to 97% [42] when the patient used a teledermoscope.

There were 4 studies that reported an acceleration in the management of patients when malignancy was suspected [29,34,36,40]. Börve reported reduced delays in obtaining an appointment (delays shorter than 2 days compared with delays longer than 80 days), a reduced delay for surgical management (36 days vs 85 days), and a lower Breslow index at the time of the diagnosis [29]. Hue reported that patients with a highly suspicious lesion were asked to return within less than 10 days [36]. There were 3 studies [35,36,40] that reported a decrease (40%, 53%, and 74%, respectively) in the proportion of patients referred to a dermatologist, whereas Ford reported a slight increase in referrals (an increase of 2.11 per 1000 patients) [34].

The proportion of uninterpretable images due to their poor quality was, on average, less than 20% [29,40,42]. Only Massone et al (2007) reported a higher percentage of poor-quality images of 70% [41]. However, only a minority of authors provided information on the modalities for picture taking. A total of 4 studies specified the size of the pictures from 1024×766 to 2592×1224 pixels. Following contact with the authors, the following information on the modalities were collected, that is, the number of pictures could vary from 1 to 12 pictures per lesion [33,39], pictures were taken at a distance of 10 to 30 cm from the skin [29,40], authors reported taking one close-up picture and another of the surrounding area [36,40,41], and specified the lighting conditions ("strong light"

[39], "day light" [41], "maximal light" [29]), neutral background [29,40], use of the zoom [41] or macro mode [40,41], and use of the autofocus [29,38,41]). There were 4 authors who reported the time required to take the picture that is, ranging from a few seconds to less than 4 min [29,34-36]. The photographer was a professional in most cases, that is, either a dermatologist [29,38,41], a general practitioner [35], or another professional [32,40]; however, 6 studies reported that the picture was taken by the patient or a family member [31,33,37,39,42]. The notion of encrypting data was not approached systematically. Following individual contact, 8 authors reported encryption of the data, either through the app or through anonymity [29,31,32,34-36,39,40,42]. A total of 8 studies used email for transferring the data [29,31,33-35,37,39,42], and 7 used an encrypted platform [29,32,34-36,40,42].

Analysis of bias is provided in Table 3. The risk of bias related to patient selection was high, as patients who participated were either volunteers or chosen by doctors in consultation, but no study was based on a random sample. In 2 studies, the same dermatologist participated in both, the store-and-forward teledermatology procedure and the face-to-face clinical evaluation, with insufficient washout, so that the risk of bias related to the test index was high. The final analysis of the studies did not include all the recruited patients so that the risk of bias of flow and timing was high. Applicability was good.

Automated Smartphone Apps

A total of 12 studies assessed the performance of automated smartphone apps [43-54] (Table 2), that is, 1 study compared the conclusions of an automated app with the conclusions of a dermatologist who conducted a face-to-face examination [51], 1 study compared the conclusions of an automated app with the conclusions of a teledermatologist [52], 7 studies analyzed the images of an already classified data bank [43-49], and the 3 last ones compared the conclusion of an automated app, both with the pathological report (after excision of the lesion) and with the conclusion of teledermatologists [50,53,54].

Among the 5 studies based on taking a photograph [51-54], 4 tested iPhones [51-54], 1 tested other brands of telephones [52], and the brand of the telephone was not specified in the last study [50]. The photographer was a dermatologist [53,54] or another professional [50-52].

Participants who were recruited in studies assessing automated smartphone apps were highly selected. Among the 8 studies based on photographs issued from a database, the proportion of melanoma ranged from 26.7% [44] to 100% [45]. Among the 5 studies based on taking a photograph, 2 studies included patients from a primary-care setting recruited during an opportunistic screening campaign conducted in general practice [53,54]. For the other 3 studies that included patients, the patients had already consulted a dermatologist [50-52]. The prevalence of melanomas in the related populations ranged from 1.8% [52,54] to 28.1% [50].

The performance of the automated smartphone apps were assessed by referring either to their capacity to classify the lesions at risk [43,45,47,49-54], or to their diagnostic capacity [44,46,48] (Table 2). The references used could be either the histology results in 5 studies [43,45,47-49], the teledermatologist's conclusion in 2 studies [50,52], or the dermatologist (face-to-face) conclusion in 3 studies [51,53,54]. For 2 studies, the authors referred to histology-based diagnosis without describing how they obtained the histopathological conclusion [44,46].

A total of 5 studies assessed the apps used in real conditions [50-54]. In 4 studies, the images came from medical [43,45,49] or commercial [47,48] data banks, whereas no details on the photograph data banks were provided for 2 studies [44,46]. The sensitivity ranged from 7% to 87% [48,49], and the specificity ranged from 9% to 100% [50,52]. Only 2 studies described the area under the curve [44,48], providing a better comparison of results. One study reported a kappa coefficient of concordance between the opinion of the app and that of a dermatologist [52].

None of the studies that assessed automated smartphone apps reported an impact on the patient's medical-care course.

Analysis provided by automated smartphone apps were made difficult by ulcerated, blood stained, speckled or tanned areas, the presence of hair, or several lesions on the same photograph. There were 5 studies that had a proportion ranging from 11% to 30% of the lesions that could not be analyzed because of technical problems other than the problems related to the quality of the initial photograph [45,49-51,53]. There were 4 studies based on apps which reported that the time required to analyze the pictures was less than 10 seconds [44,46-48]. The notion of encrypting data was not approached systematically. Following individual contact, 2 authors [52-54] reported encryption of the data, either through the app or through anonymity.

The analysis of bias is provided in Table 4. For studies based on photographs from databank, references were unknown and the risks of bias for patient selection were high. Applicability related to the patient selection was highly concerning. Automated apps had no information on the diagnosis so that the risk of bias related to the test index was low. In 3 studies, pictures were modified before intervention so that the applicability related to the index test was not good. All the photographs were not analyzed so that the risk of bias related to flow and timing was high.

Table 3. Studies based on store-and-forward teledermatology procedures. Risk of bias assessment according to quality assessment of diagnostic accuracy 2nd edition (QUADAS-2).

Authors	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Store-and-forward teledermatology without teledermoscopy							
Boyce et al, 2011 [31]	High	Low	Low	High	Low	Low	Low
Lamel et al, 2011 [32]	High	Low	Low	Low	Low	Low	Low
Janda et al, 2014 [33]	High	Low	High	Low	Low	Low	Low
Store-and-forward teledermatology that included teledermoscopy							
Ford et al, 2015 [34]	Low	Low	Low	High	Low	Low	Low
Börve et al, 2013 [29]	High	Low	Low	Low	High	Low	Low
Börve et al, 2015 [35]	High	Low	Low	High	Low	Low	Low
Hue et al, 2016 [36]	High	Low	Low	High	Low	Low	Low
Janda et al, 2013 [37]	High	Low	Low	High	Low	Low	Low
Kroemer at al, 2011 [38]	High	High	Low	High	Low	Low	Low
Manahan et al, 2015 [39]	High	Low	Low	High	Low	Low	Low
Markun et al, 2017 [40]	High	High	Low	High	Low	Low	Low
Massone et al, 2007 [41]	High	Low	Low	Low	Low	Low	Low
Wu et al, 2015 [42]	High	Low	Low	High	Low	High	Low

Table 4. Studies based on automated smartphone apps. Risk of bias assessment according to quality assessment of diagnostic accuracy 2nd edition (QUADAS-2).

Authors	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Photographs issued from a database							
Doukas et al, 2012 [44]	High	Low	Unclear	Unclear	High	Low	Unclear
Do et al, 2014 [43]	High	Low	Low	Unclear	High	High	Low
Ferrero et al, 2013 [45]	High	Low	Low	High	High	Low	Low
Ramlakhan et al, 2011 [46]	High	Low	Unclear	High	High	Low	Unclear
Wadhawan et al, 2011a [47]	High	Low	Low	High	High	High	Low
Wadhawan et al, 2011b [48]	High	Low	Low	High	High	High	Low
Wolf et al, 2013 [49]	High	Low	Low	High	High	High	Low
Photographs taken of patients in real condition							
Dorairaj et al, 2017 [50]	High	Low	Low	High	High	Low	Low
Maier et al, 2015 [51]	High	Low	Low	High	High	Low	Low
Ngoo et al, 2017 [52]	High	Low	Low	High	Low	Low	Low
Robson et al, 2012 [53]	High	Low	Unclear	High	Low	Low	Low
Thissen et al, 2017 [54]	High	Low	Unclear	Low	Low	Low	Low

Discussion

Store-and-forward teledermatology opens several perspectives, that is, it accelerates the care course, and various studies were performed in primary care populations. However, the concordance between the conclusion of a teledermatologist and the conclusion of a dermatologist who conducts a face-to-face examination depended on the study (the kappa coefficient range was .20 to .84, median κ=.60). The use of a dermoscope may improve the concordance (the kappa coefficient range was .29 to .87, median κ=.74). Regarding automated smartphone apps, the major concerns are their low sensitivity, the lack of

assessment in clinical practice conditions, and the lack of assessment in primary care populations.

In a study recently published in *Nature*, Esteva et al reported that an artificial neuronal network had a better capacity to recognize melanomas than a dermatologist [56]. However, the automated apps available on a smartphone in 2017 do not provide such expertise. Our review shows that the existing automated smartphone apps are unreliable. Some apps have a low diagnostic sensitivity that may induce false negatives and erroneously reassure patients who may then not consult a specialist [50]. Certificates do not guarantee good diagnostic performance [54]. Greater control by administrative authorities is necessary [45,57,58].

This literature review suggests that teledermatology decreases the delays in the management of melanoma lesions [34-36,40,59], reduces the referrals to a dermatologist by avoiding unnecessary consultations [35,36,40] and limits the number of patients lost from follow-up [36,60]. Moreno-Ramirez et al confirmed this point during an experimental teledermatology program without a smartphone [61]. Other authors have not been as optimistic and suggested that the many false negatives and easy access to a dermatologist's opinion may increase the number of secondary consultations [34].

Large variations in the kappa concordance coefficient might be related to the range of melanoma prevalence (depending on the recruited population). On the one hand, this result should lead to multiple studies in the general population to assess the performance of the procedures in a primary care setting. On the other hand, this review emphasizes that the publication of data related to up-to-date technologies is a challenge in teledermatology [20,62]; it is notable that up to 70% of the photographs were of poor quality in a study performed in 2007 [41], whereas all photographs were interpretable in 2016 [54].

The importance of training the person taking the photograph has been underlined [33,35-37,40], especially when a dermoscope is used [63]. Today, it is surprising to note that no standards for taking photographs exist [62,64]. Our review identified a few characteristics related to the quality of the photographs, such as several views of the same lesion, close-up and distant pictures, use of the autofocus and macro mode, a neutral background, and good lighting. The homogenization of practices based on these quality criteria is required to obtain better study reproducibility.

This review recalls that to give an opinion solely based on photographs, whatever their quality, is a challenge for dermatologists [29,41,63]. For example, the absence of palpation is one of the limitations of teledermatology. Thissens's study noted the need to obtain supplementary clinical information [54].

General practitioners and patients are likely to omit suspect lesions [31,33,39]. This limitation, which has been described in general practice [13,65], exists in teledermatology, that is, general practitioners could miss up to 30% of melanomas [66]. Another difficulty is the omission of specific areas that are either difficult to access, such as hair, wounds, or the ear [51,54], or those considered sensitive (genitalia) [67,68].

Store-and-forward teledermatology is well-accepted by patients and caregivers [59,69,70]. Patients report that one limitation of the apps is the difficulty of not having any human contact [20]. For both procedures, a limitation is the loss of the face-to-face patient-physician relationship, which may be critical since a melanoma is diagnosed early, at a severe stage, or in the aging population [59,69-71]. Positive perspectives may improve compliance to referrals and possibilities for the patient to participate actively in his or her health [24].

Confidentiality, security, and traceability of data exchange are the major ethical and legal stakes [20,63,72]. Although the abusive use of clinical photographs has become an increasing preoccupation of health fund organizations [73], only 30% of patients worry about the future of their photographs [74], and our review reported that only slightly more than half of the authors encrypted their data. Although 60% of specialists continue to store photographs of their patients in their personal mobile phones [72,74-77], one perspective could be to develop recourse to encrypt the medical-image libraries [62].

This review is original because of its specific focus on (1) melanoma early detection (mortality issues are not comparable for other dermatological pathologies) [20,62,78,79], (2) a primary care perspective (the sensitivity and specificity of a test depend on the prevalence of the disease), and (3) the use of a smartphone, that is, a tool implemented worldwide at low cost (not comparable to other expensive videoconference procedures) without limiting the study selection to apps [20,78-80]. However, this study had several weaknesses. First, this review had a large scope because we hypothesized that evidence might be sparse—the heterogeneity in study designs, the populations, the end points, the references used, and the presentations of the results made data comparison difficult. Second, the review was only based on MEDLINE. Third, the selection bias was high, and the prevalence of melanoma, which ranged from 0% to 100%, depended on the population studied. Fourth, the material used differed from one study to the next and from year to year, hence introducing a bias in evaluation.

Our review confirmed the absence of evidence of the safety and efficacy of smartphone medical apps. In contrast, our review found evidence that store-and-forward teledermatology using smartphones may affect patients' care courses, delays in obtaining a dermatologist consultation, and patients' referral to secondary care. Further research is required to determine the quality criteria, as there was major variability among the studies.

Acknowledgments

The authors wish to thank Sasha Mann, who helped draft the manuscript. Financial support was provided by a grant from the Ministry of Health. The funding source had no role in the design or conduct of the study; the collection, management, analysis,

or interpretation of the data; the preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Authors' Contributions

CR conceived the study, was responsible for its design and supervision, and was responsible for drafting the manuscript. SH participated in the design of the study, managed the selection of the included study, participated in data collection, and helped draft the manuscript. JR participated in the design of the study, managed the selection of the included study, participated in data collection, and helped draft the manuscript. GQ and BD provided administrative and technical support and revised the manuscript critically for important intellectual content. JMN obtained the grant for the corresponding research program, was responsible for the research project, and helped draft the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2012. GLOBOCAN 2012: Estimated cancer incidence, mortality, and prevalence worldwide in 2012 URL: http://globocan.iarc.fr/Pages/fact_sheets_population.aspx [accessed 2018-03-07] [WebCite Cache ID 6xkCQWnGK]
2. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Cancer Stat Facts: Melanoma of the Skin URL: https://seer.cancer.gov/statfacts/html/melan.html [accessed 2018-03-07] [WebCite Cache ID 6xkDNTH9K]
3. Curiel-Lewandrowski C, Kim CC, Swetter SM, Chen SC, Halpern AC, Kirkwood JM, Melanoma Prevention Working Group—Pigmented Skin Lesion Sub-Committee. Survival is not the only valuable end point in melanoma screening. J Invest Dermatol 2012 May;132(5):1332-1337 [FREE Full text] [doi: 10.1038/jid.2012.3] [Medline: 22336950]
4. Geller AC, Elwood M, Swetter SM, Brooks DR, Aitken J, Youl PH, et al. Factors related to the presentation of thin and thick nodular melanoma from a population-based cancer registry in Queensland Australia. Cancer 2009 Mar 15;115(6):1318-1327 [FREE Full text] [doi: 10.1002/cncr.24162] [Medline: 19189368]
5. Van Durme DJ, Ferrante JM, Pal N, Wathington D, Roetzheim RG, Gonzalez EC. Demographic predictors of melanoma stage at diagnosis. Arch Fam Med 2000 Jul;9(7):606-611. [Medline: 10910307]
6. Richard MA, Grob JJ, Avril MF, Delaunay M, Gouvernet J, Wolkenstein P, et al. Delays in diagnosis and melanoma prognosis (I): the role of patients. Int J Cancer 2000;89(3):271-279. [Medline: 10861504]
7. Grange F, Barbe C, Aubin F, Lipsker D, Granel-Brocard F, Velten M, et al. Clinical and sociodemographic characteristics associated with thick melanomas: a population-based, case-case study in France. Arch Dermatol 2012;148(12):1370-1376. [Medline: 22986518]
8. Davis JL, Buchanan KL, Katz RV, Green BL. Gender differences in cancer screening beliefs, behaviors, and willingness to participate: implications for health promotion. Am J Mens Health 2012;6(3):211-217. [Medline: 22071507]
9. Montella M, Crispo A, Grimaldi M, De Marco MR, Ascierto PA, Parasole R, et al. An assessment of factors related to tumor thickness and delay in diagnosis of melanoma in Southern Italy. Prev Med. ? 2002;35(3):271-277. [Medline: 12202070]
10. Baumert J, Plewig G, Volkenandt M, Schmid-Wendtner MH. Factors associated with a high tumour thickness in patients with melanoma. Br J Dermatol 2007;156(5):938-944. [Medline: 17381454]
11. Mandalà M, Imberti GL, Piazzalunga D, Belfiglio M, Lucisano G, Labianca R, et al. Association of socioeconomic status with Breslow thickness and disease-free and overall survival in stage I-II primary cutaneous melanoma. Mayo Clin Proc 2011;86(2):113-119. [Medline: 21282485]
12. Rat C, Quereux G, Monegier du Sorbier M, Gaultier A, Bonnaud-Antignac A, Khammari A, et al. Patients at elevated risk of melanoma: individual predictors of non-compliance to GP referral for a dermatologist consultation. Prev Med 2014;64:48-53. [Medline: 24726740]
13. Richard MA, Grob JJ, Avril MF, Delaunay M, Gouvernet J, Wolkenstein P, et al. Delays in diagnosis and melanoma prognosis (II): the role of doctors. Int J Cancer 2000;89(3):280-285. [Medline: 10861505]
14. Uhlenhake E, Brodell R, Mostow E. The dermatology work force: A focus on urban versus rural wait times. J Am Acad Dermatol 2009;61(1):17-22. [Medline: 19427058]
15. Roetzheim RG, Lee JH, Ferrante JM, Gonzalez EC, Chen R, Fisher KJ, et al. The influence of dermatologist and primary care physician visits on melanoma outcomes among medicare beneficiaries. J Am Board Fam Med 2013;26(6):637-647. [Medline: 24204060]
16. Stitzenberg KB, Thomas NE, Dalton K, Brier SE, Ollila DW, Berwick M, et al. Distance to diagnosing provider as a measure of access for patients with melanoma. Arch Dermatol 2007 Aug;143(8):991-998 [FREE Full text] [doi: 10.1001/archderm.143.8.991] [Medline: 17709657]
17. Tsang MW, Resneck JS. Even patients with changing moles face long dermatology appointment wait-times: a study of simulated patient calls to dermatologists. J Am Acad Dermatol 2006 Jul;55(1):54-58. [doi: 10.1016/j.jaad.2006.04.001] [Medline: 16781292]

18. WHO Group Consultation on Health Telematics. A health telematics policy in support of WHO's Health-for-all strategy for global health development: Report of the WHO Group Consultation on Health Telematics, 11-16 December, Geneva, 1997 URL: http://apps.who.int/iris/bitstream/10665/63857/1/WHO_DGO_98.1.pdf [accessed 2018-03-07] [WebCite Cache ID 6xkDf1dOo]
19. World Health Organization. Geneva, Switzerland; 2010. Telemedicine: opportunities and developments in member states: Report on the second Global survey on eHealth URL: http://apps.who.int/iris/bitstream/handle/10665/44497/9789241564144_eng.pdf?sequence=1&isAllowed=y
20. Tensen E, van der Heijden JP, Jaspers MW, Witkamp L. Two decades of teledermatology: current status and integration in national healthcare systems. Curr Dermatol Rep 2016;5:96-104 [FREE Full text] [doi: 10.1007/s13671-016-0136-7] [Medline: 27182461]
21. Kozera EK, Yang A, Murrell DF. Patient and practitioner satisfaction with tele-dermatology including Australia's indigenous population: a systematic review of the literature. Int J Womens Dermatol 2016 Sep;2(3):70-73 [FREE Full text] [doi: 10.1016/j.ijwd.2016.06.004] [Medline: 28492013]
22. Resneck J, Kimball AB. The dermatology workforce shortage. J Am Acad Dermatol 2004 Jan;50(1):50-54. [doi: 10.1016/j.jaad.2003.07.001] [Medline: 14699364]
23. Armstrong AW, Dorer DJ, Lugn NE, Kvedar JC. Economic evaluation of interactive teledermatology compared with conventional care. Telemed J E Health 2007 Apr;13(2):91-99. [doi: 10.1089/tmj.2006.0035] [Medline: 17489695]
24. Coates SJ, Kvedar J, Granstein RD. Teledermatology: from historical perspective to emerging techniques of the modern era: part I: history, rationale, and current practice. J Am Acad Dermatol 2015 Apr;72(4):563-74; quiz 575. [doi: 10.1016/j.jaad.2014.07.061] [Medline: 25773407]
25. van der Heijden JP, de Keizer NF, Bos JD, Spuls PI, Witkamp L. Teledermatology applied following patient selection by general practitioners in daily practice improves efficiency and quality of care at lower cost. Br J Dermatol 2011 Nov;165(5):1058-1065. [doi: 10.1111/j.1365-2133.2011.10509.x] [Medline: 21729026]
26. May C, Giles L, Gupta G. Prospective observational comparative study assessing the role of store and forward teledermatology triage in skin cancer. Clin Exp Dermatol 2008 Nov;33(6):736-739. [doi: 10.1111/j.1365-2230.2008.02850.x] [Medline: 18681874]
27. Warshaw EM, Hillman YJ, Greer NL, Hagel EM, MacDonald R, Rutks IR, et al. Teledermatology for diagnosis and management of skin conditions: a systematic review. J Am Acad Dermatol 2011 Apr;64(4):759-772. [doi: 10.1016/j.jaad.2010.08.026] [Medline: 21036419]
28. Senior K. Smart phones: new clinical tools in oncology? Lancet Oncol 2011 May;12(5):429-430. [doi: 10.1016/S1470-2045(11)70116-4] [Medline: 21536224]
29. Börve A, Terstappen K, Sandberg C, Paoli J. Mobile teledermoscopy-there's an app for that!. Dermatol Pract Concept 2013 Apr;3(2):41-48 [FREE Full text] [doi: 10.5826/dpc.0302a05] [Medline: 23785643]
30. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
31. Boyce Z, Gilmore S, Xu C, Soyer HP. The remote assessment of melanocytic skin lesions: a viable alternative to face-to-face consultation. Dermatology 2011;223(3):244-250. [doi: 10.1159/000333363] [Medline: 22095005]
32. Lamel SA, Haldeman KM, Ely H, Kovarik CL, Pak H, Armstrong AW. Application of mobile teledermatology for skin cancer screening. J Am Acad Dermatol 2012 Oct;67(4):576-581. [doi: 10.1016/j.jaad.2011.11.957] [Medline: 22243769]
33. Janda M, Loescher LJ, Banan P, Horsham C, Soyer HP. Lesion selection by melanoma high-risk consumers during skin self-examination using mobile teledermoscopy. J Am Med Assoc Dermatol 2014 Jun;150(6):656-658. [doi: 10.1001/jamadermatol.2013.7743] [Medline: 24522284]
34. Ford JA, Pereira A. Does teledermatology reduces secondary care referrals and is it acceptable to patients and doctors?: A service evaluation. J Eval Clin Pract 2015 Aug;21(4):710-716. [doi: 10.1111/jep.12373] [Medline: 25903046]
35. Börve A, Dahlén Gyllencreutz J, Terstappen K, Johansson Backman E, Aldenbratt A, Danielsson M, et al. Smartphone teledermoscopy referrals: a novel process for improved triage of skin cancer patients. Acta Derm Venereol 2015 Feb;95(2):186-190 [FREE Full text] [doi: 10.2340/00015555-1906] [Medline: 24923283]
36. Hue L, Makhloufi S, Sall N'Diaye P, Blanchet-Bardon C, Sulimovic L, Pomykala F, et al. Real-time mobile teledermoscopy for skin cancer screening targeting an agricultural population: an experiment on 289 patients in France. J Eur Acad Dermatol Venereol 2016 Jan;30(1):20-24. [doi: 10.1111/jdv.13404] [Medline: 26568080]
37. Janda M, Loescher LJ, Soyer HP. Enhanced skin self-examination: a novel approach to skin cancer monitoring and follow-up. J Am Med Assoc Dermatol 2013 Feb;149(2):231-236. [doi: 10.1001/jamadermatol.2013.1218] [Medline: 23426490]
38. Kroemer S, Frühauf J, Campbell TM, Massone C, Schwantzer G, Soyer HP, et al. Mobile teledermatology for skin tumour screening: diagnostic accuracy of clinical and dermoscopic image tele-evaluation using cellular phones. Br J Dermatol 2011 May;164(5):973-979. [doi: 10.1111/j.1365-2133.2011.10208.x] [Medline: 21219286]
39. Manahan MN, Soyer HP, Loescher LJ, Horsham C, Vagenas D, Whiteman DC, et al. A pilot trial of mobile, patient-performed teledermoscopy. Br J Dermatol 2015 Apr;172(4):1072-1080. [doi: 10.1111/bjd.13550] [Medline: 25418126]

40. Markun S, Scherz N, Rosemann T, Tandjung R, Braun RP. Mobile teledermatology for skin cancer screening: a diagnostic accuracy study. Medicine (Baltimore) 2017 Mar;96(10):e6278 [FREE Full text] [doi: 10.1097/MD.0000000000006278] [Medline: 28272243]
41. Massone C, Hofmann-Wellenhof R, Ahlgrimm-Siess V, Gabler G, Ebner C, Soyer HP. Melanoma screening with cellular phones. PLoS One 2007 May 30;2(5):e483 [FREE Full text] [doi: 10.1371/journal.pone.0000483] [Medline: 17534433]
42. Wu X, Oliveria SA, Yagerman S, Chen L, DeFazio J, Braun R, et al. Feasibility and efficacy of patient-initiated mobile teledermoscopy for short-term monitoring of clinically atypical nevi. J Am Med Assoc Dermatol 2015 May;151(5):489-496. [doi: 10.1001/jamadermatol.2014.3837] [Medline: 25629626]
43. Do TT, Zhou Y, Zheng H, Cheung NM, Koh D. Early melanoma diagnosis with mobile imaging. Conf Proc IEEE Eng Med Biol Soc 2014;2014:6752-6757. [doi: 10.1109/EMBC.2014.6945178] [Medline: 25571546]
44. Doukas C, Stagkopoulos P, Kiranoudis CT, Maglogiannis I. Automated skin lesion assessment using mobile technologies and cloud platforms. Conf Proc IEEE Eng Med Biol Soc 2012;2012:2444-2447. [doi: 10.1109/EMBC.2012.6346458] [Medline: 23366419]
45. Ferrero NA, Morrell DS, Burkhart CN. Skin scan: a demonstration of the need for FDA regulation of medical apps on iPhone. J Am Acad Dermatol 2013 Mar;68(3):515-516. [doi: 10.1016/j.jaad.2012.10.045] [Medline: 23394920]
46. Ramlakhan K, Shang Y. A mobile automated skin lesion classification system. Presented at: 23rd IEEE International Conference on tools with artificial intelligence (ICTAI); November 7-9, 2011; Boca Raton, FL, USA. [doi: 10.1109/ICTAI.2011.29]
47. Wadhawan T, Situ N, Lancaster K, Yuan X, Zouridakis G. SkinScan©: a portable library for melanoma detection on handheld devices. Proc IEEE Int Symp Biomed Imaging 2011 Mar 30;2011:133-136 [FREE Full text] [doi: 10.1109/ISBI.2011.5872372] [Medline: 21892382]
48. Wadhawan T, Situ N, Rui H, Lancaster K, Yuan X, Zouridakis G. Implementation of the 7-point checklist for melanoma detection on smart handheld devices. Conf Proc IEEE Eng Med Biol Soc 2011;2011:3180-3183 [FREE Full text] [doi: 10.1109/IEMBS.2011.6090866] [Medline: 22255015]
49. Wolf JA, Moreau JF, Akilov O, Patton T, English 3rd JC, Ho J, et al. Diagnostic inaccuracy of smartphone applications for melanoma detection. J Am Med Assoc Dermatol 2013 Apr;149(4):422-426 [FREE Full text] [doi: 10.1001/jamadermatol.2013.2382] [Medline: 23325302]
50. Dorairaj JJ, Healy GM, McInerney A, Hussey AJ. Validation of a melanoma risk assessment smartphone application. Dermatol Surg 2017 Feb;43(2):299-302. [doi: 10.1097/DSS.0000000000000916] [Medline: 28165352]
51. Maier T, Kulichova D, Schotten K, Astrid R, Ruzicka T, Berking C, et al. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. J Eur Acad Dermatol Venereol 2015 Apr;29(4):663-667. [doi: 10.1111/jdv.12648] [Medline: 25087492]
52. Ngoo A, Finnane A, McMeniman E, Tan JM, Janda M, Soyer HP. Efficacy of smartphone applications in high-risk pigmented lesions. Australas J Dermatol 2017 Feb 27:1-1. [doi: 10.1111/ajd.12599] [Medline: 28240347]
53. Robson Y, Blackford S, Roberts D. Caution in melanoma risk analysis with smartphone application technology. Br J Dermatol 2012 Sep;167(3):703-704. [doi: 10.1111/j.1365-2133.2012.11046.x] [Medline: 22762381]
54. Thissen M, Udrea A, Hacking M, von Braunmuehl T, Ruzicka T. mHealth App for risk assessment of pigmented and nonpigmented skin lesions-a study on sensitivity and specificity in detecting malignancy. Telemed J E Health 2017 May 31;23(12):948-954. [doi: 10.1089/tmj.2016.0259] [Medline: 28562195]
55. Bristol University. QUADAS 2 URL: https://www.bristol.ac.uk/media-library/sites/quadas/migrated/documents/background-doc.pdf [accessed 2018-03-07] [WebCite Cache ID 6xkTldflI]
56. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 28;546(7660):686. [doi: 10.1038/nature22985] [Medline: 28658222]
57. Kassianos AP, Emery JD, Murchie P, Walter FM. Smartphone applications for melanoma detection by community, patient and generalist clinician users: a review. Br J Dermatol 2015 Jun;172(6):1507-1518. [doi: 10.1111/bjd.13665] [Medline: 25600815]
58. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
59. Horsham C, Loescher LJ, Whiteman DC, Soyer HP, Janda M. Consumer acceptance of patient-performed mobile teledermoscopy for the early detection of melanoma. Br J Dermatol 2016 Dec;175(6):1301-1310. [doi: 10.1111/bjd.14630] [Medline: 27037999]
60. Bezalel S, Fabri P, Park HS. Implementation of store-and-forward teledermatology and its associated effect on patient access in a veterans affairs dermatology clinic. J Am Med Assoc Dermatol 2015 May;151(5):556-557. [doi: 10.1001/jamadermatol.2014.5272] [Medline: 25671336]
61. Moreno-Ramírez D, Argenziano G. Teledermatology and mobile applications in the management of patients with skin lesions. Acta Derm Venereol 2017 Jul 05:1-1 [FREE Full text] [doi: 10.2340/00015555-2718] [Medline: 28676881]
62. Ashique KT, Kaliyadan F, Aurangabadkar SJ. Clinical photography in dermatology using smartphones: an overview. Indian Dermatol Online J 2015;6(3):158-163 [FREE Full text] [doi: 10.4103/2229-5178.156381] [Medline: 26009708]

63. The Scottish Centre for Telehealth and Telecare, British Association of Dermatologists, British Dermatological Nursing Group, Primary Care Dermatology Society. 2013. Quality standards for teledermatology using Store And Forward Images URL: https://www.bad.org.uk/shared/get-file.ashx?itemtype=document&id=794
64. Quigley EA, Tokay BA, Jewell ST, Marchetti MA, Halpern AC. Technology and technique standards for camera-acquired digital dermatologic images: a systematic review. J Am Med Assoc Dermatol 2015 Aug;151(8):883-890. [doi: 10.1001/jamadermatol.2015.33] [Medline: 25970844]
65. Rat C, Grimault C, Quereux G, Dagorne M, Gaultier A, Khammari A, et al. Proposal for an annual skin examination by a general practitioner for patients at high risk for melanoma: a French cohort study. Br Med J Open 2015 Jul 29;5(7):e007471 [FREE Full text] [doi: 10.1136/bmjopen-2014-007471] [Medline: 26224016]
66. Gendreau JL, Gemelas J, Wang M, Capulong D, Lau C, Bratten DM, et al. Unimaged melanomas in store-and-forward teledermatology. Telemed J E Health 2017 Jun;23(6):517-520. [doi: 10.1089/tmj.2016.0170] [Medline: 27929365]
67. Lakdawala N, Fontanella D, Grant-Kels JM. Ethical considerations in dermatologic photography. Clin Dermatol 2012;30(5):486-491. [doi: 10.1016/j.clindermatol.2011.06.017] [Medline: 22902218]
68. Leger MC, Wu T, Haimovic A, Kaplan R, Sanchez M, Cohen D, et al. Patient perspectives on medical photography in dermatology. Dermatol Surg 2014 Sep;40(9):1028-1037. [doi: 10.1097/01.DSS.0000452632.22081.79] [Medline: 25099296]
69. Spinks J, Janda M, Soyer HP, Whitty JA. Consumer preferences for teledermoscopy screening to detect melanoma early. J Telemed Telecare 2016 Jan;22(1):39-46. [doi: 10.1177/1357633X15586701] [Medline: 26026184]
70. Mounessa JS, Chapman S, Braunberger T, Qin R, Lipoff JB, Dellavalle RP, et al. A systematic review of satisfaction with teledermatology. J Telemed Telecare 2017 Jan 01:1357633X17696587. [doi: 10.1177/1357633X17696587] [Medline: 28350281]
71. Hall S, Murchie P. Can we use technology to encourage self-monitoring by people treated for melanoma? A qualitative exploration of the perceptions of potential recipients. Support Care Cancer 2014 Jun;22(6):1663-1671. [doi: 10.1007/s00520-014-2133-3] [Medline: 24510193]
72. Kunde L, McMeniman E, Parker M. Clinical photography in dermatology: ethical and medico-legal considerations in the age of digital and smartphone technology. Australas J Dermatol 2013 Aug;54(3):192-197. [doi: 10.1111/ajd.12063] [Medline: 23713892]
73. Mahar PD, Foley PA, Sheed-Finck A, Baker CS. Legal considerations of consent and privacy in the context of clinical photography in Australian medical practice. Med J Aust 2013 Jan 21;198(1):48-49. [Medline: 23330771]
74. Anyanwu CO, Lipoff JB. Smartphones, photography, and security in dermatology. J Am Acad Dermatol 2015 Jan;72(1):193-195. [doi: 10.1016/j.jaad.2014.09.035] [Medline: 25497925]
75. Chan N, Charette J, Dumestre DO, Fraulin FO. Should 'smart phones' be used for patient photography? Plast Surg (Oakv) 2016;24(1):32-34 [FREE Full text] [Medline: 27054136]
76. Burns K, Belton S. Clinicians and their cameras: policy, ethics and practice in an Australian tertiary hospital. Aust Health Rev 2013 Sep;37(4):437-441. [doi: 10.1071/AH12039] [Medline: 23777890]
77. Abbott LM, Magnusson RS, Gibbs E, Smith SD. Smartphone use in dermatology for clinical photography and consultation: current practice and the law. Australas J Dermatol 2017 Feb 28:1-1. [doi: 10.1111/ajd.12583] [Medline: 28247404]
78. Snoswell C, Finnane A, Janda M, Soyer HP, Whitty JA. Cost-effectiveness of store-and-forward teledermatology: a systematic review. J Am Med Assoc Dermatol 2016 Jun 01;152(6):702-708. [doi: 10.1001/jamadermatol.2016.0525] [Medline: 27074289]
79. Cook SE, Palmer LC, Shuler FD. Smartphone mobile applications to enhance diagnosis of skin cancer: a guide for the rural practitioner. W V Med J 2015;111(5):22-28. [Medline: 26521532]
80. Chao E, Meenan CK, Ferris LK. Smartphone-based applications for skin monitoring and melanoma detection. Dermatol Clin 2017 Oct;35(4):551-557. [doi: 10.1016/j.det.2017.06.014] [Medline: 28886812]

Abbreviations

QUADAS-2: quality assessment of diagnostic accuracy 2nd edition

Edited by G Eysenbach; submitted 15.11.17; peer-reviewed by L Ferris, M Janda; comments to author 27.12.17; revised version received 09.01.18; accepted 10.01.18; published 13.04.18.

Please cite as:
Rat C, Hild S, Rault Sérandour J, Gaultier A, Quereux G, Dreno B, Nguyen JM
Use of Smartphones for Early Detection of Melanoma: Systematic Review
J Med Internet Res 2018;20(4):e135
URL: http://www.jmir.org/2018/4/e135/
doi:10.2196/jmir.9392
PMID:29653918

Original Paper

Uptake of Tailored Text Message Smoking Cessation Support in Pregnancy When Advertised on the Internet (MiQuit): Observational Study

Joanne L Emery[1], PhD; Tim Coleman[2], MD; Stephen Sutton[1], PhD; Sue Cooper[2], PhD; Jo Leonardi-Bee[3], PhD; Matthew Jones[2], PhD; Felix Naughton[4], PhD

[1]Behavioral Science Group, Institute of Public Health, University of Cambridge, Cambridge, United Kingdom

[2]Division of Primary Care, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[3]Division of Epidemiology and Public Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[4]School of Health Sciences, University of East Anglia, Norwich, United Kingdom

Corresponding Author:
Joanne L Emery, PhD
Behavioral Science Group
Institute of Public Health
University of Cambridge
Forvie Site
Cambridge Biomedical Campus
Cambridge, CB2 0SR
United Kingdom
Phone: 44 0 1223 330355
Email: jle40@medschl.cam.ac.uk

Abstract

Background: Smoking in pregnancy is a major public health concern. Pregnant smokers are particularly difficult to reach, with low uptake of support options and few effective interventions. Text message–based self-help is a promising, low-cost intervention for this population, but its real-world uptake is largely unknown.

Objective: The objective of this study was to explore the uptake and cost-effectiveness of a tailored, theory-guided, text message intervention for pregnant smokers ("*MiQuit*") when advertised on the internet.

Methods: Links to a website providing *MiQuit* initiation information (texting a short code) were advertised on a cost-per-click basis on 2 websites (Google Search and Facebook; £1000 budget each) and free of charge within smoking-in-pregnancy webpages on 2 noncommercial websites (National Childbirth Trust and NHS Choices). Daily budgets were capped to allow the Google and Facebook adverts to run for 1 and 3 months, respectively. We recorded the number of times adverts were shown and clicked on, the number of *MiQuit* initiations, the characteristics of those initiating *MiQuit*, and whether support was discontinued prematurely. For the commercial adverts, we calculated the cost per initiation and, using quit rates obtained from an earlier clinical trial, estimated the cost per additional quitter.

Results: With equal capped budgets, there were 812 and 1889 advert clicks to the *MiQuit* website from Google (search-based) and Facebook (banner) adverts, respectively. *MiQuit* was initiated by 5.2% (42/812) of those clicking via Google (95% CI 3.9%-6.9%) and 2.22% (42/1889) of those clicking via Facebook (95% CI 1.65%-2.99%). Adverts on noncommercial webpages generated 53 clicks over 6 months, with 9 initiations (9/53, 17%; 95% CI 9%-30%). For the commercial websites combined, mean cost per initiation was £24.73; estimated cost per additional quitter, including text delivery costs, was £735.86 (95% CI £227.66-£5223.93). Those initiating *MiQuit* via Google were typically very early in pregnancy (median gestation 5 weeks, interquartile range 10 weeks); those initiating via Facebook were distributed more evenly across pregnancy (median gestation 16 weeks, interquartile range 14 weeks).

Conclusions: Commercial online adverts are a feasible, likely cost-effective method for engaging pregnant smokers in digital cessation support and may generate uptake at a faster rate than noncommercial websites. As a strategy for implementing *MiQuit*, online advertising has large reach potential and can offer support to a hard-to-reach population of smokers.

(J Med Internet Res 2018;20(4):e146) doi:10.2196/jmir.8525

KEYWORDS

smoking cessation; pregnancy; internet; telemedicine; public health; social media

Introduction

Background

In developed countries, smoking during pregnancy is a leading preventable cause of adverse prenatal outcomes, including miscarriage [1], stillbirth [2,3], and prematurity [4]. It is also associated with a wide range of infant health problems [5]. In the United Kingdom, around 11% of women are estimated to smoke throughout pregnancy [6], but rates rise considerably with increasing social deprivation [6,7], standing at around 5 times higher in the most deprived women than in the least [7]. Children born to smokers are also more likely to become smokers themselves [8]. Thus, smoking in pregnancy not only puts great financial burden on health services but also perpetuates and exacerbates health inequalities. Reducing its prevalence is a public health priority [9].

Most pregnant smokers want to quit [10], and effective interventions exist to help them [11,12]. Specialist Stop Smoking Services in England offer free pregnancy cessation support with proven efficacy [12]; however, uptake is low [13], with convenience and concerns about being judged reported as barriers to access [14]. In addition to addressing these barriers, research efforts have also been focused on developing effective and cost-effective "distance" alternatives that will appeal to pregnant smokers and be used sufficiently to yield a public health benefit. Self-help cessation support appeals to pregnant smokers [15], and delivering self-help by mobile phone text messaging may be helpful for this group, given its low cost, convenience, anonymity, and wide reach potential, with mobile phone ownership high across the socioeconomic spectrum [16]. Systematic review evidence shows that self-help cessation interventions for pregnant smokers can be effective [17] and that mobile phone–based cessation interventions are effective for nonpregnant smokers [18].

MiQuit Intervention for Pregnant Smokers

We have developed a low-cost, tailored, text-messaging intervention specifically for pregnant smokers ("*MiQuit*") [19]. *MiQuit* is feasible to deliver and highly acceptable to pregnant smokers [19], and a recent randomized controlled trial (RCT; n=407) found that offering *MiQuit* in addition to usual care shows promising efficacy and cost-effectiveness [20]. As *MiQuit* is fully automated and user-initiated, women can start using it without the need for any health professional involvement, thus minimizing potential implementation costs. However, little is currently known about the likely real-world uptake of *MiQuit* should this become routinely available to pregnant smokers, or what the best implementation strategies might be to maximize its reach and initiate users into support as cost-effectively as possible. The public health impact of an intervention depends crucially on its real-world uptake, as well as its efficacy, but evaluations of smoking cessation interventions have largely neglected to estimate this [21].

Using the Internet to Offer Cessation Support

The internet has obvious potential as a tool for reaching pregnant smokers and enrolling them into cessation programs, and evidence suggests that a digital intervention may particularly appeal to smokers offered cessation support through digital media [22]. Interventions can be promoted to potential users on the internet through commercial search-engine and banner-based (pop-up) adverts, as well as noncommercial websites. A previous real-world study in an antenatal setting estimated that 3% to 4% of pregnant smokers initiated *MiQuit* after a brief promotional leaflet was placed into their maternity booking pack without any introduction or endorsement from a health professional [23]. If pregnant women will initiate *MiQuit* after reading a brief leaflet, it seems likely that they may do so after reading an online advert. Offering *MiQuit* to pregnant smokers through search engines, in particular, might reach them earlier in pregnancy than they would typically be targeted in antenatal settings, thus maximizing the benefits of quitting to the fetus. Search-based adverts could also present an opportunity to offer support to women when they are motivated to quit, with cohort evidence suggesting that repeated quit attempts may be made throughout pregnancy [24]. As a tool for recruiting smokers into research trials of digital interventions, studies of nonpregnant groups [25-30] suggest that commercial online advertising can achieve high participant yield rapidly [26,27] and can recruit traditionally hard-to-reach smoker populations at relatively low cost [29,30]. However, although studies typically show similar participant characteristics and retention rates for smokers recruited through online versus traditional means [26-28], with the notable exception of younger age in those recruited via social media [27], others have found lower quitting confidence and lower study completion rates among smokers recruited to trials via the internet [31].

We are aware of only 2 previous studies to explore real-world uptake of digital smoking cessation support, rather than recruitment rates to cessation trials, as a consequence of online advertising, both of which targeted nonpregnant smokers [32,33]. To our knowledge, no published studies have explored uptake of cessation support among pregnant smokers via the internet. In addition, we know little about the characteristics of pregnant smokers who can be encouraged to take up digital interventions over the internet. In this study, therefore, we investigate whether pregnant smokers will initiate the *MiQuit* intervention after seeing paid-for or free online advertising; in addition, we monitor the costs incurred and, using a previously obtained estimate for *MiQuit* efficacy [20], assess the extent to which commercial online advertising might be cost-effective. Finally, we document the extent to which users engage with the support program, including the discontinuation rate, and describe key characteristics, exploring differences between pregnant smokers initiating *MiQuit* via different online routes.

Methods

Design

This was an evaluation of the uptake of a digital (text-messaging) intervention advertised on the internet. Uptake rates of *MiQuit* were monitored while it was advertised via 4 concurrent online settings: 2 commercial websites and 2 noncommercial webpage links.

MiQuit Cessation Support for Pregnant Smokers

MiQuit provides a 12-week program of automated, theory-guided, interactive support for quitting smoking in pregnancy, delivered by text message. Support is tailored to 12 baseline user characteristics plus name, gestation, and smoking status, the latter collected at 3 and 7 weeks by text message. Tailoring characteristics include nicotine dependence, partner's smoking status, and confidence, motivation, and determination to quit [19,20,23]. Women initiate *MiQuit* by texting a short code number. They are then invited to complete 12 baseline tailoring questions, including the option of setting a quit date, either by text or by website. Those tailoring by website must answer all other 11 questions. Those tailoring by text are given the option of answering either 6 or 12 tailoring questions but can stop responding at any point. If no tailoring questions are answered using either route, then generic support is delivered. *MiQuit* delivers 0-2 scheduled daily texts ("push" support), including motivational messages; advice about quit attempt preparation, managing cravings, or trigger situations; and information about fetal development and how smoking affects it. Those setting a quit date receive extra support oriented around their nominated date. Users can access on-demand, "pull" support for combatting cravings ("HELP"), returning to abstinence after a lapse ("SLIP"), or for distraction ("QUIZ"). The support lasts for 12 weeks unless discontinued prematurely by the user sending a "STOP" message. *MiQuit* texts are free to receive. Sending the initiation text and any subsequent texts sent by the user are either free or cost the user's standard text message rate, depending on their phone "bundle."

Web-Based Advertising Campaign

Overview of Advertising Methods

With the aim of reaching as many pregnant smokers as possible, we chose 2 commercial advertisers with very large reach potential: Google AdWords [34] (search-based) and Facebook Ads [35] (banner). To identify UK websites most likely to appear as a result of internet searches for smoking-in-pregnancy keywords, 2 search engines (Google and Bing) were used to identify the top webpages returned for the phrase "quit smoking in pregnancy" and close variants. The National Health Service (NHS) website ("NHS Choices") and the National Childbirth Trust (NCT) website, whose "smoking in pregnancy" webpages [36,37] were consistently close to the top of the search results, agreed to place free-of-charge, text-only links to *MiQuit* on these webpages.

Advert Content

Key points made in the adverts were that *MiQuit* is smoking cessation support by text message; *MiQuit* is for pregnant smokers; *MiQuit* is NHS supported; and *MiQuit* is free to receive. Separate adverts were created for each of the 4 online settings, with input from a Patient and Public Involvement representative, keeping the text as similar as possible between adverts given their character or space limits (Multimedia Appendix 1).

MiQuit Sign-Up Website

Clicking on any of the 4 adverts led directly to a *MiQuit* sign-up website that provided further information about *MiQuit* and how to initiate it. Each advert led to a separate website clone with a different short code number, enabling us to isolate traffic and initiations from each source. Those wanting to initiate *MiQuit* had to navigate to the "sign-up" page, click on the "sign-up" button, and submit a response to a question asking where they first heard about *MiQuit* ("submissions"). The latter acted as a check that women had not reached the website through other means than our online adverts, such as through the recommendation of a health professional. They were then presented with the short code number on a webpage, with instructions to text the word "QUIT" to the number to begin support. The 4 short codes were not promoted anywhere outside of the 4 cloned websites. To ensure that the websites would not appear in the results of search engines, we added the "disallow" command on the websites' robots.txt file, which requests Web robots not to scan the websites. This was checked periodically to ensure the most commonly used search engines complied with this request.

Commercial Advert Settings

Google AdWords displays a brief, 3-line, text-only advert when advertiser-specified keywords are typed into Google Search. Facebook Ads display a text and image banner advert, unsolicited, to a specified demographic (eg, by age, gender, location, and interests), potentially multiple times per person. We added an image of a pregnant smoker, used elsewhere for promoting *MiQuit* [23], to the text for the Facebook advert. Detailed descriptions of Google and Facebook advertising can be found elsewhere [29,30] but, with both, costs depend on competition from other advertisers. We used a cost-per-click option for both adverts. As we could find no similar studies among pregnant smokers to inform how expenditure would translate into initiations, we set a capped budget of £1000 for each. We restricted both adverts to the United Kingdom, but put no time of day or day of week limits on their scheduling. Google keyword phrases specified were "smoking in pregnancy," "stop smoking in pregnancy," "dangers of smoking in pregnancy," and close variants. Estimated search traffic was relatively low for these; so, broad-match keywords, which permit any combination of the words comprising the phrase, were added to widen reach. On the basis of estimated search traffic and click costs for our keywords (provided by Google), we set a daily budget of £33 for the Google advert, hence a campaign duration of 1 month. For Facebook, we restricted our advert to females aged 16-45 years and specified "pregnancy" and "childbirth" as interests. On the basis of the estimated click costs for our target audience (provided by Facebook), we set a daily budget of £10 for the Facebook advert, hence a campaign duration of 3 months.

Free Links

A brief, text-only advert was displayed, permanently for 6 months, on both the NCT and NHS Choices webpage on smoking in pregnancy, under an "external links" section. These had low screen visibility compared with the 2 commercial adverts.

Procedure

The 4 adverts were run concurrently, beginning late May 2015, either until their budget ended (commercial adverts) or for 6 months (free links). Initiations were permitted up to 3 months after each advert ceased to be shown. The 2 commercial adverts were monitored closely throughout the campaigns. The performance metrics these supply (shown below) were compiled on a weekly basis. Numbers of *MiQuit* submissions, initiations, and discontinuations were compiled weekly for all 4 sources.

Measures and Analyses

Advert Performance and Uptake of MiQuit

The 2 commercial advertisers supplied a variety of reach and cost metrics (Textbox 1). The 2 noncommercial websites, NHS Choices and NCT, each provided the number of unique visits to their smoking in pregnancy webpage, where our advert was located. We used Google Analytics on the landing webpages of the *MiQuit* sign-up website for these noncommercial adverts to determine the number of clicks they received.

For all 4 cloned *MiQuit* websites, the *MiQuit* server recorded the number of times the initiation short code was accessed (submissions), whether submissions were from a desktop or mobile phone, and the number of *MiQuit* initiations. As in similar online uptake studies among nonpregnant smokers [32,33], uptake rate was calculated as the percentage of *MiQuit* initiations out of the total number of clicks on an advert to the *MiQuit* sign-up website. Data are presented as frequencies and percentages, with 95% Wilson CIs. For the commercial adverts, we also estimated the number of initiations per individual reached and calculated the cost per initiation. For the purposes of analyses, it was assumed that each advert click, submission, and initiation represented one individual, and that each Google advert impression represented one individual reached. Facebook provided the number of individuals reached by the advert ("people served"), given that an individual can be targeted repeatedly.

Estimated Cost-Effectiveness

We estimated the likely incremental cost per additional quitter (also known as the "Incremental Cost-Effectiveness Ratio") of both initiating women into *MiQuit* via commercial online advertising and delivering the support by summing the mean commercial advert cost and mean cost of sending the texts in this study (£0.035 per text at time of study), divided by the incremental quit rate found in a recent RCT of the *MiQuit* intervention (3.46%) [20]. In this RCT, pregnant smokers (n=407) were recruited from 16 antenatal clinics in England via face-to-face contact and, after responding to tailoring questions by telephone to a researcher, were randomized to receive either *MiQuit* added to usual NHS smoking cessation care or usual care alone. The quit rate for prolonged, biochemically-validated abstinence in the *MiQuit* group was 5.42% (1.96% for usual care) and the odds ratio, adjusted for site and gestation, was 2.70 (95% CI 0.93-9.35) for *MiQuit* over usual care [20]. This is the best estimate yet produced for the likely efficacy of *MiQuit*, although it has limited precision. To determine the impact of uncertainty, we bootstrapped 1000 times our incremental quit rate and cost to estimate the 95% CIs for the cost per additional quitter [38]. In addition, there was a fixed annual running cost, shared across all users, of approximately £760, consisting of a virtual reply number (£99), Web hosting with domain name (£240), and short code (£420). This was not included in the cost-effectiveness analysis as a per-person cost could not be calculated: the annual number of users is currently an unknown quantity.

User Engagement and Characteristics

MiQuit server data were used to assess engagement with the support program, including rates of tailoring question completion, quit date setting, use of "pull-support" features, and discontinuations (sending a "STOP" message). Data are presented as frequencies and percentages, with 95% Wilson CIs presented for key measures (discontinuation rate and quit date setting). Key behavioral characteristics of those initiating *MiQuit* were taken from their responses to the tailoring questions, answered by Web or text. Characteristics were compared between those initiating *MiQuit* via different online sources, where numbers permitted, using Mann-Whitney *U* tests (continuous data) and Fisher exact tests (categorical data).

Textbox 1. Advert reach and cost metrics supplied by the 2 commercial advertisers.

- Impressions—number of times the advert was displayed.
- People served (Facebook only)—number of individuals the advert was displayed to.
- Impression share (Google only)—proportion of times the advert was displayed when a relevant keyword search was made. This shows the number of impressions achievable given an unlimited budget.
- Clicks—number of times the advert was clicked on. A click took the user directly to the *MiQuit* sign-up website, so the number of clicks equates to the number of website visits.
- Cost per click—mean cost incurred for a single advert click.
- Proportion of impressions and clicks by device type (desktop or mobile).
- Mean screen position (Google only)—mean location of advert impressions on the Google search results page (1=top of screen).

Ethical Approval

Advice was sought from the National Research Ethics Central Queries Service as to whether the study should be classed as research requiring ethical review, and they confirmed that no ethical review was required. The study was conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki, 1996; the Principles of Good Clinical Practice; and the Department of Health Research Governance Framework for Health and Social Care, 2005. Participants were able to withdraw from the *MiQuit* support at any time.

Results

Advert Performance and Uptake of MiQuit

Most commercial advert clicks came from mobile devices rather than desktops (Google: 560/812, 69.0%; Facebook: 1883/1889, 99.68%). Of those who accessed the initiation short code, 94.1% (301/320) did so from a mobile device (Google: 110/121, 90.9%; Facebook: 184/187, 98.4%; NHS Choices: 1/4, 25%; NCT: 6/8, 75%).

Figure 1 shows the flow of the targeted populations, through each online route, into initiation of *MiQuit* support. The Google advert was shown 29,022 times in its 1-month duration. Given our impression share of 70% for mobile-based searches and 50% for desktop-based searches, we estimated that over 46,000 Google searches had been made for our keywords, in the United Kingdom, during this time. The mean position of the Google advert from the top of the screen was 1.0 for mobiles and 1.2 for desktops (highly visible). The Facebook advert was shown to 248,618 broadly targeted women during its 3-month duration (mean 2.4 times each) and also had high visibility on screen (Multimedia Appendix 1). In 6 months, approximately 40,000 unique visits were made to the NHS Choices and NCT smoking-in-pregnancy webpages containing the *MiQuit* link, but the proportions who scrolled down to where the links were placed on these is unknown.

In total, 2754 individuals clicked on 1 of the 4 adverts to the *MiQuit* sign-up website (Google n=812, Facebook n=1889, NHS Choices n=33, NCT n=20). For Google, assuming 1 advert impression per person, this amounted to 2.80% of those served the advert (812/29,022) and, for Facebook, 0.76% of those served the advert (1889/248,618). For the NHS Choices and NCT websites, the numbers of clicks on the *MiQuit* links amounted to 0.09% (33/38,352) and 1.43% (20/1402), respectively, of the numbers of unique visits to the webpages containing the link.

MiQuit was initiated by 93 individuals in total, with the 2 commercial campaigns each yielding 42 initiations and the free links 9 in total. For our uptake rate calculation, the percentage who subsequently initiated *MiQuit* after clicking on an advert to the *MiQuit* website was 3.38% (93/2754, 95% CI 2.76%-4.12%); Google 5.2% (42/812, 95% CI 3.9%-6.9%), Facebook 2.22% (42/1889, 95% CI 1.65%-2.99%); NHS Choices 9% (3/33; 95% CI 3%-24%), NCT 30% (6/20; 95% CI 15%-52%). One in 691 Google advert impressions resulted in a *MiQuit* initiation (1 in 5919 women targeted by Facebook). One in 12,784 visits to the NHS Choices webpage, and 1 in 234 visits to the NCT webpage, resulted in an initiation.

Commercial Advertising Costs and Estimated Cost-Effectiveness

The Facebook campaign cost £1000, whereas the Google campaign cost £1077, including £75 credited free to our account. Table 1 shows a breakdown of these costs in terms of how far *MiQuit* was accessed or activated. The mean cost per advert click to the *MiQuit* website was £1.33 for the Google advert and £0.53 for the Facebook advert. Both campaigns yielded equal initiations for their budget. Cost per *MiQuit* initiation was £25.64 via Google, £23.81 via Facebook, and £24.73 across both campaigns. Table 1 also shows the estimated cost per additional quitter of initiating pregnant smokers into *MiQuit* via commercial online advertising. Using the cost per initiation of the commercial adverts (£24.73) plus the mean cost of sending the *MiQuit* texts to those initiating support here (£2.73) gave a point estimate of £793.64 per additional quitter to both initiate and deliver the support. The mean cost per additional quitter from the bootstrap was £735.86 (95% CI £227.66-£5223.93).

Engagement and Disengagement With MiQuit

A total of 53 of the 93 initiators in this study (57%, 95% CI 47%-67%) set at least 1 quit date with *MiQuit* during the program, including at baseline. Moreover, 63 (68%) of the initiators chose to answer all 12 tailoring questions by website (with quit date noncompulsory). Of those who chose to answer by text message instead (n=27), 16 (59%) responded to at least the first 6 questions. Only 3 initiators (3%) answered no tailoring questions and received generic support. “Pull” support features (“HELP,” “SLIP,” and “QUIZ” requests) were used by 35 (38%) initiators.

A total of 34 of the 93 initiators in this study (37%, 95% CI 27%-47%) stopped the 86-day program prematurely (mean days into program 18.6, SD 21.3). Although not formally tested, discontinuation rates, tailoring question completion rates, and use of interactive program features did not appear to differ between those taking up *MiQuit* via the 4 different online routes.

Characteristics of Those Initiating MiQuit

Table 2 shows key characteristics of those who initiated *MiQuit* via online advertising (n=93) and statistically compares those who initiated *MiQuit* via Google versus Facebook; numbers were insufficient to compare data from the 2 free links (shown combined). Readiness to quit smoking appeared high among initiators in this study, with 70 out of 93 (78%) seriously planning to quit within the next 2 weeks. Gestation differed substantially between women who initiated support via Google versus Facebook; no other characteristic differed significantly between them. Those from Google were typically very early in pregnancy, with 49% reporting a baseline gestation of 4 or 5 weeks (median 5 weeks; interquartile range [IQR] 10); those from Facebook were distributed more widely across pregnancy (median 16 weeks, IQR 14, Mann-Whitney *U* test $P<.001$).

Figure 1. Flow diagram: *MiQuit* advert reach and initiation of support.

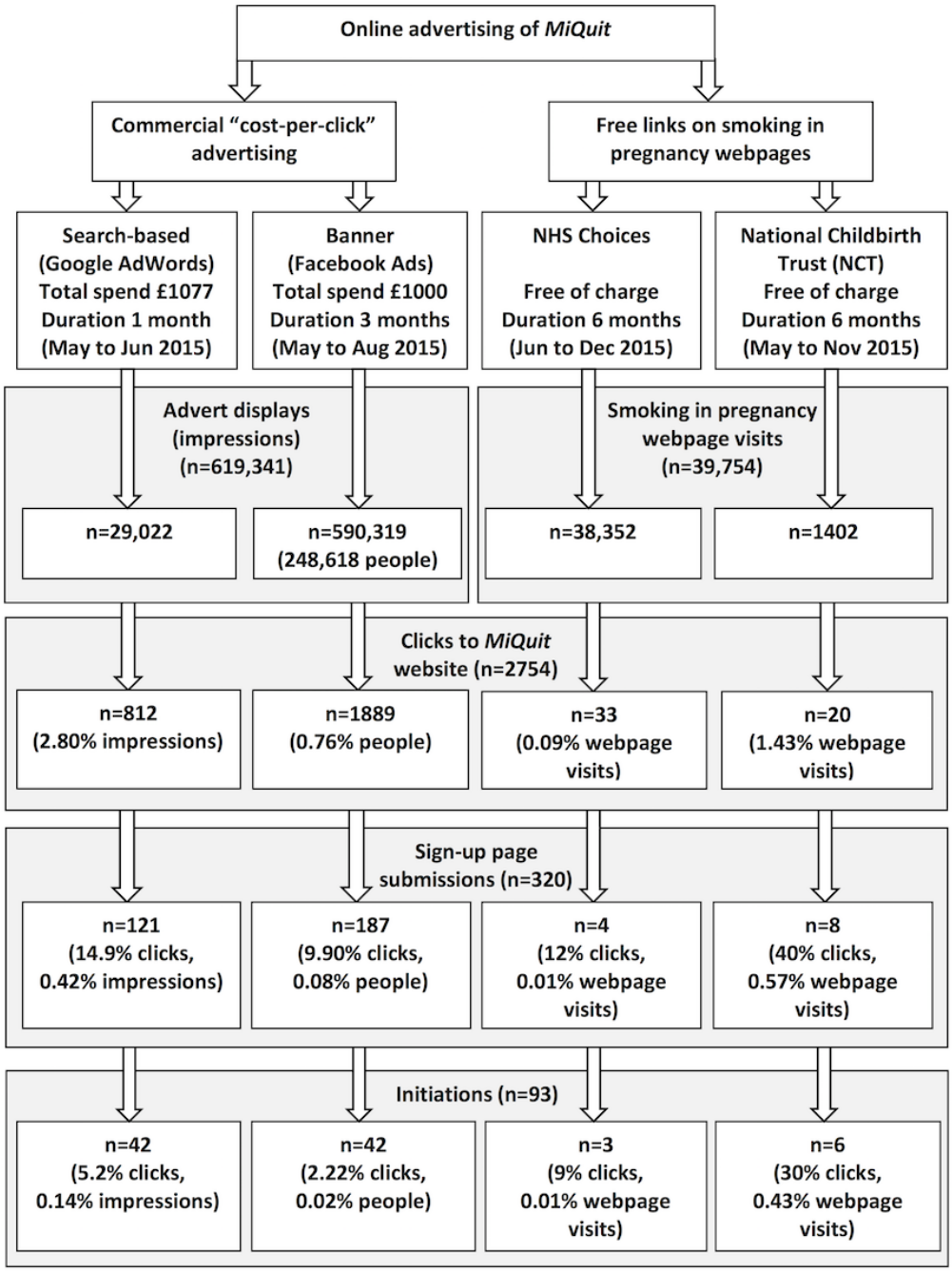


Table 1. Costs and estimated cost-effectiveness of the commercial online adverts.

Advertising campaign	Cost per:			Estimated cost per additional quitter[a]
	Advert click, leading to the *MiQuit* website	Short code obtained	*MiQuit* initiation	
Google AdWords (spend £1077)[b]	£1.33 (n=812)	£8.90 (n=121)	£25.64 (n=42)	£741.04
Facebook Ads (spend £1000)	£0.53 (n=1889)	£5.35 (n=187)	£23.81 (n=42)	£688.15
Both campaigns (spend £2077)[b]	£0.77 (n=2701)	£6.74 (n=308)	£24.73 (n=84)	£714.74

[a]On the basis of an incremental quit rate of 3.46% in the *MiQuit* randomized controlled trial (RCT) [20] (prolonged, validated abstinence).

[b]Including £75 credited free to our account by Google as a welcome offer.

Table 2. Baseline characteristics of pregnant smokers initiating *MiQuit* via online advertising.

Baseline characteristic	Total (n=93)	Google AdWords (n=42)	Facebook Ads (n=42)	Free links (total)[a] (n=9)	*P* value[b]
Gestation (weeks)					**<.001**
Mean (SD)	12.3 (8.5)	9.1 (7.2)	16.7 (8.5)	7.1 (4.1)	
Median (1st Q, 3rd Q)	9.5 (5, 18)	5 (4, 14)	16 (10, 24)	6 (4, 8)	
Min, max	2, 32	2, 28	4, 32	4, 17	
Valid n (%)	90 (96.8)	41 (97.6)	40 (95.2)	9 (100)	
Are you seriously planning to quit? n (%)					**.30**
Within the next 2 weeks	70 (78)	35 (83)	28 (72)	7 (78)	
Within the next 30 days	15 (17)	6 (14)	7 (18)	2 (22)	
Within the next 3 months	3 (3)	0 (0)	3 (8)	0 (0)	
No	2 (2)	1 (2)	1 (2)	0 (0)	
Valid n (%)	90 (97)	42 (100)	39 (93)	9 (100)	
Confidence to quit for remainder of pregnancy, n (%)					**.36**
Not at all	21 (26)	6 (17)	13 (35)	2 (25)	
A little	18 (23)	8 (23)	7 (19)	3 (38)	
Moderately	29 (36)	15 (43)	12 (32)	2 (25)	
Very much	11 (14)	6 (17)	4 (11)	1 (13)	
Extremely	1 (1)	0 (0)	1 (3)	0 (0)	
Valid n (%)	80 (86)	35 (83)	37 (88)	8 (89)	
Cigarettes per day now, n (%)					**.33**
1-3	6 (8)	5 (14)	1 (3)	0 (0)	
4-5	16 (20)	6 (17)	9 (25)	1 (13)	
6-10	23 (29)	9 (26)	12 (33)	2 (25)	
11-15	19 (24)	9 (26)	8 (22)	2 (25)	
16-20	12 (15)	4 (11)	6 (17)	2 (25)	
21+	3 (4)	2 (6)	0 (0)	1 (13)	
Valid n (%)	79 (85)	35 (83)	36 (86)	8 (89)	
Heaviness of Smoking Index[c] , n (%)					**.24**
Very low	19 (24)	6 (17)	12 (33)	1 (13)	
Low to moderate	33 (42)	18 (51)	12 (33)	3 (38)	
Moderate	24 (30)	9 (26)	12 (33)	3 (38)	
High	3 (4)	2 (6)	0 (0)	1 (13)	
Valid n (%)	79 (85)	35 (83)	36 (86)	8 (89)	
Partner's smoking status, n (%)					**.30**
Smoker	43 (61)	16 (53)	22 (69)	5 (63)	
Nonsmoker or no partner	27 (39)	14 (47)	10 (31)	3 (38)	
Valid n (%)	70 (75)	30 (71)	32 (76)	8 (89)	

[a]Data for NCT (n=6) and NHS Choices (n=3) were combined because of small numbers.

[b]Facebook vs Google *P* value. Tested via Mann-Whitney *U* (continuous) or Fisher Exact test (frequencies).

[c]Heaviness of Smoking Index was based on the sum of scores from 2 items of the Fagerström Test of Cigarette Dependence [39]: cigarettes per day (1-10=score of 0, 11-20=1, 21-30=2, >30=3) and time to first cigarette after waking (>1 hour=0, 31-60 min=1, 6-30 min=2, within 5 min=3). A combined score of 0-2=very low dependence, 3=low to moderate dependence, 4=moderate dependence, and 5-6=high dependence.

Discussion

Principal Findings

When a low-cost, text messaging, pregnancy smoking cessation support program (*MiQuit*) was advertised via the internet, with no other form of promotion or recommendation, an overall uptake (initiation) rate of 3.4% was seen among those who clicked on any of the 4 adverts to the *MiQuit* website. Commercial adverts, which yielded the vast majority of initiations in this study, cost, on average, £24.73 per initiation. Although the initiation rate was higher among those who reached the *MiQuit* website via free webpage links, the total number of initiations generated through these was much lower than the number of initiations generated by commercial advertising and occurred over a longer time frame. Behavioral characteristics appeared similar for those initiating *MiQuit* from different online sources. User engagement appeared high as over half of initiators set a quit date with the system, and approximately two-thirds continued with *MiQuit* until the end of the 12-week program.

Strengths and Limitations

To our knowledge, this is the first study to explore the feasibility of using free and paid-for online advertising to reach pregnant smokers and promote their uptake of cessation support. We have shown that a significant minority of pregnant smokers are willing to initiate an automated text messaging intervention when offered this via the internet; given the high reach of the internet, this could translate into substantial numbers of pregnant smokers supported to quit. Unlike many previous evaluations of online advertising [25], we were able to trace the source of all those who arrived at the *MiQuit* website and all who initiated support, allowing us to compare uptake rates for different online strategies. By quantifying each step of the uptake process, from viewing an advert through to initiating support, we have explored a spectrum of engagement [40] with the offer of support and identified steps where barriers to uptake might be removed. Importantly, we tracked individuals' behavior beyond support initiation, investigating their engagement with the program (quit date setting, use of "pull" support messages, tailoring question completion) as well as discontinuations (sending a "STOP" message). We were also able to use previous RCT data, with assumptions, to estimate the likely cost-effectiveness of *MiQuit* if implemented via commercial online adverts.

There are several limitations to this study, including some general challenges for real-world uptake studies. As with many studies recruiting people to interventions without person involvement, we cannot be sure that those who took up support were the intended group (ie, pregnant smokers), although our adverts and website made it clear that *MiQuit* is for pregnant smokers and almost all provided their gestation. It is also possible that adverts were clicked on out of curiosity by those who were neither pregnant nor smoking, particularly among women targeted by banner adverts on Facebook, or that adverts were passed on to others (eg, friends) by those originally targeted. In addition, it proved difficult to estimate the numbers of pregnant smokers exposed to each advert as a denominator: the number of pregnant smokers shown our Facebook advert is unknown and it is probable that not all Google searches for our keywords were made by pregnant smokers. Nonetheless, the high number of keyword searches made within the United Kingdom in a month (over 46,000) suggests that there may be many opportunities for engaging with this population on the internet. Given the low visibility of the 2 free links, our use of the number of visits to the webpages on which these were placed as a denominator may have greatly underestimated the proportions initiating *MiQuit* out of those exposed. Difficulties in specifying accurate and comparable denominators have been noted previously in uptake studies [23,41]. However, to health providers, having accurate information on the numbers of *MiQuit* initiations and on the cost of providing this is of the greatest importance.

It is important to be aware of assumptions inherent to the cost-effectiveness estimates presented. Our cost-effectiveness analyses assumed that, when initiated via the internet, *MiQuit* would have a similar quit rate to that observed in women recruited to an RCT from antenatal clinics. However, we found differences in characteristics between women in this study and those recruited to the RCT (discussed below), and it is possible that the intervention could vary in effectiveness for different groups of women. Both Facebook and Google had relatively opaque criteria used to determine their advertising charges; there is no guarantee that similar charges would be made for identical adverts in the future. Charges depend on concurrent competition from other advertisers, so are unlikely to replicate exactly from one campaign to another even if all other parameters are held constant. It is probably best to view our cost-effectiveness estimates as indicative rather than definitive; however, optimization of this type of "programmatic" advertising would likely reduce costs significantly.

Findings in Context

Compared with pregnant smokers who were recruited to a large RCT of *MiQuit* via antenatal clinics in a previous study [20], women who initiated *MiQuit* via online advertising in this study had higher readiness to quit, with 78% (vs 32% RCT) seriously planning to quit within the next 2 weeks and 57% (vs 19% RCT) sending a quit date to *MiQuit* during the program, including at baseline. This suggests that online initiators may be more likely to make a quit attempt [42,43]. Conversely, online initiators appeared to be more nicotine-dependent than RCT recruits, with 34% (vs 14% RCT) classed as "moderate" or "high" dependence, and this appears to be a key determinant of failure to quit during pregnancy [43-46]. It is, therefore, possible that quit rates might be lower among individuals who engage with *MiQuit* via the internet compared with those who did so after being recruited to a trial in an NHS setting, but this is currently speculative. Others have not found heavier smoking among online recruits [31].This is an important avenue for future evaluation: *MiQuit* could have different effects depending on how it is implemented and, therefore, who makes use of it. Previous research has found high readiness to quit in smokers recruited to RCTs by search-based [29], but not banner, online advertising methods [28]; however, in this study, readiness to quit was similar between those initiating *MiQuit* via Google and Facebook.

A total of 37% of initiators in this study stopped the 12-week program prematurely (sent a "STOP" message). This was notably higher than the discontinuation rate among trial participants, where 13% randomized to the *MiQuit* condition discontinued the support [20], though more similar to the discontinuation rate when *MiQuit* was offered in an NHS real-world context, by leaflet (46%) [23]. It is possible that those receiving *MiQuit* as part of an RCT felt obliged to continue with it because of being a trial participant and receiving human contact as part of their involvement. Previous *MiQuit* research highlights that discontinuations are made for a variety of reasons, most of which are not related to irritation or dissatisfaction [19], and a separate study indicates that discontinuation can be an indicator of increased engagement in smoking cessation behavior [47]. In other ways, online initiators appeared more engaged with *MiQuit* than RCT participants, with 3 times as many sending a quit date to the system and almost twice as many (38% vs 21%) sending a "pull" support message.

Although our Facebook advert generated activations throughout pregnancy, those who initiated *MiQuit* via Google were often early in gestation (around 50% within their first 5 weeks). Adverts attached to internet search engines may, therefore, be a useful way to reach women when they are first pregnant and looking for support, and could potentially maximize health benefits by encouraging abstinence for more of their pregnancy. Currently, the earliest used cessation interventions tend to target pregnant smokers at their antenatal booking appointment, at around 8-12 weeks' gestation.

We have shown that uptake of *MiQuit* via online advertising is feasible; our previous real-world study showed uptake of *MiQuit* to be feasible when offered via leaflets in maternity booking packs without health professional promotion. This suggests that there are promising routes to initiating pregnant smokers into support systems such as *MiQuit*, without the need for health professional involvement, in both clinical and nonclinical settings. We are aware of no other studies that have investigated using the internet for offering real-world cessation support to pregnant smokers, although 2 previous studies have explored this among nonpregnant smokers in the United States [32,33]. Our overall uptake rate (3.4%) was lower than that found among all smokers offered a national Web-based cessation program via commercial search-based and banner online adverts (6.8%) [32]; however, it was similar to the uptake rate found when banner adverts were used to promote Web-based cessation support specifically to Latino smokers, another hard-to-reach group (2.8%) [33]. Our average cost per initiation (£24.73) compares favorably with costs reported to initiate general smokers into real-world digital cessation support commercially via the internet (mean $35) [32] and very favorably with costs reported to initiate a hard-to-reach group via online banner methods (mean $209.34) [33]. Facebook may be more cost-effective than other banner-based methods for targeting specific populations, given that adverts can be restricted to a particular demographic [48].

Using an efficacy estimate of *MiQuit* from a previous RCT, we estimated a cost per additional quitter of £735.86 to initiate pregnant smokers into *MiQuit* through paid adverts, including text message delivery costs but excluding development costs. We did not include a formal cost-utility analysis here, and the caveats discussed in our limitations must be noted, but this is encouraging compared with costs reported for other smoking cessation interventions that are effective and cost-effective in pregnancy. For example, financial incentives are highly cost-effective, with a cost per additional quitter of £1127 [49].

Of those who clicked to the *MiQuit* website, a much greater percentage (11.6%) obtained the short code to initiate *MiQuit* than subsequently texted it to do so (3.4%). Obtaining the short code required a number of extra steps after clicking on an advert and landing on the *MiQuit* website, suggesting that these women were serious about taking up *MiQuit*. There may thus be potential for increasing uptake substantially, at no extra advertising cost, by ameliorating the drop-off between clicking to accept the support and texting to initiate it. Having to text a short code may be a barrier to initiating support for a number of reasons, including lack of credit among pay-as-you-go phone users, suspicion of hidden charges, and needing to act outside of the website; enabling women to sign up anonymously without the need to text a short code might increase uptake. Website content, tone, and appearance are also potential targets; clearly labeling the website as an NHS service might also reduce women's barriers to sending an initiation text.

A definitive evaluation is planned for *MiQuit*. If it is shown to be effective, as our earlier trial suggests is likely [20], then an assessment of its efficacy in an online setting may be warranted. In this study, free-of-charge webpage links yielded relatively few initiations but might have performed better if given greater visibility. High initiation rates were found for women clicking from these links, suggesting that they were well targeted despite having lower reach than the commercial adverts, and pregnant smokers may have been more likely to initiate *MiQuit* if reaching it from a recognized health source. There may therefore be scope for future work to promote *MiQuit* via such websites. Future work could aim to minimize search engine–based advertising costs by investigating which specific "smoking in pregnancy" keyword phrases are associated with support initiations; this is possible if support is initiated by a webpage click rather than an external action such as texting a short code. Finally, it is important to establish whether uptake of text-based cessation support among pregnant smokers affects their uptake of traditional cessation support or whether it attracts those who would otherwise try to quit alone, if at all.

Conclusions

Commercial online advertising appears to be a promising method for initiating pregnant smokers into text message–based cessation support. Free and commercial adverts prompting pregnant smokers to click to a sign-up website resulted in an initiation rate of 3.4%. Search-based commercial advertising was able to reach women earlier in pregnancy than interventions delivered in clinical settings seem able to achieve, and those who initiated support in this study had high readiness to quit. Commercial online advertising to pregnant smokers is likely to be cost-effective and can probably be made more so. Given that pregnant smokers' uptake of traditional support is low, it is

important to find successful strategies for offering them effective alternatives.

Acknowledgments

The authors would like to thank Susanna Mountcastle for her Patient and Public Involvement input on the adverts. They would also like to thank Lucy Watson (Google AdWords), Sanjima DeZoysa and Karen Bryant (National Childbirth Trust Media Centre), and the NHS Choices Service Desk team.

This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Reference Number RP-PG-0109-10020). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. JE, TC, SS, SC, JLB, and FN are members of the UK Centre for Tobacco and Alcohol Studies (UKCTAS), a UKCRC Public Health Research Centre of Excellence. UKCTAS receives core funding from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the Department of Health, under the auspices of the UK Clinical Research Collaboration. JE, TC, SS, SC, JLB, and MJ are members of the NIHR School for Primary Care Research. JE received salary funding from the NIHR School for Primary Care Research (SPCR). FN received salary funding from a fellowship from the Society for the Study of Addiction. The research was also supported by the NIHR Collaboration for Leadership in Applied Health Research and Care East Midlands at NHS Nottingham City CCG.

Authors' Contributions

FN and SS designed and developed the *MiQuit* intervention. SS further developed and adapted the intervention for use in this study, and FN designed the *MiQuit* sign-up website. TC, FN, and SC conceived the study, and FN is the study's chief investigator. JE, FN, TC, SC, and JLB designed the study, drafted and edited the protocol, and designed the advert text. JE conducted the study and performed the statistical analysis, MJ performed the health economic analysis, and JLB was the study statistician. JE, FN, and TC drafted the manuscript. All authors read, commented on, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Advert appearance on screen.

[PDF File (Adobe PDF File), 1MB - jmir_v20i4e146_app1.pdf]

References

1. Pineles BL, Park E, Samet JM. Systematic review and meta-analysis of miscarriage and maternal exposure to tobacco smoke during pregnancy. Am J Epidemiol 2014 Apr 01;179(7):807-823 [FREE Full text] [doi: 10.1093/aje/kwt334] [Medline: 24518810]
2. Flenady V, Koopmans L, Middleton P, Frøen JF, Smith GC, Gibbons K, et al. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. Lancet 2011 Apr 16;377(9774):1331-1340. [doi: 10.1016/S0140-6736(10)62233-7] [Medline: 21496916]
3. Högberg L, Cnattingius S. The influence of maternal smoking habits on the risk of subsequent stillbirth: is there a causal relation? BJOG 2007 Jun;114(6):699-704. [doi: 10.1111/j.1471-0528.2007.01340.x] [Medline: 17516961]
4. Moore E, Blatt K, Chen A, Van Hook J, DeFranco EA. Relationship of trimester-specific smoking patterns and risk of preterm birth. Am J Obstet Gynecol 2016 Dec;215(1):109.e1-109.e6 [FREE Full text] [doi: 10.1016/j.ajog.2016.01.167] [Medline: 26827877]
5. Cnattingius S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. Nicotine Tob Res 2004 Apr;6(Suppl 2):S125-S140. [doi: 10.1080/14622200410001669187] [Medline: 15203816]
6. NHS Digital. 2016. Statistics on women's smoking status at time of delivery in England, Quarter 3, October to December 2015 URL: http://content.digital.nhs.uk/catalogue/PUB20219 [accessed 2017-07-15] [WebCite Cache ID 6ryPzobcF]
7. NHS Digital. 2012. Infant feeding survey - UK 2010 URL: http://content.digital.nhs.uk/catalogue/PUB08694 [accessed 2017-07-15] [WebCite Cache ID 6ryQA5Po7]
8. Leonardi-Bee J, Jere ML, Britton J. Exposure to parental and sibling smoking and the risk of smoking uptake in childhood and adolescence: a systematic review and meta-analysis. Thorax 2011 Oct;66(10):847-855. [doi: 10.1136/thx.2010.153379] [Medline: 21325144]
9. HM Government. 2011. Healthy lives, healthy people: A tobacco control plan for England URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213757/dh_124960.pdf [accessed 2017-07-15] [WebCite Cache ID 6ryQRTVQY]

10. Ussher M, West R, Hibbs N. A survey of pregnant smokers' interest in different types of smoking cessation support. Patient Educ Couns 2004 Jul;54(1):67-72. [doi: 10.1016/S0738-3991(03)00197-6] [Medline: 15210262]
11. Chamberlain C, O'Mara-Eves A, Oliver S, Caird JR, Perlen SM, Eades SJ, et al. Psychosocial interventions for supporting women to stop smoking in pregnancy. Cochrane Database Syst Rev 2013 Oct 23(10):CD001055 [FREE Full text] [doi: 10.1002/14651858.CD001055.pub4] [Medline: 24154953]
12. Bauld L, Bell K, McCullough L, Richardson L, Greaves L. The effectiveness of NHS smoking cessation services: a systematic review. J Public Health (Oxf) 2010 Mar;32(1):71-82. [doi: 10.1093/pubmed/fdp074] [Medline: 19638397]
13. NHS Digital. 2016. Statistics on NHS stop smoking services: England, April 2015 to March 2016 URL: http://content.digital.nhs.uk/catalogue/PUB21162 [accessed 2017-07-15] [WebCite Cache ID 6ryQZ0yGR]
14. Ussher M, Etter JF, West R. Perceived barriers to and benefits of attending a stop smoking course during pregnancy. Patient Educ Couns 2006 Jun;61(3):467-472. [doi: 10.1016/j.pec.2005.06.021] [Medline: 16098707]
15. Ussher M, West R, Hibbs N. A survey of pregnant smokers' interest in different types of smoking cessation support. Patient Educ Couns 2004 Jul;54(1):67-72. [doi: 10.1016/S0738-3991(03)00197-6] [Medline: 15210262]
16. Office for National Statistics. 2017. Percentage of households with durable goods by income group and household composition, UK, financial year ending 2016 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure/datasets/percentageofhouseholdswithdurablegoodsbyincomegroupandhouseholdcompositionuktablea46 [accessed 2017-07-15] [WebCite Cache ID 6ryQjNUSn]
17. Naughton F, Prevost AT, Sutton S. Self-help smoking cessation interventions in pregnancy: a systematic review and meta-analysis. Addiction 2008 Apr;103(4):566-579. [doi: 10.1111/j.1360-0443.2008.02140.x] [Medline: 18339103]
18. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
19. Naughton F, Prevost AT, Gilbert H, Sutton S. Randomized controlled trial evaluation of a tailored leaflet and SMS text message self-help intervention for pregnant smokers (MiQuit). Nicotine Tob Res 2012 May;14(5):569-577. [doi: 10.1093/ntr/ntr254] [Medline: 22311960]
20. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction 2017 Jul;112(7):1238-1249 [FREE Full text] [doi: 10.1111/add.13802] [Medline: 28239919]
21. Marcano Belisario JS, Bruggeling MN, Gunn LH, Brusamento S, Car J. Interventions for recruiting smokers into cessation programmes. Cochrane Database Syst Rev 2012;12:CD009187. [doi: 10.1002/14651858.CD009187.pub2] [Medline: 23235672]
22. Balmford J, Borland R, Benda P, Howard S. Factors associated with use of automated smoking cessation interventions: findings from the eQuit study. Health Educ Res 2013 Apr;28(2):288-299 [FREE Full text] [doi: 10.1093/her/cys104] [Medline: 23107931]
23. Naughton F, Cooper S, Bowker K, Campbell K, Sutton S, Leonardi-Bee J, et al. Adaptation and uptake evaluation of an SMS text message smoking cessation programme (MiQuit) for use in antenatal care. Br Med J Open 2015 Oct 22;5(10):e008871 [FREE Full text] [doi: 10.1136/bmjopen-2015-008871] [Medline: 26493459]
24. Pickett KE, Wakschlag LS, Dai L, Leventhal BL. Fluctuations of maternal smoking during pregnancy. Obstet Gynecol 2003 Jan;101(1):140-147. [Medline: 12517659]
25. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
26. Heffner JL, Wyszynski CM, Comstock B, Mercer LD, Bricker J. Overcoming recruitment challenges of web-based interventions for tobacco use: the case of web-based acceptance and commitment therapy for smoking cessation. Addict Behav 2013 Oct;38(10):2473-2476 [FREE Full text] [doi: 10.1016/j.addbeh.2013.05.004] [Medline: 23770645]
27. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]
28. Stanczyk NE, Bolman C, Smit ES, Candel MJ, Muris JW, de Vries H. How to encourage smokers to participate in web-based computer-tailored smoking cessation programs: a comparison of different recruitment strategies. Health Educ Res 2014 Feb;29(1):23-40 [FREE Full text] [doi: 10.1093/her/cyt104] [Medline: 24287402]
29. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. Nicotine Tob Res 2006 Dec;8(Suppl 1):S35-S41. [Medline: 17491169]
30. Ramo DE, Rodriguez TM, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
31. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]

32. Graham AL, Milner P, Saul JE, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. J Med Internet Res 2008;10(5):e50 [FREE Full text] [doi: 10.2196/jmir.1001] [Medline: 19073542]
33. Graham AL, Fang Y, Moreno JL, Streiff SL, Villegas J, Muñoz RF, et al. Online advertising to reach and recruit Latino smokers to an internet cessation program: impact and costs. J Med Internet Res 2012;14(4):e116 [FREE Full text] [doi: 10.2196/jmir.2162] [Medline: 22954502]
34. Google. Google AdWords. URL: https://adwords.google.com/intl/en_uk/home/ [accessed 2017-07-15] [WebCite Cache ID 6ryPjWpdJ]
35. Facebook. Facebook Ads URL: https://www.facebook.com/business/products/ads [accessed 2018-03-20] [WebCite Cache ID 6y3p8w3DU]
36. NHS Choices. Stop smoking in pregnancy URL: http://www.nhs.uk/conditions/pregnancy-and-baby/pages/smoking-pregnant.aspx [accessed 2017-07-15] [WebCite Cache ID 6ryMV6LqX]
37. National Childbirth Trust. Smoking during pregnancy URL: https://www.nct.org.uk/pregnancy/smoking-during-pregnancy [accessed 2018-03-20] [WebCite Cache ID 6y3pSuJDf]
38. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials. 2nd ed. Oxford: Oxford University Press; 2014:9780199685028.
39. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for nicotine dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [Medline: 1932883]
40. Platt T, Platt J, Thiel DB, Kardia SL. Facebook advertising across an engagement spectrum: a case example for public health communication. JMIR Public Health Surveill 2016 May 30;2(1):e27 [FREE Full text] [doi: 10.2196/publichealth.5623] [Medline: 27244774]
41. Graham AL, Bock BC, Cobb NK, Niaura R, Abrams DB. Characteristics of smokers reached and recruited to an internet smoking cessation trial: a case of denominators. Nicotine Tob Res 2006 Dec;8(Suppl 1):S43-S48 [FREE Full text] [Medline: 17491170]
42. Vangeli E, Stapleton J, Smit ES, Borland R, West R. Predictors of attempts to stop smoking and their success in adult general population samples: a systematic review. Addiction 2011 Dec;106(12):2110-2121. [doi: 10.1111/j.1360-0443.2011.03565.x] [Medline: 21752135]
43. Emery JL, Sutton S, Naughton F. Cognitive and behavioral predictors of quit attempts and biochemically-validated abstinence during pregnancy. Nicotine Tob Res 2017 May 01;19(5):547-554. [doi: 10.1093/ntr/ntw242] [Medline: 28403458]
44. Schneider S, Huy C, Schütz J, Diehl K. Smoking cessation during pregnancy: a systematic literature review. Drug Alcohol Rev 2010 Jan;29(1):81-90. [doi: 10.1111/j.1465-3362.2009.00098.x] [Medline: 20078687]
45. Vaz LR, Leonardi-Bee J, Aveyard P, Cooper S, Grainge M, Coleman T, SNAP trial team. Factors associated with smoking cessation in early and late pregnancy in the smoking, nicotine, and pregnancy trial: a trial of nicotine replacement therapy. Nicotine Tob Res 2014 Apr;16(4):381-389 [FREE Full text] [doi: 10.1093/ntr/ntt156] [Medline: 24127265]
46. Riaz M, Lewis S, Coleman T, Aveyard P, West R, Naughton F, et al. Which measures of cigarette dependence are predictors of smoking cessation during pregnancy? Analysis of data from a randomized controlled trial. Addiction 2016 Sep;111(9):1656-1665 [FREE Full text] [doi: 10.1111/add.13395] [Medline: 26997495]
47. Balmford J, Borland R. How do smokers use a smoking cessation text messaging intervention? Nicotine Tob Res 2014 Dec;16(12):1586-1592. [doi: 10.1093/ntr/ntu111] [Medline: 25031312]
48. 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;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
49. Boyd KA, Briggs AH, Bauld L, Sinclair L, Tappin D. Are financial incentives cost-effective to support smoking cessation during pregnancy? Addiction 2016 Feb;111(2):360-370. [doi: 10.1111/add.13160] [Medline: 26370095]

Abbreviations

NHS: National Health Service
NCT: National Childbirth Trust
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 21.07.17; peer-reviewed by S Ferguson, M Frandsen; comments to author 09.08.17; revised version received 23.08.17; accepted 23.08.17; published 19.04.18.

Please cite as:
Emery JL, Coleman T, Sutton S, Cooper S, Leonardi-Bee J, Jones M, Naughton F
Uptake of Tailored Text Message Smoking Cessation Support in Pregnancy When Advertised on the Internet (MiQuit): Observational Study
J Med Internet Res 2018;20(4):e146
URL: http://www.jmir.org/2018/4/e146/
doi:10.2196/jmir.8525
PMID:29674308

Original Paper

A Mobile Health App–Based Postnatal Educational Program (Home-but not Alone): Descriptive Qualitative Study

Shefaly Shorey[1], PhD; Yen Yen Yang[1], RN; Cindy-Lee Dennis[2], PhD

[1]Alice Lee Centre for Nursing Studies, National University of Singapore, Singapore, Singapore

[2]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Shefaly Shorey, PhD
Alice Lee Centre for Nursing Studies
National University of Singapore
Level 2, Clinical Research Centre, Block MD 11
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011294 ext 1294
Fax: 65 6776 7135
Email: nurssh@nus.edu.sg

Abstract

Background: The postnatal period poses numerous challenges for new parents. Various educational programs are available to support new parents during this stressful period. However, the usefulness of educational programs must be evaluated to ascertain their credibility.

Objective: The aim of this descriptive, qualitative study was to explore the views of parents of newborns with regard to the content and delivery of a mobile health (mHealth) app–based postnatal educational program.

Methods: A qualitative semistructured interview guide was used to collect data from 17 participants who belonged to the intervention group of a randomized controlled trial. The intervention, a 4-week-long access to a mHealth app–based educational program, was evaluated. The interviews were conducted in English and at the participants' homes. Thematic analysis was used to analyze the data. The Consolidated Criteria for Reporting Qualitative Research checklist was used to report the findings.

Results: The interviews revealed 4 main themes: (1) positive features of the mHealth app, (2) advice from midwives, (3) experiences gained from using the mHealth app, and (4) recommendations for the future. The participants evaluated the educational program to be a good source of information that was tailored to the local context. The different modes of delivery, including audio and video, accentuated the accessibility of information. The parents evaluated that the facilitator of the featured communication platform, a midwife, provided trustworthy advice. Belongingness to a virtual community beyond the hospital endowed the parents the confidence that they were not alone and were supported by other parents and health care professionals.

Conclusions: According to the parents, the mHealth app–based educational program was helpful in supporting a multi-ethnic sample of parents during the postnatal period. This insight indicates that the program could be implemented in a wide community of parents in the postnatal period. The helpfulness of the educational program is a testament of the potential benefits of using telemedicine among new parents postnatally. Resources can also be dedicated toward extending the duration of access to the app beyond 1 month and developing relevant content for parents across the perinatal period.

(J Med Internet Res 2018;20(4):e119) doi:10.2196/jmir.9188

KEYWORDS

parents; postnatal care; mobile applications; midwifery; nurse midwives; nursing

Introduction

Background

The postnatal period is a trying period for new parents as they adapt to the rigorous demands of parenthood [1]. During this period, parents seek to acquire new skills and find new ways of restoring balance in their lives [2]. Mothers experience challenging issues relating to health of the baby, breastfeeding, and varied sleeping patterns, which may lead to a myriad of emotions comprising self-doubt [3], anxiety disorders [4], and postpartum depression [5-8]. Although maternal morbidities have been extensively documented, fathers also experience high levels of stress while caring for the baby and adapting to their new role as a father [9]. Ineffective management of these stress levels could lead to depression [10], which may have adverse implications on child development, mother-infant interaction, as well as conjugal and family relationships [6,11]. Therefore, mitigating the development of such affective conditions in new parents through effective postnatal care and support is important.

Current Status of Postnatal Care

Postnatal care is crucial in enhancing maternal and child well-being; however, studies have established that postnatal care from hospitals is suboptimal [12-14]. Most health services provide early discharge, that is, reduced postnatal length of stay for new mothers to reduce health care costs while sustaining the quality of care [15]. However, this practice has raised some concerns in relation to postnatal education by health care professionals, maternal confidence in transitioning to the home setting, breastfeeding, and postnatal depression [14,16,17]. New mothers also expressed great dissatisfaction with the quality of postnatal care, including insufficient time for queries to be asked, information overload within a short period of time, and an inadequate amount of midwifery care [18,19].

With current research showing the need to improve the quality of postnatal care, various interventions that aim to address parental concerns have been developed. Such interventions include domiciliary visits, clinic visits, and telephone follow-ups [20]. Home visits may be effective in providing tailored care and increasing patient satisfaction [12]. However, these measures are less feasible for health care services to adopt and implement due to logistics, such as manpower and cost [20]. Furthermore, clinic follow-up visits and telephone follow-ups usually take place on a monthly basis, which may not be solely sufficient in providing assistance to new parents during their transition to parenthood [21].

Use of Technology in Postnatal Care

The use of technology has become a cost-effective avenue to deliver quality health care [22]. The application of information technology in enhancing an individual's access to health care has been regarded as a promising innovation that might revolutionize health care [23]. With the advancements of technology over the decades, controlling technological innovations is important to enhance the efficiency and continuity of care during the postnatal period, thereby enhancing the satisfaction of the postnatal care provided [24]. Available literature on the use of mobile health (mHealth) apps in the postnatal period is scarce [24-27]. In addition, such literature is focused on service utilization and limited to low-middle-income countries [25-27]. The development and evaluation of mHealth app–based educational programs are encouraged from stakeholder and user perspectives [25-27].

Therefore, this qualitative study evaluated the mHealth app–based educational program catered for multi-ethnic new parents in Singapore. The focus was on exploring the parents' perceptions of the content, strengths, and suggestions for improvements of the program.

Methods

Procedures and Participants

A descriptive qualitative study design was used. Semistructured interviews were conducted following the Consolidated Criteria for Reporting Qualitative Research. Participants were recruited from an intervention group of a randomized controlled trial [28]. The randomized controlled trial aimed to test the effectiveness of the mHealth app–based educational program among new parents. The participants were randomly divided into 2 groups, that is, control (receiving standardized care) and intervention (receiving educational program via the mHealth app and standardized care). English-speaking parents were invited to participate based on their parenting self-efficacy scores after the intervention. Parents with high and low parenting self-efficacy scores were purposively invited to participate in the process evaluation interviews. However, all the participants from the intervention group had parenting self-efficacy scores above the median score of 20 [28]. Therefore, all participants (50 couples, 2 fathers, and 1 mother) who completed the intervention (n=103) were approached, and 21 participants (5 couples, 7 fathers, and 4 mothers) agreed to participate subsequently. Moreover, 4 participants (3 fathers and 1 mother) subsequently refused to participate for varied reasons, such as being overseas or busy. Finally, 17 participants (5 couples, 4 fathers, and 3 mothers) were interviewed. Data saturation was achieved with 15 participants, and 2 additional interviews were conducted to ensure that no new findings emerged from the interviews.

Intervention

Participants in the intervention group gained access to the *Home-but not Alone* mHealth app–based educational program for 4 weeks after discharge from the hospital with routine care. Routine care comprised care and support provided by nurses and midwives throughout the participants' hospital stay, as well as a follow-up appointment with their obstetricians [12]. Various elements of the mHealth app–based educational program include a discussion forum that enabled parents to relay photographs or messages and have their queries and concerns addressed by a midwife once within 24 hours; an extensive information resource comprising audios, videos, and PDF documents on new born, maternal, and paternal care; and daily notifications received by the parents regarding their babies' important milestones and needs. Asynchronous mode of communication was used by the midwife to answer parental inquiries once a day. The educational program follows a conceptual framework based on Bandura's self-efficacy theory [29]. The intervention

sought to be a form of social support in enhancing parenting self-efficacy and emotional well-being through reducing depressive symptoms and increasing overall parenting satisfaction. The conceptual framework is presented in the study protocol [30].

Data Collection

After 4 weeks of access to the mHealth app–based educational program, parents were spoken to via a telephone call and were asked to participate in a qualitative semistructured face-to-face interview. The purpose was to understand their perceptions and feedback regarding the content, mode of delivery, and the effectiveness of the intervention. As both parents were given the opportunity to utilize the mHealth app, they were both eligible and invited for participation in the interview. Parents were informed that all interviews would be audiorecorded. A written, informed consent was obtained for their participation in the interviews. Finally, 5 couples, 4 fathers, and 3 mothers (n=17) participated in the interview. The interviews were conducted by the coauthor who was trained to conduct these interviews by the corresponding author. The interviews occurred at the participants' homes, and the participants were reminded of the interview 1 day in advance via a text message. The interviews lasted between 30 and 45 min with an average time of 35 min and were carried out in English. For couples, both fathers and mothers were interviewed at the same time as most of them requested it to be that way; however, they were reminded throughout the interview to remain objective and share their personal accounts of using the mHealth app. Anecdotal notes on observational data were collected, and both fathers and mothers shared equally about their experiences in using the mHealth app. Data collection lasted from May to July 2016.

A semistructured interview guide was developed to evaluate the perceptions of the educational program delivered via the mHealth app. The guide was developed with reference to the conceptual framework of the main study [30] and was evaluated by an obstetrician, 2 midwives, and a couple from the pilot study. Probing questions were created to attain a comprehensive perspective on the effectiveness of the mHealth app. Table 1 represents the interview guide. One pilot interview was conducted to test the interview guide and the interview process, which was excluded from the final data analysis. All 17 interviews were audiorecorded and transcribed verbatim.

Ethical Considerations

Ethical approval was acquired from the Institutional Review Board (National Health Group Domain Specific Review Board, Ref: 2015/01250) of the study hospital. Relevant information about the research study was comprehensively explained to the participants. Following which, written informed consent was obtained from each participant. Participation in this research was strictly voluntary, and confidentiality was adhered to.

Data Analysis

Thematic analysis was performed, in which the entire dataset was reviewed and the following 3 phase analysis were used: (1) preparation of data for analysis, (2) generation of initial codes, and (3) development of respective themes and subthemes.

The coauthor interviewed the participants and transcribed the audio data verbatim without editing any grammatical errors, which is useful in retaining distinct features of the participants' language, keeping it relevant to the local context [31]. The final transcripts were examined several times to gain familiarity with the dataset. The 2 authors were independently involved in the analysis of the content, development of the codes, and revelation of the eventual themes. An initial understanding of the content of the entire dataset was established.

Subsequently, the color-coding method was used to identify different concepts, forming the initial codes. These excerpts were then extracted and placed into a tabular format, entailing themes and subthemes within a new document. The authors reviewed the themes for likeness and differences before deciding the overarching themes. To achieve confirmability and objectivity in the analysis of data, the various themes and subthemes were discussed extensively through several meetings between the 2 authors. Any discrepancy was discussed with the third author. The field notes were often referred to, and constant comparative analyses were performed. Member checking was also established with the participants after each interview by allowing them access to their transcripts to ensure that the intended meanings behind their verbatim quotes were accurately captured. The final findings were shared with the participants, and they had no further inputs on the derived themes and subthemes.

Rigor

To enhance credibility and trustworthiness of the study, reflexivity was maintained throughout the data collection and analysis [32]. Field notes were taken to capture nuances in nonverbal responses from the participants and to enhance the authenticity of the verbal quotes. The coauthor maintained a diary and constantly self-reflected before and after the interview to maintain accuracy in reporting the participants' emic views during the data collection. The purpose of these reflections was to consciously acknowledge one's values, assumptions, and goals toward the study topic so that the researchers can clarify their belief systems and subjectivities toward participants' responses. A semistructured interview guide was developed based on a theoretical framework and was evaluated by the experts and the parents.

Table 1. Process evaluation semistructured guide. mHealth: mobile health.

Number	Probing question
1	What were some of the education or services that you received in the hospital to better prepare you for participation in your baby's care?
2	What other types of services have you participated in or received apart from those provided in the hospital?
3	Did you find the *Home-but not Alone* mHealth app useful? Probe: If yes, how? If no, why not?
4	What are the strengths of the *Home-but not Alone* mHealth app? Probe: Do you think the duration of the access to the mHealth app was sufficient?
5	What were the main difficulties that you faced while using the *Home-but not Alone* mHealth app? Probe: Can you describe any technical difficulties you might have faced? What would you want to improve about the app?
6	Do you think it was appropriate to introduce the mHealth app after childbirth? Probe: Would you want to have the app made available for your future babies? Do you think such an app should be made available during pregnancy too?
7	Do you think use of the *Home-but not Alone* app has enhanced your parenting confidence? Probe: If yes, how? If no, why not?
8	Do you think use of the *Home-but not Alone* app has enhanced your parenting satisfaction? Probe: If yes, how? If no, why not?
9	Were you satisfied with the added support you received in terms of the mHealth app? Probe: If yes, how? If no, why not?
10	Did you feel emotionally supported and stable with the additional support you received from the mHealth app? Probe: If yes, how? If no, why not?
11	Would you recommend the use of the *Home-but not Alone* mHealth app to others?

Results

Sample Characteristics

A total of 5 couples, 4 fathers, and 3 mothers were interviewed. Their ages ranged between 26 and 42 years. Among them, 41% (7/17) were Chinese, 29% (5/15) were Malay, 6% (1/17) were Indian, and 24% (4/17) were of other ethnicities. The majority were degree holders and full-time employed staff at various organizations. The majority (13/17, 86%) of participants had a monthly household income of more than SG $6000. Nearly half of the participants attended antenatal classes before delivery, and 71% (12/17) experienced a normal vaginal delivery. All the participants were married. A description of the participants' characteristics is displayed in Multimedia Appendix 1.

Four major themes emerged from the thematic analysis of the interviews: (1) positive features of the mHealth app, (2) advice from the midwife, (3) experiences gained from using the mHealth app, and (4) recommendations for the future. The themes and subthemes are summarized in Textbox 1.

Theme 1: Positive Features of the mHealth App

Most of the parents found that the mHealth app was a good informational resource that catered to the local context and new-generation parents, and that the information provided was tailored to individualized needs, easy to access, and allowed the recap of information.

Subtheme 1: Good Informational Resource

Many parents utilized the mHealth app as a point of reference to clarify any doubts they faced when performing routine baby care tasks. The participants compared the mHealth app with a library, regarding it as a useful informational tool in preparing them for any issues that they might potentially experience during the first few weeks:

> *It's like a library for me…I can just view it to see what new parents face at this point of time.* [Participant 3]

Subtheme 2: Cater to Local Context and New Generation Parents

Many parents discerned the mHealth app to contain localized information. Thus, the app is more contextually relevant as opposed to other global apps. These data include the matrices used (cm and kg vs feet and pounds) and the discussion of common issues babies encounter (Asian vs Caucasian):

> *I think the app is better, because, I say, even the matrix uses cm and kg...I would say the app is more appropriate and suitable for Asian mummies, especially in Singapore.* [Participant 8]

Subtheme 3: Tailored to Individualized Needs

The parents felt that being able to acquire answers from the midwife and having their queries answered through the discussion forum provided a clearer and more individualized answer compared with the standard answers found on web:

> *Yes, [any answer found] on web is vague and very generic [in their] answer. For example, each baby should be gaining this certain amount of weight when you are pregnant, so you like, okay, [but] when you go to the doctor, [you realize,] oh no, your baby is*

not gaining as much as what [the] web recourses have been saying…but with this app, we received focused answers to our queries. [Participant 11]

Subtheme 4: Various Features—Audio, Video, Push Notifications, and Discussion Forum

The mHealth app catered to the participants' individual preferences of learning styles through the different modes of educational materials. Some mothers found the audio files more convenient while breastfeeding. Others perceived the videos to be an easier way to learn hands-on skills. A good mix of both theory (PDF files and audios) and practical (videos and the discussion forum) learning existed, which the parents could benefit from:

...but you know when you breastfeed, sometimes you feel tired, you want to [be] lying down, but breastfeeding and lying down is sometimes [a] problem. So, I realized that there are audios for each of the PDF file. So, I just stuck my earphone [in], and listened to it. [Participant 8]

Subtheme 5: Easy to Access and Allows Recap and Recall

Majority of the parents expressed that the interface of the mHealth app was user-friendly, making it easy to acquire necessary information through their phones. Many participants also revealed that the mHealth app facilitated their recapping and recalling of essential information required in taking care of their baby:

...because, when you learn at the hospital, you only see [it] one time. If you watch the video, you can recap. [Participant 7]

Theme 2: Advice From the Midwife

The participants underlined the advice given by the midwife to be reliable, reassuring, prompt, and that it also facilitated their decision-making process.

Textbox 1. Themes and subthemes of the findings.

Positive features of the mobile health (mHealth) app

1. Good informational resource
2. Cater to local context and new generation parents
3. Tailored to individualized needs
4. Varied features: audio, video, push notifications, and the discussion forum
5. Easy to access and allows recap and recall

Advice from the midwife

1. Reliable guidance
2. Reassuring advice
3. Prompt replies
4. Facilitation of the decision-making process

Experiences gained from using the mHealth app

1. Provides continuity of care
2. Is like a friend
3. Integrates a community of people who are undergoing a similar phase in their lives, that is, parenthood
4. Enhances confidence
5. Enhances satisfaction
6. Receptiveness

Recommendations for future

1. Reduce technical hiccups and enhance technical aspects
2. Extend the duration of use of the mHealth app
3. Encourage engagement with the mHealth app
4. Widen educational topics

Subtheme 1: Reliable Guidance

Nearly all the participants felt that they can easily trust the advice provided by the midwife as opposed to acquiring information on web where sources may not be as credible:

> *...being able to get the response you knew you could trust such as a midwife…Er…giving us advices, you could trust that.* [Participant 10]

Subtheme 2: Reassuring Advice

The advice given by the midwife often left parents reassured and empowered to overcome the issues they encountered. The participants could even resonate with the advice given for queries other than that of their own:

> *I think the best part about it is the personal touch…from the midwife herself. When you ask questions, and know that the midwife will always reply with a 'you're doing good mummy keep it up!' [laughs] Yea, so I think it is encouraging…even though it is not me who sent the question. But when you read [it]…you know? That's nice.* [Participant 17]

Subtheme 3: Prompt Replies

The participants were also content with the efficiency of the midwife's replies to their queries, which were useful in alleviating their worries within a relatively short period of time:

> *The questions we were asking also, they were answered very promptly. I think that was quite amazing. I think it was…not even within 24 hours, but I think within 12 hours…we will get a reply also…* [Participant 12]

Subtheme 4: Facilitation of the Decision-Making Process

Several parents felt that the midwife provided detailed explanations, thereby giving them improved understanding of their situation. This initiative allowed them to make informed decisions regarding the care of their baby:

> *...after seeking midwife's advice, that's when I decided to bring the baby for his jaundice check…because of the answers that have been given to us, lah. So, it reassured me to seek help.* [Participant 3]

Theme 3: Experiences Gained From Using the mHealth App

The experiences gained by most participants entailed the continuity of care received, the integration of a community of people undergoing a similar phase in their lives, and the increased confidence and satisfaction of taking care of their babies.

Subtheme 1: Provides Continuity of Care

The continuous support received by the parents via the mHealth app was gratifying because it facilitated easy transition from the hospital setting to their home. Many participants also felt comforted and secure in having their queries answered by other experienced parents:

> *I think what is good about the app is [that] you have the mothers and fathers for users, user-based answering the questions, but then in 9 out of the 10, the moderator [midwife] comes on and either clarifies all the answers other people [have] given, and adds an extra professional part as well.* [Participant 9]

Several parents felt that the app acted as a support mechanism between the doctor's follow-up sessions, which typically lasted between a week to a month, thus providing comfort and reassurance to the parents regarding their concerns:

> *I really feel that…not to worry…too much. Because you can only meet the doctor at [a] certain point in time, right? So, in the meantime, we didn't have any information. So we were glad that we had the message.* [Participant 5]

Subtheme 2: Is Like a Friend

The mHealth app was described as a friend to new parents by providing much-needed support during their transition phase, reassuring them that they were not alone:

> *I say, "friend"…Because sometimes you were alone, then…I went to the mobile app a lot to see what other mums said.* [Participant 14]

Subtheme 3: Integrates a Community of People Undergoing a Similar Phase in Their Lives, That Is, Parenthood

The majority of the parents emphasized that the mHealth app helped them to realize that the issues they were facing were commonly experienced by many parents, thereby putting their problems into perspective and allowing them to feel better. Most participants appraised the expediency of learning from the experiences of other parents that were shared within the discussion forum or by using their own experiences to help answer other parents' queries:

> *So when we have this app and you know there's another parent…the person answering is really answering related directly to the question. And the question is linkable to us…We felt others are also going through the same things…So I think that was the most helpful about the app.* [Participant 17]

The participants exhibited mutual empathy, which is known to augment prosocial behaviors, such as supporting one another. This value was commonly manifested in the form of multiparous parents being willing to share their past experiences to alleviate the worries and concerns of their primiparous counterparts:

> *I have been through this…I know how it feels to be a new mother…So, if I [can] confirm [that] I know the answer, then I will try to help out by answering the queries.* [Participant 13]

Subtheme 4: Enhances Confidence

The educational materials and having their queries answered by the midwife reassured many participants that they were doing the right thing for their baby:

> *Because when I asked the question and then I've got [a] response [that] I was comfortable knowing, doing the right thing. I can continue that way, and you feel more confident when you are making [the] right decisions.* [Participant 10]

Subtheme 5: Enhances Satisfaction

Many parents expressed a sense of satisfaction in relation to their parenting role with the support from the mHealth app and felt they were more adept in playing their parenting role:

> *Er…I think, with the increased confidence level, then you'll be more satisfied…Because you feel you understand your child better! And you know when this happens…you know roughly what to do.* [Participant 17]

Subtheme 6: Receptiveness

All the participants were receptive to using the mHealth app in the future if they have another child and were forthcoming to recommend this mHealth app to their friends who were expecting:

> *The Home-but not Alone, I am going to recommend because [it is] very focused on [the] first month of the support. We need it.* [Participant 5]

Theme 4: Recommendations for Future

Many parents disclosed that certain technical hiccups should be resolved and enhancements should be added in the mHealth app. In addition, the duration of use of the mHealth app needs to be extended, the usability of the mHealth app has to be promoted, and additional educational topics included.

Subtheme 1: Reduce Technical Hiccups and Enhance Technical Aspects

Some participants felt that the language in the video was not clearly audible due to the enunciation and suggested for the addition of subtitles. The majority of the parents highlighted the need to add notifications to their queries because checking repeatedly for any replies to their raised queries was time-consuming. Furthermore, many participants found locating specific information within the discussion forum very tedious because questions were not organized into topics:

> *Yea, because now I only look based on the latest post, but not what is actually relevant…I will have to open each one [discussion thread] to read and see if there's anybody talking about breastfeeding.* [Participant 14]

Subtheme 2: Extend the Duration of use of the mHealth App

Participants felt the need to extend the duration of the mHealth app by introducing the mHealth app during the antenatal period, preferably during the third trimester to prepare themselves:

> *I think during the third trimester…Yea…towards the third trimester, I think you're more ready and prepared to actually absorb all the knowledge.* [Participant 4]

Most participants also conveyed their interest in using the mHealth app for more than a month after the baby was born and requested for information to be provided for important milestones within the first 6 months of their baby's growth, such as at 3 and 6 months:

> *I think maybe the first six months…Because it's the first vital…you know…stages! After that, I think they can start crawling, recognizing you proper[ly], so it shouldn't be so much of a problem.* [Participant 17]

Subtheme 3: Encourage Engagement With the mHealth App

Participants felt the need to promote members of the group to be active users of the mHealth app and to ask questions. They believed that the number of questions asked translated into how much they can learn from:

> *...but I think it would have been even more useful if more people shared…because different people had their [own] way of coping.* [Participant 12]

Subtheme 4: Widen Educational Topics

Fathers faced concerns regarding their roles in supporting their wives emotionally during this time and requested for such relevant topics. Many parents also felt that the current topics found in the mHealth app mostly catered to first-time parents. They felt that the mHealth app would be more beneficial if it included topics relevant to multiparous parents, such as dealing with older siblings and babies:

> *...because some of the other topics I have already read about or did research on my first baby. But, now that it is my second baby, it might be useful to have some topics for parents that are already parents.* [Participant 16]

Discussion

Principal Findings

The participants reported numerous strengths of the app. Many parents expressed that it was convenient and found that it encapsulates a wide range of educational information, which allowed for an easy search and retrieval of information independently. Such information was useful as a general informative resource or for more explicit information that may address their specific concerns. A previous research [33] found that mothers preferred midwives to provide information in a nondirective form to allow them the opportunity of making informed decisions independently. Similarly, the app facilitated parents in their decision making to a certain extent and encouraged them to act autonomously. This function is in accordance with Bandura's theory of self-efficacy [29], in which learning independently and building one's own experiences is one way to attain the mastery of such experiences, thereby augmenting parenting self-efficacy [24].

The multifaceted approach regarding the delivery of information through PDF, video, or audio files augmented the expediency of the app and enhanced parenting self-efficacy. Mothers were able to breastfeed and listen to audio recordings, and the parents

reported that they were only able to observe demonstrations on baby care tasks once within the hospital; hence, the videos came in as a handy tool for recapping and recalling practical tasks as opposed to written information. A previous study acknowledged that the memory for visual depictions is greater than that for verbal or written content [24]. Moreover, the learning theory [34] asserts that learning is more enhanced when the content is organized as opposed to being fragmented. Moreover, according to Bandura [29], vicarious encounters can instill in parents a belief that they too have the ability to carry out the task at hand, which can lead to great parenting self-efficacy.

The participants also expressed their satisfaction toward the app in that it provided a means of access to support from a health care professional. Receiving guidance from a health care professional is a vital facet in parents developing a sense of security during the postnatal period [24,35]. The answers to their queries by the midwife were regarded as trustworthy, prompt, and reassuring. This finding corroborates with a previous study that regarded midwives as a point of reference when making decisions [2]. It implies the usefulness of asynchronous communication in alleviating concerns for new parents in their transition into parenthood.

The app was also beneficial in creating a community in which parents could learn from the issues and concerns brought up by other parents as well as receive advice from those with experience. As previously noted [33], mothers found sharing their experiences with one another and gaining reassurance and support beneficial. Although sharing with a professional would generate a clinical or scientific reasoning on what was being physically experienced by them, sharing with other mothers produced a communal response of the emotional experience itself. Although advice from other experienced parents was easily available, it was constantly reviewed by the midwife and, in case of any discrepancy, the advice was corrected. In general, experienced parents' advice was appropriate and practical, and overall, parents also appreciated a conducive milieu to express their concerns and felt validated.

The app was successful because most participants no longer felt alone but emotionally supported by both the midwife and other parents. Sharing their experiences with other parents who are undergoing similar encounters can create a sense of belonging within a particular community as well as enhance their social support. Both factors are essential in enhancing parents' psychological well-being [23]. Similarly, previous studies [36,37] demonstrated that the midwife offers emotional support to parents as opposed to providing support that is solely informative or didactic, which is commonly observed among health care professionals. The finding that parents can feel supported through the mHealth app implies that personal encounters or direct communication is no longer the only way for individuals to gain emotional support; moreover, with the rise in the utilization of technology and social media, information technology might be an alternative means of extending emotional support [23].

All participants demonstrated their keenness to use the app in future childbearing or to recommend it to their friends who might be expecting, showing their receptiveness and satisfaction toward using this app. This receptivity may be attributed to the essence of anonymity that comes with using information technology, enabling parents to ask questions openly without feeling embarrassed about their uncertainty [24].

The participants felt more confident in performing parenting tasks because they felt adequately guided through the videos and the support they received from the midwife and other parents. Most participants also expressed that the midwife was very encouraging and reassuring in her feedback, which instilled in them the confidence to perform the tasks at hand. According to Bandura's theory of self-efficacy [29], verbal persuasion is another way to enhance self-efficacy. Individuals who are continually reminded that they harness the ability to become proficient at various tasks will be more likely to muster up the courage and effort to work toward achieving it [34,38].

Strengths and Limitations

This study is the first to evaluate the technology-based postnatal educational program qualitatively among the multiethnic parents in Singapore. The qualitative approach provided the in-depth perspectives of new parents in receiving the mHealth app–based educational program. This study also has few limitations. It may not be representative of the entire population because a high proportion of the participants were of high economic status (employed, educated, and had high income). In addition, the study focused only on English-speaking parents. Future studies should include multilingual participants with differing socio-economic backgrounds. Moreover, only parents' perspectives were captured in the usefulness of the program. Future research could explore the views of the involved midwives or health care professionals regarding the feasibility or acceptability of such a program in their professional practice.

Conclusions

Technological advancement and social media usage provide the impetus for health care to invest in the development of information technology to render accessible and efficient care. This research provides insights into the use of technology in providing support to new parents during the postnatal period. Specifically, it details parents' perceptions on the content and effectiveness of the mHealth app on their satisfaction and well-being. Future studies should continue to evaluate the use of information technology, especially across the perinatal period.

Acknowledgments

The authors would like to thank the National University of Singapore for the start-up grant. The authors would also like to thank the National University Health System, Medical Publications Support Unit, for their assistance in the language editing of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic details of participants who participated in process evaluation interviews.

[PDF File (Adobe PDF File), 31KB - jmir_v20i4e119_app1.pdf]

References

1. Hunter LP, Rychnovsky JD, Yount SM. A selective review of maternal sleep characteristics in the postpartum period. J Obstet Gynecol Neonatal Nurs 2009;38(1):60-68. [doi: 10.1111/j.1552-6909.2008.00309.x] [Medline: 19208049]
2. Razurel C, Bruchon-Schweitzer M, Dupanloup A, Irion O, Epiney M. Stressful events, social support and coping strategies of primiparous women during the postpartum period: a qualitative study. Midwifery 2011 Apr;27(2):237-242. [doi: 10.1016/j.midw.2009.06.005] [Medline: 19783333]
3. Brown SJ, Davey M, Bruinsma FJ. Women's views and experiences of postnatal hospital care in the Victorian Survey of Recent Mothers 2000. Midwifery 2005 Jun;21(2):109-126. [doi: 10.1016/j.midw.2004.09.006] [Medline: 15878426]
4. Ross LE, McLean LM. Anxiety disorders during pregnancy and the postpartum period: a systematic review. J Clin Psychiatry 2006 Aug;67(8):1285-1298. [Medline: 16965210]
5. Dennis CL, Chung-Lee L. Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review. Birth 2006 Dec;33(4):323-331. [doi: 10.1111/j.1523-536X.2006.00130.x] [Medline: 17150072]
6. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: a review. Infant Behav Dev 2010 Feb;33(1):1-6 [FREE Full text] [doi: 10.1016/j.infbeh.2009.10.005] [Medline: 19962196]
7. McLearn KT, Minkovitz CS, Strobino DM, Marks E, Hou W. Maternal depressive symptoms at 2 to 4 months post partum and early parenting practices. Arch Pediatr Adolesc Med 2006 Mar;160(3):279-284. [doi: 10.1001/archpedi.160.3.279] [Medline: 16520447]
8. O'hara MW, Swain AM. Rates and risk of postpartum depression—a meta-analysis. Int Rev Psychiatry 2009 Jul 11;8(1):37-54. [doi: 10.3109/09540269609037816]
9. Gao L, Chan SW, Mao Q. Depression, perceived stress, and social support among first-time Chinese mothers and fathers in the postpartum period. Res Nurs Health 2009 Feb;32(1):50-58. [doi: 10.1002/nur.20306] [Medline: 18980171]
10. Goodman JH. Paternal postpartum depression, its relationship to maternal postpartum depression, and implications for family health. J Adv Nurs 2004 Jan;45(1):26-35. [Medline: 14675298]
11. Ramchandani PG, Psychogiou L, Vlachos H, Iles J, Sethna V, Netsi E, et al. Paternal depression: an examination of its links with father, child and family functioning in the postnatal period. Depress Anxiety 2011 Jun;28(6):471-477 [FREE Full text] [doi: 10.1002/da.20814] [Medline: 21506206]
12. Shorey S, Chan SW, Chong YS, He HG. Predictors of maternal parental self-efficacy among primiparas in the early postnatal period. West J Nurs Res 2015 Dec;37(12):1604-1622. [doi: 10.1177/0193945914537724] [Medline: 24906360]
13. Fenwick J, Butt J, Dhaliwal S, Hauck Y, Schmied V. Western Australian women's perceptions of the style and quality of midwifery postnatal care in hospital and at home. Women Birth 2010 Mar;23(1):10-21. [doi: 10.1016/j.wombi.2009.06.001] [Medline: 19632912]
14. Ong SF, Chan WS, Shorey S, Chong YS, Klainin-Yobas P, He HG. Postnatal experiences and support needs of first-time mothers in Singapore: a descriptive qualitative study. Midwifery 2014 Jun;30(6):772-778. [doi: 10.1016/j.midw.2013.09.004] [Medline: 24161493]
15. Martell LK. From innovation to common practice: perinatal nursing pre-1970 to 2005. J Perinat Neonatal Nurs 2006;20(1):8-16; quiz 17. [Medline: 16508453]
16. Shorey S, Chan SW, Chong YS, He HG. A randomized controlled trial of the effectiveness of a postnatal psychoeducation programme on self-efficacy, social support and postnatal depression among primiparas. J Adv Nurs 2015 Jun;71(6):1260-1273. [doi: 10.1111/jan.12590] [Medline: 25496615]
17. Naser E, Mackey S, Arthur D, Klainin-Yobas P, Chen H, Creedy DK. An exploratory study of traditional birthing practices of Chinese, Malay and Indian women in Singapore. Midwifery 2012 Dec;28(6):e865-e871. [doi: 10.1016/j.midw.2011.10.003] [Medline: 22079015]
18. Bahadoran P, Mohseni M, Abedi H. Mothers' experiences of postpartum period: A phenomenological study. Iranian Journal of Nursing and Midwifery Research 2008;13(3):110-115 [FREE Full text]
19. Schmied V, Cooke M, Gutwein R, Steinlein E, Homer C. Time to listen: strategies to improve hospital-based postnatal care. Women Birth 2008 Sep;21(3):99-105. [doi: 10.1016/j.wombi.2008.04.002] [Medline: 18579461]
20. Shaw E, Levitt C, Wong S, Kaczorowski J, McMaster University Postpartum Research Group. Systematic review of the literature on postpartum care: effectiveness of postpartum support to improve maternal parenting, mental health, quality of life, and physical health. Birth 2006 Sep;33(3):210-220. [doi: 10.1111/j.1523-536X.2006.00106.x] [Medline: 16948721]

21. Tahir NM, Al-Sadat N. Does telephone lactation counselling improve breastfeeding practices? A randomised controlled trial. Int J Nurs Stud 2013 Jan;50(1):16-25. [doi: 10.1016/j.ijnurstu.2012.09.006] [Medline: 23084438]
22. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. Br Med J 2000 Jun 3;320(7248):1517-1520 [FREE Full text] [Medline: 10834899]
23. Carissoli C, Villani D, Riva G. An emerging model of pregnancy care: The introduction of new technologies in maternal wellbeing. In: Integrating Technology in Positive Psychology Practice. United States of America: IGI Global; 2016:162-192.
24. Danbjørg DB, Wagner L, Kristensen BR, Clemensen J. Intervention among new parents followed up by an interview study exploring their experiences of telemedicine after early postnatal discharge. Midwifery 2015 Jun;31(6):574-581. [doi: 10.1016/j.midw.2015.02.007] [Medline: 25765743]
25. Watterson J, Walsh J, Madeka I. Using mHealth to improve usage of antenatal care, postnatal care, and immunization: a systematic review of the literature. Biomed Res Int 2015;2015:153402. [doi: 10.1155/2015/153402] [Medline: 26380263]
26. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016 May;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
27. Feroz A, Perveen S, Aftab W. Role of mHealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. BMC Health Serv Res 2017 Nov 07;17(1):704 [FREE Full text] [doi: 10.1186/s12913-017-2664-7] [Medline: 29115992]
28. Shorey S, Lau Y, Dennis C, Chan YS, Tam WW, Chan YH. A randomized-controlled trial to examine the effectiveness of the 'Home-but not Alone' mobile-health application educational programme on parental outcomes. J Adv Nurs 2017 Apr 07;73(9):2103-2117. [doi: 10.1111/jan.13293]
29. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
30. Shorey S, Ng YP, Danbjørg DB, Dennis CL, Morelius E. Effectiveness of the 'Home-but not Alone' mobile health application educational programme on parental outcomes: a randomized controlled trial, study protocol. J Adv Nurs 2017 Jan;73(1):253-264. [doi: 10.1111/jan.13151] [Medline: 27650320]
31. Holloway I, Galvin K. Qualitative research in nursing and healthcare. 4th ed. Chichester, West Sussex, UK: Wiley-Blackwell; 2016.
32. Glaser BG, Strauss AL, Strutzel E. The discovery of grounded theory: strategies for qualitative research. Nurs Res 1968;17(4):364. [doi: 10.1097/00006199-196807000-00014]
33. Seefat-van TA, Nieuwenhuijze M, Korstjens I. Women want proactive psychosocial support from midwives during transition to motherhood: a qualitative study. Midwifery 2011 Feb;27(1):e122-e127. [doi: 10.1016/j.midw.2009.09.006] [Medline: 19931954]
34. Suter P, Suter WN, Johnston D. Theory-based telehealth and patient empowerment. Popul Health Manag 2011 Apr;14(2):87-92. [doi: 10.1089/pop.2010.0013] [Medline: 21241182]
35. Elek SM, Hudson DB, Bouffard C. Marital and parenting satisfaction and infant care self-efficacy during the transition to parenthood: the effect of infant sex. Issues Compr Pediatr Nurs 2003;26(1):45-57. [Medline: 12623734]
36. Melender H. Fears and coping strategies associated with pregnancy and childbirth in Finland. J Midwifery Womens Health 2002;47(4):256-263. [Medline: 12138933]
37. Webster J, Linnane JW, Dibley LM, Hinson JK, Starrenburg SE, Roberts JA. Measuring social support in pregnancy: can it be simple and meaningful? Birth 2000 Jun;27(2):97-101. [Medline: 11251486]
38. Reece SM, Harkless G. Self-efficacy, stress, and parental adaptation: applications to the care of childbearing families. J Fam Nurs 2016 Jul 24;4(2):198-215. [doi: 10.1177/107484079800400206]

Abbreviations

mHealth: Mobile Health
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 16.10.17; peer-reviewed by H Miller, M Ashford, A Mehralizade; comments to author 07.12.17; revised version received 21.01.18; accepted 23.01.18; published 19.04.18.

Please cite as:
Shorey S, Yang YY, Dennis CL
A Mobile Health App–Based Postnatal Educational Program (Home-but not Alone): Descriptive Qualitative Study
J Med Internet Res 2018;20(4):e119
URL: http://www.jmir.org/2018/4/e119/
doi:10.2196/jmir.9188
PMID:29674314

Review

Self-Harm, Suicidal Behaviours, and Cyberbullying in Children and Young People: Systematic Review

Ann John[1], FFPH; Alexander Charles Glendenning[1], MSc; Amanda Marchant[1], MSc; Paul Montgomery[2], DPhil; Anne Stewart[3], FRCPsych; Sophie Wood[1], MSc; Keith Lloyd[1], FRCPsych; Keith Hawton[4], FMedSci

[1]Population Psychiatry, Suicide and Informatics, Medical School, Swansea University, Swansea, United Kingdom

[2]Department of Social Policy, Sociology and Criminology, School of Social Policy, University of Birmingham, Birmingham, United Kingdom

[3]Oxford Central Child and Adolescent Mental Health Services, Oxford Health NHS Foundation Trust, Oxford, United Kingdom

[4]Centre for Suicide Research, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Ann John, FFPH
Population Psychiatry, Suicide and Informatics
Medical School
Swansea University
Institute of Life Science 2
Swansea, SA2 8PP
United Kingdom
Phone: 44 01792602568 ext 2568
Email: A.John@swansea.ac.uk

Abstract

Background: Given the concerns about bullying via electronic communication in children and young people and its possible contribution to self-harm, we have reviewed the evidence for associations between cyberbullying involvement and self-harm or suicidal behaviors (such as suicidal ideation, suicide plans, and suicide attempts) in children and young people.

Objective: The aim of this study was to systematically review the current evidence examining the association between cyberbullying involvement as victim or perpetrator and self-harm and suicidal behaviors in children and young people (younger than 25 years), and where possible, to meta-analyze data on the associations.

Methods: An electronic literature search was conducted for all studies published between January 1, 1996, and February 3, 2017, across sources, including MEDLINE, Cochrane, and PsycINFO. Articles were included if the study examined any association between cyberbullying involvement and self-harm or suicidal behaviors and reported empirical data in a sample aged under 25 years. Quality of included papers was assessed and data were extracted. Meta-analyses of data were conducted.

Results: A total of 33 eligible articles from 26 independent studies were included, covering a population of 156,384 children and young people. A total of 25 articles (20 independent studies, n=115,056) identified associations (negative influences) between cybervictimization and self-harm or suicidal behaviors or between perpetrating cyberbullying and suicidal behaviors. Three additional studies, in which the cyberbullying, self-harm, or suicidal behaviors measures had been combined with other measures (such as traditional bullying and mental health problems), also showed negative influences (n=44,526). A total of 5 studies showed no significant associations (n=5646). Meta-analyses, producing odds ratios (ORs) as a summary measure of effect size (eg, ratio of the odds of cyber victims who have experienced SH vs nonvictims who have experienced SH), showed that, compared with nonvictims, those who have experienced cybervictimization were OR 2.35 (95% CI 1.65-3.34) times as likely to self-harm, OR 2.10 (95% CI 1.73-2.55) times as likely to exhibit suicidal behaviors, OR 2.57 (95% CI 1.69-3.90) times more likely to attempt suicide, and OR 2.15 (95% CI 1.70-2.71) times more likely to have suicidal thoughts. Cyberbullying perpetrators were OR 1.21 (95% CI 1.02-1.44) times more likely to exhibit suicidal behaviors and OR 1.23 (95% CI 1.10-1.37) times more likely to experience suicidal ideation than nonperpetrators.

Conclusions: Victims of cyberbullying are at a greater risk than nonvictims of both self-harm and suicidal behaviors. To a lesser extent, perpetrators of cyberbullying are at risk of suicidal behaviors and suicidal ideation when compared with nonperpetrators. Policy makers and schools should prioritize the inclusion of cyberbullying involvement in programs to prevent traditional bullying. Type of cyberbullying involvement, frequency, and gender should be assessed in future studies.

(J Med Internet Res 2018;20(4):e129) doi:10.2196/jmir.9044

KEYWORDS

cyberbullying; bullying; self-injurious behavior; suicide; suicide, attempted; suicidal ideation

Introduction

Cyberbullying

Bullying is an aggressive, intentional act carried out by a group or an individual repeatedly and over time against a victim who cannot easily defend himself or herself. Traditionally, bullying could be direct—physical, verbal, or relational (eg, social exclusion)— or indirect (eg, rumor spreading) [1]. However, with the advent of electronic communication (eg, social media and instant messaging) via the internet and mobile phones, *cyberbullying* has emerged. This can be similarly defined, with the addition that it occurs via electronic forms of contact [2]. As the harassment of victims takes place electronically, the manner and timings in which they are targeted, as well as how they cope in response, and the proximity of relationships between victims and perpetrators, are uniquely different compared with traditional bullying. Cyberbullying victimization tends to occur at a later age, around 14 years, when children spend more time on their mobile phones [3] and social networking sites [4]. Perpetrators of cyberbullying have a degree of anonymity not possible in traditional bullying, and the potential exposure and embarrassment of the victim is on a larger scale. It is possible to victimize a peer within their own home or elsewhere at any time of day or night, and should they remove themselves from the site, the messages often accumulate. This presents new challenges for individuals, families, schools, professionals, researchers, and policy makers.

The adverse impact of bullying on children and young people's lives, be they victim, perpetrator, or both, has long been recognized [1]. Being bullied is often associated with mental health problems (including depressive symptomatology), self-harm (SH), and suicidal behaviors [5-9]. A meta-analysis [7] found that traditional bullying was associated with general anxiety, with an odds ratio (OR) of 2.55 (95% CI 1.28-3.83), and depression, with OR 6.22 (95% CI 3.11-9.33). School bullying (less than weekly) has been shown to be associated with suicidal ideation (OR 2.79, 95% CI 1.64-4.75) and suicide attempt (OR 2.66, 95% CI 1.58-4.47) [9]. Some studies have found over 85% of those involved in cyberbullying are also involved in traditional bullying and have suggested that health issues associated with cyberbullying involvement are mediated through traditional bullying [10]. The reported prevalence of cyberbullying involvement varies widely across countries. This reflects societal factors, stigma, and also differing interpretations of "repeatedly and over time." Estimates indicate that between 15% and 35% of young people have been victims of cyberbullying and between 10% and 20% of individuals admit to having cyberbullied others [11].

Previous Literature on Cyberbullying and Self-Harm and Suicidal Behaviors

Four previous systematic reviews [12-15] have demonstrated an association between cyberbullying involvement and SH or suicidal behaviors. They included a maximum of 5 studies each; 8 in total between them with only 6 studies eligible for meta-analysis [11,16-22]. The study by Daine et al [12], which included 2 papers on this topic, concluded that cyberbullying involvement was one of the most significant negative aspects of the influence of the internet on SH but that this was an area of research still in its infancy. Given the rapid expansion of evidence in the field, the apparent rise in prevalence of SH [23], and the changing nature of electronic communication in young people, it is timely to reassess the literature.

The aim of this study was to systematically review the current evidence examining the association between cyberbullying involvement (as victim, perpetrator, or both) and SH and suicidal behaviors in children and young people (younger than 25 years). Where possible, we aimed to meta-analyze data on the associations.

Methods

Search Strategy

A protocol for this review was registered with PROSPERO (ID: CRD42017056487). This review was conducted in compliance with the guidelines for Meta-Analyses and Systematic Reviews of Observational Studies [24] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [25].

A literature search was conducted for all studies published in English between January 1, 1996, and February 3, 2017. The databases searched included the Cochrane Library, Medical Literature Analysis and Retrieval System Online, PROSPERO, PsycINFO, PubMed, and Scopus. Additional searches were conducted in health improvement sources (eg, Health Evidence Canada), topic-specific websites (eg, American Association of Suicidology), and meta-search engines (Google). Gray literature was further explored by contacting experts in the field for any unreported or ongoing studies.

The following terms were searched in free text or keywords:

> "Automutilation," "Distress*," "Emotion*," "nssi," "((oneself or myself or self) adj2 (cut* or harm* or hurt* or kill or injur* or mutilat*))," "(psychological adj (stress or distress))," "SIB', 'Suicid*," "Aol," "Askfm," "Bebo," "blog*," "chat room* OR chatroom*," "cyber*," "discussion forum," "e-communi*," "e-material*," "Facebook," "google*," "hashtag," "image sharing," "Instagram," "instant messag*," "internet*," "live chat," "live journal*," "meme," "MSN," "Myspace," "on line OR online," "photo sharing," "Pinterest," "podcast*," "social network*," "spam*," "troll*," "Tumblr," "tweet*," "Twitter," "video sharing," "vine," "virtual*," "vlog*," "web*," and "YouTube."

The following terms were searched alongside the following database subject headings:

Medical subject headings: "self-injurious behavior," "stress, psychological," "blogging," "electronic mail," "internet," "social media," "social networking," "bullying," "adolescent," "child," "students," and "young adult."

Health Management Information Consortium: "attempted suicide," "self harm," "suicide pacts," "suicide," "bullying," "cyberspace," "internet," "internet websites," "intranet," and "world wide web."

PsycInfo: "attempted suicide," "self destructive behaviour," "self injurious behavior," "suicidal ideation," "suicide prevention," "cyberbullying," and "adolescent attitudes."

Excerpta Medica dataBASE: "automutilation," "suicidal behaviour," "suicide," "bullying," "internet," "social network," "adolescent," "child," and "young."

Study Selection

Inclusion and Exclusion Criteria

Included studies were those which examined any association between cyberbullying involvement (victimization or perpetration) and SH or suicidal behaviors and included empirical data in a sample aged younger than 25 years. The criteria used to determine eligibility for inclusion in the study were based on the study by Daine et al [12]. These are shown in Multimedia Appendix 1.

Screening

A stepwise screening process was performed independently by two researchers (AJ and AG). Initially, title and duplication manual screening was conducted. Titles with no relevance to the study were excluded. Any disagreements between reviewers' categorizations were put forward for abstract review. In the second stage, the remaining titles and abstracts were screened. Reference lists of review articles and included articles were manually screened for relevant studies. Studies were forwarded to the third stage of full-text article screening if they met the inclusion criteria or a decision could not be made on title and abstract alone. Any disagreements were resolved by consensus with a third reviewer (AM).

Included papers that investigated cyberbullying involvement (victimization or perpetration) and SH or suicidal behaviors and reported empirical data in a sample under the age of 25 years were forwarded for detailed analysis of their methodology and content.

Study Quality and Data Extraction

The quality of included papers was assessed independently by two reviewers (AG and AM) using the Critical Appraisal Skills Programme [26]. This program assesses multiple aspects of each paper in detail, including study design, representativeness of sample, bias, aspects of data collection, use of validated outcome measures, and whether conclusions reflect results. A quality rating of low, medium, or high was obtained using these quality standards as per Daine et al [12].

A data extraction sheet, developed by Daine et al [12], was adapted and used to record specific findings; identify themes; and ascertain potential biases, limitations, and weaknesses. Data were extracted independently by two reviewers (AG and AM). In particular, data on the type of bullying involvement (victim or perpetrator) and types of outcome measures were recorded. The latter were measures of SH, suicide, and other suicidal behaviors including suicide attempts and suicidal ideation, as recorded in the study. SH was defined as an intentional act of nonfatal self-injury or self-poisoning, regardless of intent or motivation [27]. Suicidal behaviors included thoughts of suicide or suicidal ideation, suicide plans, and suicide attempts [28]. Other outcomes such as mental health problems and traditional bullying were only extracted if it was possible to distinguish their relationship to cyberbullying involvement and SH or suicidal behaviors. Any disagreements at this stage were resolved by consensus with a third reviewer (AJ). Data were stored in Statistical Package for the Social Sciences (SPSS) version 22 (IBM Corp).

Prevalence

All available data were collected on the prevalence rates of cybervictimization for the total study population. Weighted prevalence rates were then calculated based on the differing populations of each study [29]. Individual study weights were calculated using each study population total with respect to the sum total of all study populations eligible for inclusion. Subsequently, it was possible to calculate a simple weighted average for the overall prevalence. Studies that did not include data on prevalence or where the prevalence was based on one sex only were excluded from the overall weighted calculation.

Meta-Analysis

Studies with outcomes relating to SH, suicidal behaviors, suicide attempts, and suicidal ideation were assessed for suitability for meta-analysis. Decisions on the appropriateness of meta-analysis were based on consistency of outcome measures and level of heterogeneity between studies.

The common effect size index, the log of the OR, was used in the meta-analysis. Other types of effect sizes were transformed into this before the analysis. Inclusion criteria for the effect size index were based on the recommendations of Borenstein et al [30]. Reanalyses of raw data or conversions were performed only when necessary. Studies that did not include measures of precision with their results, that is, a corresponding CI or *P* value, were excluded from meta-analysis as these are required to calculate corresponding variances [31]. Final results were transformed from the log of the OR to the OR for presentation. The OR is here defined as the ratio between the odds of an individual who is involved in cyberbullying having experienced SH or suicidal behaviors and the odds of an individual who is not involved in cyberbullying having experienced SH or suicidal behaviors.

Where a study presented more than one effect size eligible for a meta-analysis, the most appropriate measure to maintain homogeneity of outcomes was included, for example, "suicide attempt" was chosen over "suicide attempt requiring medical treatment." However, where it was not possible to make such a distinction between two eligible outcomes (eg, female and male populations), the effect sizes were combined as an average

based on the recommendations of Borenstein et al [30]. If a study presented results in such a way that it was not possible to disaggregate outcomes of interest from other measures not considered in this review (eg, combined with mental health problems or with other forms of bullying), then the study was excluded from meta-analysis. Where two or more studies based on the same study population were eligible for meta-analysis, the study with the greatest sample size was included. Where a study presented results in terms of a range of frequencies (eg, "rarely," "sometimes," and "often"), the best average fit was chosen, that is, "sometimes" would be chosen over "rarely" or "often" for inclusion. Further details of the method are available in Multimedia Appendix 2.

Meta-analysis was performed using Matlab R2015a. The DerSimonian and Laird random-effects model was employed. Forest plots, summary effect sizes, CIs, *P* values, and measures of heterogeneity in the form of the *Q*- and I^2-statistics were calculated. The I^2-statistic was interpreted as per Higgins et al [32]: low ($25\% \leq I^2 < 50\%$), moderate ($50\% \leq I^2 < 75\%$), and high ($I^2 \geq 75\%$).

Meta-regressions, sensitivity analyses, and funnel plots were conducted to assess the effects of potential confounders, where relevant, and publication bias, where the number of eligible studies allowed for a robust assessment. The methods used are described in Multimedia Appendix 2.

Results

In total, 153 citations were identified by all electronic searches. A flowchart of the results of the search strategy and screening process is detailed in Figure 1.

Description of Included Studies

A total of 33 articles were eligible for inclusion in the review and forwarded for data extraction, comprising 26 independent studies and 156,384 individual participants. In total, 19 studies were from the United States; 7 from Canada; 1 each from Belgium, the Netherlands, Taiwan, Hong Kong, South Korea, and Australia; and 1 study that was conducted based on data from a combination of 24 different European nations [33]. All papers were based on observational studies: 28 cross-sectional based on survey data, 3 case-control studies, 1 cohort study [34], and 1 ecological study [33]. Multimedia Appendix 3 summarizes the aims, quality ratings, and findings of the included articles.

Study Populations

Of the 26 independent studies (33 articles), 20 were based on unique populations, whereas 6 independent studies (13 articles) had populations shared by at least one other article. Those articles that shared study populations were as follows: Schenk et al [22,35]; Bauman et al [18] and Romero et al [36]; Alavi et al [37] and Roberts et al [38]; Cénat et al [39] and Hébert et al [40]; Hay and Meldrum [16] and Hay et al [41]; and Messias et al [42], Reed et al [43], and Kindrick et al [44]. Further details of these study populations are available in Multimedia Appendix 4.

Excluding duplicate populations [35-37,39,41,43,44], the total number of unique participants was 156,384, with a mean of 6015 and median of 2243 individuals per study. Most studies included both female and male participants (often not reported separately). However, 2 studies included one sex only [36,45]. The youngest reported mean participant age was 12.5 years [34], whereas the oldest was 20.0 years [22,35].

Cyberbullying Involvement

Cybervictimization was analyzed in 25 included studies [16,17,20,22,33,34,37-44,46-56]; 7 studies examined both cybervictimization and cyberbullying perpetration [11, 18,19,21,36,45,57], and 1 study investigated cyberbullying perpetration, but also included those who were both victims and perpetrators [35]. Inclusion in one of these groups was most commonly assessed by a participant's yes or no response to a single question. For example, 7 studies used a question from the Youth Risk Behavior Survey: "During the past 12 months, have you ever been electronically bullied? (include being bullied through email, chat rooms, instant messaging, websites, or texting)." A total of 29 studies were based on self-report questionnaires, 2 on researcher-completed ones [33,52], and 2 on retrospective reviews of patients' medical records [37,38].

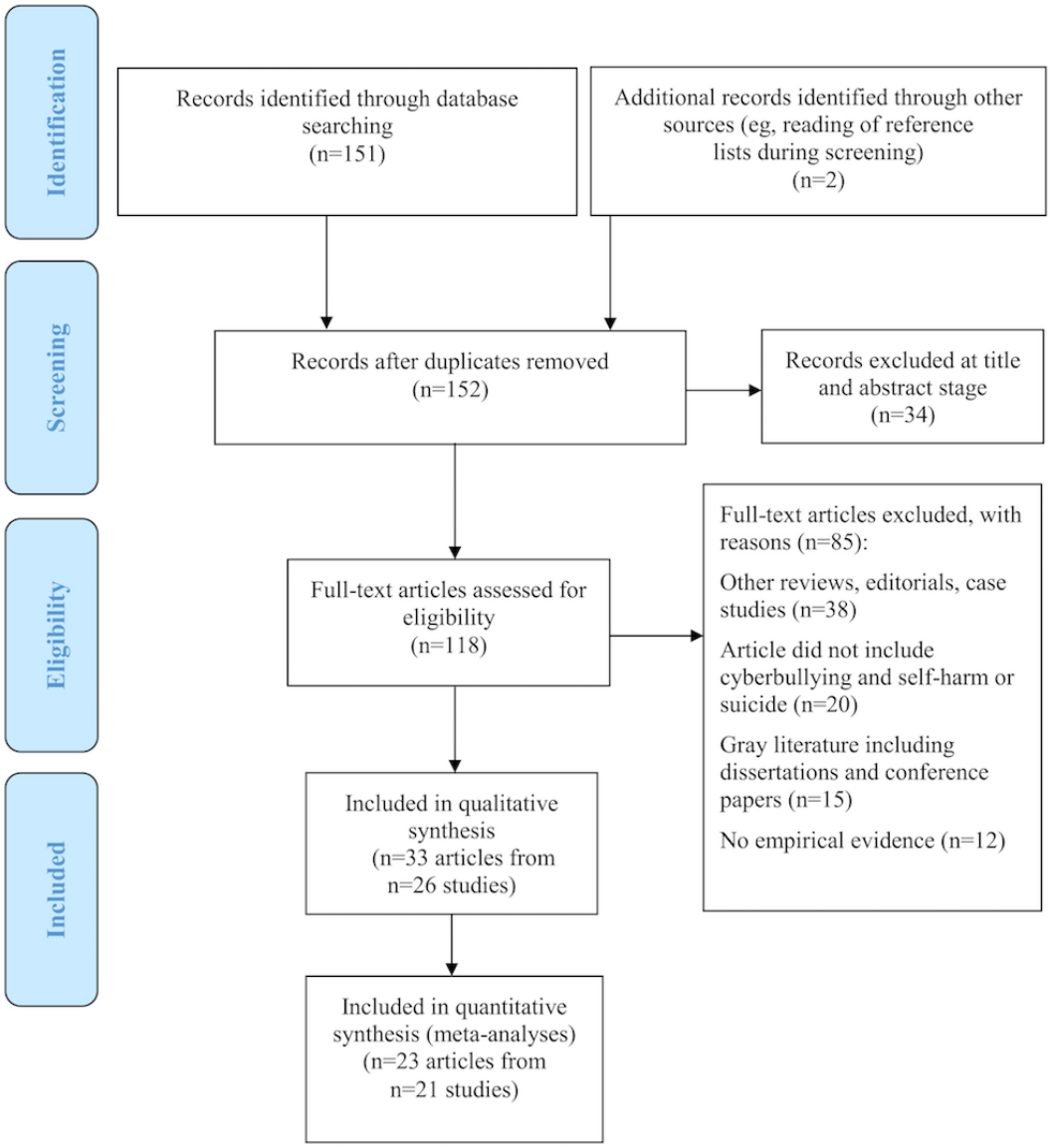
Only 4 studies [11,19-21] reported the medium through which cyberbullying (victimization and perpetration) occurred. The three most commonly experienced forms of victimization reported by Hinduja and Patchin [11] were as follows: "email" (18.3%), "instant message" (16.0%), and "MySpace" (14.2%), whereas the most common forms of perpetration were as follows: "posted something online about another person to make others laugh" (23.1%), "sent someone a computer text message to make them angry or to make fun of them" (13.7%), and "took a picture of someone and posted it online without their permission" (12.1%). Goebert et al [20] reported wide ranges of the medium of cybervictimization used across different Asian and Pacific Islander ethnic groups, for example, "text" (18.5%-27.8%). No direct associations could be calculated between medium of cyberbullying involvement and the SH or suicidal outcomes because of the way the data were collected and presented. Multimedia Appendix 5 displays the measures used in all 33 studies.

The findings of Elgar et al [49] and Kodish et al [50] suggested that the health consequences of cybervictimization are not completely attributable to its co-occurrence with face-to-face bullying. Similarly, the correlations reported in Fu et al [33] between cybervictimization and unnatural child deaths were independent of traditional bullying.

Prevalence of Cybervictimization

On the basis of 20 eligible studies (116,433 individuals), 12.6% (95% CI 12.4%-12.7%) of individuals had experienced cybervictimization. Weighted prevalence and prevalence by study are shown in Multimedia Appendix 6.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart displaying the different stages of the screening process.



Outcomes

Measures of SH were assessed in 19 articles (14 independent studies), suicidal behaviors in 32 articles (25 independent studies), suicide attempts in 16 articles (12 independent studies), and suicidal ideation in 27 articles (20 independent studies). One study [52] measured "thoughts of self-harm." No studies included death by suicide as an outcome, but 1 study [33] explored the association between cybervictimization and unnatural child death, which included suicides, accidental deaths, and death by assault. Outcomes and measures used are included in Multimedia Appendix 5.

Associations Between Cyberbullying Involvement and Self-Harm and Suicidal Behaviors

A total of 25 articles reported a positive association or negative influence of cyberbullying involvement (victimization or perpetration) on SH and suicidal behaviors; 3 [33,40,53] found negative influences in analyses in which the cyberbullying, SH, or suicidal behaviors measures had been combined with other measures; and 5 [19,21,36,52,56] found no significant association (2 of these articles [19,21] reported the proportion of cyber victims who had experienced SH and suicidal behaviors, but not that of nonvictims, meaning that no association could be determined).

Six meta-analyses were conducted (Table 1). Figure 2 displays the forest plots of the meta-analyses relating to cybervictimization, and Figure 3 displays those relating to cyberbullying perpetration. Further details of the measures included in all meta-analyses are available in Multimedia Appendix 4.

Table 1. Results of all meta-analyses performed. OR: odds ratio.

Cyberbullying group	Measure	*k*	n	OR (95 % CI)	ln OR (95 % CI)	*z*	P_z	T^2	*Q*	P_Q	I^2 (%)
Victimization	Self-harm	11	85,967	2.35 (1.65-3.34)	0.85 (0.50-1.21)	4.76	<.001	0.30	174.92	<.001	94.28
Victimization	Suicidal behaviors	21	116,616	2.10 (1.73-2.55)	0.74 (0.55-0.94)	7.45	<.001	0.15	256.94	<.001	92.22
Victimization	Suicide attempt	10	85,541	2.57 (1.69-3.90)	0.94 (0.52-1.36)	4.41	<.001	0.39	171.48	<.001	94.75
Victimization	Suicidal ideation	16	103,774	2·15 (1.70-2.71)	0.76 (0.53-1.00)	6.39	<.001	0.17	157.62	<.001	90.48
Perpetration	Suicidal behaviors	5	4062	1.21 (1.02-1.44)	0.19 (0.02-0.37)	2.19	.03	0.02	14.36	.006	72.14
Perpetration	Suicidal ideation	4	3811	1.23 (1.10-1.37)	0.21 (0.10-0.32)	3.75	<.001	0.00	3.91	.27	23.35

Cybervictimization and Self-Harm

A total of 11 independent studies [11,16,18,20,42,48,50,54,55] (n=85,967) were eligible for meta-analysis of the association between cybervictimization and SH (Figure 2). A total of 7 articles were rated high quality, and 4 were rated medium. The meta-analysis produced OR 2.35 (95% CI 1.65-3.34).

Cybervictimization and Suicidal Behaviors

An empirical association between cybervictimization and suicidal behaviors was identified in 32 articles. Of these, 11 were of high quality, 16 medium, and 5 low. Regression coefficients ranged from beta=.15 (P<.01) for suicide risk [50] to beta=.97 (P<.001) for suicidal behavior [51]. ORs ranged from 1.73 (95% CI 1.26-2.38) for suicide attempt [54] to 6.32 (95% CI 1.44-8.69) for suicidal ideation [47]. Schenk et al [22] (medium quality) applied a χ^2 goodness-of-fit producing χ^2_{2138}=9.1 (P=.03) when the frequencies of suicidal planning and attempts between cyber victims and controls were compared. Five papers found no significant association between cybervictimization and measures of suicidal behaviors [19,21,36,52,56].

A total of 21 studies [11,16-18,20,22,34,38,39,42,45-52,54,55,57], with 116,616 participants, were included in the meta-analysis (Figure 2). Of these, 9 studies were rated high quality, 11 medium, and 1 low [38]. A number of studies were excluded from meta-analysis as a subsample of another study or for being ineligible [19,21,35-37,40,41,43,44,53,56]. The meta-analysis produced OR 2.10 (95% CI 1.73-2.55).

Cybervictimization and Suicide Attempt

A total of 10 studies [11,17,18,20,42,48-50,54,55] with 85,541 participants were eligible for inclusion in meta-analysis for this association (Figure 2). Of these, 7 studies were rated high quality and 3 as medium quality. The summary effect size of the association between cybervictimization and suicide attempt was OR 2.57 (95% CI 1.69-3.90).

Cybervictimization and Suicidal Ideation

A total of 16 studies [11,16,17,22,34,38,39,42,46,47,49] with 103,774 participants were included in the meta-analysis for this association (Figure 2). Of these, 7 studies were rated high quality, 7 medium quality, and 2 low quality. The summary effect size for this meta-analysis was OR 2.15 (95% CI 1.70-2.71).

Cyberbullying Perpetration and Suicidal Behaviors

The association between cyberbullying perpetration and suicidal behaviors was examined in 6 papers [11,18,35,36,45,57] (5 independent studies [11,35,36,45,57] with 4062 participants). Of the 5 studies included in the meta-analysis, 1 study was rated high quality, 3 medium, and 1 low. Combination of effect sizes was again applied where appropriate. The summary effect size for this association was OR 1.21 (95% CI 1.02-1.44).

Cyberbullying Perpetration and Suicide Attempt

Three articles [11,18,36] examined this association. One [11] found an OR of 1.49 (P<.05). Bauman et al [18] reported a direct effect of beta=.14 (P<.05) for males only, whereas a study based on a subsample of its population [36] found no significant effect. Meta-analysis was not conducted for the association between cyberbullying perpetration and suicide attempt as only 2 studies would be included.

Cyberbullying Perpetration and Suicidal Ideation

A total of 4 studies [11,35,36,57], with 3811 participants, were included in this meta-analysis. Of these, 1 was rated high quality, 2 medium, and 1 low. A summary effect size of OR 1.23 (95% CI 1.10-1.37) was produced for this association.

Heterogeneity between studies was both high and statistically significant in all cybervictimization meta-analyses. Heterogeneity was moderate and significant for the association between cyberbullying perpetration and suicidal behaviors, but nonsignificant for that of suicidal ideation. All calculated values of I^2 are displayed in Table 1.

Further Analyses

Two meta-regressions were performed: the first, for prevalence of traditional victimization against effect size for cybervictimization and suicidal behaviors, returned a standardized coefficient of beta=−.84. The second was for prevalence of traditional victimization against effect size for cybervictimization and suicidal ideation, producing a coefficient of beta=−.89. Both results were significant to the P<.001 level. This means that with increasing prevalence of traditional victimization comes a decrease in study effect size for the association between cybervictimization and suicidal behaviors, as well as that of cybervictimization and suicidal ideation.

Figure 2. Forest plot of cybervictimization meta-analyses. OR: odds ratio.

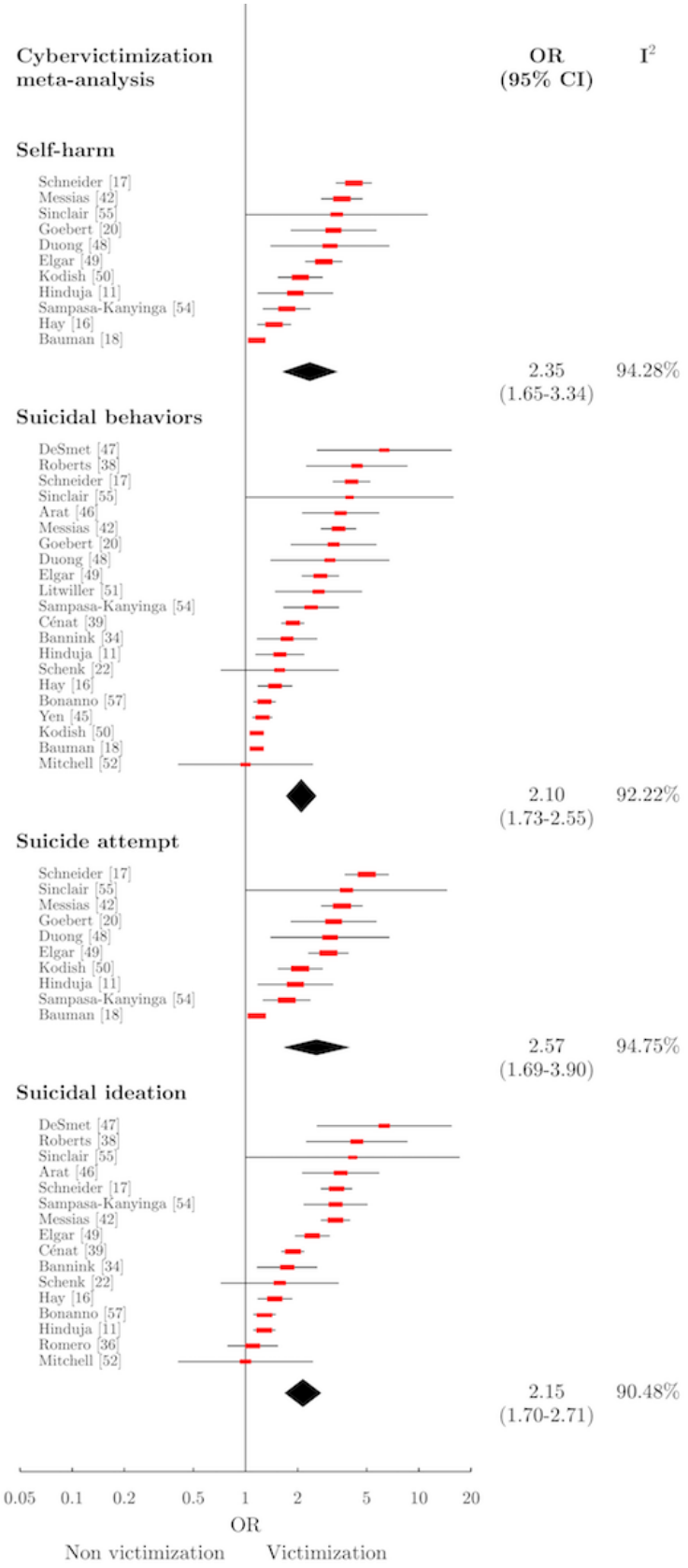
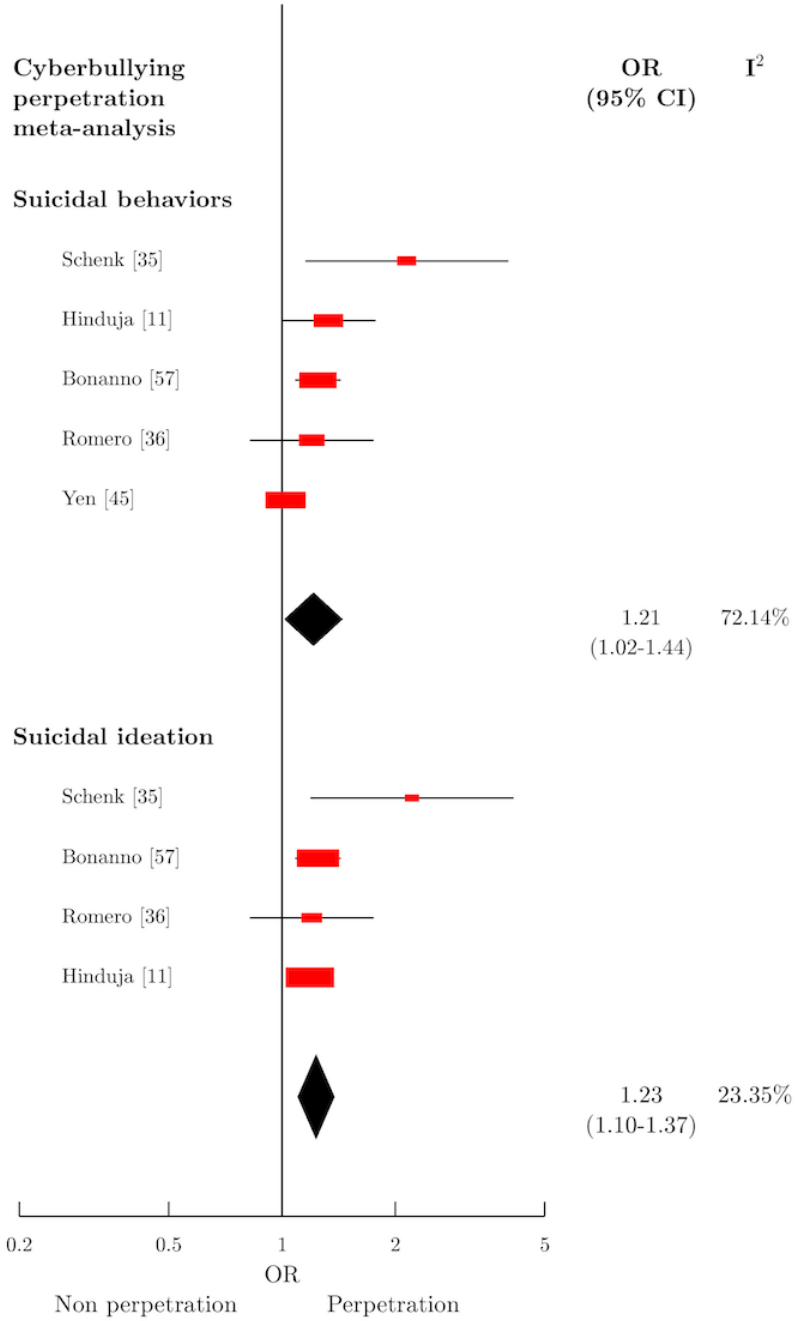


Figure 3. Forest plot of cyberbullying perpetration meta-analyses. OR: odds ratio.



A sensitivity analysis was conducted based solely on articles using a school-based sample. When compared with the results of the original meta-analyses, no significant difference was observed. We also conducted a sensitivity analysis based on articles that reported cybervictimization and traditional victimization separately, or which controlled for traditional victimization in their analyses. For each of the four cybervictimization meta-analyses, this produced greater ORs than those of the original meta-analyses, for example, SH produced OR 3.09 (95% CI 2.36-4.04) compared with OR 2.35 (95% CI 1.65-3.34), whereas suicidal behaviors produced OR 2.35 (95% CI 1.56-3.54) compared with OR 2.10 (95% CI 1.73-2.55). Fewer articles were included in each meta-analyses when restricted in this way (*k*=5 for SH, *k*=8 for suicidal behaviors, *k*=5 for suicide attempt, and *k*=6 for suicidal ideation).

Funnel plots were created for the cybervictimization and suicidal behaviors meta-analysis and cyberbullying perpetration and suicidal behaviors meta-analysis. These displayed no clear signs of publication bias.

More details regarding the results of the meta-regressions and sensitivity analyses, as well as figures for the funnel plots are available in Multimedia Appendix 4.

Discussion

Principal Findings

A total of 26 independent studies (33 articles) were included in this review covering a population of 156,384 children and young people younger than 25 years. A total of 20 independent studies (25 articles) found positive associations (negative influences) between cyberbullying victimization and SH or suicidal behaviors, or between cyberbullying perpetration and suicidal behaviors. One article [33] found an association (negative influence) between cybervictimization and unnatural child death (which included suicide). Two further articles [40,53], where the cyberbullying, SH, or suicidal behaviors measures had been combined with other measures, found negative influences. No significant associations were reported in 5 articles [19,21,36,52,56]. No positive influences of cyberbullying involvement were reported.

These associations were quantified in 6 meta-analyses: those who have experienced cybervictimization are 2.35 times as likely to SH, 2.10 times as likely to exhibit suicidal behaviors, 2.57 times more likely to attempt suicide, and 2.15 times more likely to have suicidal thoughts than nonvictims. Cyberbullying perpetrators were 1.21 times more likely to exhibit suicidal behaviors and 1.23 times more likely to experience suicidal ideation than nonperpetrators.

These findings were comparable with those found for traditional victimization in previous studies. One meta-analysis [13] reported ORs of 2.23 (95% CI 2.10-2.37) for the association between traditional victimization and suicidal ideation and OR 2.55 (95% CI 1.95-3.34) for suicide attempt. Another recent meta-analysis [58] also reported elevated odds of suicidal ideation and suicide attempt for victims of traditional bullying, with ORs of 1.77 (95% CI 1.56-2.02) and OR 2.13 (95% CI 1.66-2.73), respectively.

Only 5 of our eligible articles [17,33,42,48,50] presented results for cyberbullying independently of traditional bullying, with the relative contributions of both types of bullying impossible to determine in the majority of cases because results for cyberbullying involvement did not preclude the simultaneous occurrence of traditional bullying. We performed sensitivity analyses based on those articles that reported cyberbullying or traditional victimization separately or those that controlled for traditional victimization. In each case, ORs were greater than those of the original related meta-analyses, this suggests that cybervictimization could have a greater effect on SH and suicidal behaviors than does traditional victimization. It should be noted, however, that these sensitivity analyses were based on significantly fewer articles. Other evidence was also found suggesting that the effect of cybervictimization on SH and suicidal behaviors acted independently of its co-occurrence with traditional bullying [33,49,50], as well as some evidence of a cumulative effect [17,42,44,56] on SH and suicidal behaviors for the two types of bullying, although this was not seen in all studies [50]. We performed a meta-regression that showed that effect size decreased for cybervictimization and suicidal behaviors as the prevalence of traditional victimization increased. This was also seen for effect sizes in cybervictimization and suicidal ideation. This suggests that cybervictimization has a large effect on its own, but is less important when there are higher levels of traditional victimization present. It is possible that cyberbullying, which enables perpetrators to remain anonymous, has changed and extended the characteristics of both victims and perpetrators. Individuals who, in the nonvirtual world, were unlikely to be victimized as they were able to respond in person may be more vulnerable online where perpetrators may not be identified and are possibly emboldened in a way that they would not be face-to-face. Two recent cross-sectional studies have tried to shed some light on this issue, with one [59] finding that only 1% of adolescents reported being pure cyber victims, whereas the other [60] found traditional bullying to be far more common than cyberbullying, which should be taken into account when interpreting our findings.

Twenty-five new articles were identified since the previous systematic reviews and meta-analyses [12-15]. Van Geel et al [13] found that cybervictimization was more strongly related to suicidal ideation (OR 3.12, 95% CI 2.40-4.05) than in our meta-analysis; however, we included 13 more studies. Across studies, the weighted cybervictimization prevalence was calculated as 12.6% (95% CI 12.4%-12.7%). This is lower than estimates reported in some individual studies [11] of between 15% and 35%. A recent review of cyberbullying highlighted that variability in reported prevalence across studies was dependent on time frames and frequency of cyberbullying used in questions [61].

Limitations

All included studies were observational in design and prone to bias (eg, recall and ecological fallacy). No conclusions can be drawn regarding causality and temporality from cross sectional or case-control studies. Indeed, the possibility of reverse causality (ie, that SH or suicidal behaviors influence cyberbullying involvement) is currently not accounted for in the literature because of these limitations in study design. However, such study designs are appropriate, as manipulation of the level of exposure to cyberbullying involvement would be unethical [12]. We were unable to calculate the prevalence of both cyberbullying perpetration, SH, and suicidal behaviors because of a lack of relevant information.

Definitions of *cyberbullying* varied across studies. Few authors conducted their analyses by frequency of cyberbullying involvement. There is a lack of agreement among researchers as to the exact concept being researched, that is, electronic bullying (email and texts) and online internet harassment, as well as levels of repetition required and intentions of perpetrators. This has been highlighted in a recent review of cyberbullying [61] as the reason why such a broad range of instruments are used to assess cyberbullying. Although these issues may reflect the changing nature of communication technology and its use by young people, it does have an impact on case ascertainment. Future studies should do more to clarify the type of bullying under consideration, including expanding on the different mediums and modes of cyberbullying. For example, by making it clear if texts, social media, or emails are received (medium) and whether these were sent or received by

just one individual or by an entire group (mode). This would allow for more precise analysis of whether these aspects had differential effects on outcomes, enabling those working in bullying prevention programs to equip children and young people with more closely tailored strategies for dealing with cybervictimization. We excluded one study [62] that asked "Have you ever felt hurt by a message you have seen on the internet or on a mobile website?" as it did not necessarily imply that the individual had "felt hurt" more than once, nor that the message was directed toward them.

Validated questionnaires were rarely used to determine SH and suicidal behaviors. Often these were assessed through the response to a single question from self-report surveys. Even where a more detailed methodology was employed, SH and suicidal status were often dichotomized, with all levels of severity grouped together (attempts, plans, and thoughts) to create outcomes such as suicidality and suicidal behaviors [35,45,50,51]. This will have had an impact on effect sizes. In particular, we were unable to distinguish the severity of suicidal ideation. Similarly for the SH and suicidal behavior meta-analyses, where it was not possible to isolate a single measure from two or more eligible measures, we computed a combined average with corresponding variance [11,17,18,42,46,49,54,55]. Although the pragmatic approach we employed was in keeping with both methodological guidance [30] and the literature [35,45,50,51], we acknowledge it may have had an impact on effect sizes in these meta-analyses and may appear reductive. It should be noted however that this was done for only a small proportion of papers per meta-analysis. We also computed a combined average when included studies presented results for females and males separately [18]. Only 1 ecological study assessed deaths [33] and found that countries with higher rates of cybervictimization were more likely to have higher incidence of unnatural child death, which included suicide. However, causality cannot be inferred from this type of study design, and there may be many other influences on deaths, including suicide. Few articles employed any statistical strategies to reduce bias because of confounding factors. Only 4 articles [22,35,47,53] used matching or propensity scoring while conducting their analyses, and this should be taken into account while interpreting our findings.

Sex differences in SH and suicidal behaviors are well recognized [63], yet, only 6 papers examined this with regard to cyberbullying [18,34,39,41,43,56], with conflicting results. We were unable to allow for the effects of other confounders in our meta-analyses (eg, past history of mental disorders or traditional bullying or suicidal behaviors), as these were either not reported, or reported in a way that did not allow us to distinguish them. Similarly, current traditional bullying involvement or mental health issues were presented in such a way that we could not distinguish whether individuals were suffering from more than one outcome (eg, whether cyber victims were suffering from both SH and mental health problems) or often one or more type of bullying. That we were unable to fully account for these two important factors in our analysis should be considered as perhaps the biggest limitation of our review. However, this is a reflection of current literature that should be addressed in future, ideally longitudinally designed, studies.

Some studies were excluded from certain meta-analyses where it was not possible to calculate the standard error for the effect size given the data presented [18,56]. As expected, the random-effects model indicated that variation in effects between studies existed. The high degree of heterogeneity (I^2) observed between studies may reflect the varying victimization populations (case definitions and ascertainment), settings, and methodologies used in the calculation of individual effect sizes. Sensitivity analysis was performed for articles in which school-based samples were used. This revealed no significant change in effect sizes or heterogeneity across any of the outcomes examined and suggests that there exists some unexplained variance other than setting in this instance.

Heterogeneity of studies was particularly high for articles examining cyberbullying perpetration, and the quality of research was lower than that of cybervictimization. Only 8 articles investigated cyberbullying perpetration, with just 1 rated as high quality. It should be noted that the high degree of heterogeneity between studies (for both cybervictimization and perpetration) is a major limitation of the review. We acknowledge that the interpretation of summary OR statistics can be problematic. ORs and the relative risk diverge only when there are large effects (a twofold or threefold increase in risk) for groups already at a large initial risk. However it should be noted that where this occurs the interpretation is the same: these are large effects. We have presented CIs with our summary measures to address this issue.

We used funnel plots to investigate whether there were any signs of publication bias in our review. These gave no indication of bias, though it must be noted that the plot for cyberbullying perpetration was based on a low number of articles (k=5). However, despite conducting an extensive search, we cannot rule out the effects of publication bias on our results, and we only included English language studies. We attempted to address this by the breadth of our search; we identified many new studies compared with previously published reviews.

Implications

This study highlights the significant impact that cyberbullying involvement can have on children and young people. Cybervictimization is a risk factor for SH and suicidal behaviors as is, to a lesser extent, cyberbullying perpetration for suicidal behaviors and ideation. Cyberbullying involvement should be considered by policy makers who implement bullying prevention (in addition to traditional bullying) and safe internet use programs. School, family, and community programs that promote appropriate use of technology are important. Prevention of cyberbullying should be included in school antibullying policies, alongside broader concepts such as digital citizenship, online peer support for victims, how an electronic bystander might appropriately intervene, and more specific interventions such as how to contact mobile phone companies and internet service providers to block, educate, or identify users. Suicide prevention and intervention is essential within any comprehensive antibullying program and should incorporate a whole-school approach to include awareness raising and training for staff and pupils.

A strong link between being a cyber victim and a perpetrator was found in some studies [18,36,45] This duality can particularly put males at higher risk of depression and suicidal behaviors [18]. These vulnerabilities should be recognized at school so that cyberbullying behaviors are seen not as disciplinary issues but as an opportunity to support vulnerable young people. Antibullying programs and protocols should address the needs of both victims and perpetrators. School exclusion might contribute to an individual's sense of isolation. The relationship between cyber victims and suicidal behaviors appears robust. It may be that the persistent and pervasive nature of cybervictimization may lead to feelings of hopelessness, which are associated with suicidal behaviors in adolescents [64]. Students who are cybervictimized are less likely to report and seek help than those victimized by more traditional means [65,66]. Therefore it is important for staff to encourage help-seeking in relation to this issue.

Clinicians working with children and young people and assessing mental health issues should routinely ask about experiences of cyberbullying. The impact of cyberbullying should be included in the training of child and adolescent mental health professionals. Children and young people involved in cyberbullying should be screened for common mental disorders and SH.

The quality of study design, methods, and reporting in future studies needs improvement. Only a third of included articles (11 of 33) were rated as high quality, with 17 rated medium and 5 rated low. Validated psychometric instruments should be used to assess the suicidal status of individuals wherever possible to increase reliability and the ability to make comparisons across populations. More detailed analysis of the medium of cyberbullying (eg, via phones or instant messaging) should be explored to investigate any differences in populations and impact. The ability to distinguish media would support the development of targeted prevention strategies.

Finally, researchers should investigate the mechanisms by which mental disorders such as anxiety and depression mediate the link between cyberbullying involvement and SH and suicide. One study [18] found that depression mediated the link between cybervictimization and suicide attempts for females only, whereas perpetration was a direct predictor of suicide attempts for males only. This suggests that gender specific strategies for prevention and intervention may be helpful. Further research exploring the mechanisms of these associations is required. For cyberbullying perpetrators, a statement intended as a joke with no harm intended, may have unforeseen consequences with resultant guilt [18]. This association could be explored more deeply in mixed-methods and qualitative studies to gain a deeper understanding. Our review included no such studies. Future studies could also collect information from parents, peers, and teachers.

It is important to identify protective factors for children and young people exposed to cyberbullying. Although some studies [49] explored the moderating effects of social support (family and peers) on cyberbullying and mental health problems, no studies explored this in relation to SH and suicidal behaviors. Other protective factors may include aspects of resilience, such as internal locus of control and self-esteem [11].

Conclusions

In conclusion, our review suggests that cyber victims are at greater risk of both SH and suicidal behaviors and, to a lesser extent, perpetrators of cyberbullying are at greater risk of suicidal behaviors than those with no cyberbullying involvement. The evidence base in this field has grown rapidly, but it is clear that the quality of future studies needs improvement. This research area would benefit from a clear definition of cyberbullying, assessed in longitudinal studies using validated assessments of SH and suicidal behaviors. Cyberbullying type, frequency, and gender should be explored. This is important to support policymakers, teachers, parents, clinicians, and others working with young people to make informed decisions in the safeguarding of children and young people.

Acknowledgments

KH is a National Institute for Health Research Senior Investigator. AJ and KL are codirectors of the Cochrane Satellite for Suicide and Self-Harm Prevention. The authors would like to thank Chukwudi Okolie and Public Health Wales Evidence Service for support in conducting electronic searches. Health and Care Research Wales funded this study (Grant number SC-14-11). The funder has no role in study design, results, or interpretation of findings.

Authors' Contributions

AJ and KH conceived the study; all authors designed the study; AG and AM conducted the literature searches, AG, AM, and AJ extracted the data; all authors interpreted study findings; AJ and AG drafted the manuscript; and all authors critically reviewed the manuscript and approved the final submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Inclusion and exclusion criteria.

[PDF File (Adobe PDF File), 30KB - jmir_v20i4e129_app1.pdf]

Multimedia Appendix 2

Further details of methods.

[PDF File (Adobe PDF File), 366KB - jmir_v20i4e129_app2.pdf]

Multimedia Appendix 3

Summary table displaying the study type, mean age, quality score, and results of studies included in the review.

[PDF File (Adobe PDF File), 334KB - jmir_v20i4e129_app3.pdf]

Multimedia Appendix 4

Further details of results.

[PDF File (Adobe PDF File), 760KB - jmir_v20i4e129_app4.pdf]

Multimedia Appendix 5

Summary of outcome measures used in studies included in review.

[PDF File (Adobe PDF File), 89KB - jmir_v20i4e129_app5.pdf]

Multimedia Appendix 6

Cybervictimization weighted prevalence rates and prevalence rates by study.

[PDF File (Adobe PDF File), 64KB - jmir_v20i4e129_app6.pdf]

References

1. Olweus D. Victimization by peers: antecedents and long-term outcomes. In: Rubin KH, Asendorpf JB, editors. Social Withdrawal, Inhibition, and Shyness in Childhood. New York, London: Lawrence Erlbaum Associates; 1993:315-341.
2. Smith PK, Mahdavi J, Carvalho M, Fisher S, Russell S, Tippett N. Cyberbullying: its nature and impact in secondary school pupils. J Child Psychol Psychiatry 2008 Apr;49(4):376-385. [doi: 10.1111/j.1469-7610.2007.01846.x] [Medline: 18363945]
3. Suzuki K, Asaga R, Sourander A, Hoven CW, Mandell D. Cyberbullying and adolescent mental health. Int J Adolesc Med Health 2012;24(1):27-35. [doi: 10.1515/ijamh.2012.005] [Medline: 22909909]
4. Kowalski RM, Limber SP. Electronic bullying among middle school students. J Adolesc Health 2007 Dec;41(6 Suppl 1):S22-S30. [doi: 10.1016/j.jadohealth.2007.08.017] [Medline: 18047942]
5. Arseneault L, Bowes L, Shakoor S. Bullying victimization in youths and mental health problems: 'much ado about nothing'? Psychol Med 2010 May;40(5):717-729. [doi: 10.1017/S0033291709991383] [Medline: 19785920]
6. Gini G, Pozzoli T. Association between bullying and psychosomatic problems: a meta-analysis. Pediatrics 2009 Mar;123(3):1059-1065. [doi: 10.1542/peds.2008-1215] [Medline: 19255040]
7. Hawker DS, Boulton MJ. Twenty years' research on peer victimization and psychosocial maladjustment: a meta-analytic review of cross-sectional studies. J Child Psychol Psychiatry 2000 May;41(4):441-455. [Medline: 10836674]
8. Kim YS, Leventhal BL, Koh YJ, Boyce WT. Bullying increased suicide risk: prospective study of Korean adolescents. Arch Suicide Res 2009;13(1):15-30. [doi: 10.1080/13811110802572098] [Medline: 19123106]
9. Brunstein Klomek A, Marrocco F, Kleinman M, Schonfeld IS, Gould MS. Bullying, depression, and suicidality in adolescents. J Am Acad Child Adolesc Psychiatry 2007 Jan;46(1):40-49. [doi: 10.1097/01.chi.0000242237.84925.18] [Medline: 17195728]
10. Juvonen J, Gross EF. Extending the school grounds?--Bullying experiences in cyberspace. J Sch Health 2008 Sep;78(9):496-505. [doi: 10.1111/j.1746-1561.2008.00335.x] [Medline: 18786042]
11. Hinduja S, Patchin JW. Bullying, cyberbullying, and suicide. Arch Suicide Res 2010;14(3):206-221. [doi: 10.1080/13811118.2010.494133] [Medline: 20658375]
12. Daine K, Hawton K, Singaravelu V, Stewart A, Simkin S, Montgomery P. The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. PLoS One 2013;8(10):e77555 [FREE Full text] [doi: 10.1371/journal.pone.0077555] [Medline: 24204868]
13. van Geel M, Vedder P, Tanilon J. Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: a meta-analysis. JAMA Pediatr 2014 May;168(5):435-442. [doi: 10.1001/jamapediatrics.2013.4143] [Medline: 24615300]

14. Hamm MP, Newton AS, Chisholm A, Shulhan J, Milne A, Sundar P, et al. Prevalence and effect of cyberbullying on children and young people: a scoping review of social media studies. JAMA Pediatr 2015 Aug;169(8):770-777. [doi: 10.1001/jamapediatrics.2015.0944] [Medline: 26098362]
15. Kowalski RM, Giumetti GW, Schroeder AN, Lattanner MR. Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. Psychol Bull 2014 Jul;140(4):1073-1137. [doi: 10.1037/a0035618] [Medline: 24512111]
16. Hay C, Meldrum R. Bullying victimization and adolescent self-harm: testing hypotheses from general strain theory. J Youth Adolesc 2010 May;39(5):446-459. [doi: 10.1007/s10964-009-9502-0] [Medline: 20072852]
17. Schneider SK, O'Donnell L, Stueve A, Coulter RW. Cyberbullying, school bullying, and psychological distress: a regional census of high school students. Am J Public Health 2012 Jan;102(1):171-177. [doi: 10.2105/AJPH.2011.300308] [Medline: 22095343]
18. Bauman S, Toomey RB, Walker JL. Associations among bullying, cyberbullying, and suicide in high school students. J Adolesc 2013 Apr;36(2):341-350. [doi: 10.1016/j.adolescence.2012.12.001] [Medline: 23332116]
19. Cassidy W, Jackson M, Brown KN. Sticks and stones can break my bones, but how can pixels hurt me? Students' experiences with cyber-bullying. Sch Psychol Int 2009 Jul 31;30(4):383-402. [doi: 10.1177/0143034309106948]
20. Goebert D, Else I, Matsu C, Chung-Do J, Chang JY. The impact of cyberbullying on substance use and mental health in a multiethnic sample. Matern Child Health J 2011 Nov;15(8):1282-1286. [doi: 10.1007/s10995-010-0672-x] [Medline: 20824318]
21. Price M, Dalgleish J. Cyberbullying: experiences, impacts and coping strategies as described by Australian young people. Youth Stud Aust 2010;29(2):51.
22. Schenk AM, Fremouw WJ. Prevalence, psychological impact, and coping of cyberbully victims among college students. J Sch Violence 2012 Jan;11(1):21-37. [doi: 10.1080/15388220.2011.630310]
23. McManus S, Bebbington P, Jenkins R, Brugha T. 2016. Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014 URL: https://digital.nhs.uk/catalogue/PUB21748 [WebCite Cache ID 6tFPEAUfb]
24. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. J Am Med Assoc 2000 Apr 19;283(15):2008-2012. [Medline: 10789670]
25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
26. Critical Appraisal Skills Programme . 2017. Critical Appraisal Skills Programme Checklists URL: http://www.casp-uk.net/casp-tools-checklists [accessed 2017-09-05] [WebCite Cache ID 6tFOnnqaZ]
27. Welsh Government. 2015. Suicide and Self Harm Prevention Strategy for Wales, 2015-2020 Talk to me 2 URL: http://gov.wales/docs/dhss/publications/150716strategyen.pdf [accessed 2017-09-05] [WebCite Cache ID 6tFQ03Awf]
28. Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S. Suicide and suicidal behavior. Epidemiol Rev 2008 May 14;30(1):133-154. [doi: 10.1093/epirev/mxn002]
29. Thomas SL, Heck RH. Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. Res High Educ 2001;42(5):517-540. [doi: 10.1023/A:1011098109834]
30. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. Hoboken, New Jersey: John Wiley & Sons; 2011.
31. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Chichester: The Cochrane Collaboration and John Wiley & Sons; 2011.
32. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
33. Fu KW, Chan CH, Ip P. Exploring the relationship between cyberbullying and unnatural child death: an ecological study of twenty-four European countries. BMC Pediatr 2014 Jul 30;14:195 [FREE Full text] [doi: 10.1186/1471-2431-14-195] [Medline: 25079144]
34. Bannink R, Broeren S, van de Looij–Jansen PM, de Waart FG, Raat H. Cyber and traditional bullying victimization as a risk factor for mental health problems and suicidal ideation in adolescents. PLoS One 2014;9(4):e94026 [FREE Full text] [doi: 10.1371/journal.pone.0094026] [Medline: 24718563]
35. Schenk AM, Fremouw WJ, Keelan CM. Characteristics of college cyberbullies. Comput Human Behav 2013 Nov;29(6):2320-2327. [doi: 10.1016/j.chb.2013.05.013]
36. Romero AJ, Wiggs CB, Valencia C, Bauman S. Latina teen suicide and bullying. Hisp J Behav Sci 2013 Mar 15;35(2):159-173. [doi: 10.1177/0739986312474237]
37. Alavi N, Roberts N, Sutton C, Axas N, Repetti L. Bullying victimization (being bullied) among adolescents referred for urgent psychiatric consultation: prevalence and association with suicidality. Can J Psychiatry 2015 Oct;60(10):427-431. [doi: 10.1177/070674371506001003] [Medline: 26720189]
38. Roberts N, Axas N, Nesdole R, Repetti L. Pediatric emergency department visits for mental health crisis: prevalence of cyber-bullying in suicidal youth. Child Adolesc Soc Work J 2016 Feb 18;33(5):469-472. [doi: 10.1007/s10560-016-0442-8]

39. Cénat JM, Blais M, Hébert M, Lavoie F, Guerrier M. Correlates of bullying in Quebec high school students: The vulnerability of sexual-minority youth. J Affect Disord 2015 Sep 01;183:315-321 [FREE Full text] [doi: 10.1016/j.jad.2015.05.011] [Medline: 26047959]
40. Hébert M, Cénat JM, Blais M, Lavoie F, Guerrier M. Child sexual abuse, bullying, cyberbullying, and mental health problems among high schools students: a moderated mediated model. Depress Anxiety 2016 Jul;33(7):623-629 [FREE Full text] [doi: 10.1002/da.22504] [Medline: 27037519]
41. Hay C, Meldrum R, Mann K. Traditional bullying, cyber bullying, and deviance: a general strain theory approach. J Contemp Crim Justice 2010 Mar 21;26(2):130-147. [doi: 10.1177/1043986209359557]
42. Messias E, Kindrick K, Castro J. School bullying, cyberbullying, or both: correlates of teen suicidality in the 2011 CDC Youth Risk Behavior Survey. Compr Psychiatry 2014 Jul;55(5):1063-1068 [FREE Full text] [doi: 10.1016/j.comppsych.2014.02.005] [Medline: 24768228]
43. Reed KP, Nugent W, Cooper RL. Testing a path model of relationships between gender, age, and bullying victimization and violent behavior, substance abuse, depression, suicidal ideation, and suicide attempts in adolescents. Child Youth Serv Rev 2015 Aug;55:128-137. [doi: 10.1016/j.childyouth.2015.05.016]
44. Kindrick K, Castro J, Messias E. Sadness, suicide, and bullying in Arkansas: results from the Youth Risk Behavior Survey 2011. J Ark Med Soc 2013 Oct;110(5):90-91. [Medline: 24383197]
45. Yen CF, Chou WJ, Liu TL, Ko CH, Yang P, Hu HF. Cyberbullying among male adolescents with attention-deficit/hyperactivity disorder: prevalence, correlates, and association with poor mental health status. Res Dev Disabil 2014 Dec;35(12):3543-3553. [doi: 10.1016/j.ridd.2014.08.035] [Medline: 25241113]
46. Arat G. Emerging protective and risk factors of mental health in Asian American students: findings from the 2013 Youth Risk Behavior Survey. Vulnerable Child Youth Stud 2015 May 12;10(3):192-205. [doi: 10.1080/17450128.2015.1045437]
47. DeSmet A, Deforche B, Hublet A, Tanghe A, Stremersch E, De Bourdeaudhuij I. Traditional and cyberbullying victimization as correlates of psychosocial distress and barriers to a healthy lifestyle among severely obese adolescents-a matched case-control study on prevalence and results from a cross-sectional study. BMC Public Health 2014 Mar 05;14:224 [FREE Full text] [doi: 10.1186/1471-2458-14-224] [Medline: 24593118]
48. Duong J, Bradshaw C. Associations between bullying and engaging in aggressive and suicidal behaviors among sexual minority youth: the moderating role of connectedness. J Sch Health 2014 Oct;84(10):636-645. [doi: 10.1111/josh.12196] [Medline: 25154527]
49. Elgar FJ, Napoletano A, Saul G, Dirks MA, Craig W, Poteat VP, et al. Cyberbullying victimization and mental health in adolescents and the moderating role of family dinners. JAMA Pediatr 2014 Nov;168(11):1015-1022. [doi: 10.1001/jamapediatrics.2014.1223] [Medline: 25178884]
50. Kodish T, Herres J, Shearer A, Atte T, Fein J, Diamond G. Bullying, depression, and suicide risk in a pediatric primary care sample. Crisis 2016 May;37(3):241-246. [doi: 10.1027/0227-5910/a000378] [Medline: 27040126]
51. Litwiller BJ, Brausch AM. Cyber bullying and physical bullying in adolescent suicide: the role of violent behavior and substance use. J Youth Adolesc 2013 May;42(5):675-684. [doi: 10.1007/s10964-013-9925-5] [Medline: 23381779]
52. Mitchell KJ, Wells M, Priebe G, Ybarra ML. Exposure to websites that encourage self-harm and suicide: prevalence rates and association with actual thoughts of self-harm and thoughts of suicide in the United States. J Adolesc 2014 Dec;37(8):1335-1344. [doi: 10.1016/j.adolescence.2014.09.011] [Medline: 25313930]
53. Roh BR, Yoon Y, Kwon A, Oh S, Lee SI, Ha K, et al. The structure of co-occurring bullying experiences and associations with suicidal behaviors in Korean adolescents. PLoS One 2015;10(11):e0143517 [FREE Full text] [doi: 10.1371/journal.pone.0143517] [Medline: 26619356]
54. Sampasa-Kanyinga H, Roumeliotis P, Xu H. Associations between cyberbullying and school bullying victimization and suicidal ideation, plans and attempts among Canadian schoolchildren. PLoS One 2014;9(7):e102145 [FREE Full text] [doi: 10.1371/journal.pone.0102145] [Medline: 25076490]
55. Sinclair KO, Bauman S, Poteat VP, Koenig B, Russell ST. Cyber and bias-based harassment: associations with academic, substance use, and mental health problems. J Adolesc Health 2012 May;50(5):521-523. [doi: 10.1016/j.jadohealth.2011.09.009] [Medline: 22525118]
56. Turner MG, Exum ML, Brame R, Holt TJ. Bullying victimization and adolescent mental health: General and typological effects across sex. J Crim Justice 2013 Jan;41(1):53-59. [doi: 10.1016/j.jcrimjus.2012.12.005]
57. Bonanno RA, Hymel S. Cyber bullying and internalizing difficulties: above and beyond the impact of traditional forms of bullying. J Youth Adolesc 2013 May;42(5):685-697. [doi: 10.1007/s10964-013-9937-1] [Medline: 23512485]
58. Moore SE, Norman RE, Suetani S, Thomas HJ, Sly PD, Scott JG. Consequences of bullying victimization in childhood and adolescence: A systematic review and meta-analysis. World J Psychiatry 2017 Mar 22;7(1):60-76 [FREE Full text] [doi: 10.5498/wjp.v7.i1.60] [Medline: 28401049]
59. Wolke D, Lee K, Guy A. Cyberbullying: a storm in a teacup? Eur Child Adolesc Psychiatry 2017 Aug;26(8):899-908 [FREE Full text] [doi: 10.1007/s00787-017-0954-6] [Medline: 28188382]
60. Przybylski AK, Bowes L. Cyberbullying and adolescent well-being in England: a population-based cross-sectional study. Lancet Child & Adolesc Health 2017 Sep;1(1):19-26. [doi: 10.1016/S2352-4642(17)30011-1]

61. Garett R, Lord LR, Young SD. Associations between social media and cyberbullying: a review of the literature. Mhealth 2016;2:46 [FREE Full text] [doi: 10.21037/mhealth.2016.12.01] [Medline: 28293616]
62. Katsumata Y, Matsumoto T, Kitani M, Takeshima T. Electronic media use and suicidal ideation in Japanese adolescents. Psychiatry Clin Neurosci 2008 Dec;62(6):744-746 [FREE Full text] [doi: 10.1111/j.1440-1819.2008.01880.x] [Medline: 19068014]
63. Hawton K. Sex and suicide: gender differences in suicidal behaviour. Br J Psychiatry 2018 Jan 2;177(06):484-485. [doi: 10.1192/bjp.177.6.484]
64. Hawton K, Saunders KE, O'Connor RC. Self-harm and suicide in adolescents. Lancet 2012 Jun 23;379(9834):2373-2382. [doi: 10.1016/S0140-6736(12)60322-5] [Medline: 22726518]
65. Agatston PW, Kowalski R, Limber S. Students' perspectives on cyber bullying. J Adolesc Health 2007 Dec;41(6 Suppl 1):S59-S60. [doi: 10.1016/j.jadohealth.2007.09.003] [Medline: 18047946]
66. Dooley JJ, Gradinger P, Strohmeier D, Cross D, Spiel C. Cyber-victimisation: the association between help-seeking behaviours and self-reported emotional symptoms in Australia and Austria. Aust J Guid Couns 2012 Feb 23;20(02):194-209. [doi: 10.1375/ajgc.20.2.194]

Abbreviations

OR: odds ratio
SH: self-harm

Edited by G Eysenbach; submitted 29.09.17; peer-reviewed by G Portzky, H Miller; comments to author 26.10.17; revised version received 21.12.17; accepted 23.01.18; published 19.04.18.

Please cite as:
John A, Glendenning AC, Marchant A, Montgomery P, Stewart A, Wood S, Lloyd K, Hawton K
Self-Harm, Suicidal Behaviours, and Cyberbullying in Children and Young People: Systematic Review
J Med Internet Res 2018;20(4):e129
URL: http://www.jmir.org/2018/4/e129/
doi:10.2196/jmir.9044
PMID:29674305

Original Paper

Solution to Detect, Classify, and Report Illicit Online Marketing and Sales of Controlled Substances via Twitter: Using Machine Learning and Web Forensics to Combat Digital Opioid Access

Tim Mackey[1*], MAS, PhD; Janani Kalyanam[2*], MA, PhD; Josh Klugman[3], BS; Ella Kuzmenko[3], BA; Rashmi Gupta[4], BE, MBA

[1]Division of Infectious Disease and Global Public Health, Department of Anesthesiology, School of Medicine, University of California San Diego, La Jolla, CA, United States

[2]Department of Electrical and Computer Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, CA, United States

[3]IBM Global Business Services, Washington, DC, United States

[4]Centers for Disease Control and Prevention, Atlanta, GA, United States

[*]these authors contributed equally

Corresponding Author:
Tim Mackey, MAS, PhD
Division of Infectious Disease and Global Public Health
Department of Anesthesiology, School of Medicine
University of California San Diego
8950 Villa La Jolla Drive, A124
La Jolla, CA, 92037
United States
Phone: 1 9514914161
Email: tmackey@ucsd.edu

Abstract

Background: On December 6 and 7, 2017, the US Department of Health and Human Services (HHS) hosted its first Code-a-Thon event aimed at leveraging technology and data-driven solutions to help combat the opioid epidemic. The authors—an interdisciplinary team from academia, the private sector, and the US Centers for Disease Control and Prevention—participated in the Code-a-Thon as part of the prevention track.

Objective: The aim of this study was to develop and deploy a methodology using machine learning to accurately detect the marketing and sale of opioids by illicit online sellers via Twitter as part of participation at the HHS Opioid Code-a-Thon event.

Methods: Tweets were collected from the Twitter public application programming interface stream filtered for common prescription opioid keywords in conjunction with participation in the Code-a-Thon from November 15, 2017 to December 5, 2017. An unsupervised machine learning–based approach was developed and used during the Code-a-Thon competition (24 hours) to obtain a summary of the content of the tweets to isolate those clusters associated with illegal online marketing and sale using a biterm topic model (BTM). After isolating relevant tweets, hyperlinks associated with these tweets were reviewed to assess the characteristics of illegal online sellers.

Results: We collected and analyzed 213,041 tweets over the course of the Code-a-Thon containing keywords codeine, percocet, vicodin, oxycontin, oxycodone, fentanyl, and hydrocodone. Using BTM, 0.32% (692/213,041) tweets were identified as being associated with illegal online marketing and sale of prescription opioids. After removing duplicates and dead links, we identified 34 unique “live” tweets, with 44% (15/34) directing consumers to illicit online pharmacies, 32% (11/34) linked to individual drug sellers, and 21% (7/34) used by marketing affiliates. In addition to offering the “no prescription” sale of opioids, many of these vendors also sold other controlled substances and illicit drugs.

Conclusions: The results of this study are in line with prior studies that have identified social media platforms, including Twitter, as a potential conduit for supply and sale of illicit opioids. To translate these results into action, authors also developed a prototype wireframe for the purposes of detecting, classifying, and reporting illicit online pharmacy tweets selling controlled substances illegally to the US Food and Drug Administration and the US Drug Enforcement Agency. Further development of solutions based on these methods has the potential to proactively alert regulators and law enforcement agencies of illegal opioid sales, while also making the online environment safer for the public.

(J Med Internet Res 2018;20(4):e10029) doi:10.2196/10029

KEYWORDS

online pharmacies; drug abuse; opioid abuse; machine learning; unsupervised machine learning; prescription drug misuse

Introduction

National Opioid Crisis

It is estimated that 90 Americans die daily by overdosing on opioids, a staggering figure highlighting the human toll of this public health crisis that continues to escalate [1]. Since the year 2000, an estimated 300,000 lives have been claimed by the opioid epidemic, which has expanded beyond nonmedical use of prescription opioids into transition of use to heroin addiction and deaths occurring from illicitly manufactured synthetic opioids (such as fentanyls and their analogues [2-5]). Additionally, the US Centers for Disease Control and Prevention (CDC) estimates that the annual economic losses from this crisis equate to US $78.5 billion because of the costs of health care, addiction treatment, the criminal justice system, and lost productivity [6,7]. Rising death tolls and growing economic burden (with CDC reporting a quadrupling of deaths attributable to prescription opioids since 1999) have prompted certain state jurisdictions and President Donald Trump to declare the opioid crisis a public health emergency [8,9].

Responses to tackle the opioid epidemic have occurred at both the state and federal level, largely focused on actions aimed at reducing inappropriate prescribing, expanding opioid treatment and prevention programs (including access to the opioid antagonist and rapidly reversing overdose drug naloxone), establishing prescription drug monitoring programs, preventing drug diversion, and even the use of litigation against pharmaceutical companies [5,8,10-13]. These approaches largely fit into the five major priority areas outlined by the US Department of Health and Human Services (HHS) to combat the opioid crisis, which include improving access to treatment and recovery services, promoting use of overdose-reversing drugs, strengthening public health surveillance, enhancing research on pain and addiction, and supporting better practices for pain management. These strategic goals are being carried out through investments in science, training, various mitigation strategies, community-based activities, efforts to change prescribing and management practices, and policy making [13-15].

Seeking to further catalyze efforts around HHS' five-part opioid strategy through technology and innovation, on December 6 and 7, 2017, the agency hosted an Opioid Code-a-Thon event that brought together over 300 participants to develop data-driven solutions to combat the opioid epidemic in three challenge tracks (treatment, usage, and prevention tracks; see Table 1) [16]. The Code-a-Thon, the first of its kind for HHS, also involved partnership with several data science providers, organizations, and platform sponsors, including Socrata, Tableau, IEEE Standards Association, and Google. The event provided participants access to a hosted data portal that included deidentified datasets from HHS, federal, state, and local governments, and the private industry (eg, datasets such as CDC WONDER—Multiple Causes of Death, Medicare Part D Prescribing Data—Centers for Medicare & Medicaid Services, National Survey on Drug Use and Health—Substance Abuse and Mental Health Services Administration, and Medical Expenditure Panel Survey—Agency for Healthcare Research and Quality), which were used by participants to build data visualizations, interpret data in new ways, build analysis tools, and propose broader solutions.

In total, over 50 teams competed and pitched their ideas through two rounds of judging, with nine selected for a final round of presentations. In the end, three winners were awarded prizes of US $10,000 each to further develop and implement their solutions [16]. The coauthors of this paper, comprising an interdisciplinary public-private team with members from UC San Diego—School of Medicine and Jacobs School of Engineering, the CDC, and IBM, participated as team Ryan Haight, as part of the prevention track (a track that asked for solutions designed to predict and analyze the supply and movement of legal and illicit opioids) and were selected as one of nine finalists (for recorded video of Code-a-Thon presentation visit the HHS YouTube video [17]). In this paper, we describe the opioid challenge we addressed, the methods and solution we used to address the challenge, the results of our analysis, and provide a discussion on our approach to move this innovation forward in an attempt to address the digital dangers of opioid abuse via social media and illicit online sellers.

Opioid Challenge: Illicit Online Marketing of Sale of Opioids Direct-to-Consumer via Twitter

We named our Code-a-Thon team after an 18-year old adolescent from San Diego, California, who overdosed and died after purchasing the prescription opioid Vicodin from a no prescription internet seller in February 2001 [18,19]. The death of Ryan Haight eventually led to passage of federal legislation, the 2008 Ryan Haight Online Pharmacy Consumer Protection Act (RHA), aimed at curbing illegal diversion of controlled substances by making it a federal crime to purchase, sell, or import controlled substances online without a valid prescription as currently enforced by the US Drug Enforcement Agency (DEA) [18].

However, since the enactment of RHA, internet technologies have experienced rapid growth, with an estimated 84% and 65% of adult Americans using the internet and social media respectfully [20]. Furthermore, a survey conducted by the US Food and Drug Administration (FDA) of adults who had made purchases online found 23% had purchased a prescription medication online [21]. Reflecting this trend of increasing internet adoption and use of internet for health information–seeking and e-commerce, online pharmacies have also rapidly proliferated with an estimated 30,000 legitimate and illegal sites (though it is estimated that 96% of these sites fail to adhere to legal and safety requirements) in existence [22-24].

Table 1. Summary of US Department of Health and Human Services (HHS) Opioid Code-a-Thon challenge tracks.

Track name	Description	Winning team
Prevention track	Solutions used to better predict and analyze the supply and movement of legal and illicit opioids	Visionist Inc, solution to assess unmet need for takeback programs at pharmacies
Treatment track	Solutions to improve access to effective treatment and recovery services	Origigami Innovations, developed model for real-time tracking of overdoses
Usage track	Solutions to help identify at-risk populations and their underlying risk characteristics of opioid misuse or abuse	Visionist Inc, solution to assess unmet need for takeback programs at pharmacies

Of these tens-of-thousands of cyber-pharmacies, the internet security firm LegitScript estimates that approximately 9% sell controlled substances, with a separate report by the National Association of Boards of Pharmacy (NABP), which reviewed more than 11,000 online pharmacy websites, estimating that 13% illegally dispensed controlled substances [22,25,26]. Results from these studies also confirm findings from US government investigations, including a 2004 study by the US Government Accountability Office (GAO), where investigators were able to purchase OxyContin, Percocet, and Vicodin from no prescription online pharmacies, and a recent bipartisan report detailing how Senate investigators were able to easily purchase fentanyl online and have product shipped through the mail [27-29].

Supplementing reports from the FDA, NABP, the GAO, and Senate investigations, a number of published research studies have also found that prescription opioids are readily marketed and sold online, both pre and post RHA [30-35]. This includes recent studies establishing a link between illicit online pharmacies that use social media channels (primarily Twitter) to market and directly sell opioids and other controlled substances (including fentanyls) [34-37]. Growing social media popularity among consumers and the fact that social media platforms are generally less regulated compared with other parts of the Web, such as indexed search engine results, may be driving these trends (eg, in 2011, Google was fined US $500 million by the US Department of Justice for knowingly allowing illegal ads for fraudulent pharmacies, including those selling controlled substances, and has since instituted certain policies for prevention [38]). The connection between illicit access and social media is particularly concerning given that many social media platforms are popular among young adults (a recent 2017 study reported that 30.85% of young adults used Twitter), a population at specific and increasing risk to prescription opioid addiction [39,40]. This clear risk to patient safety and the need for technology-based solutions has also recently caught the attention of the US Congress, with bipartisan letters from senators Chuck Grassley and Diane Feinstein sent to Google, Microsoft, Yahoo!, and Pinterest on February 15, 2018, warning them of the how their platforms are facilitating the online sale of illicit narcotics [41].

Hence, recognizing the need for innovative solutions leveraging advances in infoveillance, big data, machine learning, and Web forensic analysis, our team developed a method to detect marketing and sale of controlled substances via Twitter by online sellers. The main component of our solution was its use and establishing the viability of a proof-of-concept protocol for an unsupervised machine-learning algorithm to detect illicit online opioid seller tweets. We also created a prototype wireframe of a Web application to detect, classify, and report results for potential use by stakeholders such as the DEA, FDA, pharmaceutical manufacturers, and consumer patient safety groups.

Methods

Overview

The analysis for this study was conducted in two distinct phases including (1) Code-a-Thon challenge assessment and (2) “big data” analysis using machine learning of a Twitter dataset. We describe the design of our prototype Web application solution with a wireframe demo in the Results and Discussion section. The first phase involved coordinating with one of the coauthors of this manuscript, who is currently a CDC Entrepreneur in Residence, to scope out an appropriate challenge problem that fit the specific objectives and appropriate track of the Code-a-Thon (ie, addressing an under recognized threat in illegal opioid supply and access), identify the relevant datasets needed to address the challenge (ie, collecting a Twitter dataset associated with opioids precompetition), and organize team registration and logistics associated with Code-a-Thon participation. The second phase comprised data analysis conducted during the Code-a-Thon as described below. As this study involved the collection and analysis of existing publicly available data, it did not require institutional review board approval, nor was that required for participation in the Code-a-Thon.

Data Collection and Analysis

After determining the challenge problem in partnership with CDC, we proceeded to collect messages (ie, tweets) published on Twitter over a period of approximately 20 days from November 15, 2017 to December 5, 2017 (the day before the start of the Code-a-Thon). The public streaming application programming interface (API) available from Twitter was used with certain preselected keywords that were a combination of International Nonproprietary Names and brand names of commonly abused opioids. Our final keyword list contained the terms codeine, fentanyl, hydrocodone, oxycodone, Oxycontin, Percocet, and Vicodin.

Upon commencement of the competition, we used a machine learning–based protocol to isolate word groupings associated with tweets that mentioned marketing and purported sale of prescription opioid drugs as has been carried out in prior studies by the first and second author [34,35,37]. To identify relevant tweets related to our challenge problem (ie, “signal” data) in

large volumes of Twitter data (in the hundreds of thousands) vs nonrelevant data (ie, "noise" that contains opioid-related keywords but do not relate to online promotion or sale), the application of machine learning is critical to achieve scale and comprehensive analysis in a reasonable time frame compared with approaches solely using manual annotation by human coders.

Specifically, unsupervised methods such as topic models prove to be useful in obtaining a summary of the underlying themes present in large text corpora. As we wanted to use methods that are able to capitalize on all the volume of our data collection, we chose unsupervised topic modeling–based methods over other approaches, such as those primarily based on analyzing hashtag co-occurrences [42]. We used a model called the biterm topic model (BTM) designed to detect themes and patterns in corpora of short texts (such as tweets), which we have previously used to examine prescription drug abuse behavior and online marketing and access (see Figure 1 for summary of this methodology) [35,43]. BTM was chosen because it is specifically designed to work in scenarios where the length of the documents or messages are short.

BTM (in its "learning" phase) first detected a preconfigured number of themes from the filtered dataset of tweets containing prescription opioid keywords. This produced a set of topics (or word groupings) that are thematic summaries of the contents of the entire set of tweets. The resulting word groupings are used to inform the next steps in the methodology—which are either identification of themes to detect signal associated with illegal marketing and sales of prescription opioids and to eliminate noise (discarding tweets that are deemed as "noisy" to isolate signal tweets) and then reapplying BTM on the smaller subset that has been filtered for noise until thematic saturation is achieved.

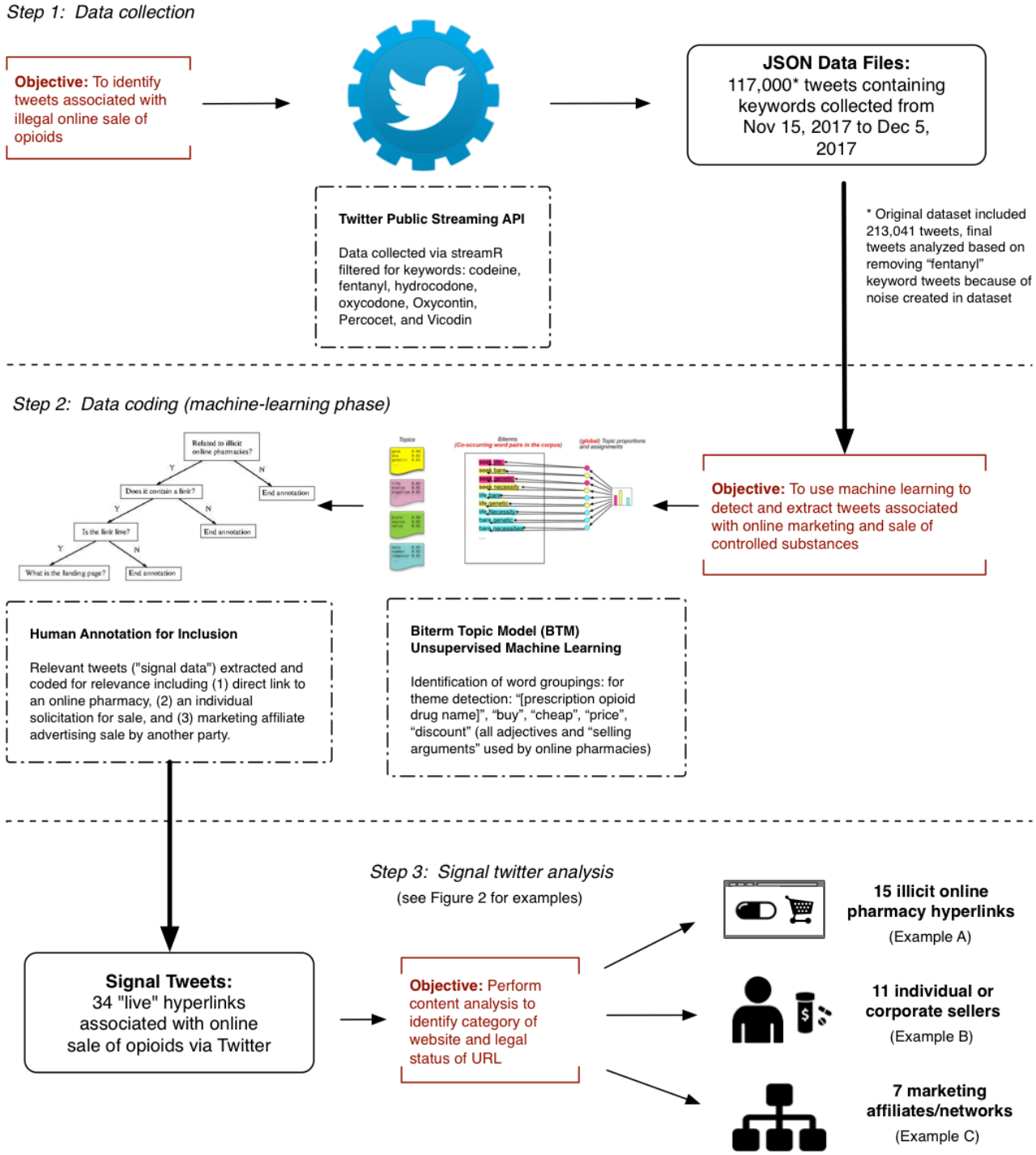
However, to ensure that BTM was the most appropriate method for this study, we also performed experiments with three other topic models: (1) Latent Dirichlet allocation, (2) Nonnegative Matrix Factorization, (3) and Kernel k-means. After each run of each model, we calculated both the cluster purity and perplexity scores for each of the topics and obtained the average across all clusters. The cluster purity measures how tightly knit each cluster of documents is, and hence, a higher purity is better. The perplexity measures how good a language model is at predicting words with a lower perplexity score being an indication of a good language model. On the basis of these tests, BTM scored higher across both metrics for different cluster numbers. Hence, we felt confident that using BTM for this particular data source had high performance compared with other topic models.

Once the learning phase of BTM produced a set of themes (or word groupings), each of the themes was manually annotated to identify "signal" themes clearly associated with prescription opioid marketing, distribution, and sale. Those themes that were marked as relevant (eg, a theme that would typically marked as being "relevant" is one with a combination of words including "[prescription opioid drug name]," "buy," "cheap," "price," and "discount," where these adjectives and "selling arguments" are identified as used by online sellers) were then extracted for further analysis to identify specific characteristics of the seller or marketplace by retrieving the tweets that were highly correlated with them using what is called the "inference" phase of BTM [44]. Within this subset of tweets, any false positives were first manually eliminated by two of our coauthors (TM and EK) by discarding those tweets whose content did not have a clear indication that the tweet was about the sale or promotion of a prescription opioid.

For example, as we narrowly focus on tweets purportedly offering online sale of opioids, only those tweets with hyperlinks contained in the message of the tweet or other contact information were considered. This produced a narrower group of tweets that contained hyperlinks to external websites or contract information for analysis. Hyperlinks were further manually coded to determine if the link was still active (ie, still redirecting to an external website vs a "dead" link that failed to redirect to a working website or produced an error code), whether it marketed or purported to directly sell an opioid product, and classified according to the nature and type of seller (eg, illicit online pharmacy, individual seller, or marketing affiliate) [24]. TM and EK coded for false positive tweets and the characteristics of links or websites independently and achieved a high intercoder reliability for results (k= 0.94). Discrepancies were resolved through reevaluation and consensus.

For all "signal tweets" that were specifically categorized as illegal online pharmacies, we also cross-referenced the URLs of these websites with LegitScript's external database that includes a legal classification. LegitScript legal classification is based on its own assessment of whether the website is (1) "rogue": vendor engaged in illegal, unsafe, or misleading activity, (2) "unapproved": vendor with a problem of regulatory compliance or risk in one or more jurisdictions, (3) "unverified": not subject to LegitScript review or monitoring, or (4) "legitimate": passed LegitScript certification criteria [34,45]. LegitScript classification queries offer another layer of verification regarding an online pharmacy's legal status and can help to confirm that the sites present high risk to consumers. We also reviewed WHOIS data to determine the internet protocol (IP) address and registered owner location for links classified as online pharmacies.

Figure 1. Summary of study methodology. API: application programming interface.



Results

Main Results

We collected data for approximately 20 days, resulting in a total of 213,041 tweets used for our analysis. The number of preconfigured themes was set to 100 during our BTM learning phase. During our first run of BTM on the entire dataset, we observed that a significant number of the word groupings were related to fentanyl, all of which were news-related themes (see examples in Table 2). This can be attributed to the high volume of news events surrounding illicit fentanyl before the time of the competition during the data collection phase. News essentially adds noise to the dataset (ie, it does not contain conversations related to online sales)—hence, we discarded fentanyl from the data and were left with approximately 117K tweets. Applying BTM on this smaller dataset gave clear signals of certain selling argument word groupings with words such as "buy," "online," "cheap," "free," "shipping," and the name of a prescription opioid. This demonstrates that fentanyl-related tweets added noise to the dataset by suppressing signal data (see examples in Table 3).

Table 2. Some example themes that were obtained from the first round of biterm topic model (BTM).

Example theme	Word groupings
Example theme—1	fentanyl, deaths, dangers, nations, drugs, china, learn, surge, deals
Example theme—2	fentanyl, man, police, charges, faces, arrest, brockton, bust, Calgary
Example theme—3	fentanyl, overdose, suspected, teen, life, drugs, dead, death, Vancouver

Table 3. Some example themes that were obtained from the data after fentanyl-related tweets were removed (italics denotes relevant "signal" theme).

Example theme	Word groupings
Example theme—1	cannabis, legalization, respect, completely, opinions, varying, merits
Example theme—2	drug, war, opposition, benches, friends, conservative, political, dead, props
Example theme—3	*buy, online, pill, free, cheap, find, shipping, generic*

Signal Twitter Data Analysis

Using the inference phase of BTM, we retrieved tweets that were most correlated and second most correlated to selling argument word groupings 0.32% (692/213,041). Of the total 692 tweets retrieved, 68.8% (476/692) contained hyperlinks. After manually coding tweets associated with these hyperlinks, we removed all duplicates (eg, retweets and identical tweets) and also any tweets with dead links. From these tweets, 23 unique Twitter user accounts were identified, with one account generating over 100K total tweets (containing both relevant and nonrelevant content) in 1 year, a characteristic common to social bot or spam accounts, a component of data analysis important in determining the source of content and whether it is biased [46]. All the other accounts examined had approximately 1000 tweets/year or lower and appeared to be human operated Twitter accounts. These user accounts generated a total of 34 live tweets with hyperlinks (followed by an estimated 1800 Twitter users during the duration of the data collection period). These tweets were then classified into one of three online seller categories: (1) an online pharmacy (defined as a website that purports to be an online pharmacy storefront, operates an e-commerce shopping cart where products can be checked out, paid and shipped directly to a consumer); (2) an individual seller or drug dealer (defined as a user offering the direct sale of prescription opioids via email, phone, or other direct contact solicitation); and (3) a marketing affiliates (defined as a website that hosts links to other websites that directly sell controlled substances; see Multimedia Appendix 1). Of the live hyperlinks coded at the Code-a-Thon, 44% (15/34) were online pharmacies, 32% (11/34) were individual sellers, 21% (7/34) were marketing affiliates, and 1 was purportedly from the Twitter handle of the darknet site AlphaBay (operated via the Tor network) linking to a reddit community page. The authenticity of the final result is questionable given that AlphaBay was shut down by law enforcement officials in July 2017.

The first category of 15 coded tweets consisted of 10 distinct live hyperlinks to illicit online pharmacies (see select examples in Multimedia Appendix 1, example A). All of these websites sold opioids with "no prescription" in combination with other drug products including other controlled substances (eg, barbiturates, Xanax, Ketamine, fentanyl patch, and codeine cough syrup), nonopioid prescription drugs therapeutic classes (eg, injectable steroids; antidepressants; weight loss drugs including Sibutramine, which has been removed from several markets; hormones; contraception; and erectile dysfunction drugs), recreational drugs (eg, cannabis), and some sites that also sold illicit drugs (eg, bath salts, MDMA, cocaine, heroin, and methamphetamines). One site purportedly accepted payment via cryptocurrencies bitcoin and ether (the cryptocurrency for the blockchain platform Ethereum). When cross-referencing for LegitScript status, 40% (6/15) were identified as "rogue," with the remaining having no information available (ie, likely not detected or included in LegitScript's database). Some online pharmacy tweets also used interesting selling arguments, including advertising Black Friday sales in reference to the day after Thanksgiving holiday in the United States, when retail discount sales are heavily marketed. Interestingly, when examining purported geographic location of IP addresses and website registered owners, the majority reported addresses in the United States.

The second category of 11 tweets appeared to originate from individual sellers using Twitter accounts that advertised the direct sale of opioids using hashtags (eg, #buy, #sell, #buypainmeds, #drugsforsale, #opioids, #painmeds, and #[controlled substance names]) and generally included contact information for the seller including an email address (in most cases a Google gmail account) and a phone number to call, text, or contact via the messaging application WhatsApp (phone numbers were mostly US 10-digit phone numbers, though phone numbers for UK country code +44 were also observed, indicating that these individuals may be located in these countries; see Multimedia Appendix 1, example B.) These tweets generally included pictures of prescription opioids and other drugs shown in an individual's hand or displayed next to a pill bottle. Some Twitter individual accounts also linked to an external webpage (including WordPress domains), blogs, or online classified ad services. The sellers generally included in their tweets or external webpages a list of all drugs they offered to sell, often comprising a mix of controlled substances, other prescription drugs, and illicit drugs.

The third category consisted of seven tweets that included links to marketing affiliate websites and networks that were hosted on their own distinct domains or used other blog sites (eg, blogspot). These sites did not directly sell opioids but included information on how to purchase illegally from other sites and

hosted hyperlinks that redirected traffic directly to an online pharmacy engaged in that activity (see Multimedia Appendix 1, example C). One of these sites included a frequently asked questions page that stated, "No, we are not sell any pills or medication. We are only provide medical information" and included several Web banners or advertisements to illicit online sellers, including one that used a fake banner claiming it was FDA approved.

Web Application Wireframe Prototype Solution

In parallel with our Twitter data collection and analysis and to better demonstrate the potential real-world application of our BTM methodology to detect, classify, and report online pharmacies marketing the sale of controlled substances, we developed a wire frame solution using the prototyping tool for Web and mobile apps, Justinmind [47]. We based our conceptual model of our Web app solution on simplicity, streamlined navigation, and user interaction with a task focus. The Web app wireframe was designed to implement three primary functions: (1) data collection and detection of tweets marketing the sale of opioids online using BTM assisted by human interpretation, (2) HTML classification of websites to determine if they are illicit online pharmacies, and (3) automating a script to report to the FDA and DEA about detection of illicit online pharmacy for further regulatory action (see Figure 2 with screenshots of the justinmind wireframe solution developed and as presented during the Code-a-Thon). Targeted end users for the wireframe solution, which could operate as a hosted Web application or software, include pharmaceutical manufacturers, government agencies and regulators (FDA, DEA, and state Boards of Pharmacy), nongovernmental groups (NABP), consumer advocacy groups (the Alliance for Safe Online Pharmacies, ASOP, and the Center for Safe Internet Pharmacies), and potentially a modified version for the public to report suspect websites.

The first component of this solution included a webpage with a query function that included a date range for data collection and the ability to enter prescription opioid keywords for filtering of tweets from the public streaming API. The second page exports tweets filtered for selected keywords and then specifically outputs signal tweets highly correlated with selling arguments associated with online pharmacies. This step could also include an output of Twitter handles or accounts that have interacted with this high-risk content (eg, Twitter users that are followers, have retweeted, or favorited identified signal tweets) for possible targeted countermarketing, health communication regarding potential risk, and also generate data for potential social network analysis. The third page would then output screenshots of hyperlinks associated with signal tweets for human inspection, while also reporting the IP geographic location of the website (using WHOIS data) and the LegitScript legal classification of the URL if available (LegitScript operates a fee-based API for their information). This page would also include a "report" button under each website screen capture. User selection of violating websites would generate a final page detailing what URLs had been reported to the FDA and DEA using an automated script that would fill in the necessary information fields on the FDA's "Reporting Unlawful Sales of Medical Products on the internet" webpage [48] and the DEA's "Report Submission Form for Suspected Unlawful Sales of Pharmaceutical Drugs on the internet [49]. This prototype was presented during our finalist presentation at the Code-a-Thon, but because of time considerations, was only a click-thru but not fully functioning demo. However, illicit online pharmacy website results generated in this study were reported to both the DEA and FDA by manually filling out the online reporting tools. Once this information is reported, DEA and FDA exercise their own discretion on how to pursue enforcement against potential violations of federal law.

Figure 2. Design elements and screenshots of prototype wireframe solution.

Discussion

Principal Findings

Similar to prior studies that have examined the use of Twitter by illicit online sellers to market and sell prescription opioids, our primary finding indicates that the overall volume of tweets directly engaged in this illegal activity is relatively low compared with the entire corpus of tweets collected that contained common opioid substance abuse product keywords [34,35,37,43]. This indicates that though this activity presents clear risk to the consumer and could contribute to diversion of opioids via convenient online access, the occurrence of these tweets are not widespread compared with other more commonly occurring conversations, including news reports about the opioid crisis, tweets about opioid abuse behavior self-reported by users, and other content that includes opioid keywords but which we would classify as noise per the aims of this study. However, some interesting results emanating from this study also provide us with key information on emerging trends regarding social media–enabled opioid abuse and methodological considerations when attempting to identify this type of content.

First, a major strength of this study compared with prior studies that conducted analysis on larger volumes of data was the short duration and immediate processing of the data collection and analysis (data collection was over a period of 2 weeks and immediately following it was the analysis over a 48-hour period). More specifically, previous studies involved a data collection and analysis process that was separated by more than 1 year [35,37,43,50]. By collecting and analyzing data immediately, fewer dead links where URLs were no longer active or website landing pages were no longer being used by owners were detected. Hence, our study demonstrates that a near real-time infoveillance approach, which collects and analyzes data over a short duration (eg, less than 30 days) can result in better detection and higher quality data related to illegal activities of online pharmacies and sellers. This approach could inure benefits to regulators and law enforcement officials, as Twitter accounts, illegal websites, and marketing affiliates could have their content taken down more quickly to mitigate potential exposure to consumers and accompanying patient safety risk. This would allow for proactive detection, particularly important given that illicit online pharmacies do not maintain a consistent presence on the internet and often change URLs frequently [24,51,52].

Immediate data coding and analysis also allowed us to detect our second category of tweets characterized as individuals who market and sell prescription opioids via open solicitations that had not been detected in prior studies. These digital drug dealers openly tweet that they can sell prescription opioids and other illicit substances directly to the public and that they can be contacted through a simple email or phone number. Oftentimes, they purport to validate availability of drugs offered by including a picture of their products included in the tweet and also use hashtags (#) to curate and target their marketing messages. These digital sellers sometimes represented themselves as individuals and in other cases as a company (primarily a name that represented itself as a pharmacy or pharmaceutical company). These results are alarming as they represent a potential new strategy where traditional “street” dealers may simply use Twitter in an effort to broadly market and extend their services to a wider and more diverse customer base that they would otherwise not have access to. The range of followers for these accounts varied, with observed ranges as few as 24 followers to as high as 989 followers, though generally these accounts were observed to have less than 100 followers. At the time of this writing, many of these accounts were still active on Twitter.

Additionally, a tweet we detected linked to a reddit community associated with the now defunct AlphaBay dark Web storefront leads us to unanswered questions and areas for future research. Specifically, this tweet raises the question of examining the interaction between social media, the open or clear Web accessible to the public, the dark Web accessed by clients such as tor or Onion, and illegal access to opioids. Other studies examining popular dark Web store fronts have shown that controlled substances and illicit drugs are actively sold on these sites to both consumers and potentially at wholesale to other distributors or dealers [53-55]. Hence, future studies should expressly examine the interaction between these different online ecosystems to determine their role in drug diversion and consumer purchasing.

Finally, we note that in addition to certain limitations, this study has some inherent weaknesses given that it examined a single social media platform: Twitter. According to the Pew Research Center, Twitter ranks at only #7 out of 8 popular social media platforms, with approximately 24% of US adult user share (compared with #1 YouTube at 73% and #2 Facebook at 68%), though use increases to 45% among users aged 18 to 24 year [56]. Hence, to get a more complete view of how controlled substances are marketed and sold to consumers online, future studies adopting machine-learning approaches used here for Twitter combined with the use of other methods, including deep learning (primarily for image recognition) and multiple modalities (such as simultaneously looking at additional data features), will need to be tailored to the types of data and user interactions occurring on other social media platforms (such as Instagram and Facebook) [57]. Additionally, lessons from prior studies that have examined how the internet has been used to circumvent the ban on purchasing illicit substances and how it has been used by minors to purchase alcohol and tobacco will also be informative to future technology-based and regulatory efforts against this illegal online opioid activity [58-63].

Limitations

There are certain limitations associated with the study results generated and as reported in the Code-a-thon. Specifically, our sample of tweets was filtered for a select group of commonly abused prescription opioid drugs using their International Nonproprietary Names and brand names. For example, we did not collect street or slang names of these drugs (eg, oxy, roxies, percs, and vikes), as our prior studies that have examined both types of keywords indicated that illicit online pharmacies do not use these nonspecific words for marketing or sales, and most uses of street drug names are user behavioral–related [43]. However, further confirmatory analysis is needed to validate that online pharmacies consistently do not use street name terms

for selling purposes. Additionally, we first cleaned the dataset before analysis to exclude non-English language tweets, which may further limit the generalizability of our sample of tweets containing opioid keywords. Furthermore, there are other strategies outside of topic modeling that can be used to learn about topical dynamics and trends in social media data, such as examining word and hashtag co-occurrences [42,64]. We chose not to use hashtag analysis as only one percent of all tweets in our dataset had hashtags in the content and less than one percent of signal tweets relevant to illicit online pharmacies used hashtags. For signal data that either included a direct hyperlink to an online pharmacy or marketing affiliates that redirected to live websites, content analysis was reviewed at a specific point of time after the tweets were collected and analyzed using BTM. Though tweets were coded right after the data collection was completed, it is possible that the content residing on hyperlinked content and/or the online pharmacy's website or domain may have changed from the exact date of data collection as websites often update and change content. Additionally, our exclusion of "fentanyl" from our analysis may have removed content related to the illicit online sale of fentanyls via online pharmacies, user forums, and individual drug sellers but was necessary to further refine our results to detect signal tweets with clear selling argument word topic groups. A recent study by the first and second author conducted on an older 2015 fentanyl Twitter dataset can inform future analysis examining this drug class specifically [37]. The validity of WHOIS geographic data for online pharmacies reviewed is also unclear. Though many online pharmacies listed an IP address or registered owner address in the United States, their actual server location and/or physical business location or registration could be falsely entered or masked by a privacy internet service provider company. Finally, we note that the RHA and DEA explicitly have rules and regulations that make it illegal to purchase controlled substances online and have no clarifying guidance or exemptions in relation to conducting test purchases for research purposes. As researchers reside in US jurisdiction, we were unable to actually purchase controlled substances from online pharmacies and test them for authenticity.

Conclusions

This study and our participation in the HHS Code-a-Thon established the viability of an important methodology to detect illicit online sales of opioids that are marketed via Twitter. Importantly, the machine-learning approach, which represents the core technology for our proposed solution and proof-of-concept deployed during the Code-a-Thon, is scalable and can be done relatively quickly following big data collection. This allows us to more rapidly detect illicit online sellers and classify their marketing characteristics. In fact, the vast majority of online sellers detected in this study remain active on social media and the Web at the time of this writing. Though the machine-learning component of this study is relatively mature, with testing of this algorithm and approach now published in four separate studies, the translation of the innovation to an easy-to-use, accessible, and largely automated solution is still at an early stage [34,35,37]. Though our prototype wireframe demonstrates the potential extension of the BTM machine-learning algorithm into a Web application that could be used by key stakeholders such as the FDA and DEA, it does not have the functionality or integration of different data sources to be considered a minimally viable product (MVP). To take this next step, funding with the primary aim of translating this research into MVP phase and eventual production and scale-up is needed (such as our recent award of an NIH NIDA 2017 "Start a SUD Startup" challenge, which provides small awards for startups related to substance abuse disorders with the aim of transitioning companies to a successful NIDA Small Business Innovation Research grant; see [65] for a recorded video discussing Code-a-Thon solution submitted as part of NIH NIDA 2017 "Start a SUD Startup" challenge grant). Additionally, we would need to automate data collection and backend analysis of Twitter data with integration via a Web application or software, while also developing solutions using natural processing language to automate classification of hyperlinks suspected as engaged in the sale of prescription opioids, techniques that have been explored in prior studies [23,66,67]. Finally, automated scripts that generate information needed for standardized reporting of results to the FDA and DEA via their online Web forms would also need to be developed. Despite these challenges, results from this study are useful and can inform regulators, law enforcement, public health officials, and the public about current and changing trends regarding supply, access, and distribution of illicit opioids. Technology, such as the big data and machine-learning approaches used in this study, will be critical components of any strategy to combat the opioid epidemic, an approach that HHS through its Code-a-Thon has begun to catalyze.

Acknowledgments

The authors would like to thank ASOP for their support of prior work associated with developing a methodology to detect and characterize illicit online pharmacies using social media platforms. The authors also would like to thank the Robert Wood Johnson Foundation New Connections Program for financial support associated with publication fees (paper processing fees) to publish this study.

TM received funding for a pilot grant unrelated to this study from ASOP, a 501(c)(4) social welfare organization engaged on the issue of illicit online pharmacies. TM also received travel support from the FDA for an invited accredited educational activity on drug safety and surveillance, which included discussion regarding the results of this study.

Authors' Contributions

JK conducted data collection for the study. All authors contributed to the formulation, analysis, drafting, completion, and approval of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of illicit online pharmacies, individual sellers, and marketing affiliates (examples A, B, and C).

[PNG File, 4MB - jmir_v20i4e10029_app1.png]

References

1. Rudd RA, Seth P, David F, Scholl L. Increases in drug and opioid-involved overdose deaths - United States, 2010-2015. MMWR Morb Mortal Wkly Rep 2016 Dec 30;65(5051):1445-1452 [FREE Full text] [doi: 10.15585/mmwr.mm655051e1] [Medline: 28033313]
2. Green TC, Gilbert M. Counterfeit medications and fentanyl. JAMA Intern Med 2016 Oct 01;176(10):1555-1557. [doi: 10.1001/jamainternmed.2016.4310] [Medline: 27533891]
3. Frank RG, Pollack HA. Addressing the fentanyl threat to public health. N Engl J Med 2017 Feb 16;376(7):605-607. [doi: 10.1056/NEJMp1615145] [Medline: 28199808]
4. Drug Enforcement Agency. DEA. Counterfeit prescription pills containing fentanyls: a global threat URL: https://www.dea.gov/docs/Counterfeit%20Prescription%20Pills.pdf [accessed 2018-02-05] [WebCite Cache ID 6x0pOPV8Z]
5. Haffajee RL, Mello MM. Drug companies' liability for the opioid epidemic. N Engl J Med 2017;377(24):2305. [Medline: 29236640]
6. Bonnie R, Kesselheim A, Clark D. Both urgency and balance needed in addressing opioid epidemic: a report from the National Academies of Sciences, Engineering, and Medicine. J Am Med Assoc 2017 Aug 01;318(5):423-424. [doi: 10.1001/jama.2017.10046] [Medline: 28715551]
7. Florence CS, Zhou C, Luo F, Xu L. The economic burden of prescription opioid overdose, abuse, and dependence in the United States, 2013. Med Care 2016 Dec;54(10):901-906. [doi: 10.1097/MLR.0000000000000625] [Medline: 27623005]
8. Rutkow L, Vernick JS. Emergency legal authority and the opioid crisis. N Engl J Med 2017 Dec 28;377(26):2512-2514. [doi: 10.1056/NEJMp1710862] [Medline: 29140760]
9. Centers for Disease Control and Prevention. CDC. Understanding the Epidemic: Drug Overdose URL: http://www.cdc.gov/drugoverdose/epidemic/ [accessed 2018-02-05] [WebCite Cache ID 6x0pPLyld]
10. Felter C. Council on Foreign Relations. The US Opioid Epidemic URL: https://www.cfr.org/backgrounder/us-opioid-epidemic [accessed 2018-02-05] [WebCite Cache ID 6x0pNgqqL]
11. Doyon S, Aks SE, Schaeffer S. Expanding access to naloxone in the United States. J Med Toxicol 2014 Dec;10(4):431-434 [FREE Full text] [doi: 10.1007/s13181-014-0432-1] [Medline: 25316516]
12. Barnett ML, Gray J, Zink A, Jena AB. Coupling policymaking with evaluation - the case of the opioid crisis. N Engl J Med 2017 Dec 14;377(24):2306-2309. [doi: 10.1056/NEJMp1710014] [Medline: 29236636]
13. Volkow ND, McLellan AT. Opioid abuse in chronic pain--misconceptions and mitigation strategies. N Engl J Med 2016 Mar 31;374(13):1253-1263. [doi: 10.1056/NEJMra1507771] [Medline: 27028915]
14. National Institute on Drug Abuse. 2018 Jan 8. Opioid Overdose Crisis URL: https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis [accessed 2018-02-05] [WebCite Cache ID 6x0pMVg19]
15. Volkow ND, Collins FS. The role of science in addressing the opioid crisis. N Engl J Med 2017 Jul 27;377(4):391-394. [doi: 10.1056/NEJMsr1706626] [Medline: 28564549]
16. US Department of Health & Human Services. HHS. HHS Opioid Code-a-Thon URL: https://www.hhs.gov/challenges/code-a-thon/index.html [accessed 2018-02-05] [WebCite Cache ID 6x0pQgefy]
17. YouTube. HHS Code-a-Thon finalist presentations URL: https://www.youtube.com/watch?v=7B7jWbsktC0 [accessed 2018-04-21] [WebCite Cache ID 6yq3YEHlM]
18. Mackey TK, Liang BA, Strathdee SA. Digital social media, youth, and nonmedical use of prescription drugs: the need for reform. J Med Internet Res 2013 Jul 26;15(7):e143 [FREE Full text] [doi: 10.2196/jmir.2464] [Medline: 23892156]
19. Liang BA, Mackey T. Searching for safety: addressing search engine, website, and provider accountability for illicit online drug sales. Am J Law Med 2009;35(1):125-184. [Medline: 19534258]
20. Perrin A. Pew Internet. Social Media Usage: 2005-2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [accessed 2018-02-05] [WebCite Cache ID 6x0pLsRwl]
21. US Food and Drug Administration. FDA. Consumer Updates - Buying Medicines Online? Be Wary URL: https://www.fda.gov/Drugs/ResourcesForYou/Consumers/BuyingUsingMedicineSafely/BuyingMedicinesOvertheInternet/BeSafeRxKnowYourOnlinePharmacy/default.htm [accessed 2018-02-05] [WebCite Cache ID 6x0pKuhhe]

22. LegitScript. SafeMedsOnline. 2016 Jan 14. The Internet Pharmacy-Market in 2016 URL: http://safemedsonline.org/wp-content/uploads/2016/01/The-Internet-Pharmacy-Market-in-2016.pdf [accessed 2018-04-19] [WebCite Cache ID 6ynpzLNml]
23. Mackey TK, Nayyar G. A review of existing and emerging digital technologies to combat the global trade in fake medicines. Expert Opin Drug Saf 2017 May;16(5):587-602. [doi: 10.1080/14740338.2017.1313227] [Medline: 28349715]
24. Mackey TK, Nayyar G. Digital danger: a review of the global public health, patient safety and cybersecurity threats posed by illicit online pharmacies. Br Med Bull 2016 Jun;118(1):110-126 [FREE Full text] [doi: 10.1093/bmb/ldw016] [Medline: 27151957]
25. The National Association of Boards of Pharmacy. NABP. Internet Drug Outlet Identification Program: Progress Report for State and Federal Regulators URL: https://nabp.pharmacy/wp-content/uploads/2016/08/Internet-Drug-Outlet-Report-August-2017.pdf [accessed 2018-02-05] [WebCite Cache ID 6x0puZ0h1]
26. Clark F. Rise in online pharmacies sees counterfeit drugs go global. Lancet 2015 Oct 03;386(10001):1327-1328. [doi: 10.1016/S0140-6736(15)00394-3] [Medline: 26460765]
27. United States General Accounting Office. GAO. Internet Pharmacies: Some Pose Safety Risks for Consumers URL: https://www.gao.gov/new.items/d04820.pdf [accessed 2018-02-05] [WebCite Cache ID 6x0pGD09U]
28. United States General Accounting Office. GAO. Internet Pharmacies: Federal Agencies and States Face Challenges Combating Rogue Sites, Particularly Those Abroad URL: https://www.gao.gov/assets/660/655751.pdf [accessed 2018-02-05] [WebCite Cache ID 6x0pFS7tN]
29. Nixon R. The New York Times. Online Sales of Illegal Opioids from China Surge in US URL: https://www.nytimes.com/2018/01/24/us/politics/senate-investigation-china-mail-opioids.html [accessed 2018-02-05] [WebCite Cache ID 6x0pEZEeA]
30. Boyer EW, Wines Jr JD. Impact of Internet pharmacy regulation on opioid analgesic availability. J Stud Alcohol Drugs 2008 Sep;69(5):703-708 [FREE Full text] [Medline: 18781245]
31. Forman RF. Availability of opioids on the Internet. J Am Med Assoc 2003 Aug 20;290(7):889. [doi: 10.1001/jama.290.7.889] [Medline: 12928464]
32. Forman RF, Woody GE, McLellan T, Lynch KG. The availability of web sites offering to sell opioid medications without prescriptions. Am J Psychiatry 2006 Jul;163(7):1233-1238. [doi: 10.1176/appi.ajp.163.7.1233] [Medline: 16816229]
33. Forman RF, Block LG. The marketing of opioid medications without prescription over the Internet. J Public Policy Mark 2006 Sep;25(2):133-146. [doi: 10.1509/jppm.25.2.133]
34. Katsuki T, Mackey TK, Cuomo R. Establishing a link between prescription drug abuse and illicit online pharmacies: analysis of Twitter data. J Med Internet Res 2015 Dec 16;17(12):e280 [FREE Full text] [doi: 10.2196/jmir.5144] [Medline: 26677966]
35. Mackey TK, Kalyanam J, Katsuki T, Lanckriet G. Twitter-Based Detection of Illegal Online Sale of Prescription Opioid. Am J Public Health 2017 Dec;107(12):1910-1915. [doi: 10.2105/AJPH.2017.303994] [Medline: 29048960]
36. Mackey TK, Liang BA. Global reach of direct-to-consumer advertising using social media for illicit online drug sales. J Med Internet Res 2013 May 29;15(5):e105 [FREE Full text] [doi: 10.2196/jmir.2610] [Medline: 23718965]
37. Mackey TK, Kalyanam J. Detection of illicit online sales of fentanyls via Twitter. F1000Res 2017;6:1937 [FREE Full text] [doi: 10.12688/f1000research.12914.1] [Medline: 29259769]
38. The United States Department of Justice. Google Forfeits $500 Million Generated by Online Ads & Prescription Drug Sales by Canadian Online Pharmacies URL: https://www.justice.gov/opa/pr/google-forfeits-500-million-generated-online-ads-prescription-drug-sales-canadian-online [accessed 2018-02-05] [WebCite Cache ID 6x0p7SWtm]
39. Abbasi J. Emergency department opioid misuse diagnoses increasing in adolescents and young adults. J Am Med Assoc 2017 Dec 26;318(24):2416-2417. [doi: 10.1001/jama.2017.16586] [Medline: 29214309]
40. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social media use and access to digital technology in US young adults in 2016. J Med Internet Res 2017 Jun 07;19(6):e196 [FREE Full text] [doi: 10.2196/jmir.7303] [Medline: 28592394]
41. US Senate. Colleagues Urge Tech Companies to Clamp Down on Illegal Online Drug Sales and Advertising URL: https://www.grassley.senate.gov/news/news-releases/grassley-feinstein-colleagues-urge-tech-companies-clamp-down-illegal-online-drug [accessed 2018-04-19] [WebCite Cache ID 6ynqGbVPs]
42. Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98-2017 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
43. Kalyanam J, Katsuki T, R G Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav 2017 Feb;65:289-295. [doi: 10.1016/j.addbeh.2016.08.019] [Medline: 27568339]
44. Orizio G, Rubinelli S, Schulz PJ, Domenighini S, Bressanelli M, Caimi L, et al. “Save 30% if you buy today”. Online pharmacies and the enhancement of peripheral thinking in consumers. Pharmacoepidemiol Drug Saf 2010 Sep;19(9):970-976. [doi: 10.1002/pds.2007] [Medline: 20652863]

45. Mackey TK, Aung P, Liang BA. Illicit Internet availability of drugs subject to recall and patient safety consequences. Int J Clin Pharm 2015 Dec;37(6):1076-1085. [doi: 10.1007/s11096-015-0154-8] [Medline: 26148859]
46. Lee K, Caverlee J, Webb S. The social honeypot project: protecting online communities from spammers. New York, New York, USA: ACM; 2010 Presented at: Proceedings of the 19th international conference on World Wide Web WWW '10; April 26-30, 2010; Raleigh, North Carolina, USA.
47. JustinMind. URL: https://www.justinmind.com/ [accessed 2018-04-21] [WebCite Cache ID 6yq2QkIxk]
48. US Food and Drug Administration. Reporting Unlawful Sales of Medical Products on the Internet URL: https://www.fda.gov/Safety/ReportaProblem/ucm059315.htm [accessed 2018-04-21] [WebCite Cache ID 6yq2dkHb5]
49. US Department of Justice, Drug Enforcement Administration, Diversion Control Division. Report Submission Form for Suspected Unlawful Sales of Pharmaceutical Drugs on the Internet URL: https://apps.deadiversion.usdoj.gov/webforms/jsp/umpire/umpireForm.jsp [accessed 2018-04-21] [WebCite Cache ID 6yq2ul7XP]
50. Kalyanam J, Mackey TK. Detection and Characterization of Illegal Marketing and Promotion of Prescription Drugs on Twitter. 2017 Presented at: NIPS 2017 Workshop on Machine Learning for Health (ML4H); December 8, 2017; Long Beach, CA p. 1-5 URL: https://arxiv.org/abs/1712.00507v1
51. Fittler A, Bősze G, Botz L. Evaluating aspects of online medication safety in long-term follow-up of 136 Internet pharmacies: illegal rogue online pharmacies flourish and are long-lived. J Med Internet Res 2013 Sep 10;15(9):e199 [FREE Full text] [doi: 10.2196/jmir.2606] [Medline: 24021777]
52. Orizio G, Merla A, Schulz PJ, Gelatti U. Quality of online pharmacies and websites selling prescription drugs: a systematic review. J Med Internet Res 2011 Sep 30;13(3):e74 [FREE Full text] [doi: 10.2196/jmir.1795] [Medline: 21965220]
53. Van Hout MC, Hearne E. New psychoactive substances (NPS) on cryptomarket fora: an exploratory study of characteristics of forum activity between NPS buyers and vendors. Int J Drug Policy 2017 Dec;40:102-110. [doi: 10.1016/j.drugpo.2016.11.007] [Medline: 28027812]
54. Broséus J, Rhumorbarbe D, Mireault C, Ouellette V, Crispino F, Décary-Hétu D. Studying illicit drug trafficking on Darknet markets: structure and organisation from a Canadian perspective. Forensic Sci Int 2016 Jul;264:7-14. [doi: 10.1016/j.forsciint.2016.02.045] [Medline: 26978791]
55. Pergolizzi Jr JV, LeQuang JA, Taylor Jr R, Raffa RB, NEMA Research Group. The “Darknet”: The new street for street drugs. J Clin Pharm Ther 2017 Dec;42(6):790-792. [doi: 10.1111/jcpt.12628] [Medline: 28921578]
56. Smith A, Anderson M. Pew Internet. Social Media Use in 2018 URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2018-04-17] [WebCite Cache ID 6ykw0G0Gc]
57. Kalyanam J, Mackey TK. A review of digital surveillance methods and approaches to combat prescription drug abuse. Curr Addict Rep 2017 Sep 18;4(4):397-409. [doi: 10.1007/s40429-017-0169-4]
58. Mackey TK, Miner A, Cuomo RE. Exploring the e-cigarette e-commerce marketplace: identifying Internet e-cigarette marketing characteristics and regulatory gaps. Drug Alcohol Depend 2015 Nov 01;156:97-103. [doi: 10.1016/j.drugalcdep.2015.08.032] [Medline: 26431794]
59. Allem J, Ayers JW, Althouse BM, Williams R. When a ban really is not a ban: internet loopholes and Djarum flavoured cigarettes in the USA. Tob Control 2016 Dec;25(4):489-490 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052309] [Medline: 26085123]
60. Williams RS, Derrick J, Phillips KJ. Cigarette sales to minors via the internet: how the story has changed in the wake of federal regulation. Tob Control 2017 Jul;26(4):415-420. [doi: 10.1136/tobaccocontrol-2015-052844] [Medline: 27413060]
61. Williams R, Derrick J, Ribisl K. Electronic cigarette sales to minors via the internet. JAMA Pediatr 2015 Mar;169(3):e1563 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.63] [Medline: 25730697]
62. Williams RS, Schmidt A. The sales and marketing practices of English-language internet alcohol vendors. Addiction 2014 Mar;109(3):432-439. [doi: 10.1111/add.12411] [Medline: 24219065]
63. Williams RS, Ribisl KM. Internet alcohol sales to minors. Arch Pediatr Adolesc Med 2012 Sep 01;166(9):808-813. [doi: 10.1001/archpediatrics.2012.265] [Medline: 22566520]
64. Kalyanam J, Mantrach A, Saez-Trumper D, Vahabi H, Lanckriet G. Leveraging social context for modeling topic evolution. In: KDD '15. New York, New York, USA: ACM; 2015 Presented at: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '15; August 10-13, 2015; Sydney, NSW, Australia p. 517-526.
65. YouTube. SUD Challenge URL: https://www.youtube.com/watch?v=XZaOqgHVn3Q [accessed 2018-04-21] [WebCite Cache ID 6yq3qZSIC]
66. Corona I, Contini M, Ariu D, Giacinto G, Roli F, Lund M. PharmaGuard: Automatic identification of illegal search-indexed online pharmacies. : IEEE; 2015 Presented at: 2015 IEEE 2nd International Conference on Cybernetics (CYBCONF); June 24-26, 2015; Gdynia, Poland p. 324-329. [doi: 10.1109/CYBConf.2015.7175954]
67. Simpson SS, Adams N, Brugman CM, Conners TJ. Detecting novel and emerging drug terms using natural language processing: a social media corpus study. JMIR Public Health Surveill 2018 Jan 08;4(1):e2 [FREE Full text] [doi: 10.2196/publichealth.7726] [Medline: 29311050]

Abbreviations

API: application programming interface
ASOP: Alliance for Safe Online Pharmacies
BTM: biterm topic model
CDC: Centers for Disease Control and Prevention
DEA: Drug Enforcement Agency
FDA: Food and Drug Administration
GAO: Government Accountability Office
HHS: US Department of Health and Human Services
IP: internet protocol
MVP: minimally viable product
NABP: National Association of Boards of Pharmacy
RHA: Ryan Haight Online Pharmacy Consumer Protection Act

Edited by G Eysenbach; submitted 05.02.18; peer-reviewed by E Ferrara, A Fittler, JP Allem; comments to author 22.02.18; revised version received 27.03.18; accepted 28.03.18; published 27.04.18.

Please cite as:
Mackey T, Kalyanam J, Klugman J, Kuzmenko E, Gupta R
Solution to Detect, Classify, and Report Illicit Online Marketing and Sales of Controlled Substances via Twitter: Using Machine Learning and Web Forensics to Combat Digital Opioid Access
J Med Internet Res 2018;20(4):e10029
URL: http://www.jmir.org/2018/4/e10029/
doi:10.2196/10029
PMID:29613851

Original Paper

Characterization of Patient Interest in Provider-Based Consumer Health Information Technology: Survey Study

Joseph Featherall[1], BS; Brittany Lapin[2,3], PhD; Alexander Chaitoff[1], MPH; Sonia A Havele[4], BA; Nicolas Thompson[2,3], MS; Irene Katzan[1,2], MS, MD

[1]Cleveland Clinic Lerner College of Medicine, Cleveland Clinic, Cleveland, OH, United States

[2]Neurological Institute Center for Outcomes Research & Evaluation, Cleveland Clinic, Cleveland, OH, United States

[3]Lerner Research Institute, Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, United States

[4]School of Medicine, Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Irene Katzan, MS, MD
Neurological Institute Center for Outcomes Research & Evaluation
Cleveland Clinic
Mail Code S80
9500 Euclid Avenue
Cleveland, OH, 44195
United States
Phone: 1 216 636 5860
Fax: 1 216 636 2061
Email: katzani@ccf.org

Abstract

Background: Consumer health information technology can improve patient engagement in their health care and assist in navigating the complexities of health care delivery. However, the consumer health information technology offerings of health systems are often driven by provider rather than patient perspectives and inadequately address patient needs, thus limiting their adoption by patients. Consideration given to patients as stakeholders in the development of such technologies may improve adoption, efficacy, and consumer health information technology resource allocation.

Objective: The aims of this paper were to measure patient interest in different health system consumer health information technology apps and determine the influence of patient characteristics on consumer health information technology interest.

Methods: Patients seen at the Cleveland Clinic Neurological Institute were electronically surveyed on their interest in using different consumer health information technology apps. A self-efficacy scale, Patient Health Questionnaire-9 depression screen, and EuroQol 5 dimensions health-related quality of life scale were also completed by patients. Logistic regression was used to determine the influence of patient characteristics on interest in consumer health information technology in the categories of self-management, education, and communication.

Results: The majority of 3852 patient respondents had an interest in all technology categories assessed in the survey. The highest interest was in apps that allow patients to ask questions of providers (3476/3852, 90.24%) and to schedule appointments (3211/3839, 83.64%). Patient interest in consumer health information technology was significantly associated with greater depression symptoms, worse quality of life, greater health self-efficacy, and smartphone ownership (P<.001 for all listed).

Conclusions: Patients should be viewed as active stakeholders in consumer health information technology development and their perspectives should consistently guide development efforts. Health systems should consider focusing on consumer health information technologies that assist patients in scheduling appointments and asking questions of providers. Patients with depression should also be considered for targeted consumer health information technology implementation. Health self-efficacy is a valid predictor of consumer health information technology interest and may play a role in the utilization of consumer health information technologies. Health systems, broadly, should put forth greater effort to understand the needs and interests of patients in the consumer health information technology development process. Consumer health information technology design and implementation may be improved by understanding which technologies patients want.

(J Med Internet Res 2018;20(4):e128) doi:10.2196/jmir.7766

KEYWORDS

consumer health informatics; medical informatics; self efficacy; self-management; telemedicine; patient-centered care

Introduction

Background

Health care is undergoing enormous transformation in pursuit of achieving the triple aim: enhanced individual experience of care, improved health of populations, and reduced per capita costs of care [1,2]. Consumer health information technology (CHIT) has been defined in a number of ways [3-5] but has recently been well described by Tao et al as consumer-centered electronic tools, technologies, apps, or systems that are interacted with directly by health consumers to provide them with data, information, recommendations, or services for promotion of health and health care [6]. Such technologies are a key component of the patient-centered transformation of health care systems [5,7,8].

Implementation of CHIT is often driven by provider perspectives and available technologies rather than by the examined needs and motivations of users [5,9]. Of the largest 100 US hospitals, 66 provide mobile apps, but only 2% of patients at these hospitals are using them—an indication that health systems inadequately address the interests of patients in the CHIT design process [10]. Despite benefits in transparency and access associated with freely allowing patients to view their health data, health systems provide only limited data [11]. Focus groups have demonstrated that patients have innovative and useful perspectives as to how CHIT can improve health and ease patient workload [12-14]. Consumer perception of benefit, convenience, and integration into daily life is necessary for the successful implementation of CHIT and should be incorporated into the CHIT development process [4,15]. CHIT has been shown to improve health outcomes, but successful implementation requires that the needs of patients, that is, end users, are adequately met [3]. Overall, there is increasing recognition that health systems should systematically gather patient perspectives and use them to inform the design of CHIT [16].

The adoption of CHIT is influenced by a number of patient characteristics, but further investigation is necessary to fully characterize the CHIT needs of specific patient populations for optimum engagement and adoption [17]. For example, previous explorations have found patient portal adoption is associated with ethnic, educational, and cultural factors [18]; digital health is not used by the fastest-growing age segment of the US population, the elderly [19]; and patients who are sicker are more likely to search the internet for information and use this information during a provider visit [20]. However, these past studies have been limited in their scope, often exploring only how consumers react to already existing single CHIT solutions and not investigating the broad range of technology interests of patients. Further understanding of the technologies that patients want and the characteristics of these patients can prospectively guide strategies for providers and health systems investing in CHIT development.

Broadly, provider-based CHITs can be classified into one of 3 categories: patient-provider communications, education, and self-management technologies. Patient-provider communication technologies, including telecare, secure messaging, Web visits, and scheduling apps have demonstrated promise [21-25]. These apps improve patient convenience in accessing their health care teams. Recent data demonstrate that deficits in Web-based communication services may actually drive patients away from health systems [10]. Patient-centered prioritization of integrating communication technologies into the health system can be aided by quantifying and responding to the needs and interests of patients rather than basing technology development on data from historical technology adoption or qualitative data from small focus groups.

Electronic access to accurate patient educational resources is another established category of CHIT. Technologies that deliver educational programs to patients have demonstrated improvements in diverse outcomes, such as medication compliance [26,27], preprocedure anxiety (iPod-based modules) [28], postoperative perception of pain (educational website) [29], and plasma cholesterol (computer-based dietary workbook) [30]. During 2012, 59% of Americans searched for health-related information on the Web [31]; a smaller, more recent study indicates that internet health information seeking is global and likely increasing [32]. The use of electronic education tools in patient care varies according to patient access to the internet and patient awareness of digital education resources [33,34]. Knowledge of other patient factors that impact the use of educational resources can help design, development, and implementation strategies for education-related CHIT.

Self-management technologies encompass a third category of CHITs and have been shown to improve outcomes and reduce cost [35,36]. These include reminder systems, physical activity trackers, physiologic data (such as blood pressure or glucose) monitors, and apps that record self-reported health status over time. Despite the growing literature on this subject, limited information currently exists on how these technologies should be designed and which patients desire such programs [37-39]. Moreover, whether or not patients desire these technologies to be offered by their health care providers, as compared with third party CHIT offerings, has yet to be explored.

In addition to gaps in the literature with regard to specific patient characteristics associated with preference for each classification of CHIT, guidance on which types of CHIT apps are more likely to be adopted by patients is not well established [40]. Despite the "perceived utility" of a CHIT being a very reliable predictor of acceptance of technology, the utility of the technology to patients is addressed in very few studies [3]. There have been survey studies regarding the factors driving acceptance of CHIT; however, these studies investigate why patients choose to accept a single technology and did not ask, "In which technologies do patients have interest [39,41-43]?"

Objectives

The aims of this study are as follows: (1) to quantify patients' interests in CHIT by surveying patients seen in the ambulatory clinics of the Cleveland Clinic Neurological Institute and (2) to examine the influence of patient demographics, health-related quality of life, health self-efficacy, and depression on their interest in 3 categories of CHIT: patient-provider communication, educational resources, and self-management tools. This information can be used to better target the development and implementation of provider-based CHIT apps of greatest importance to patients.

Methods

Survey Design

The CHIT interest survey was designed to cover technologies that are common candidates for offerings by health systems. The technologies were grouped into 3 main categories: education, communication with health care providers, and self-management. Questions were developed by a multidisciplinary team of clinicians, researchers, and patients. Questions were evaluated by the team for content, ease of understanding, usefulness, and comprehensiveness in an iterative process and reviewed by the Neurological Institute Patient Advisory Committee for content and clarity [44]. The survey asked patients to rank their interest in various CHIT apps using a Likert scale with the following measurements: "not at all interested" (0), "not very interested" (1), "neutral" (3), "somewhat interested" (3), and "very interested" (4) (Multimedia Appendix 1). The survey was piloted, and response distributions were evaluated by the team to ensure no significant ceiling or floor effects. Internal reliability was assessed by Cronbach alpha.

To determine the impact of patient self-efficacy and current use of technology on interest in CHIT, the survey included a validated 5-item health self-efficacy questionnaire developed by Lee et al [45], as well as questions on current use of smartphones and tablet computers. The self-efficacy questionnaire included the following statements: (1) "I am confident I can have a positive effect on my health," (2) "I have set some definite goals to improve my health," (3) "I have been able to meet the goals I set for myself to improve my health," (4) "I am actively working to improve my health," and (5) "I feel that I am in control of how and what I learn about my health." Patients rated their level of agreement with these statements using a Likert scale with the following measurements: "disagree very much"(0), "disagree"(1), "neutral"(2), "agree"(3), and "agree very much"(4). The health self-efficacy score was calculated by summing the responses to the 5 questions; thus, scores of "0" and "20" indicate the lowest and highest possible health self-efficacy scores, respectively.

Patient Selection and Data Collection

Through the Knowledge Program patient-entered data collection system, patients complete the electronic questionnaires before their visits to the ambulatory clinics of the 3 pediatric and 14 adult condition–based centers of the Cleveland Clinic Neurological Institute [46]. These centers manage patients with neurological, neurosurgical, rehabilitation, and mental health conditions. Patients were given the option to complete the questionnaires through a MyChart patient portal (Epic, Epic Systems, Verona, Wisconsin) up to 5 days before their visits. Patients not completing questionnaires beforehand completed them using tablet computers in the clinic lobby at the time of check-in for their appointment. In addition to disease-specific instruments, all patients completed the EuroQol 5 Dimensions (EQ-5D) and Patient Health Questionnaire-9 (PHQ-9) as part of their regular questionnaire set. The EQ-5D is a commonly used 5-item generic scale that measures health-related quality of life [47]. Scores are transformed into utility weights derived from the general population and range from −0.109 (state worse than death) to 1.00 (best possible health). The PHQ-9 is a 9-item depression screen that ranges from 0 to 27, with higher scores indicating more depressive symptoms [48].

Between January 2013 and December 2015, patients seen in the 14 adult clinics who completed their questionnaires through MyChart were asked to complete the CHIT survey at the end of their regular questionnaire set. The CHIT survey was administered in each center until completed by at least 200 patients per center. As patient volume is variable across centers, a minimum response from 200 patients per center was established to ensure an adequate representation of patients. The survey was administered primarily through MyChart to avoid workflow disruptions at the time of the visit. To assess potential bias introduced by the survey delivery method, additional surveys were completed by patients seen in the Spine Center in the clinic lobby using tablet computers. Patient demographics and selected comorbid conditions (diabetes, atrial fibrillation, cancer, asthma, depression, congestive heart failure, coronary artery disease, chronic renal insufficiency, and hypertension) were identified from encounter diagnoses, problem list, or medical history sections of the electronic health record. Approximate household income was estimated based on 2010 census data by zip code. Patients 18 years of age and younger were excluded from the study. The study was approved by the Cleveland Clinic Institutional Review Board.

Statistical Analysis

Individual survey item responses were summarized using frequency count with percentage. Internal reliability of self-efficacy and the CHIT survey were assessed using Cronbach alpha. Patient characteristics are presented as mean with SD or median with interquartile range for continuous variables, and frequency count with percentage for categorical variables. For the purpose of developing a predictive model, potential uses of CHIT were organized into 3 categories: education (educational offerings and online discussion forum), communication with providers (booking appointments, asking provider questions, and entering medical history), and self-management (reminder systems, tracking physiologic data, tracking physical activity, recording lifestyle information, comparing health to others, and tracking health status information). Given the distribution of patient response, with very few patients indicating "not at all interested," high interest was defined by the multidisciplinary team as patients answering "very interested" to one or more questions within a category, whereas patients not responding

with at least one “very interested” in the category were defined as low interest. A sensitivity analysis was conducted to define high-interest in a category as a response of “very interested” or “somewhat interested” to every question within that category. Low-interest for the sensitivity analysis was defined as a response of “neutral,” “not very interested,” or “not at all interested” to every question within the category.

Logistic regression models were constructed to determine univariate predictors of interest in the 3 CHIT domains. Independent predictors of the 3 outcomes were assessed using multivariable logistic regression models. Due to issues of multicollinearity, an indicator variable for any comorbidity (excluding depression) was included in the multivariable models in lieu of specific comorbidities, and smartphone ownership was included over tablet ownership as determined a priori. All potential interactions were evaluated within the final models at $P<.05$.

A subset analysis was conducted to assess potential bias introduced by surveying patients via the MyChart patient medical record portal. The 3 CHIT domain outcomes as well as individual question mean Likert scores for Spine Center patients completing the survey via MyChart were compared with those of Spine Center patients completing the survey in the lobby of the clinic using chi-squared test and Mann-Whitney *U* test, respectively.

Additionally, univariate and multivariable logistic regression models for the 3 outcomes were constructed as described above within Spine Center patients, including the location of survey completion as a covariate, in order to test for completion location influence on patient interest. Statistical significance was established throughout at an alpha level of .05. All statistical analyses were conducted using SAS version 9.4 (SAS Institute Inc, Cary, NC).

Results

Patient Characteristics

The study cohort consisted of 3852 patients who completed a survey: 3735 in MyChart and 117 in the Spine Center lobby. The mean age of the study cohort was 57.0 (SD 14.8) years (Table 1). The majority of the cohort was white (3115/3353, 92.90%), female (2351/3852, 61.03%), and married (2204/3351, 65.77%). Respondents had an average of 2 comorbid conditions, with the most common being hypertension (1465/2615, 56.02%). The majority of patients reported owning a smartphone (2640/3799, 69.49%) and a tablet computer (2103/3809, 55.21%).

Table 1. Clinical characteristics of the study cohort (N=3852).

Demographics	Value
Age (years), mean (SD)	57.0 (14.8)
Female, n (%)	2351 (61.03)
Race (N=3353), n (%)	
White	3115 (92.90)
African American	196 (5.85)
Other	42 (1.25)
Marital status (N=3351), n (%)	
Married	2204 (65.77)
Single	748 (22.32)
Divorced/widowed	399 (11.91)
Household income (in US $)[a], N=3326, median (Q1-Q3)[b]	54,578 (44,371-66,360)
Any comorbid condition, N=2615, n (%)	2087 (79.81)
Number of comorbidities[c], median (Q1-Q3)	2 (1-3)
Health status, median (Q1-Q3)	
PHQ-9[d], N=3434	6 (2-11)
EQ-5D[e], N=3483	0.78 (0.60-0.83)
Health efficacy score, N=3791	15 (12-17)
Personal mobile devices, n (%)	
Smartphone ownership (N=3799)	2640 (69.49)
Tablet ownership (N=3809)	2103 (55.21)

[a]Median household income by zip code.

[b]Q1: first quartile, Q3: third quartile.

[c]Comorbid conditions include diabetes, atrial fibrillation, cancer, coronary artery disease, chronic renal insufficiency, and hypertension.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]EQ-5D: EuroQol 5 Dimensions.

Health Self-Efficacy

Internal reliability of this scale was high (Cronbach alpha=.845). The majority of patients responded affirmatively to all self-efficacy items, except the ability to meet health goals. Regarding meeting health goals, only 1663 out of 3815 (43.59%) respondents indicated that they agreed with the health efficacy statement. Regarding ability to have a positive effect on one's health, 3274 out of 3825 (85.59%) patients indicated perceived self-efficacy in this area (1597/3825, 41.75%, "Agree" and 1677/3825, 43.84%, "Agree very much").

Consumer Health Information Technology Interest

Internal reliability of the CHIT survey was good, with Cronbach alpha of .832 for education, .762 for self-management, and .898 for communication. The majority of patients indicated interest in all of the CHIT categories asked in the survey (Figure 1). Patients expressed the greatest interest in the communication category, specifically booking appointments (2227/3839, 58.01%, "very interested" and 984/3839, 25.63%, "interested") and asking their provider questions (2570/3852, 66.72%, "very interested" and 906/3852, 23.52%, "interested").

There were no categories in which a majority of subjects reported being disinterested in a service potentially offered by health technologies. However, greater than half of subjects reported neutral interest or less when asked about their interest in a technology that acted as a reminder system (2207/3845, 57.40%) or that would allow subjects to compare their health with others (1958/3848, 50.88%).

The majority of patients (2729/3838, 71.10%) were very interested in one of the items comprising the domain of communication, with 50.50% (1938/3838) being very interested in the domain of self-management and over a third of patients (1362/3843, 33.44%) being very interested in one of the items comprising the domain of education.

In the unadjusted analysis, patients were significantly more likely to be "very interested" in all 3 domains if they were female, African American (as compared with Caucasian), single (as compared with married), owned a smartphone, and had higher depression and health self-efficacy scores. They were less likely to be "very interested" in CHIT if they were older, had better health-related quality-of-life, or if they had coronary artery disease or hypertension (data not shown).

After adjustment for patient characteristics, however, only higher depression ratings, higher health self-efficacy scores, and smartphone ownership remained significant predictors of patient interest in all 3 technology categories: education, communication, and self-management (Table 2). Health-related quality of life was also an independent predictor of lack of interest in education technologies (odds ratio [OR] 0.92, 95% CI 0.86-0.98 per 0.1 increase), and older age was an independent predictor of lack of interest in self-management technologies (OR 0.89, 95% CI 0.81-0.97 per decade increase). No interaction effects were found.

Figure 1. Response distribution to consumer health information technology interest survey questions. Response distributions (n) are shown for the level of interest to each question on the health information technology survey. Questions are grouped by category: education, communication, and self-management.

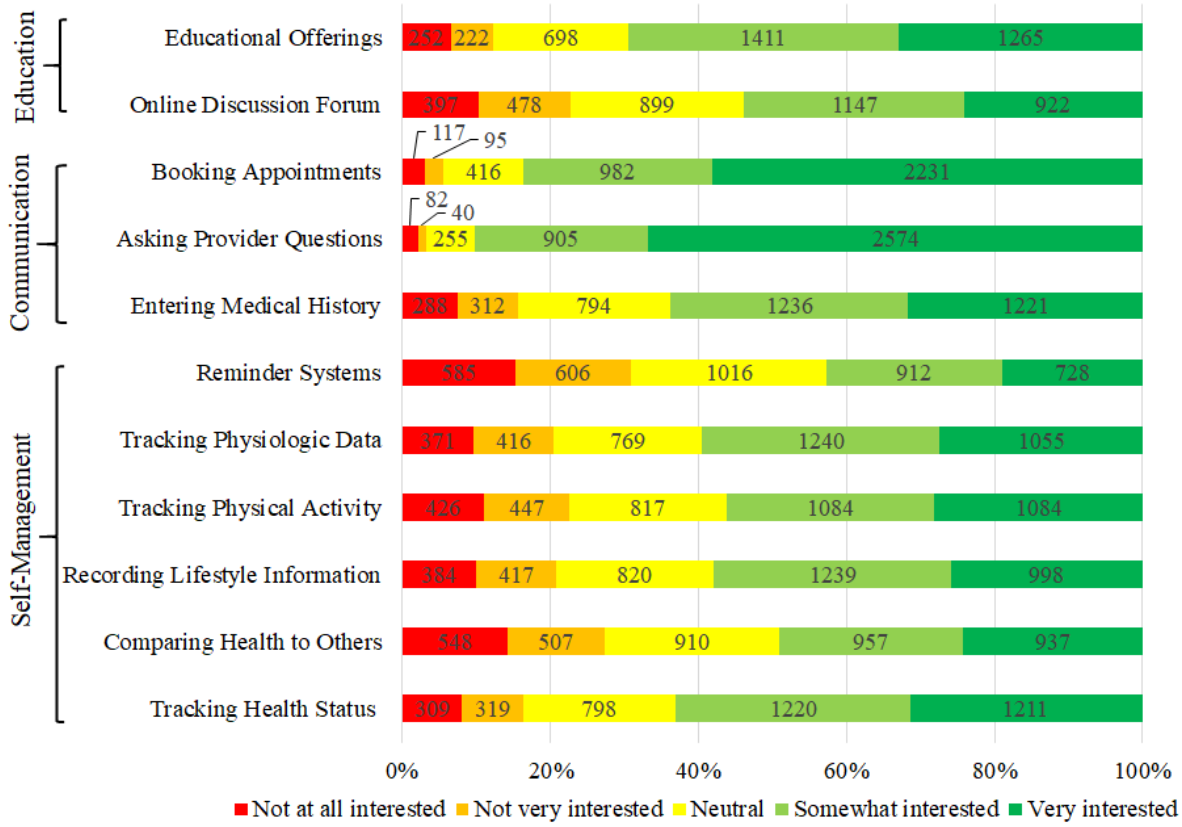


Table 2. Multivariable predictors of consumer health information technology (CHIT) outcomes: education, communication, and self-management.

Predictors	Education		Communication		Self-management	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (per decade)	0.98 (0.89-1.07)	.60	0.98 (0.89-1.08)	.67	0.89 (0.81-0.97)	.01
Female Gender	0.97 (0.78-1.22)	.80	1.20 (0.96-1.50)	.12	0.99 (0.80-1.22)	.89
Race						
White	Reference		Reference		Reference	
African American	1.46 (0.96-2.22)	.08	1.12 (0.70-1.77)	.64	1.48 (0.97-2.26)	.07
Other	1.48 (0.55-3.94)	.44	2.16 (0.61-7.65)	.23	1.33 (0.51-3.50)	.56
Marital status						
Married	Reference		Reference		Reference	
Single	1.12 (0.85-1.48)	.43	0.85 (0.64-1.41)	.28	1.02 (0.78-1.33)	.90
Divorced/widowed	1.29 (0.95-1.76)	.11	0.80 (0.58-1.09)	.15	1.09 (0.81-1.46)	.58
Income (per US $10k)	0.99 (0.93-1.06)	.81	1.06 (0.99-1.13)	.09	1.04 (0.98-1.10)	.26
PHQ-9[b]	1.04 (1.01-1.06)	.003	1.04 (1.02-1.07)	<.001	1.05 (1.03-1.08)	<.001
EQ-5D[c] Index (per 0.1)	0.92 (0.86-0.98)	.01	0.98 (0.92-1.05)	.58	0.97 (0.91-1.03)	.32
Health self-efficacy score	1.20 (1.16-1.25)	<.001	1.12 (1.08-1.16)	<.001	1.15 (1.11-1.19)	<.001
Smart phone ownership	1.83 (1.43-2.34)	<.001	1.75 (1.39-2.22)	<.001	1.77 (1.41-2.21)	<.001
Any comorbidity[d]	0.82 (0.62-1.07)	.14	1.10 (0.83-1.48)	.49	1.05 (0.81-1.37)	.72

[a]OR: odds ratio.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]EQ-5D: Euroqol 5 Dimensions.

[d]Comorbid conditions include diabetes, atrial fibrillation, cancer, coronary artery disease, chronic renal insufficiency, and hypertension.

Sensitivity Analysis

Analyses were replicated after defining the outcome of interest in CHIT categories as a response of "somewhat interested" or "very interested" for every question within each category. Using this definition, 50.42% (1938/3843) of patients expressed interest in the education questions, 58.52% (2246/3838) in the communication questions, and 22.00% (844/3838) in the self-management questions. The independent predictors of interest in education and self-management remained the same as those presented in Table 2. Predictors for communication also remained the same, with the addition of decreasing age (OR 0.89, 95% CI 0.81-0.97 per decade increase).

Subset Analysis

To assess whether completion location biased the study results, 117 out of 904 (12.9%) patients seen in the Spine Center completed the technology survey in the lobby, and results were compared with 787 out of 904 (87.1%) patients who completed the survey using MyChart before their Spine Center visit. In the unadjusted analysis, there were no significant differences in interest in CHIT regarding education or self-management (data not shown). Regarding communication questions, 513 of the 787 patients (65.2%) who completed the survey via MyChart indicated they were "very interested" in asking their provider questions compared with 64 of the 117 (54.7%) waiting room responders (*P*=.03). Similarly, 441 of 782 (56.4%) MyChart respondents expressed they were "very interested" in booking appointments versus 50 of the 117 (42.7%) patients in the waiting room (*P*=.006). After adjustment for patient characteristics, survey completion through MyChart was not a significant predictor of interest in communication (OR 1.81, 95% CI 0.85-3.85; *P*=.12).

Discussion

Principal Findings

CHIT is a promising strategy to enhance individual experience of care and improve the health of populations. CHIT has been demonstrated to engage patients, enhance clinical interventions, and improve health outcomes [49,50]. Focusing on CHIT approaches that are of greatest interest to patients would allow better allocation of resources by health care institutions that are struggling to contain costs and provide superior care in the current environment of reducing reimbursements. Our study found that over half of patient respondents are interested in CHIT apps that provide education, methods to communicate with healthcare providers, and self-management tools. The greatest interest is in CHIT apps that allow patients to ask questions of their health providers and to schedule appointments. CHIT apps that were of lowest interest to respondents were reminder systems and those that provided the ability to compare health status with others with the same condition.

Historically, the design of CHIT has been largely guided by provider attributes, financial incentives, and provider's

perceptions of patient needs rather than the needs and interests of patients themselves [5,17,51-53]. While a recent global survey indicated that only 8% of hospital-produced mobile apps offer the ability to book appointments [10], this study indicates that 83.65% (3213/3841) of current health system patients are interested in this functionality (Figure 1). Current estimates indicate that 7% of patients have changed their health systems because of deficits in online services [10]. This study is currently the largest survey of patient interest in provider-based CHIT and addresses a broader range of technologies than any previous surveys, providing actionable information for health systems working to address the provider-based CHIT gap.

Quality of Life (EQ-5D) and Education Interest

In this study, patients with worse health-related quality of life, as assessed by EQ-5D, had a greater interest in educational technologies than patients with better quality of life. This is consistent with prior literature that sicker patients are more likely to search for health information [20]. When targeting CHIT strategies for patients with poor quality of life, education-related CHIT apps may be especially beneficial.

Depression (PHQ-9) and Overall Interest

Patients with increasingly severe depression, as assessed with the PHQ-9 scale, demonstrated increasing interest in all 3 CHIT categories: education, communication, and self-management. It has been established that patients with depression may prefer treatment that does not require face-to-face interaction [54]. Recent trials of mobile apps have also demonstrated efficacy in managing depression [55-57]. Although, in the past decade, there has been a sharp increase in depression and the depression CHIT market, the large majority of these apps are neither evidence-based nor affiliated with any established medical institution [58]. Depression, despite being a leading global cause of disability, remains underdiagnosed and undertreated [59]. The high interest of patients with depression in this study and the lack of widely available evidence-based CHITs for depression behoove providers to carefully consider this population for focused CHIT development efforts.

Health Self-Efficacy and Overall Interest

General self-efficacy, the belief in one's competence to cope with a broad range of stressful or challenging demands, correlates well with self-regulation, coping strategies, and well-being [60]. Health, nutrition, physical exercise, and alcohol resistance self-efficacies are valid predictors of constructive health-related behaviors [61-63]; health self-efficacy, specifically, has been shown to influence health information–seeking behaviors [45,64]. Perceived difficulty of use decreases patient acceptance of CHITs [3]. CHITs are often promoted as mechanisms for increasing patient "empowerment," "engagement," "activation," and so on [65], but little data exist on the association between self-efficacy and interest in using CHIT [3]. Our data suggest that the underlying attitudes of patients toward their health influence the desire to use such technologies. Health self-efficacy may explain the underlying attitudes that assist patients in overcoming the learning curve required to adopt CHITs. Such a concept has practical implications in that such underlying attitudes may explain some of the variability in the efficacy of CHITs observed in trials within different populations [6]. Developers may also consider targeting high self-efficacy users for increasing the likelihood of a successful CHIT implementation.

Smartphone Ownership and Overall Interest

Mobile health apps are poised to become a determining factor in restructuring old health care services and systems still based on physical relationships between patients and providers [66,67]. The data from this study indicate that smartphone use is a strong predictor of CHIT interest. Supporting this finding is a recent patient survey, in which 86% of respondents demonstrated interest in using a mobile app to improve their health [68]. Despite widespread enthusiasm and public interest, national survey data demonstrate that mobile apps offered by providers were lower rated by users and are not aligned with patient wants or needs [10]. Due to their ease of use, smartness, accessibility, mobility, and connectivity, smartphones are an effective and consumer-oriented technology platform; however, health systems consistently lag behind in mobile technology offerings [69]. These data, together, suggest that there exists a large opportunity for providers to improve their mobile offerings, focusing on usability and the functionalities patients desire.

Limitations

This study has some limitations. The survey was administered to patients seen within the Cleveland Clinic Neurological Institute. The study sample was largely white (3115/3852, 92.90%), with a moderately high median annual household income (US $54,578); centers with different demographics may have different findings. However, these clinics see patients with a broad range of conditions, spanning from back pain to neurodegenerative disorders to psychiatric conditions. Moreover, this study assesses the CHIT interest of current health care users rather than the population-based surveys that are often used to assess attitudes and preferences, and thus, may more accurately reflect the preferences of current patients. Another limitation is the deployment of surveys via the MyChart patient portal, introducing possible responder bias as respondents using the patient portal may be more apt to use health technologies. However, there was no significant difference in patient interest in using CHIT between respondents receiving the survey via the patient portal and the sample of respondents completing the survey in the clinic lobby.

Future Directions

This study provides insight into the interests and needs of current health system patients. This study also demonstrates how such information can be used to model patient interest to target technology offerings of providers. Future studies should include the development of more sophisticated predictive models; continuing to identify subpopulations defined by disease states, preferences, and need attributes will further optimize CHIT implementations.

Conclusions

Health care is under pressure to improve outcomes and decrease cost; CHITs provide a promising solution. CHIT developers often pay minimal attention to patients' motivations, needs, and psychosocial characteristics [70]. Such oversights may lead to

underutilization of provider-based CHIT. In order to solve this problem, patients should be recognized as full stakeholders in health information systems and their perspectives should be systematically incorporated into development efforts [71]. Due to high patient interest, CHIT developers and health system clinician leaders should consider focusing CHIT development toward smartphone platforms and developing CHIT apps that allow patients to schedule appointments and to ask questions of their providers [72,73]. Patients with depression have a high interest in CHITs and thus may benefit from provider offerings of evidence-based CHITs. Health self-efficacy is an independent predictor of CHIT interest and should be considered in future attempts to explain CHIT adoption and efficacy patterns.

Acknowledgments

The authors wish to thank Ya-Ling Lu, MLIS, PhD, Staff Informationist, National Institutes of Health Library for reviewing the literature.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Consumer health information technology interest survey.

[PDF File (Adobe PDF File), 30KB - jmir_v20i4e128_app1.pdf]

References

1. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769 [FREE Full text] [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
2. Sheikh A, Sood HS, Bates DW. Leveraging health information technology to achieve the "triple aim" of healthcare reform. J Am Med Inform Assoc 2015 Jul;22(4):849-856 [FREE Full text] [doi: 10.1093/jamia/ocv022] [Medline: 25882032]
3. Or CK, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Aug;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
4. Improving Consumer Health IT Application Development: Lessons From Other Industries URL: https://healthit.ahrq.gov/sites/default/files/docs/citation/background_report_final.pdf [WebCite Cache ID 6pKb4Diis]
5. Eysenbach G. Consumer health informatics. Br Med J 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [Medline: 10864552]
6. Tao D, Wang T, Wang T, Liu S, Qu X. Effects of consumer-oriented health information technologies in diabetes management over time: a systematic review and meta-analysis of randomized controlled trials. J Am Med Inform Assoc 2017 Sep 01;24(5):1014-1023. [doi: 10.1093/jamia/ocx014] [Medline: 28340030]
7. Agarwal R, Khuntia J. 2009. Personal Health Information Management and the Design of Consumer Health Information Technology URL: https://healthit.ahrq.gov/sites/default/files/docs/page/09-0075-EF_1.pdf [WebCite Cache ID 6pKa4bHX8]
8. Wolff JL, Darer JD, Larsen KL. Family caregivers and consumer health information technology. J Gen Intern Med 2016 Jan;31(1):117-121 [FREE Full text] [doi: 10.1007/s11606-015-3494-0] [Medline: 26311198]
9. Knight EP, Shea K. A patient-focused framework integrating self-management and informatics. J Nurs Scholarsh 2014 Mar;46(2):91-97. [doi: 10.1111/jnu.12059] [Medline: 24354997]
10. Kalis B, Froelich J, Burke A, Warren M. Accenture. 2016. Losing Patience: Why Healthcare Providers Need to Up Their Mobile Game URL: https://www.accenture.com/t20160629T221243__w__/us-en/_acnmedia/PDF-24/Accenture-Losing-Patience.pdf [accessed 2017-03-29] [WebCite Cache ID 6pKas3SIK]
11. Mandl KD, Kohane IS. Time for a patient-driven health information economy? N Engl J Med 2016 Jan 21;374(3):205-208 [FREE Full text] [doi: 10.1056/NEJMp1512142] [Medline: 26789868]
12. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How can eHealth technology address challenges related to multimorbidity? Perspectives from patients with multiple chronic conditions. J Gen Intern Med 2015 Feb 18;30(8):1063-1070. [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]
13. Walker J, Ahern DK, Le LX, Delbanco T. Insights for internists: "I want the computer to know who I am". J Gen Intern Med 2009 Jun;24(6):727-732 [FREE Full text] [doi: 10.1007/s11606-009-0973-1] [Medline: 19412641]
14. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You get reminded you're a sick person": personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
15. Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008 Nov(175):1-1422. [Medline: 19408968]

16. Abelson J, Wagner F, DeJean D, Boesveld S, Gauvin F, Bean S, et al. Public and patient involvement in health technology assessment: a framework for action. Int J Technol Assess Health Care 2016 Jan;32(4):256-264. [doi: 10.1017/S0266462316000362] [Medline: 27670693]
17. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the Science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
18. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
19. Levine DM, Lipsitz SR, Linder JA. Trends in seniors' use of digital health technology in the United States, 2011-2014. J Am Med Assoc 2016 Aug 2;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]
20. Houston TK, Allison JJ. Users of Internet health information: differences by health status. J Med Internet Res 2002 Dec;4(2):E7 [FREE Full text] [doi: 10.2196/jmir.4.2.e7] [Medline: 12554554]
21. Kroenke K, Theobald D, Wu J, Norton K, Morrison G, Carpenter J, et al. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. J Am Med Assoc 2010 Jul 14;304(2):163-171 [FREE Full text] [doi: 10.1001/jama.2010.944] [Medline: 20628129]
22. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 1;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
23. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
24. Tsiligianni I, Kocks J, Tzanakis N, Siafakas N, van der Molen T. Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. Prim Care Respir J 2011 Sep;20(3):257-268 [FREE Full text] [doi: 10.4104/pcrj.2011.00029] [Medline: 21472192]
25. Or CK, Tao D, Wang H. The effectiveness of the use of consumer health information technology in patients with heart failure: a meta-analysis and narrative review of randomized controlled trials. J Telemed Telecare 2017 Jan;23(1):155-166. [doi: 10.1177/1357633X15625540] [Medline: 26759365]
26. Lewis D. Computer-based approaches to patient education: a review of the literature. J Am Med Inform Assoc 1999;6(4):272-282 [FREE Full text] [Medline: 10428001]
27. Denizard-Thompson NR, Singh S, Stevens SR, Miller DP, Wofford JL. iPod™ technology for teaching patients about anticoagulation: a pilot study of mobile computer-assisted patient education. Prim Health Care Res Dev 2012 Jan;13(1):42-47. [doi: 10.1017/S146342361100034X] [Medline: 21854695]
28. Shaw MJ, Beebe TJ, Tomshine PA, Adlis SA, Cass OW. A randomized, controlled trial of interactive, multimedia software for patient colonoscopy education. J Clin Gastroenterol 2001 Feb;32(2):142-147. [Medline: 11205650]
29. Goldsmith DM, Safran C. Using the Web to reduce postoperative pain following ambulatory surgery. Proc AMIA Symp 1999:780-784 [FREE Full text] [Medline: 10566466]
30. Clark M, Ghandour G, Miller NH, Taylor CB, Bandura A, DeBusk RF. Development and evaluation of a computer-based system for dietary management of hyperlipidemia. J Am Diet Assoc 1997 Feb;97(2):146-150. [doi: 10.1016/S0002-8223(97)00040-0] [Medline: 9020241]
31. Fox S, Duggan M. Pew Internet. 2013. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [WebCite Cache ID 6pKcBWOlW]
32. Moreland J, French TL, Cumming GP. The prevalence of online health information seeking among patients in Scotland: a cross-sectional exploratory study. JMIR Res Protoc 2015;4(3):e85 [FREE Full text] [doi: 10.2196/resprot.4010] [Medline: 26177562]
33. Horton KM, Garland MR, Fishman EK. The Internet as a potential source of information about radiological procedures for patients. J Digit Imaging 2000 Feb;13(1):46-47 [FREE Full text] [Medline: 10696601]
34. Robinson C, Flowers CW, Alperson BL, Norris KC. Internet access and use among disadvantaged inner-city patients. J Am Med Assoc 1999 Mar 17;281(11):988-989. [Medline: 10086432]
35. Coye MJ, Haselkorn A, DeMello S. Remote patient management: technology-enabled innovation and evolving business models for chronic disease care. Health Aff (Millwood) 2009;28(1):126-135 [FREE Full text] [doi: 10.1377/hlthaff.28.1.126] [Medline: 19124862]
36. Panagioti M, Richardson G, Small N, Murray E, Rogers A, Kennedy A, et al. Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. BMC Health Serv Res 2014;14:356 [FREE Full text] [doi: 10.1186/1472-6963-14-356] [Medline: 25164529]
37. Pearson M, Mattke S, Shaw R, Ridgely M. 2007. Patient Self-Management Support Programs: An Evaluation URL: http://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/final-reports/ptmgmt/ptmgmt.pdf [WebCite Cache ID 6pKcNDXxt]
38. Lyles C, Schillinger D, Sarkar U. Connecting the dots: health information technology expansion and health disparities. PLoS Med 2015 Jul;12(7):e1001852 [FREE Full text] [doi: 10.1371/journal.pmed.1001852] [Medline: 26172977]

39. Huygens MW, Vermeulen J, Swinkels IC, Friele RD, van Schayck OC, de Witte LP. Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes. BMC Health Serv Res 2016 Jul 08;16:232 [FREE Full text] [doi: 10.1186/s12913-016-1484-5] [Medline: 27391471]
40. 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]
41. Wang SH. Web-based medical service: technology attractiveness, medical creditability, information source, and behavior intention. J Med Internet Res 2017 Aug 02;19(8):e285 [FREE Full text] [doi: 10.2196/jmir.8114] [Medline: 28768608]
42. Cranen K, Groothuis-Oudshoorn CG, Vollenbroek-Hutten MM, IJzerman MJ. Toward patient-centered telerehabilitation design: understanding chronic pain patients' preferences for web-based exercise telerehabilitation using a discrete choice experiment. J Med Internet Res 2017 Jan 20;19(1):e26 [FREE Full text] [doi: 10.2196/jmir.5951] [Medline: 28108429]
43. Jenssen BP, Mitra N, Shah A, Wan F, Grande D. Using digital technology to engage and communicate with patients: a survey of patient attitudes. J Gen Intern Med 2016 Jan;31(1):85-92. [doi: 10.1007/s11606-015-3517-x] [Medline: 26385117]
44. Hickam D, Totten A, Berg A, Rader K, Goodman S, Newhouse R. 2013. The PCORI Methodology Report URL: http://www.pcori.org/research-results/research-methodology [accessed 2017-03-29] [WebCite Cache ID 6pKcSm0FJ]
45. Lee SY, Hwang H, Hawkins R, Pingree S. Interplay of negative emotion and health self-efficacy on the use of health information and its outcomes. Communic Res 2008 Apr 14;35(3):358-381. [doi: 10.1177/0093650208315962]
46. Katzan I, Speck M, Dopler C, Urchek J, Bielawski K, Dunphy C, et al. The Knowledge Program: an innovative, comprehensive electronic data capture system and warehouse. AMIA Annu Symp Proc 2011;2011:683-692 [FREE Full text] [Medline: 22195124]
47. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care 2005 Mar;43(3):203-220. [Medline: 15725977]
48. 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] [Medline: 11556941]
49. Jha AK. Health information technology: on the cusp of healthcare transformation. Am J Manag Care 2013 Nov;19(10 Spec No):SP335-SP336 [FREE Full text] [Medline: 24511887]
50. Gibbons C, Wilson R, Samal L, Lehmann C. 2009. Impact of Consumer Health Informatics Applications URL: https://www.ahrq.gov/downloads/pub/evidence/pdf/chiapp/impactchia.pdf [WebCite Cache ID 6pKcpeMeh]
51. 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]
52. Audet A, Squires D, Doty MM. Where are we on the diffusion curve? Trends and drivers of primary care physicians' use of health information technology. Health Serv Res 2014 Feb;49(1 Pt 2):347-360 [FREE Full text] [doi: 10.1111/1475-6773.12139] [Medline: 24358958]
53. Menefee HK, Thompson MJ, Guterbock TM, Williams IC, Valdez RS. Mechanisms of communicating health information through Facebook: implications for consumer health information technology design. J Med Internet Res 2016 Aug 11;18(8):e218 [FREE Full text] [doi: 10.2196/jmir.5949] [Medline: 27515151]
54. Gustafson DH, Hawkins RP, Boberg EW, McTavish F, Owens B, Wise M, et al. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. Int J Med Inform 2002 Nov 12;65(3):169-177. [Medline: 12414016]
55. 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;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
56. Agyapong VI, Juhás M, Ohinmaa A, Omeje J, Mrklas K, Suen VY, et al. Randomized controlled pilot trial of supportive text messages for patients with depression. BMC Psychiatry 2017 Aug 02;17(1):286 [FREE Full text] [doi: 10.1186/s12888-017-1448-2] [Medline: 28768493]
57. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
58. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
59. Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? Br Med J 2014 Feb 04;348:g1253. [Medline: 24496211]
60. Luszczynska A, Scholz U, Schwarzer R. The general self-efficacy scale: multicultural validation studies. J Psychol 2005 Sep;139(5):439-457. [doi: 10.3200/JRLP.139.5.439-457] [Medline: 16285214]
61. Health-Specific Self-Efficacy Scales. 2009. URL: https://userpage.fu-berlin.de/health/healself.pdf [accessed 2017-10-03] [WebCite Cache ID 6twhI7xQR]
62. Self-Efficacy and Health Behaviours. 1995. URL: http://userpage.fu-berlin.de/gesund/publicat/conner9.htm [accessed 2017-10-03] [WebCite Cache ID 6twhQCnmM]

63. Sheeran P, Maki A, Montanaro E, Avishai-Yitshak A, Bryan A, Klein WM, et al. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: a meta-analysis. Health Psychol 2016 Nov;35(11):1178-1188. [doi: 10.1037/hea0000387] [Medline: 27280365]
64. Deng Z, Liu S. Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. Int J Med Inform 2017 Sep;105:98-109. [doi: 10.1016/j.ijmedinf.2017.05.014] [Medline: 28750916]
65. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
66. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272. [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
67. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3. [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
68. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
69. Lee J. Future of the smartphone for patients and healthcare providers. Healthc Inform Res 2016 Jan;22(1):1-2 [FREE Full text] [doi: 10.4258/hir.2016.22.1.1] [Medline: 26893944]
70. Marquard JL, Zayas-Cabán T. Commercial off-the-shelf consumer health informatics interventions: recommendations for their design, evaluation and redesign. J Am Med Inform Assoc 2012;19(1):137-142 [FREE Full text] [doi: 10.1136/amiajnl-2011-000338] [Medline: 21727206]
71. Hunt JS, Gibson RF, Whittington J, Powell K, Wozney B, Knudson S. Guide for developing an information technology investment road map for population health management. Popul Health Manag 2015 Jun;18(3):159-171. [doi: 10.1089/pop.2014.0092] [Medline: 25607932]
72. Zhao P, Yoo I, Lavoie J, Lavoie BJ, Simoes E. Web-based medical appointment systems: a systematic review. J Med Internet Res 2017 Apr 26;19(4):e134 [FREE Full text] [doi: 10.2196/jmir.6747] [Medline: 28446422]
73. Uei S, Tsai C, Kuo Y. The effect of telehealth systems and satisfaction with health expenditure among patients with metabolic syndrome. Technol Health Care 2016 Apr 29;24(Suppl 2):S527-S532. [doi: 10.3233/THC-161178] [Medline: 27163314]

Abbreviations

CHIT: consumer health information technology
EQ-5D: EuroQol 5 Dimensions
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9

Edited by G Eysenbach; submitted 30.03.17; peer-reviewed by T Irizarry, D Tao, H Devan; comments to author 10.08.17; revised version received 03.10.17; accepted 02.11.17; published 19.04.18.

Please cite as:
Featherall J, Lapin B, Chaitoff A, Havele SA, Thompson N, Katzan I
Characterization of Patient Interest in Provider-Based Consumer Health Information Technology: Survey Study
J Med Internet Res 2018;20(4):e128
URL: http://www.jmir.org/2018/4/e128/
doi:10.2196/jmir.7766
PMID:29674312

Original Paper

The Association Between Technology Use and Health Status in a Chronic Obstructive Pulmonary Disease Cohort: Multi-Method Study

Matthew Witry[1], PhD, PharmD; Alejandro Comellas[2], MD; Jacob Simmering[3], MS, PhD; Philip Polgreen[2], MPH, PhD

[1]Department of Pharmacy Practice and Science, College of Pharmacy, University of Iowa, Iowa City, IA, United States

[2]Department of Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, IA, United States

[3]Signal Center for Health Innovation, University of Iowa, Iowa City, IA, United States

Corresponding Author:
Matthew Witry, PhD, PharmD
Department of Pharmacy Practice and Science
College of Pharmacy
University of Iowa
115 S Grand Ave
S515 PHAR
Iowa City, IA, 52242
United States
Phone: 1 3193358763
Email: matthew-witry@uiowa.edu

Abstract

Background: Telemedicine and electronic health (eHealth) interventions have been proposed to improve management of chronic obstructive pulmonary disease (COPD) for patients between traditional clinic and hospital visits to reduce complications. However, the effectiveness of such interventions may depend on patients' comfort with technology.

Objective: The aim was to describe the relationship between patient demographics and COPD disease severity and the use of communication-related technology.

Methods: We administered a structured survey about the use of communication technologies to a cohort of persons in the COPDGene study at one midwestern hospital in the United States. Survey results were combined with clinical and demographic data previously collected as part of the cohort study. A subsample of patients also completed eHealth simulation tasks. We used logistic or linear regression to determine the relationship between patient demographics and COPD disease severity and reported use of communication-related technology and the results from our simulated eHealth-related tasks.

Results: A total of 686 patients completed the survey and 100 participated in the eHealth simulation. Overall, those who reported using communication technology were younger (P=.005) and had higher incomes (P=.03). Men appeared less likely to engage in text messaging (P<.001) than women. Patients who spent more time on tasks in the eHealth simulation had greater odds of a COPD Assessment Test score >10 (P=.02) and walked shorter distances in their 6-minute walk tests (P=.003) than those who took less time.

Conclusions: Older patients, patients with lower incomes, and less healthy patients were less likely to report using communication technology, and they did not perform as well on our simulated eHealth tasks. Thus, eHealth-based interventions may not be as effective in these populations, and additional training in communication technology may be needed.

(J Med Internet Res 2018;20(4):e125) doi:10.2196/jmir.9382

KEYWORDS

pulmonary disease, chronic obstructive; telemedicine; wireless technology; electronic mail; remote consultation; patient simulation; surveys

Introduction

Chronic obstructive pulmonary disease (COPD) is the third-leading cause of death in the United States and is a major cause of morbidity and mortality [1]. The management of COPD includes smoking cessation, pulmonary rehabilitation, oxygen, and pharmacologic treatments such as systemic steroids, inhalers, and periodic antibiotic treatments [2]. Diligent adherence to these therapies is associated with decreased mortality, fewer exacerbations and hospitalizations, and quality of life improvements [3-5]. Most patients, however, are nonadherent to approximately half of their therapies [6]. Barriers to COPD therapy adherence include the complexity and cost of regimens, incorrect medication knowledge and beliefs, health literacy, and cognitive deficits [4,7,8]. The management of COPD can be especially challenging in rural areas where access to care is limited. Indeed, rural patients with COPD have been found to have worse outcomes [9].

One approach for supporting patient self-management between traditional office visits, including in rural areas, has been to use telephone-based interventions, an approach often referred to as "telehealth." Such programs have used nurses or other members of the health care team to regularly monitor patient symptoms via telephone. When patient worsening is identified, actions such as the early prescribing of antibiotics or steroids can be administered [10]. However, the results of telehealth interventions are mixed [10,11].

Recently, cellular and internet-connected devices have been suggested as an alternative or extension to telehealth interventions as a way for patients and providers to engage in bidirectional electronic information exchange. These interventions have been referred to as *electronic health (eHealth)* or included under the umbrella of telehealth. For example, internet-connected devices have been used as a means to measure a patient's lung function and oxygen saturation, to provide information about a patient's clinical condition to clinicians, or provide direct recommendations to the patient [12]. Identifying interventions that increase the quality of patient-provider communication, symptom reporting, and data exchange are critical as health care systems look for strategies to decrease COPD exacerbations and hospital admissions [13]. However, studies on internet-based telehealth interventions in patients with COPD have reported mixed results as well [14-17].

Although the use of the internet among older adults is increasing, it still is lower than among younger populations, with rates of internet adoption at 58% for persons aged 65 years and older [18]. A study examining the use of secure messaging between patients and physicians found one-third of those aged 55 years and older had used secure messaging, but there were significant disparities in use due to income, race, education, and health status [19]. Barriers to internet and information technology adoption for older adults include lack of broadband access, rural living, lower ownership of internet-enabled tablets or mobile phones, physical disabilities (ie, impaired vision), the preference for assistance when interacting with a new device, and health literacy [20-23]. Patients with COPD may also have a lower level of internet use because advanced age is a risk factor for COPD.

This paper aims to report the relationship between COPD and ownership and use of various technologies and devices. Specifically, we examine the effect of COPD severity and patient demographics on not only technology use and ownership, but also completion time for a simulated eHealth task.

Methods

Overview

All data were collected from adults participating in the COPDGene multicenter cohort study at the University of Iowa Hospitals and Clinics. The COPDGene study details have been reported elsewhere [24]. Briefly, COPDGene is a multicenter observational study including current and former smokers designed to identify genetic factors associated with COPD. A total of 10,192 non-Hispanic white and African American adults aged 45 to 80 years with a minimum 10 pack-year smoking history were enrolled between January 2008 and April 2011. Participants were phenotyped by completing questionnaires, blood tests, imaging, and spirometry. During a scheduled COPDGene study visit, patients were surveyed using a structured interview conducted by a respiratory therapist research nurse either by telephone or in person. Surveys were completed between August 2014 and June 2016. Beginning in May 2015, interested patients also could participate in a brief simulation in which a set of simple eHealth tasks were performed and timed. Patients were compensated for taking part in the COPDGene cohort, but no additional compensation was provided for taking part in the survey. Those agreeing to participate in the additional eHealth simulation were offered a US $10 gift card for their time. Both the survey and eHealth simulation substudy were approved by the COPDGene group and the University of Iowa Institutional Review Board. This study was conducted in accordance with the amended Declaration of Helsinki.

Survey

The survey was administered via a structured interview. The objective was to assess the use of various communication technologies that could be used in eHealth interventions including cellular telephones, text messaging, email, and video chat (Multimedia Appendix 1). A combination of yes/no questions and several short responses were used. Interviews lasted 5 to 10 minutes and responses were recorded in RedCap (Nashville, TN, USA), a secure online database hosted at the University of Iowa and linked to the patient's COPDGene identifier.

Clinical Data

Survey data were linked to clinical and demographic data from the COPDGene study database using the study identifier. Variables included age, gender, income level, and validated measures of disease progression—the 6-minute walk test and COPD Assessment Test (CAT) [25] score. These variables were collected during their most recent COPDGene visit.

eHealth Simulation

Six months after the structured interviews began, we invited participants to engage in a series of timed eHealth simulation exercises. These tasks involved launching an app, entering an access code, and responding to two CAT items. Participants performed these tasks on a laptop computer running Microsoft Windows (Redmond, WA, USA) and two tablet computers, one running the iOS mobile operating system (Cupertino, CA, USA) and one running the Android operating system (Mountain View, CA, USA). The time required for study participants to perform each task was recorded using a digital stopwatch. The order of devices was randomized to allow for comparisons.

Analysis

To describe technology ownership and use, we calculated the percentages of ownership of computers and cell phones, as well as the use of email and video chat services, such as Skype (Redmond, WA, USA). For patients who owned a cell phone, we computed percentages of those who carry a phone regularly, own an internet-enabled mobile phone (smartphone), and use text messaging. We also calculated the percentages of those who use text messaging, given they either own a smartphone or use a cell phone regularly.

To determine if technology use differed by demographic characteristics, we estimated six logistic regression models. All models had age, sex, and income as covariates, and the outcome variables were cell phone ownership, smartphone ownership, computer ownership, use of text messaging, use of email, and use of video chat.

To determine if technology use differed by the severity of the patient's disease, we estimated three sets of models. The outcomes for each set were the result of the 6-minute walk test and dichotomized CAT score (<10, ≥10). Because the result of the 6-minute walk test is continuous, we estimated linear regression models. For the dichotomous CAT score variable, we estimated logistic regression models. For each set of models, we estimated three separate models (one for each technology): text messaging, email, and video chat. Covariates for all models were patient age, sex, income, and a measure of technology use.

We characterized technology use in three ways. First, we divided ownership and use into three levels: nonowner of relevant technology, owner/nonuser of relevant technology, and owner/user of relevant technology. Second, we considered standardized task time from the eHealth simulations. We first converted each task time into a *z* score and took the sum of the normalized task times for the use of the iPad, Android tablet, cell phone texting, and email as a direct measure of familiarity of comfort with the given technology.

Finally, we considered the ability of simple questions about technology ownership and use (eg, "Do you own a smartphone, computer, or tablet?" and the video chat measure "Do you use Skype?") to predict task performance. We estimated 12 linear regression models, one for each combination of device tested (laptop, Android tablet, iPad, and cell phone) and task (email, text messaging, and video chat). The outcome was task time, and the covariates were gender, age income, and a variable representing the use/nonuse of that task (nonowner, owner, nonuser, and user).

All statistical analysis was completed using R 3.2 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Survey Results

There were 712 persons approached to complete the structured interview survey, of these 686 (96.3%) participated and provided complete survey data. In all, 100 patients also participated in the eHealth simulation task subsample out of 256 who were approached (39.1%). Demographics and summary statistics of the full sample and the subsample who performed the eHealth simulation tasks can be found in Table 1. The mean age of participants who completed the survey was 68.7 (SD 8.2) years, and 52.2% (358/686) were female.

Table 1. Characteristics of study participants.

Characteristic	Study sample (N=686)	Task subsample (n=100)
Age (years), mean (SD)	68.7 (8.2)	65.6 (7.8)
Gender (female), n (%)	358 (52.2)	66 (66.0)
Income (US $), n (%)		
<15,000	42 (6.1)	8 (8.0)
15,000-35,000	154 (22.4)	22 (22.0)
35,000-50,000	139 (20.3)	17 (17.0)
50,000-75,000	160 (23.3)	23 (23.0)
>75,000	133 (19.4)	18 (18.0)
Missing	58 (8.5)	12 (12.0)
CAT[a] Score, mean (SD)	9.6 (6.9)	9.8 (7.8)
Six-Minute Walk, mean (SD)	1410.3 (402.7)	1442.5 (367.3)

[a]CAT: COPD Assessment Test.

Table 2. Description of technology ownership and use for the participants and mean task time for the subsample of participants engaging in the eHealth simulation tasks (N=686).

Technology use characteristic	Participants
Cell phone ownership, n (%)	645 (94.0)
Cell phone owners who..., n (%)	
Have a smartphone	307 (47.6)
Carry their phone daily	510 (79.1)
Send/receive text messages	442 (68.5)
Send/receive text messages given ownership of a smartphone	286 (93.2)
Send/receive text messages given daily carrying of a cell phone	384 (75.3)
Computer ownership, n (%)	607 (88.8)
Use email given computer ownership, n (%)	564 (92.9)
Use video chat (Skype), n (%)	195 (32.1)
Task times (simulation subsample only), mean (SD)	
iPad	76.3 (46.9)
Android	49.4 (30.6)
Laptop	52.8 (25.9)
Basic phone texting	55.0 (26.4)

Table 3. Logistic model of demographic predictors of ownership and use for the total sample.

Characteristic	Own cell phone		Own smartphone		Use text messaging		Own computer		Use email		Use video chat	
	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
Age[a]	0.89 (0.84-0.94)	<.001	0.92 (0.90-0.94)	<.001	0.87 (0.84-0.89)	<.001	0.95 (0.92-0.99)	.005	0.93 (0.89-0.97)	.002	0.99 (0.96-1.01)	.20
Male[b]	0.54 (0.26-1.10)	.08	0.81 (0.57-1.16)	.24	0.66 (0.45-0.98)	<.001	0.96 (0.57-1.60)	.87	0.59 (0.31-1.12)	.10	0.82 (0.57-1.18)	.27
Income (US $1000)												
<15	0.67 (0.19-2.37)	.53	0.86 (0.38-1.91)	.70	0.87 (0.36-2.12)	.76	0.41 (0.18-0.95)	.04	0.91 (0.23-3.57)	.89	0.98 (0.39-2.46)	.96
15-35	1.00		1.00		1.00		1.00		1.00		1.00	
35-50	1.04 (0.44-2.44)	.93	1.47 (0.86-2.51)	.15	0.94 (0.53-1.64)	.81	1.42 (0.71-2.85)	.31	1.51 (0.53-3.63)	.35	1.24 (0.69-2.22)	.46
50-75	3.19 (0.98-10.45)	.05	1.63 (0.98-2.72)	.05	1.31 (0.75-2.28)	.34	2.06 (0.97-4.40)	.06	1.50 (0.63-3.53)	.35	1.68 (0.97-2.91)	.06
>75	5.99 (1.26-28.46)	.02	3.29 (1.92-5.68)	<.001	1.95 (1.06-3.60)	.03	11.27 (2.52-50.36)	.001	8.16 (1.74-38.21)	.007	2.95 (1.70-5.14)	<.001
Declined	0.67 (0.19-2.37)	.31	0.71 (0.34-1.46)	.34	0.88 (0.43-1.81)	.72	0.94 (0.41-2.16)	.89	1.08 (0.36-3.29)	.88	0.98 (0.44-2.20)	.97

[a]Age is continuous.

[b]Male is in comparison to female.

A description of technology ownership and use is found in Table 2. Nearly all (94.0%, 645/686) participants owned a cell phone, and most (74.3%, 510/686) of those carried their phone daily. Somewhat fewer (89%, 607/686) participants owned a computer, although nearly all (92.9%, 564/607) computer owners used email. A third (195/607) used video chat. The mean length of time needed for the eHealth simulations ranged from 49.4 (SD 30.6) seconds for the Android tablet to 76.3 (SD 46.9) seconds for the iOS tablet. Demographic predictors of technology use are summarized in Table 3. Generally, increasing age was associated with lower odds of owning or using technology. Also, a high income (greater than US $75,000/year) had higher odds of owning and using technology compared to lower income groups. Notably, among cell phone owners, men had odds of texting that were 66% of those of women.

Table 4. Logistic and ordinary least squares models including standardized eHealth simulation task times predicting disease severity (n=100).

Characteristic	CAT score ≥10[a]		Distance walked[b]	
	OR (95% CI)	*P*	OR (95% CI)	*P*
Age	0.93 (0.87, 1.00)	.04	2.78 (–6.31, 11.87)	.54
Male	1.30 (0.45, 3.69)	.62	72.91 (–71.37, 217.19)	.32
Income (US $)				
<15,000	5.10 (0.72, 36.11)	.10	–310.27 (–579.59, –40.95)	.02
15,000-35,000	Reference		Reference	
35,000-50,000	0.53 (0.11, 2.66)	.43	163.26 (–56.20, 382.73)	.14
50,000-75,000	2.44 (0.63, 9.40)	.19	158.26 (–38.72, 355.24)	.11
>75,000	0.45 (0.09, 2.33)	.33	113.80 (–96.61, 323.20)	.28
Declined	3.49 (0.67, 18.22)	.13	–37.39 (–271.88, 197.10)	.75
Standardized task time	1.27 (1.04, 1.55)	.02	–40.78 (–67.41, –14.15)	.003

[a]CAT: COPD Assessment Test. CAT score is based on logistic model.

[b]The 6-minute walk is based on ordinary least squares model.

In all cases (text messaging, email, and video chat), being an owner and user was associated with a lower CAT score compared with owner/nonusers and nonowners (Multimedia Appendix 2). Multimedia Appendix 3 repeats the analysis of Multimedia Appendix 2 but considers the distance covered in a 6-minute walk instead of CAT score. A statistically significant increase in distance covered among owner/users relative to owner/nonusers of email and video chat was observed, but not for text messaging. Nonowners of a computer or a smartphone walked a longer distance than owner/nonusers of email. None of the other nonowners traveled a statistically significant distance relative to owner/nonusers.

eHealth Simulation Results

Results of a regression of standardized task time from the eHealth simulation and disease severity are reported in Table 4. Increased task time was statistically significantly associated with having a CAT score greater than 10 and recording fewer steps on the 6-minute walk test, surrogate markers for increased disease severity. For each standard deviation from the mean task time, we found a 27% increase in the odds of having a CAT score of 10 or greater. Likewise, for each standard deviation increase from the mean task time, we found a 41-unit decrease in the distance covered in the 6-minute walk test.

Lastly, after adjustment for age, sex, and income, users of video chat took less time than owner/nonusers to complete the laptop, Android tablet, iOS, and text messaging tasks: OR 10.2 (95% CI –0.5 to –20.0), OR 6.3 (95% CI –16.5 to 3.9), OR 16.4 (95% CI –33.6 to 0.8), and OR 6.3 (95% CI –37.3 to 39.5) seconds, respectively. On the other hand, nonowners took more time than owner/nonusers on these tasks: OR 58.4 (95% CI 21.8-94.9), OR 42.7 (95% CI 4.5-80.8), OR 22.7 (95% CI –42.7 to 87.1), and OR 1.1 (95% CI –16.5 to 3.9) seconds, respectively. Lastly, in Multimedia Appendix 4, we report the results of models of task time from the eHealth simulation and reported ownership/nonownership and use/nonuse of the devices. Email users were significantly faster when using a laptop or Android tablet than nonusers of email. Similarly, video chat users were significantly faster on the laptop and iOS device than nonowners.

Discussion

Our results showed that most respondents in our COPD cohort, regardless of income or age, had access to either a cell phone or personal computer, and most reported comfort using these devices. However, use and familiarity with newer technologies, such as smartphones or video chat, were less common. In general, we found that participants with more severe disease were less likely to report the use of technology. In addition, when we tested participants' ability to use technology in a simulated eHealth task, we found that patients who had more difficulty with completion of simulated eHealth tasks as evident by taking longer to complete the task were more likely to have more severe COPD as evident by having a greater odds of a CAT score greater than 10 and walking shorter distances in their 6-minute walk tests. Thus, persons with worse COPD—the very patients often targeted with eHealth-related interventions—may require more significant training and infrastructure support, such as greater incorporation of caregivers into the process. Others have made similar recommendations for bridging this digital divide with regard to patients living in rural communities, racial minorities, and persons with low health literacy [19,21-23].

We found that patients who reported owning and using video chat and email, presumably for communicating with friends and family, completed the simulated tasks of submitting answers to CAT items more quickly. With some devices, however, patients who owned the device but reported not using it had worse task times than someone who did not even own the device. Thus, ownership alone may not be a sufficient screening question before an eHealth intervention because owner/nonusers had worse task performance than owners who were more frequent users. Thus, screening questions based on both access to and use of eHealth technologies could serve as surrogates for

self-efficacy related to digital health literacy, a component of Health 2.0 skills [26].

In addition to having less advanced COPD, participants who used more and newer communication technologies were also younger and had higher incomes. Accordingly, studies and clinical services that allow for patients to self-select may be composed of relatively healthy patients and thus could make eHealth-related interventions and programs look more effective than studies with a broader sample. Conversely, studies or programs consisting of patients who are older, less affluent, or with more advanced stages of disease may be less effective because these patients are less likely to be comfortable using newer forms of communication. To help ensure effectiveness of eHealth interventions, it may be valuable to design and tailor interventions specific to these persons who are less familiar with technology because older adults, in general, prefer assistance when learning a new technology [20].

Other studies have reported findings suggesting that patients with more severe disease may use technology less often, independent of other demographic factors [27]. This finding has important implications for COPD interventions because patients with certain sociodemographic factors [28] and a history of COPD exacerbations are more likely to experience exacerbations, hospitalizations, and readmissions and therefore are the most in need of intervention. These findings, combined with our finding that demographic factors are associated with lower technology use, might imply that interventions to prevent readmissions in COPD patients may struggle to achieve the desired results based on the interplay between sociodemographics, health history, and experience with technology.

Our findings echo concerns about the relationship between eHealth interventions and health disparities. There was a significant association in most of the analyses between low income, higher disease severity, and lower technology use. Proliferating eHealth initiatives that use advanced technologies could disproportionately benefit wealthier, more native users of technologies, adding to health disparities that already favor those with more financial resources. Although we did not specifically investigate differences between urban and rural patients with COPD, rural patients may have more severe disease than their urban and suburban counterparts [9,29]. This, combined with issues of rural access to high-speed internet, may further exacerbate health disparities for rural COPD patients. Having high-speed internet access would increase someone's likelihood to have prior experience with newer communication technologies, and their comfort levels with eHealth-related care [21].

This study is subject to several limitations. Although our survey had a high participation rate, we only surveyed and tested participants within a single center in the COPDGene study. Thus, our results may not be generalizable to the population of COPD patients as a whole. Also, unmeasured characteristics may have influenced each participant's decision to participate in COPDGene and in the eHealth simulation task subsample, factors that presumably could affect their use of the technologies of interest. The majority of our participants were white, and patients belonging to different racial or ethnic groups may have different experiences with technology and barriers to fully participating in eHealth interventions. Also, the survey was only administered in English. Future studies need to investigate the use of technology in diverse populations of COPD patients. Lastly, the cross-sectional nature of this study limits the ability to make causal inferences.

Despite our limitations, we show that patients with COPD have different levels of access and experience using communications technology. With our simulated health tasks, we also showed that older patients and patients with more severe disease had more difficulty using technology. These findings demonstrate the need for education and assistance for patients who are either not as healthy or not as familiar with technology. Testing the effectiveness of new interventions should include assessment of previous technology use and familiarity.

Acknowledgments

This study received no external funding. The manuscript was approved by the COPDGene project, but they had no input on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Structured interview script.

[PDF File (Adobe PDF File), 21KB - jmir_v20i4e125_app1.pdf]

Multimedia Appendix 2

Logistic model of variables predicting the odds of a patient having a CAT score ≥10 given age, gender, income, and prior use of the given technology (N=686).

[PDF File (Adobe PDF File), 20KB - jmir_v20i4e125_app2.pdf]

Multimedia Appendix 3

Ordinary least squares model of the relationship between gender, income, and prior use of the given technology and 6-minute walk distance (N=686).

[PDF File (Adobe PDF File), 21KB - jmir_v20i4e125_app3.pdf]

Multimedia Appendix 4

Task time relative to owner/nonuser for nonowner and owner/user adjusted for age, sex, and income.

[PDF File (Adobe PDF File), 24KB - jmir_v20i4e125_app4.pdf]

References

1. National Heart Lung and Blood Institute. Morbidity & Mortality: Chart Book on Cardiovascular, Lung, and Blood Diseases. Bethesda, MD: US Department of Healthcare and Human Services, National Institutes of Health; 2012 Feb. Background data URL: https://www.nhlbi.nih.gov/files/docs/research/2012_ChartBook.pdf [accessed 2018-02-27] [WebCite Cache ID 6xYAdkck2]
2. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2013 Feb 15;187(4):347-365. [doi: 10.1164/rccm.201204-0596PP] [Medline: 22878278]
3. Vestbo J, Anderson JA, Calverley PM, Celli B, Ferguson GT, Jenkins C, et al. Adherence to inhaled therapy, mortality and hospital admission in COPD. Thorax 2009 Nov;64(11):939-943. [doi: 10.1136/thx.2009.113662] [Medline: 19703830]
4. Bourbeau J, Bartlett SJ. Patient adherence in COPD. Thorax 2008 Sep;63(9):831-838 [FREE Full text] [doi: 10.1136/thx.2007.086041] [Medline: 18728206]
5. Bourbeau J, Nault D, Dang-Tan T. Self-management and behaviour modification in COPD. Patient Educ Couns 2004 Mar;52(3):271-277. [doi: 10.1016/S0738-3991(03)00102-2] [Medline: 14998597]
6. Bender BG. Nonadherence in chronic obstructive pulmonary disease patients: what do we know and what should we do next? Curr Opin Pulm Med 2014 Mar;20(2):132-137. [doi: 10.1097/MCP.0000000000000027] [Medline: 24452102]
7. George J, Kong DC, Thoman R, Stewart K. Factors associated with medication nonadherence in patients with COPD. Chest 2005 Nov;128(5):3198-3204. [doi: 10.1378/chest.128.5.3198] [Medline: 16304262]
8. Restrepo RD, Alvarez MT, Wittnebel LD, Sorenson H, Wettstein R, Vines DL, et al. Medication adherence issues in patients treated for COPD. Int J Chron Obstruct Pulmon Dis 2008;3(3):371-384. [Medline: 18990964]
9. Abrams TE, Vaughan-Sarrazin M, Fan VS, Kaboli PJ. Geographic isolation and the risk for chronic obstructive pulmonary disease-related mortality: a cohort study. Ann Intern Med 2011 Jul 19;155(2):80-86. [doi: 10.7326/0003-4819-155-2-201107190-00003] [Medline: 21768581]
10. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2010;16(3):120-127. [doi: 10.1258/jtt.2009.090812] [Medline: 20197355]
11. Bartoli L, Zanaboni P, Masella C, Ursini N. Systematic review of telemedicine services for patients affected by chronic obstructive pulmonary disease (COPD). Telemed J E Health 2009 Nov;15(9):877-883. [doi: 10.1089/tmj.2009.0044] [Medline: 19919194]
12. van der Heijden M, Lucas PJ, Lijnse B, Heijdra YF, Schermer TR. An autonomous mobile system for the management of COPD. J Biomed Inform 2013 Jun;46(3):458-469 [FREE Full text] [doi: 10.1016/j.jbi.2013.03.003] [Medline: 23500485]
13. Centers for Medicare and Medicaid Services. 2017. Readmissions reduction program (HRRP) URL: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html [accessed 2018-02-27] [WebCite Cache ID 6xYAvIxp3]
14. Fan VS, Gaziano JM, Lew R, Bourbeau J, Adams SG, Leatherman S, et al. A comprehensive care management program to prevent chronic obstructive pulmonary disease hospitalizations: a randomized, controlled trial. Ann Intern Med 2012 May 15;156(10):673-683. [doi: 10.7326/0003-4819-156-10-201205150-00003] [Medline: 22586006]
15. Johnston NW, Lambert K, Hussack P, Gerhardsson DV, Higenbottam T, Lewis J, et al. Detection of COPD exacerbations and compliance with patient-reported daily symptom diaries using a smart phone-based information system [corrected]. Chest 2013 Aug;144(2):507-514. [doi: 10.1378/chest.12-2308] [Medline: 23519329]
16. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. BMJ 2013;347:f6070 [FREE Full text] [Medline: 24136634]
17. Alwashmi M, Hawboldt J, Davis E, Marra C, Gamble J, Abu AW. The effect of smartphone interventions on patients with chronic obstructive pulmonary disease exacerbations: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 Sep 01;4(3):e105 [FREE Full text] [doi: 10.2196/mhealth.5921] [Medline: 27589898]
18. Perrin A, Duggan M. 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 2018-02-27] [WebCite Cache ID 6udBCXi1z]

19. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. The digital divide and patient portals: Internet access explained differences in patient portal use for secure messaging by age, race, and income. Med Care 2016 Aug;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
20. Smith A. Pew Research Center. 2014 Apr 2. Older adults and technology use URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [accessed 2018-02-27] [WebCite Cache ID 6udBINoUy]
21. Wang J, Bennett K, Probst J. Subdividing the digital divide: differences in internet access and use among rural residents with medical limitations. J Med Internet Res 2011;13(1):e25 [FREE Full text] [doi: 10.2196/jmir.1534] [Medline: 21371989]
22. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
23. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
24. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. COPD 2010 Feb;7(1):32-43 [FREE Full text] [doi: 10.3109/15412550903499522] [Medline: 20214461]
25. Jones PW, Harding G, Berry P, Wiklund I, Chen W, Kline LN. Development and first validation of the COPD Assessment Test. Eur Respir J 2009 Sep;34(3):648-654 [FREE Full text] [doi: 10.1183/09031936.00102509] [Medline: 19720809]
26. 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]
27. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
28. Simmering JE, Polgreen LA, Comellas AP, Cavanaugh JE, Polgreen PM. Identifying patients with COPD at high risk of readmission. Chronic Obstr Pulm Dis 2016 Aug 29;3(4):729-738 [FREE Full text] [doi: 10.15326/jcopdf.3.4.2016.0136] [Medline: 28848899]
29. Jackson BE, Coultas DB, Suzuki S, Singh KP, Bae S. Rural-urban disparities in quality of life among patients with COPD. J Rural Health 2013 Aug;29 Suppl 1:s62-s69 [FREE Full text] [doi: 10.1111/jrh.12005] [Medline: 23944281]

Abbreviations

CAT: COPD Assessment Test
COPD: chronic obstructive pulmonary disease
eHealth: electronic health

Edited by G Eysenbach; submitted 10.11.17; peer-reviewed by B McKinstry, I Graetz; comments to author 30.11.17; revised version received 03.01.18; accepted 03.02.18; published 02.04.18.

Please cite as:
Witry M, Comellas A, Simmering J, Polgreen P
The Association Between Technology Use and Health Status in a Chronic Obstructive Pulmonary Disease Cohort: Multi-Method Study
J Med Internet Res 2018;20(4):e125
URL: http://www.jmir.org/2018/4/e125/
doi:10.2196/jmir.9382
PMID:29610113

Original Paper

Understanding Older People's Readiness for Receiving Telehealth: Mixed-Method Study

Cornelis TM van Houwelingen[1,2*], MSc, RN; Roelof GA Ettema[3,4*], PhD; Michelangelo GEF Antonietti[1*], MSc; Helianthe SM Kort[1,5*], PhD

[1]Technology for Healthcare Innovations Research Group, Research Centre for Healthy and Sustainable Living, University of Applied Sciences Utrecht, Utrecht, Netherlands

[2]Clinical Health Sciences, Faculty of Medicine, University Medical Center Utrecht, Utrecht, Netherlands

[3]Chronic Illnesses, Methodology of Applied Research Research Group, Research Centre for Healthy and Sustainable Living, University of Applied Sciences Utrecht, Utrecht, Netherlands

[4]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

[5]Building Physics and Services, Department of the Built Environment, Eindhoven University of Technology, Eindhoven, Netherlands

[*]all authors contributed equally

Corresponding Author:
Cornelis TM van Houwelingen, MSc, RN
Technology for Healthcare Innovations Research Group
Research Centre for Healthy and Sustainable Living
University of Applied Sciences Utrecht
Heidelberglaan 7
Utrecht,
Netherlands
Phone: 31 641097762
Email: thijs.vanhouwelingen@hu.nl

Abstract

Background: The Dutch Ministry of Health has formulated ambitious goals concerning the use of telehealth, leading to subsequent changes compared with the current health care situation, in which 93% of care is delivered face-to-face. Since most care is delivered to older people, the prospect of telehealth raises the question of whether this population is ready for this new way of receiving care. To study this, we created a theoretical framework consisting of 6 factors associated with older people's intention to use technology.

Objective: The objective of this study was to understand community-dwelling older people's readiness for receiving telehealth by studying their intention to use videoconferencing and capacities for using digital technology in daily life as indicators.

Methods: A mixed-method triangulation design was used. First, a cross-sectional survey study was performed to investigate older people's intention to use videoconferencing, by testing our theoretical framework with a multilevel path analysis (phase 1). Second, for deeper understanding of older people's actual use of digital technology, qualitative observations of older people executing technological tasks (eg, on a computer, cell phone) were conducted at their homes (phase 2).

Results: In phase 1, a total of 256 people aged 65 years or older participated in the survey study (50.0% male; median age, 70 years; Q1-Q3: 67-76). Using a significance level of .05, we found seven significant associations regarding older people's perception of videoconferencing. Older people's (1) intention to use videoconferencing was predicted by their performance expectancy (odds ratio [OR] 1.26, 95% CI 1.13-1.39), effort expectancy (OR 1.23, 95% CI 1.07-1.39), and perceived privacy and security (OR 1.30, 95% CI 1.17-1.43); (2) their performance expectancy was predicted by their effort expectancy (OR 1.38, 95% CI 1.24-1.52); and (3) their effort expectancy was predicted by their self-efficacy (OR 1.55, 95% CI 1.42-1.68). In phase 2, a total of 6 men and 9 women aged between 65 and 87 years participated in the qualitative observation study. Of the primary themes, 5 themes were identified that could provide greater understanding of older people's capacities and incapacities in using digital technology: (1) "self-efficacy and digital literacy," (2) "obstacles to using technology," (3) "prior experience and frequency of use," (4) "sources of support and facilitating conditions," and (5) "performance expectancy." These 5 themes recurred in all 15 observations.

Conclusions: Performance expectancy, effort expectancy, and perceived privacy and security are direct predictors of older people's intention to use videoconferencing. Self-efficacy appeared to play a role in both older people's intention to use, as well

as their actual use of technology. The path analysis revealed that self-efficacy was significantly associated with older people's effort expectancy. Furthermore, self-efficacy and digital literacy appeared to play a major role in older people's capacities to make use of digital technology.

(J Med Internet Res 2018;20(4):e123) doi:10.2196/jmir.8407

KEYWORDS

older adults; videoconferencing; technology; path analysis; observations; community-dwelling people; UTAUT; TAM; self-efficacy; digital literacy

Introduction

Background

The increasing use of digital technology in society requires that all citizens, including older people, have digital literacy. In the Netherlands, people are gradually forced to regulate tasks online, for example, banking or government-related issues, such as tax returns or passport applications. The same applies for health care, in which digital technologies are increasingly integrated, for example, in telehealth, in which health care is delivered remotely through the use of digital technology such as videoconferencing.

On the basis of the belief that telehealth can offer a solution for the increasing number of older people with a (chronic) disease and the accompanying increasing demand for care, the Dutch Ministry of Health formulated in 2014, three ambitions with regard to the use of e-health to be achieved within 5 years: (1) 80% of chronically ill patients have direct access to (parts of) their medical record, (2) 75% of chronically ill patients and vulnerable older people who are willing and able to, actual perform self-measurements, and (3) all community-dwelling patients have the possibility to communicate via videoconferencing with their health care providers [1]. These ambitions require a major change to the current health care situation, in which 93% of the care occurs face-to-face according to a recent poll [2]. These ambitions are based on the technological possibilities of telehealth; however, they raise the question of whether patients, especially older patients, are ready for this new method of care delivery; do older people intend to use videoconferencing and what capacities do they have in using digital technology?

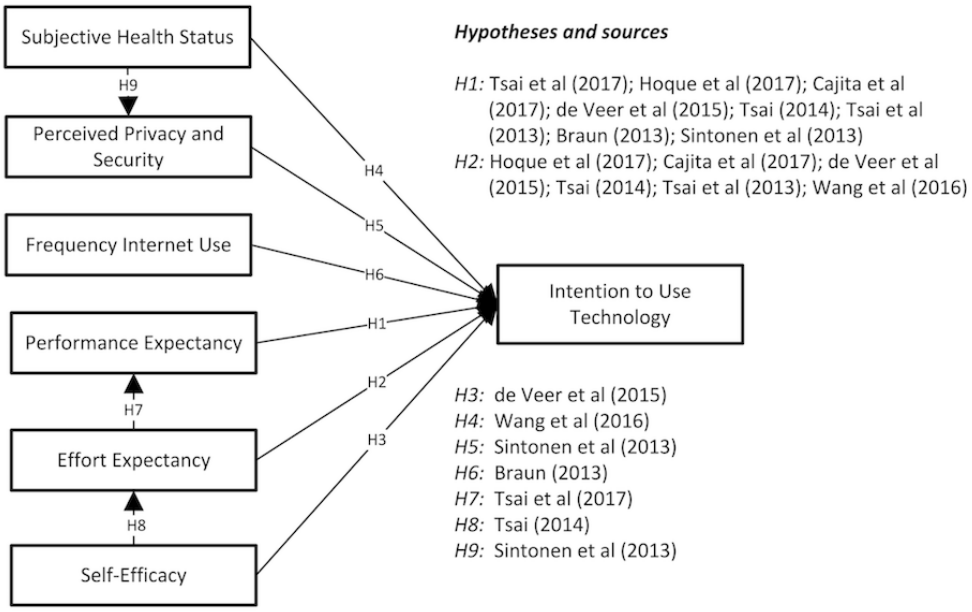
Olson and colleagues [3] showed that there is limited evidence that older adults are averse to using technology, but their frequency and choice of the type of technology often differ from younger adults. Older people are part of another technology generation than younger people and consequently raised with different types of technology (eg, television, radio, telephone) [4] than the technologies that are currently used in health care (eg, using the internet via PCs, notebooks, tablets, including videoconferencing and apps). To facilitate the use of new technologies, Holden and Karsh [5] emphasize the importance of end users receiving sufficient support to ensure that they feel confident in their ability to use these technologies. In health care, nurses have an important role in providing this technological support to patients to enable older people to receive telehealth [6].

Consequently, to support older people in the use of digital technology in health care, we must first understand their readiness to do so, by exploring the factors associated with older people's intention to use digital technologies, such as videconferencing (which is a part of telehealth). Furthermore, it is relevant to explore how older people address technology in their daily life. Several studies with regard to older people's acceptance of technology [7-10] are built on the Technology Acceptance Model (TAM) [11] and the modified version of this model, the Unified Theory of Acceptance and Use of Technology (UTAUT) [12]. The TAM [11] and UTAUT [12], however, provide neither a deep understanding of the relations and interactions between factors and nor insight into the capacities of community-dwelling older adults to use digital technology. This insight, however, is needed to understand older people's readiness to receive telehealth.

Purpose of the Study

This study aims to obtain a deeper understanding of community-dwelling older people's readiness to receive telehealth by studying older people's intention to use videoconferencing and capacities or incapacities to use digital technology in daily life as indicators. Since individual's intention to use technology can substantially differ from their actual behavior [13], both intention and actual use of technology are examined in this study. This knowledge could benefit health care professionals' abilities to assist older people in using technology and enable older people to benefit more from novel technology that supports them in aging in place. To achieve this, the following steps were taken and reported in this study: (1) literature review to build a theoretical framework of intention to use technology, (2) testing the framework, (3) collection of data on older people's capacities or incapacities to use digital technology, (4) synthesis of all results, and (5) conclusions and implications for older people's readiness to use telehealth.

Figure 1. Theoretical framework. This framework displays the factors associated with older people's intention to use technology. Each hypothesis is based on prior research, as shown. H=hypothesis.



Literature Review—Older People's Intention to Use Technology and Associated Factors

In 2017, a literature search was performed to build a theoretical framework of older people's intention to use technology and associated factors. Therefore, the search terms "older people," "technology," "intention," and "factors," as well as alternative terms, such as "seniors" and "associations," were used for a search in CINAHL, Google Scholar, PsycINFO, MEDLINE/PubMed, ScienceDirect, Scopus, and Web of Science. To scope the literature review, inclusion criteria were as follows: (1) target group with a median or mean age of 65 or older, (2) publication date less than 10 years ago, (3) peer-reviewed original research, (4) quantitative studies in which "intention to use technology" was tested as a dependent variable, and (5) studies written in English. The process and results of the literature search, in terms of search strings, number of hits, and number of selected studies, are shown in the Multimedia Appendix 1.

Theoretical Framework of Intention to Use Technology

Of the 249 studies that were found in September 2017, only 29 studies met the criteria. After duplicate studies were filtered out, 11 studies remained [8,14-23], and these 11 studies were used to build the theoretical framework on older people's intention to use technology (see Figure 1). On the basis of the 11 included studies, the theoretical framework shows six predictors of older people's intention to use technology: performance expectancy, effort expectancy, self-efficacy, subjective health status, perceived privacy and security, and frequency of internet use. The operationalization of these predictors is presented in Table 2 in the Results section. All hypotheses with regard to older people's intention to use technology and the related sources of evidence are illustrated in Figure 1.

Methods

Design

A mixed-method triangulation design [24] was used, including a cross-sectional survey study (phase 1), generating quantitative data concerning older people's intention to use videoconferencing, and an observational study (phase 2), yielding qualitative data concerning their actual use, by observing their capacities in using technology in daily life. In phase 1, we focused on "videonferencing," a relatively new technology (in health care), that is often used in telehealth services. To observe a representative sample in phase 2, we focused in this phase on more traditional, commonly used technologies. If we adhered to videoconferencing, we could have observed only those older people who use videoconferencing, which might have introduced a selection bias. Both insights, intention to use and actual use of technology, are important to understand older people's readiness to receive telehealth.

Phase 1: Cross-Sectional Study on Older People's Intention to Use Videoconferencing

Setting and Participants

Participants were recruited in September 2012 for the cross-sectional study. Older people were invited to fill out a paper version of the survey through 2 patient advocacy organizations, 2 senior social clubs, 5 health care organizations, and a senior information day in Utrecht. Additionally, to reach a large group of potential participants, a Web-based panel of approximately 2000 clients was invited to fill out an online version of the survey.

Two inclusion criteria, (1) independently living at home and (2) being 65 years of age or older, were maintained. To estimate

the required sample size, we followed the rule of thumb for multiple regression analysis, with the purpose of building a prediction framework: maintain a ratio of 10 positive cases in the dataset to 1 predictor variable in the full path analysis [25]. Since the full path analysis contains 6 predictors, a sample with at least 60 positive cases was required. The dependent variable "intention to use" (measured on a 5-point scale, with 1=totally disagree, 2=disagree, 3=neutral, 4=agree, and 5=totally agree) was used to calculate the number of positive cases by labeling participants with a response from 1 to 3 as "non-cases" and participants with a response from 4 to 5 as "positive cases," resulting in 70 positive cases.

Cross-Sectional Survey

We collected data using a survey to test the hypotheses illustrated in the theoretical framework. The outcome measured was set as "intention to use videoconferencing," aligned with the ambitions of the Dutch ministry in which the use of videoconferencing is an important part of telehealth.

The survey included items covering the following topics: demographic questions (eg, age, gender, and educational level); health-related questions (eg, health status) based on Czaja et al [26]; technology experience in daily life (with, eg, internet, computer) based on Czaja et al [27]; and older people's perception of videoconferencing (eg, performance expectancy, effort expectancy, and intention to use videoconferencing) based on Chang and Hsu [28] and Gagnon et al [29]. All constructs regarding older people's perception of their health and perception of videoconferencing were measured with multiple statements.

Table 2 lists all constructs (predictors and dependent variables) and related statements that were used to test our theoretical framework. The survey items that were based on previously developed and used questionnaires were translated from English into Dutch and cross-translated. Subsequently, the content of the survey was discussed with experts in aging and technology who were selected from our network and pilot-tested among a representative group of older people to determine the readability and comprehensibility.

Statistical Analysis

Missing values were substituted using the 5-time multiple imputation method to reduce bias [30]. The results of the statistical analysis of each of the 5 imputed dataset were pooled using Rubin's rule [31].

The internal consistency of the constructs (eg, self-efficacy, performance expectancy) was assessed with Cronbach alpha, considering Cronbach alpha values between .70 and .95 to be "good" [32]. For our multilevel regression analysis, the variable "frequency of internet use" was dichotomized, using a data-driven method to select an appropriate cut-off point. In the survey, participants were asked: "on average, how many hours per week do you use the internet?," whereby, 0=not, 1=0 to 1 hours, 2=1 to 5 hours, 3=5 to 10 hours, 4≥10 hours. The cut-off point was set at 2, meaning 0=less than 5 hours a week and 1=5 or more hours per week.

All hypotheses, as illustrated in our theoretical framework (Figure 1), were tested using a path analysis approach [33]. In the full path analysis, 4 outcome variables, that were interconnected, were tested at once (see Figure 1): Outcome 1 "intention to use videoconferencing" predicted by (1) "performance expectancy," (2) "effort expectancy," (3) "self-efficacy," (4) "subjective health status," (5) "perceived privacy and security," and (6) "frequency of internet use"; Outcome 2 "performance expectancy" predicted by "effort expectancy"; Outcome 3 "effort expectancy" predicted by "self-efficacy"; and Outcome 4 "perceived privacy and security" predicted by "perceived privacy and security." In this way, the effects of group-level predictors were deliberately confounded with the effects of the group variables, thus upholding the hierarchical structure. As such, the effects of both levels of variables could be estimated. All three outcome variables and accompanying predictors in the path analysis were based on the literature mentioned in the introduction. A significance level of .05 was used to determine whether predictors had a significant association with the dependent variable.

Additionally, starting with the full path analysis (Figure 2), a backward selection procedure was performed; at each step, the variable with the highest *P* value was excluded first. Simultaneously, at each step, the goodness of fit of the framework was examined using the Akaike Information Criterion (AIC) [34] to assess the performance of each model compared with the initial framework. Following this procedure, a final path analysis (Figure 3) was reached with only significant (significance level of .05) predictors and using the AIC as a threshold. In this final framework, the themes derived from the qualitative data were included.

We used the statistical package R (version 3.4.2; 2017-09-28; The R Foundation for Statistical Computing) for the path analysis. All other statistical analyses were performed using SPSS (IBM Corp Released 2016, IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp).

Phase 2: Observations on Using Technology in Daily Life

Qualitative Approach and Research Paradigm

Given the relative lack of exploratory studies on community-dwelling older people's capacities or incapacities to use technology, we performed a conventional content analysis [35] on the qualitative data obtained from observations of older people using digital technology at their home. Our approach was constructivist, using an interpretative phenomenological epistemology [36] based on the notion that there is not one "truth" in regard to the phenomenon of technology use.

The observations were executed by the third author and by third-year bachelor's degree-level students with backgrounds in nursing, health care management and Cesar exercise therapy. Before their observations, these students received training from our research team on how to perform observations. None of the observers were known to the participants before the observations being performed. All observations were discussed by our multidisciplinary research team. As members of the research team, the first author of this paper (CvH) has a background in

nursing, sociology, and nursing education; the second author (RE), in nursing, nursing science, and epidemiology; the third author (MA), in human movement science and nursing education; and the fourth author, in medical biology, built environment, and gerontechnology (HK).

Context, Setting, and Sampling

In 2012, between September and December 2012, we conducted the study among community-dwelling older people (65 years of age or older). The sampling started in phase 1, along with the survey, and participants were informed about the opportunity to also volunteer in phase 2 (the observations). When the participants were willing to participate in phase 2, they filled out a form with contact details, which they returned to the research team together with their completed survey. For phase 2, the same inclusion criteria were held as in phase 1: (1) independently living at home and (2) being 65 years of age or older. However, we also selected based on both "experience with a computer" and "experience without a computer," with the aim of being able to observe participants with and without computer experience. After 15 cases were observed, theoretical saturation was reached; sufficient data were collected to understand the concepts of our study.

Data Collection and Processing

To facilitate consistency over different sets of observations, a list of day-to-day technological tasks was composed before the data collection. The tasks were developed by the third and last authors during a 2-day workshop with our American research partners (details described under "Acknowledgments"), which resulted in the following 8 categories: (1) computer basics (eg, "create a new folder on your desktop"), (2) email (eg, "send an email with an attachment"), (3) use of the internet (eg, "show a map of your town"), (4) television (eg, "change the volume of your TV"), (5) mobile phone (eg, "show how to save a contact in your contact list"), (6) household (eg, "make popcorn in a microwave"), (7) health (eg, "show how to use a digital scale"), and (8) videoconferencing (eg, "show how to start videoconferencing with your nurse").

Without a specific time constraint, the observant followed this list of tasks and encouraged the older adult to accomplish the task independently. The older people could ask for a hint when they could not proceed with the task. The observant encouraged the older people to think aloud during the performance of their technological tasks. During the direct observations, notes were made using a form with space for notes for each of the tasks and blank space for other possible remarkable occurrences. These notes were used during the iterative analysis process.

All observations were audio-recorded and lasted 1 hour on average. The audio-recordings were transcribed and anonymized. All transcripts were stored, coded, and analyzed in MAXQDA (software for qualitative data analysis, 1995-2016, VERBI Software—Consult—Sozialforschung GmbH, Version 12.2.1, Berlin, Germany).

Data Analysis

Data analysis followed the steps for conventional content analysis as outlined by Hsieh and Shannon [35], a method to describe a phenomenon—in this case, the daily use of digital technology by older people. Through an iterative process of coding, by discussing findings in the light of the literature, the research team identified and described the most prevalent themes with regard to older people's day-to-day use of technology. The concepts derived from our theoretical framework in (Figure 1) were used as "sensitizing concepts," defined as "interpretive devices and as a starting point for qualitative research" [37]. In addition to the concepts derived from Figure 1, the contextual factors in the use of technology, as described by McFarland and Hamilton [38], were also used as a starting point for analysis (eg, task structure, prior experience, other's use).

Although these 2 frameworks were used, we conducted an open, inductive analysis, starting with open coding. To enhance trustworthiness, CvH and MA coded the verbatim transcripts independently, with a focus on the sensitizing concepts and the main question: "How do older people struggle with digital technology use and what supports them?" The first round of coding by two of the authors (CvH and MA) resulted in 1022 coded text segments. Then, these open codes were discussed among all authors to organize and group the codes into meaningful clusters. After this discussion, 157 text segments were considered irrelevant. The remaining 881 text segments and related codes were clustered and categorized. We searched for themes that occurred in each observation with all participants. Eventually, we achieved consensus on the primary themes observed in the data.

In the last phase, definitions for each theme were developed and provided with illustrative examples or quotations from the data. Quotations in this study were translated from Dutch into English. During the whole analysis, we kept in mind that we were looking for information that could eventually benefit nurses' in assisting older people to use technology in health care. To illustrate the qualitative results, the themes and subthemes were drawn in Figure 3.

Ethical Approval

This research was conducted following Dutch human subject regulations. Since the Dutch Medical Research Involving Human Subjects Act did not apply to either phase 1 or phase 2 of this study, no official ethical approval was required. Nevertheless, all necessary precautions were taken to protect the anonymity and confidentiality of our participants. The Dutch Medical Research Involving Human Subjects Act applies to medical research "if there is an infringement of the physical and/or psychological integrity of the subject" [39].

Cliëntenbelang Utrecht (an organization that defends the interest of health care clients) approved the study and provided access to the client panel. All participants were informed with a letter containing information about the purpose of the study. Participants were informed that their participation was voluntary, that they were free to decline or discontinue their participation at any time and that their responses were processed anonymously and only used for research purposes. No person identifying information was collected.

Results

Phase 1: Cross-Sectional Study on Older People's Intention to Use Videoconferencing

Characteristics of Study Population

In total, 288 older people filled out the questionnaire on paper or online. Of these individuals, 22 were excluded since they were younger than 65 years of age. Of the 256 cases left, 50.0% (128/256) were male and 50.0% (128/256) were female, with a median of 71 years (Q1-Q3 67-76). A minority (13.7%, 35/256) of participants had experience with videoconferencing, of whom approximately half had less than 1 year of experience, while the other half had more than 1 year. The majority (71.1%, 182/256) completed an average or high level of education. Of the 256 cases, 21.1% (54/256), missed one or more questions that were used for this study. Their missing values were substituted using the 5-time multiple imputation method. All demographic details of the participating older people are listed in Table 1.

Descriptive Results and Consistency of the Research Constructs

The internal consistency of the 6 constructs was "good," with a Cronbach alpha of more than .70 [32]. All grouped items and accompanying median scores, 1st and 3rd quartile ranges, and Cronbach alphas are presented in Table 2.

Results of the Path Analysis

Using a significance level of .05, the multilevel path analysis revealed that 5 of the 9 hypotheses regarding older people's perception of videoconferencing were supported. On level 1, older people's intention to use videoconferencing was significantly predicted by their performance expectancy (odds ratio [OR] 1.26, 95% CI 1.13-1.39), effort expectancy (OR 1.23, 95% CI 1.07-1.39), and perceived privacy and security (OR 1.30, 95% CI 1.17-1.43). In our sample, self-efficacy (OR 1.09, 95% CI 0.94-1.23), subjective health status (OR 0.90, 95% CI 0.79-1.01), and frequency of internet use (OR 1.03, 95% CI 1.42-1.68) were not significantly associated with older people's intention to use videoconferencing.

On level 2, older people's performance expectancy was predicted by their effort expectancy (OR 1.38, 95% CI 1.24-1.52). On level 3, their effort expectancy was predicted by their self-efficacy (OR 1.55, 95% CI 1.42-1.68). Our last hypothesis, on level 4, was not supported: older people's perceived privacy and security was nonsignificantly predicted by their subjective health status (OR 1.05, 95% CI 0.95-1.16). The complete path analysis and unstandardized regression coefficients, from which the ORs were derived, are illustrated in Figure 2, with intention to use videoconferencing as the main outcome variable.

Table 1. Demographic characteristics of participating older people (n=256; paper participants [n=70] and online participants [n=186]). N/A: not applicable.

Characteristics	n (%)
Gender	
Male	128 (50.0)
Female	128 (50.0)
Age by category (in years)	
65-74	182 (71.1)
75-84	67 (26.2)
>85	7 (2.7)
Median age=71 (Q1-Q3=67-76)	N/A
Experience with the use of video conferencing	
Yes	35 (13.7)
No	221 (86.3)
Highest completed educational level	
Lowest (primary education)	10 (3.9)
Low (lower secondary education)	57 (22.3)
Average (general or vocational upper secondary education)	70 (27.3)
High (postsecondary nontertiary education)	119 (46.5)

Table 2. Constructs of the path analysis: internal consistency and median scores. N/A: not applicable.

Construct and related items	Cronbach alpha[a]	Median (1st quartile-3rd quartile)
Subjective health status (predictor variable)	.87	3.0 (2.3-3.3)
1. In general, I would say my health is[b]		3.0 (2.0-3.0)
2. Compared with other people of my age, I would say my health is[c]		3.0 (2.0-3.0)
3. How satisfied are you with your present health?[c]		4.0 (3.0-4.0)
Performance expectancy (predictor and outcome variable)	.72	3.3 (3.0-4.0)
1. By using videoconferencing, I can live longer in my own home independently[d]		4.0 (3.0-4.0)
2. The use of videoconferencing will give me more freedom[d]		3.0 (3.0-4.0)
3. The use of videoconferencing will enhance my self-reliance[d]		3.0 (3.0-4.0)
Effort expectancy (predictor and outcome variable)	.85	3.8 (3.0-4.0)
1. I think videoconferencing will be clear and easy to use[d]		4.0 (3.0-4.0)
2. Videoconferencing will be easy to operate and use[d]		4.0 (3.0-4.0)
3. Videoconferencing will be easy to learn[d]		4.0 (3.0-4.0)
4. Videoconferencing will have a clear guide for operation[d]		4.0 (3.0-4.0)
Self-efficacy (predictor variable)	.77	4.0 (3.4-4.2)
1. I am confident enough to use videoconferencing[d]		4.0 (3.0-4.0)
2. Given an appropriate training, I will have the ability to use videoconferencing[d]		4.0 (3.0-4.0)
3. I possess the necessary skills to learn how to use videoconferencing[d]		4.0 (3.0-4.0)
4. I am afraid I will not learn how to use videoconferencing[e]		4.0 (4.0-5.0)
5. I think I will find it hard to acquire the necessary skills to use videconferencing[e]		4.0 (3.0-5.0)
Perceived privacy and security (predictor and outcome variable)	.79	3.3 (2.8-3.7)
1. My feeling of security is higher with the use of videoconferencing[d]		3.0 (3.0-4.0)
2. With the use of videoconferencing my feeling of security will be higher[d]		3.0 (3.0-4.0)
3. The possibility of immediate contact with a health care professional will give me a safe feeling[d]		4.0 (3.0-4.0)
4. The use of videoconferencing is confidential[d]		3.6 (3.0-4.0)
5. I will have no problems with the idea that videoconferencing consultations are saved[d]		3.0 (2.0-4.0)
6. The use of videoconferencing will not influence my feeling of privacy[d]		3.0 (2.4-4.0)
Frequency of internet use[f] (predictor variable)	N/A	N/A
Intention to use videoconferencing (outcome variable)	.76	3.5 (2.8-4.0)
1. I am willing to use videoconferencing to complement my traditional care[d]		3.0 (2.4-4.0)
2. I have the intention to use videoconferencing routinely to receive care[d]		3.0 (2.0-4.0)
3. I intend to use videoconferencing when this is necessary to receive care[d]		4.0 (3.0-4.0)
4. After an appropriate training, I am willing to use videoconferencing[d]		4.0 (3.0-4.0)

[a]Cronbach alpha between .70 and .95 is "good" [32].

[b]Likert scale ranging from 1="poor" to 5="excellent."

[c]Likert scale ranging from 1="not satisfied at all" to 5="very satisfied."

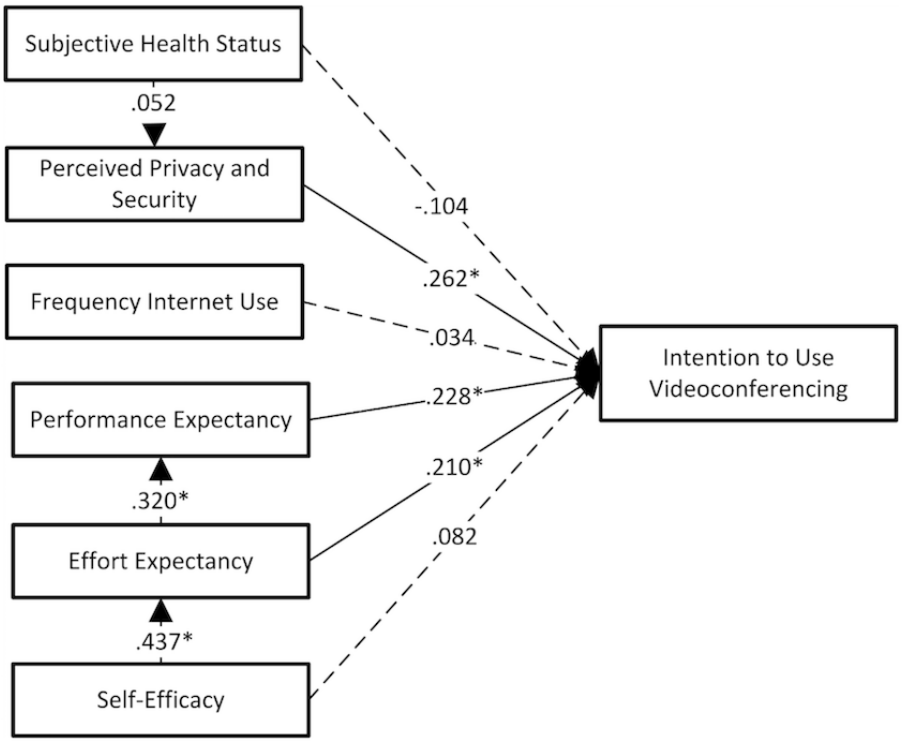
[d]Likert scale ranging from 1="totally disagree" to 5="totally agree."

[e]Likert scale ranging from 1="totally agree" to 5="totally disagree."

[f]Participants were asked: "on average, how many hours per week do you the internet?" 0: not, 1: 0-1 hours, 2: 1-5, 3: 5-10 hours, 4: >10 hours. For the

path analysis, this variable was dichotomized, using a data driven method to select an appropriate cut-off point. The cut-off point was set at 2, meaning 0=less than 5 hours a week and 1=5 or more hours per week.

Figure 2. Older people's (n=256) intention to use videoconferencing and associated factors. Unstandardized regression coefficients are shown, derived from the path analysis. Estimates were pooled from the results of the analysis of 5 imputed datasets using Rubin's rules. *Significant association, using a significance level of .05 (dotted-line indicates nonsignificance).



Phase 2: Observations on Using Technology in Daily Life

Characteristics of Observed Older People

Of the survey population of 256 older people, 16 older people volunteered to take part in phase 2 of the study: with observations conducted at their homes while they executed technological tasks. The quality of one of the audio recordings was too poor to be able to create a verbatim transcription, leaving 15 observations suitable for qualitative analysis.

Among the observed older people were 6 men and 9 women. Their age ranged from 65 to 87 years (mean=73.21, SD=6.59, 1 missing value). Of the participants, 7 had low levels of education (lower secondary education), 2 received average-level education (general or vocational upper-secondary education), and 4 completed high-level education (bachelor's degree or higher, 2 values missing). Approximately half of the participants did not have a computer (n=7), while the other half did use a computer (n=8).

Understanding Older People's Capacities to Use Technology

In all, 5 primary themes were identified that could help us understand older people's capacities and incapacities in using digital technology (ordered by frequency of occurrence): (1) "self-efficacy and digital literacy," (2) "obstacles to using technology," (3) "prior experience and frequency of use," (4) "sources of support and facilitating conditions," and (5) "performance expectancy." These 5 themes were observed among all 15 participating older people and included 865 of the 1022 coded text segments. Within these 5 primary themes, several subthemes were identified, which are described below and illustrated with exemplary quotations.

Theme 1: Self-Efficacy and Digital Literacy

In phase 2, "self-efficacy and digital literacy" was the most prevalent theme that appeared to play a role in the day-to-day use of technology by the older people we observed. "Self-efficacy" refers to an individual's belief in his or her ability to accomplish a certain task in a specific situation [40]. We observed many situations in which older people expressed low self-efficacy regarding technology use, but approximately the same number of situations occurred in which high self-efficacy was expressed. The following conversations between a man and his wife illustrate the low self-efficacy of the man and obstacles he experienced with his computer. The conversation between the man and his wife started after the participant (man) was asked to open his email, and to be able to use his email, the participants had to turn on his computer first:

> *Man: [With e-mails] I do nothing. I'm a "digital illiterate." [To his wife:] You always have a note attached, don't you? It's not there, so I know nothing.*
>
> *Wife: Can you turn it on, or not?*

Man: I don't know, something with green, whether it's the right or left button. Was it something green?

Wife: Just try it.

Man: Nothing happens.

Wife: You have to push it longer. Do you hold it the other way around?

Man: Nothing happens at all.

Wife: You do have to hold it longer, the red button.

Man: There is nothing red at all.

Wife: No, it isn't red yet, you have to hold it longer. [Participant 2, male, 81 years]

The theme of self-efficacy occurred in a variety of ways during the observations; older people believed that they were not able to accomplish certain technological tasks (low self-efficacy), but discovered that they actually were able to do so or could do so after a small suggestion on how to proceed. Further, we observed older people who could explain very clearly how an application or device worked and were proud that they possessed the right skills, for example:

I think it's already good that I am able to open my e-mail and send e-mails back. [Participant 9, female, 72 years]

Another recurrent observation was older people who kept very strictly to the things that they had learned and stayed away from abilities outside of their knowledge. For example, one participant said the following about his email application:

I never look over there, I just do everything I have learned. [Participant 2, male, 81 years]

The same participant added:

Outside of that [email application], I become nervous. [Participant 2, male, 81 years]

"Digital literacy" refers to "a large variety of complex cognitive, motor, sociological, and emotional skills, which users need in order to function effectively in digital environments" [41]. During the observations, while older people were executing technological tasks, almost all participants experienced their limited digital literacy. This limited "digital literacy" impacted their technology use in several ways: (1) the functionality of a device was only partially used since participants did not understand how to use several functions or how to use the required buttons and (2) when a new device was bought, everything had to be learned from the start, as exemplified by one of the participants:

But I notice that I'm not so good at electrical devices, so this [task] has to go very slow. [...] Yes, I remember, I was in the store and I touched it but I did not know how it worked anyway. I actually felt like a "dummy." And I was reluctant, but he [salesman] explained me how to put that thing on/off. He said, "try to do it." And there are also things that I could execute at that moment, but not anymore [once at home]. And then I have to ask again how it works. [Participant 9, female, 72 years]

Task Structure

Within the theme of self-efficacy and digital literacy, two subthemes were identified: "task structure" and "effort expectancy." Regarding the task structure, which is referred to in the literature as the extent to which a task is nonroutine and varied [38], we observed several older people who used the functionality of a device only partially and, as a result, did not benefit from all the possibilities the device offered. One male participant, for example, stated that he only *reads* emails but never responds:

In the past, I've had to type sometimes, but that's way too difficult, so I only read e-mails. As long as I have her [his wife], she does that. [Participant 2, male, 66 years]

Another participant explained that she only uses her cell phone in specific situations:

I only use it [cell phone] when I visit my son. When I sit in the train, I call my son and ask him to pick me up. But besides that, I never take my phone outside. [Participant 13, female, 70 years]

Effort Expectancy

We also gained insight into the role of "effort expectancy," defined in the literature as "the degree of ease associated with consumers' use of technology" [12]. Several participants were complaining about the nonease of use of the technologies they used while executing the technological tasks. One woman, for example, talked about the difficulty of saving a number in the contact list of her cell phone:

...[in order to save a contact] I have to search a lot, but I will get it done. Please wait, this is very illogical [...]. Very illogical. I hope future devices are smarter. [Participant 11, female, 76 years]

Another example of how effort expectancy plays a role in the use of technology came from a participant who prepared himself for executing a task with his cell phone. Seemingly easy functions can already be difficult:

I first have to turn it [cell phone] on. That's always a bit tricky. Especially my wife has difficulties finding that on button. [Participant 7, male, 74 years]

Theme 2: Hurdles to Using Technology

Obstacles

Older people experience all kinds of obstacles to using technology, also referred to in the literature as "barriers" [42,43], which are elements that hamper their use of technology. We observed obstacles in diverse categories. At first, technical obstacles presented themselves, for example, the disruption of internet service, a broken button, a slow-running computer, or a stuttering connection while videoconferencing.

Furthermore, we observed obstacles in the category "limited digital literacy," for example, unable to find the cursor (of the mouse), getting confused after updates, or not knowing how to use the internet. The third category included more personal use-related barriers, for example, prefers to read the news in the newspaper instead of on an iPad, forgets his or her password

very often, or having resistance toward social media, as expressed by one participant:

> *Wearing a personal alarm around my neck is fine with me, but [...] Facebook and whatever else there is, is another reality beyond my sensory reality.* [Participant 11, female, 76 years]

Anxiety

Additionally, in 12 cases, the subtheme "anxiety" was identified as an obstacle to using technology. McFarland and Hamilton [38] use a slightly different term, namely "computer anxiety," which they describe as "an individual's uneasiness or apprehension toward computers." During our observations, a variety of anxiety-inducing sources arose related to the use of technology, including (1) receiving spam, (2) experiencing system updates, (3) losing written text, (4) damaging a device, (5) fearing the use of technology in general, (6) fearing microwave radiation, (7) fearing inadequate privacy protection, (7) feeling unsafe using the internet, and (8) fearing online scams or cyber criminals. Regarding the last 2 obstacles mentioned, one participant expressed her fear of online banking:

> *One hears so much...things that can go wrong with online banking. I dedicated myself to, if possible, only do online banking when one of my two children is with me.* [Participant 9, female, 72 years]

All anxiety sources mentioned above hampered the participants' use of technology.

Theme 3: Prior Experience and Frequency of Use

While executing technological tasks, the theme of prior experience and frequency of use was exhibited by all participants. We observed people with much experience and little experience, as well as participants who told to have a device but reported never using it (eg, did not use their cell phone since they already had their landline telephone).

The capacities and incapacities regarding technology use seemed to be associated with older people's experience in the past and/or their frequency in use. Some participants said they were glad that they learned to work with computers during their working career. Others did not and had to learn everything from the start. Their limited experience hampered their capacity to accomplish technological tasks:

> *I really don't know how it works, I just have it [computer].* [Participant 14, female, 68 years]

In several cases, participants had forgotten how to accomplish a specific task since they reported only doing it once or twice in the past.

In contrast, more prior experience was clearly supportive:

> *This isn't really complicated, since I already have been working with that computer for 2 years now.* [Participant 6, male, 86 years]

According to the participants in this study, capacity in using technology is a matter of experience and practice.

Theme 4: Sources of Support and Facilitating Conditions

Facilitating Conditions

When participants had to overcome obstacles to technology use, they reached out to various sources of support. These sources of support are part of the "facilitating conditions," defined as "consumers' perceptions of the resources and support available to perform a behavior" [12]. A variety of sources of support came up, such as following a computer course via SeniorWeb, a very important Dutch forum according to one of the participants:

> *SeniorWeb is really important, but I wonder if people take that step. [...] For me, it's amazing to see that, myself included, my family, brothers and sisters encounter the same [obstacles].* [Participant 8, male, 65 years]

Further sources of support that were mentioned by our participants were manuals, helpdesks, installers, and persons, often including partners, friends, children, and grandchildren.

Significant Role of Children

Children and grandchildren play a significant and diverse role in the use of technology by older people; they appeared to function as a motive or incentive to start using technology, for example, since technology offers the ability to communicate more easily (and over distance). Subsequently, children help their parents in purchasing, installing, and using technology. The active support of children solves issues in the use of technology on the one hand while on the other hand, it might cause older people to maintain their lack of technology skills. When they struggle with technology, some older people wait for their children to solve it:

> *I'm not good at saving a number. My grandchildren always come to do that.* [Participant 14, female, 68 years]

Another example came from a woman who was asked to send an email to multiple persons:

> *My children once said, "just put all those names here" but I don't have a clue of the meaning of all this. [* [Participant 13, female, 70 years]

Theme 5: Performance Expectancy

Performance Expectancy

"Performance expectancy," a well-known construct in technology-acceptance theories, refers to "the degree to which using a technology will provide benefits to consumers in performing certain activities" [12]. Our participants mentioned benefits in various categories: (1) leisure, for example, playing games, reading books, and using street view; (2) increasing communication possibilities, for example, (also mentioned earlier) with family or nurses; and (3) aging in place, as illustrated by the following statement:

> *I'm already thinking of what do I need to have? What do I have to do, so in about 10 years...what do I need in order to be able to live at home as long as possible?* [Participant 8, male, 65 years]

Task-Technology Fit

Within the performance expectancy theme, a recurrent statement was that the technology must fulfill a need. This idea is close to the construct of "task-technology fit," which refers to the assumption that "performance impacts will result from task-technology fit—that is, when a technology provides features and support that 'fit' the requirements of a task" [44]. Sometimes an event occurred in the lives of our participants that caused a certain technology to suddenly fit their needs, as illustrated in the following statement:

> *This tablet...I bought it because I like to read. And now, my eyes have become so bad that I can't read books anymore [from paper]*. [Participant 9, female, 72 years]

The same participant explained her motivation to purchase a computer:

> *I had to do financial matters, and at that moment, I took a computer.*

Discussion

Principal Findings

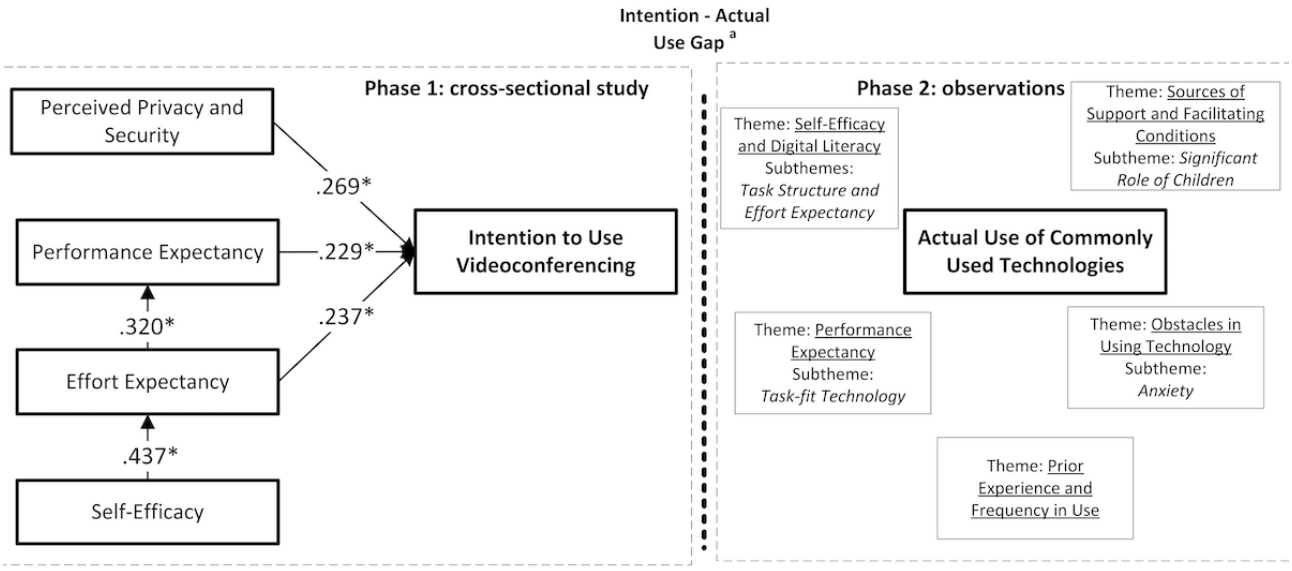
In this study, 7 significant associations regarding older people's perception of videoconferencing were found. Older people's (1) intention to use videoconferencing was predicted by their performance expectancy, effort expectancy, and perceived privacy or security; (2) their performance expectancy was predicted by their effort expectancy; and (3) their effort expectancy was predicted by their self-efficacy. In other words, whether older people intend to use videoconferencing depends on their expectations of the usefulness of this application, their expectations of how easy it is to use videoconferencing, and their confidence whether their privacy and security is protected when using videoconferencing.

Self-efficacy did not appear to be a significant predictor of older people's intention to use videoconferencing. However, the multilevel regression analysis made it possible to identify multiple associations within the path analysis and showed us that self-efficacy significantly impacts older people's effort expectancy of technology, which in turn impacts older people's intention to use videoconferencing. Since self-efficacy and effort expectancy can be quite comparable [45], we executed as a kind of sensitivity analysis the path analysis without effort expectancy, which showed a significant association between self-efficacy and intention to use technology.

Self-efficacy and digital literacy was also identified as the most prevalent theme during the observations in phase 2. Four additional themes were identified that could help us understand older people's readiness to receive telehealth: "obstacles to using technology," "prior experience and frequency in use," "sources of support and facilitating conditions," and "performance expectancy."

Two of the themes, self-efficacy and performance expectancy, were also part of our theoretical framework and path analysis on intention to use technology. Additionally, the construct effort expectancy was observed within the theme "self-efficacy and digital literacy." The qualitative results indicate that older people's use of technology is associated with the themes we found. It is interesting to test in future research whether these themes (eg, facilitating conditions, prior experience, task structure) are also associated with older people's intention to use. In our path analysis, frequency of internet use did not appear to be a significant predictor of intention to use, but perhaps (prior) experience with other types of technology does have a significant association with older people's intention to use. Figure 3 summarizes the findings of both the constructs of the path analysis (phase 1) and the themes and subthemes derived from the observations (phase 2).

Figure 3. Understanding older people's intention to and actual use of technology. A mixed-method framework of a multilevel regression path analysis (n=256) and qualitative observations (n=15). *Significant (alpha .05) associations; unstandardized regression coefficients are shown. The letter "a" denotes that this "Internet—actual use gap" was based on prior research.



Integration With Prior Research

Figure 3 illustrates how the themes found in phase 2 are related with the subthemes. On the basis of prior research, one can argue that there are more interactions within this framework to explore. Sponselee [46], for example, describes that family support positively impacts users' frequency of use. Subsequently, the frequency of use might positively impact older people's self-efficacy since performance accomplishments and successes that raise mastery expectations are seen as the strongest methods of increasing self-efficacy [47]. Another association that might be useful to explore in further research is that between facilitating conditions and obstacles. During observations, we learned that when participants had to overcome obstacles to technology use, they reached out to various sources of support, which differed from person to person depending on the level of the individuals' facilitating conditions.

Additionally, older people's motivation to start using or purchase technology can substantially differ, illustrated by one of our participant who explained her motivation to purchase a computer: "I had to do financial matters, and at that moment, I took a computer." This finding is in line with Peek [48] who concludes that improving older people's acceptance of technology requires, among other things, an understanding of the specific needs and circumstances of the targeted individual. Peek [48] also emphasizes that the acceptance of technology by older people is a dynamic process; specific events that occur in an individual's life can trigger the need of using technology.

Regarding the predictors we found, effort expectancy and performance expectancy were already known from the TAM [11] and UTAUT [12], as well as observed in other health care-related studies [8,17] and from health care providers [49]. What this study adds to the TAM [11] and UTAUT [12] is that older people's intention to use videoconferencing also can be predicted by their perceived privacy or security. Furthermore, the multilevel regression shows that effort expectancy was predicted by self-efficacy, and performance expectancy was predicted by effort expectancy. Our findings concur with those from other research studies [50] that emphasized the shortcomings of the common TAMs with regard to obtaining a deep understanding of older people's readiness for using technology. By using a mixed-method design, this study shows (in phase 2) how some of the constructs of the path analysis regarding older people's intention to use videoconferencing (ie, performance expectancy, effort expectancy, and self-efficacy) also play a role in the day-to-day situation of older people when they are using technology.

Contrary to the findings of prior research [51-53], "subjective health status" in our study was not found to be a relevant theme for older people's technology use, neither in their intention to use videoconferencing, as shown in Figure 2, nor during the observations. Moreover, the performance of the path analysis model enhanced considerably (on the basis of the AIC) after we excluded subjective health status in Figure 3. Zimmer and Chappell [54] drew a comparable conclusion. In their study, older people's self-assessed health was not significantly associated with their receptivity to new technology. Thus, caution is required when linking older people's subjective health status to their intention to use technology.

Within the theme "sources of support and facilitating conditions," the significant role of family members was identified. This observation is aligned with the prior research of Luijkx et al [55]. In this interview-study, Luijkx and colleagues emphasized the importance of including family members when implementing technology into the lives of older people and described that especially grandchildren can positively influence the acceptance of technology. Peek and colleagues [51] added that older people sometimes are afraid to burden their children with technology-related questions. This could also have played a part in one of our observations, in which an older person told us: "My children once said, 'just put all those names here' but I don't have a clue of the meaning of all this." When it comes to the role of family members, we observed an ambivalent mechanism; in accordance to prior research, family members can generate enthusiasm for using technology among older people, but at the same time, family members can also hamper the digital literacy of older people by taking over their technological tasks, which foregoes the opportunity for older people to become more skilled with using the technology.

Study Limitations and Strengths

Our sampling strategy might have been a study limitation. Since the total number of potential respondents was not known, we could not measure a response rate and may have thus missed this indicator of representativeness. Only for those respondents who were recruited via the e-panel (n=186) we could, resulting in a response rate 9.30% (186/2000), which is low [56]. In the Netherlands, only 5% of the community-dwelling older people uses videoconferencing, according to a poll in 2016 [2]. Perhaps, the lack of experience of the remaining 95% of the population hampered their enthusiasm to participate in the survey about videoconferencing.

The online respondents of our study represent the largest part of our sample (72.3%). As a result, our sample was biased by a higher percentage of internet users compared with the general Dutch population of older people, in which 74% of the 65- to 75-year-old population and 34% of the population over 75 years of age occasionally used the internet in 2012 [57]. In our sample, about 94% of the 65- to 75-year-old population and 89% of the population over 75 years of age had experience with using the internet (at least) occasionally. Additionally, 46.5% of our sample completed higher education, which does not reflect the percentage of highly educated older people in Dutch society, namely 17% in 2012 [58]. We do not know whether the interactions we found in phase 1 would have also been found if the distribution of our sample was less skewed toward highly educated older people with a relatively high amount of technology experience. The sample skewness, however, only applies to phase 1. To observe both older people who possibly already had more digital skills or technology experience and those who did not, in phase 2, we carefully selected our participants, resulting in a sample in which approximately half of the participants did not use a computer (n=7) and the other half did use a computer (n=8).

We believe that our study strength lies in the triangulation of two methods, which helped us to gain a deep understanding of the often-used constructs in technology-acceptance models. Moreover, we noted the added value of the observation method (instead of interviews) to gain an understanding of technology use. With 9 of the 15 participants, a situation occurred in which they misjudged their digital skills; they overestimated or underestimated their skills, and as a result, they could or could not complete a technology task in contrast to their prior expectations. Our method of observations was not hindered by this form of recall bias, whereas it might have spoiled our results if we had chosen a different method, such as interviews.

Implications for Practice or Education and Future Research

Education or Training

Older people's intention to use technology is directly predicted by their effort expectancy, performance expectancy, and perceived privacy or security. Furthermore, self-efficacy and digital literacy appeared to play an important role in the day-to-day use of technology by older people and increase their effort expectancy. Therefore, we recommend addressing these concepts in technology training for older people to be given by nurses or other educators. We believe that in starting with increasing older people's self-efficacy, their effort expectancy and intention to use will follow. In the literature, performance accomplishments, which are successes that raise mastery expectations, are seen as the strongest method of increasing self-efficacy [47]. As mentioned, during our observations, several participants discovered their ability to accomplish a technological task contrary to their prior expectations. In training, similar practices could be organized with the aim of giving older people the opportunity to achieve performance accomplishments. This practice will be the strongest intervention to raise their self-efficacy and as a result their intention and capacity to use technology.

The second strongest source of self-efficacy is vicarious experience, namely seeing others accomplish difficult situations [47]. During training, older people's self-efficacy will most likely increase as technological tasks are repeatedly shown to be achievable by a variety of models. Although this modeling strategy is less effective than personal accomplishment, it may be suited for training purposes by letting participants observe each other executing technological tasks.

A final thought for supporting older people in technology use comes from our observation that some of our participants kept very strictly to the skills that they had learned and became nervous about trying anything outside of their skill set. One can argue about the most appropriate way of learning: (1) providing very specific concrete instructions focused on specific applications or devices or (2) starting from more general technological competencies that could perhaps be applied to a variety of situations, applications, or devices. Hickman et al [59] show that, if the goal is to support learning, "guided attention training" works better for older people than "guided action training," in which participants are told exactly what to do at every step. More research, similar to Hickman et al [59], is needed to learn more about what approach may work best.

Above, we take the perspective that barriers to technology use are a result of a lack of self-efficacy among the end users, in this study of older people. However, the lack of self-efficacy can also be the result of an inappropriate design of the technology. Tsai and colleagues [60] showed that when a new technology is easy to use, a lack of self-efficacy was not a strong barrier for older people to use this technology. So, besides developing adequate training programs for older people, it is useful to think of designing appropriate technology that is easy to use.

Future Research

To test the suggestions above, more research with regard to older people's technology use is required. Our overarching aim was to place older people in a better position to benefit from new ways of health care provision. In this study, we gained a deep understanding of older people's day-to-day use of technology, which can be used as a basis for training development. Research into older people's beliefs regarding their capacities in using health care technology using a pretest-posttest setup, before and after a training, might be a logical next step in research.

Conclusions

This study shows that older people's intention to use videoconferencing is directly predicted by their performance expectancy, effort expectancy, and perceived privacy or security. Additionally, self-efficacy significantly impacts older people's effort expectancy, which subsequently impacts older people's performance expectancy of videoconferencing. In the day-to-day situation, older people experience all kinds of obstacles when using digital technology. Self-efficacy and digital literacy appeared to be the most important theme that plays a role in their technology use and overcoming barriers. Overcoming barriers to technology use is necessary to be able to make use of the new ways of receiving health care involving digital technology.

Acknowledgments

This study was supported by the Foundation Innovation Alliance, SIA RAAK project number: 2010-2-004 INT. Special attention goes to Ganna de Jonge, student in health care management; Linsey Pouw, student in Cesar exercise therapy; and Masokwe Sablerolles, student in Bachelor of Nursing, with all three studying at the Utrecht University of Applied Sciences. We are grateful for their assistance during the observations.

Furthermore, we are grateful to our American partner, the Center for Research and Education on Aging and Technology Enhancement, a cooperative center of the University of Miami, Georgia Institute of Technology and Florida State University, for the fruitful collaboration during the entire project.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Literature review, search terms, and hits.

[PDF File (Adobe PDF File), 35KB - jmir_v20i4e123_app1.pdf]

References

1. Schippers EI, van Rijn M. Rijksoverheid. 2014. Kamerbrief over e-health en zorgverbetering [Room letter on e-health and health care improvement] URL: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2014/07/02/kamerbrief-over-e-health-en-zorgverbetering/kamerbrief-over-e-health-en-zorgverbetering.pdf [accessed 2018-02-21] [WebCite Cache ID 6xPBmvrG3]
2. Krijgsman J, Peeters J, Waverijn G, Lettow B, Hoek L. Nivel. 2016. Omdat ik het belangrijk vind om goed voor mezelf te zorgen?: Rapportage eHealth-doelstellingen [Because I think it's important to take care of myself: Reporting eHealth Objectives] URL: https://www.nivel.nl/sites/default/files/bestanden/Rapportage_eHealth_doelstellingen_2016.pdf? [accessed 2018-02-21] [WebCite Cache ID 6rr3bpXVo]
3. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of technology: frequency of use for younger and older adults. Ageing Int 2011 Mar;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
4. Sackmann R, Winkler O. Technology generations revisited: the internet generation. Gerontechnology 2013;11(4):493-503. [doi: 10.4017/gt.2013.11.4.002.00]
5. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
6. van Houwelingen CT, Moerman AH, Ettema RG, Kort HS, Ten Cate O. Competencies required for nursing telehealth activities: a Delphi-study. Nurse Educ Today 2016 Apr;39:50-62 [FREE Full text] [doi: 10.1016/j.nedt.2015.12.025] [Medline: 27006033]
7. Heerink M, Kröse B, Wielinga BV, Evers V. Enjoyment intention to use and actual use of a conversational robot by elderly people. 2008 Presented at: Proceedings of the third ACM/IEEE International Conference on Human-Robot Interaction (HRI) 2008; March 12-15, 2008; Amsterdam, Netherlands. [doi: 10.1145/1349822.1349838]
8. de Veer AJ, Peeters JM, Brabers AE, Schellevis FG, Rademakers JJ, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
9. Pan S, Jordan-Marsh M. Internet use intention and adoption among Chinese older adults: From the expanded technology acceptance model perspective. Comput Hum Behav 2010 Sep;26(5):1111-1119. [doi: 10.1016/j.chb.2010.03.015]
10. Nayak LU, Priest L, White AP. An application of the technology acceptance model to the level of Internet usage by older adults. Univ Access Inf Soc 2010 Feb 16;9(4):367-374. [doi: 10.1007/s10209-009-0178-8]
11. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
12. Venkatesh V, Thong JY, 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 [FREE Full text]
13. Bhattacherjee A, Sanford C. The intention–behaviour gap in technology usage: the moderating role of attitude strength. Behav Inf Technol 2009 Jul;28(4):389-401. [doi: 10.1080/01449290802121230]
14. Tsai T, Chang H, Chen Y, Chang Y. Determinants of user acceptance of a specific social platform for older adults: an empirical examination of user interface characteristics and behavioral intention. PLoS One 2017;12(8):e0180102 [FREE Full text] [doi: 10.1371/journal.pone.0180102] [Medline: 28837566]
15. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
16. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017 Feb 28;32(6):E1-E7. [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
17. Tsai C. Integrating social capital theory, social cognitive theory, and the technology acceptance model to explore a behavioral model of telehealth systems. Int J Environ Res Public Health 2014 May 07;11(5):4905-4925 [FREE Full text] [doi: 10.3390/ijerph110504905] [Medline: 24810577]

18. Tsai T, Wong AM, Hsu C, Tseng KC. Research on a community-based platform for promoting health and physical fitness in the elderly community. PLoS One 2013;8(2):e57452 [FREE Full text] [doi: 10.1371/journal.pone.0057452] [Medline: 23460859]
19. Wang L, Rau PP, Salvendy G. Older adults' acceptance of information technology. Educ Gerontol 2011 Dec;37(12):1081-1099. [doi: 10.1080/03601277.2010.500588]
20. Braun MT. Obstacles to social networking website use among older adults. Comput Hum Behav 2013 May;29(3):673-680. [doi: 10.1016/j.chb.2012.12.004]
21. Wang Q, Sun X. Investigating gameplay intention of the elderly using an Extended Technology Acceptance Model (ETAM). Technol Forecast Soc 2016 Jun;107:59-68. [doi: 10.1016/j.techfore.2015.10.024]
22. Sintonen S, Immonen M. Telecare services for aging people: assessment of critical factors influencing the adoption intention. Comput Hum Behav 2013 Jul;29(4):1307-1317. [doi: 10.1016/j.chb.2013.01.037]
23. Shah SG, Barnett J, Kuljis J, Hone K, Kaczmarski R. Factors determining patients' intentions to use point-of-care testing medical devices for self-monitoring: the case of international normalized ratio self-testing. Patient Prefer Adherence 2013;7:1-14 [FREE Full text] [doi: 10.2147/PPA.S38328] [Medline: 23300344]
24. Greene JC, Caracelli VJ, Graham WF. Toward a conceptual framework for mixed-method evaluation designs. Educ Eval Policy An 1989 Jan 01;11(3):255-274. [doi: 10.3102/01623737011003255]
25. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996 Feb 28;15(4):361-387. [doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4] [Medline: 8668867]
26. Czaja SJ, Charness N, Dijkstra K, Fisk AD, Rogers WA, Sharit J. Create Center. 2006. Demographic and Background Questionnaire URL: http://create-center.gatech.edu/resources.php [accessed 2017-07-11] [WebCite Cache ID 6rs5bhmGe]
27. Czaja SJ, Charness N, Dijkstra K, Fisk AD, Rogers WA, Sharit J. Create Center. 2006. Computer and technology experience questionnaire URL: http://create-center.gatech.edu/resources.php [accessed 2017-07-11] [WebCite Cache ID 6rs5bhmGe]
28. Chang IC, Hsu HM. Predicting medical staff intention to use an online reporting system with modified unified theory of acceptance and use of technology. Telemed J E Health 2012;18(1):67-73. [doi: 10.1089/tmj.2011.0048] [Medline: 22150638]
29. Gagnon MP, Orruño E, Asua J, Abdeljelil AB, Emparanza J. Using a modified technology acceptance model to evaluate healthcare professionals' adoption of a new telemonitoring system. Telemed J E Health 2012;18(1):54-59 [FREE Full text] [doi: 10.1089/tmj.2011.0066] [Medline: 22082108]
30. Janssen KJ, Donders AR, Harrell FE, Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol 2010 Jul;63(7):721-727. [doi: 10.1016/j.jclinepi.2009.12.008] [Medline: 20338724]
31. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996 Jun;91(434):473. [doi: 10.2307/2291635]
32. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
33. Duncan OD. Path analysis: sociological examples. Am J Sociol 1966;72(1):1-16 [FREE Full text]
34. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974 Dec;19(6):716-723. [doi: 10.1109/TAC.1974.1100705]
35. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
36. Bunniss S, Kelly DR. Research paradigms in medical education research. Med Educ 2010 Apr;44(4):358-366. [doi: 10.1111/j.1365-2923.2009.03611.x] [Medline: 20444071]
37. Bowen GA. Grounded theory and sensitizing concepts. Int J Qual Meth 2016 Nov 29;5(3):12-23. [doi: 10.1177/160940690600500304]
38. McFarland DJ, Hamilton D. Adding contextual specificity to the technology acceptance model. Comput Hum Behav 2006 May;22(3):427-447. [doi: 10.1016/j.chb.2004.09.009]
39. Central Committee on Research Involving Human Subjects. Your research: does it fall under the WMO? URL: http://www.ccmo.nl/en/your-research-does-it-fall-under-the-wmo [accessed 2017-07-11] [WebCite Cache ID 6rsBFAYnK]
40. Bandura AE. Self-efficacy. In: Ramachaudran VS, editor. Encyclopedia of Human Behavior, Volume 4. New York: Academic Press; 1994.
41. Eshet-Alkalai Y. Digital literacy: a conceptual framework for survival skills in the digital era. J Educ Multimed Hypermedia 2004;13(1):93-106.
42. Gitlow L. Technology use by older adults and barriers to using technology. Phys Occup Ther Geriatr 2014 Aug 12;32(3):271-280. [doi: 10.3109/02703181.2014.946640]
43. Gatto SL, Tak SH. Computer, internet, and e-mail use among older adults: benefits and barriers. Educ Gerontol 2008 Aug 22;34(9):800-811. [doi: 10.1080/03601270802243697]
44. Goodhue DL, Thompson RL. Task-technology fit and individual performance. MIS Q 1995 Jun;19(2):213-236. [doi: 10.2307/249689]

45. Venkatesh V, Davis FD. A model of the antecedents of perceived ease of use: development and test. Decision Sci 1996 Sep;27(3):451-481. [doi: 10.1111/j.1540-5915.1996.tb00860.x]
46. Sponselee A. Acceptance and Effectiveness of Telecare Services From the End-User Perspective [Doctoral Thesis]. Wageningen: Proefschriftmaken.nl; 2013.
47. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
48. Peek S. Understanding Technology Acceptance by Older Adults Who Are Aging in Place [Doctoral Thesis]. Enschede: Ipskamp Printing; 2017.
49. van Houwelingen CT, Barakat A, Best R, Boot WR, Charness N, Kort HS. Dutch nurses' willingness to use home telehealth: implications for practice and education. J Gerontol Nurs 2015 Apr;41(4):47-56. [doi: 10.3928/00989134-20141203-01] [Medline: 25486115]
50. Berge M. Telecare acceptance as sticky entrapment: a realist review. Gerontechnology 2016;15(2):98-108. [doi: 10.4017/gt.2016.15.2.023.00]
51. Peek ST, Wouters EJ, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJ. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
52. Chen K, Chan AH. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
53. Bürmann Genannt Siggemann C, Mensing M, Classen T, Hornberg C, Terschüren C. Specific health status has an impact on the willingness to use telemonitoring: data from a 2009 health survey in North Rhine-Westphalia, Germany. Telemed J E Health 2013 Sep;19(9):692-698. [doi: 10.1089/tmj.2012.0214] [Medline: 23906307]
54. Zimmer Z, Chappell NL. Receptivity to new technology among older adults. Disabil Rehabil 1999;21(5-6):222-230. [Medline: 10381234]
55. Luijkx K, Peek S, Wouters E. “Grandma, you should do it-it's cool” older adults and the role of family members in their acceptance of technology. Int J Environ Res Public Health 2015;12(12):15470-15485 [FREE Full text] [doi: 10.3390/ijerph121214999] [Medline: 26690188]
56. Fincham JE. Response rates and responsiveness for surveys, standards, and the Journal. Am J Pharm Educ 2008 Apr 15;72(2):43 [FREE Full text] [Medline: 18483608]
57. Centraal Bureau voor de Statistiek. One third of over-75s use the internet URL: https://www.cbs.nl/en-gb/news/2013/21/one-third-of-over-75s-use-the-internet [accessed 2017-07-10] [WebCite Cache ID 6rr0iHn8N]
58. Centraal Bureau voor de Statistiek. Bevolking; hoogst behaald onderwijsniveau; geslacht, leeftijd en herkomst Society; highest completed education; gender, age and origin. URL: http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=82275NED&LA=NL [accessed 2018-02-21] [WebCite Cache ID 6rr4MBNcU]
59. Hickman JM, Rogers WA, Fisk AD. Training older adults to use new technology. J Gerontol B Psychol Sci Soc Sci 2007 Jun;62 Spec No 1:77-84. [Medline: 17565168]
60. Tsai HS, Shillair R, Cotten SR, Winstead V, Yost E. Getting grandma online: are tablets the answer for increasing digital inclusion for older adults in the US? Educ Gerontol 2015;41(10):695-709. [doi: 10.1080/03601277.2015.1048165] [Medline: 26877583]

Abbreviations

AIC: Akaike Information Criterion
OR: odds ratio
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by T Kool; submitted 21.07.17; peer-reviewed by S Peek, L Gitlow, Y Wu; comments to author 05.09.17; revised version received 28.11.17; accepted 20.12.17; published 06.04.18.

Please cite as:
van Houwelingen CTM, Ettema RGA, Antonietti MGEF, Kort HSM
Understanding Older People’s Readiness for Receiving Telehealth: Mixed-Method Study
J Med Internet Res 2018;20(4):e123
URL: http://www.jmir.org/2018/4/e123/
doi:10.2196/jmir.8407
PMID:29625950

Original Paper

A Remote Collaborative Care Program for Patients with Depression Living in Rural Areas: Open-Label Trial

Graciela Rojas[1,2], MD, Dr med; Viviana Guajardo[1,2], MSc, MD, PhD; Pablo Martínez[2,3,4]; Ariel Castro[1], MSc; Rosemarie Fritsch[1], MD; Markus Moessner[5], PhD; Stephanie Bauer[5], PhD

[1]Hospital Clínico Universidad de Chile, Facultad de Medicina, Universidad de Chile, Santiago, Chile

[2]Instituto Milenio para la Investigación en Depresión y Personalidad (MIDAP), Santiago, Chile

[3]Escuela de Psicología, Facultad de Humanidades, Universidad de Santiago de Chile, Santiago, Chile

[4]Centro de Innovación en Tecnologías de la Información para Aplicaciones Sociales (CITIAPS), Universidad de Santiago de Chile, Santiago, Chile

[5]Center for Psychotherapy Research, University Hospital Heidelberg, Heidelberg, Germany

Corresponding Author:
Pablo Martínez
Escuela de Psicología
Facultad de Humanidades
Universidad de Santiago de Chile
Tercer Piso
Av Ecuador 3650
Santiago, 9170197
Chile
Phone: 56 227182953
Email: pablo.martinezd@usach.cl

Abstract

Background: In the treatment of depression, primary care teams have an essential role, but they are most effective when inserted into a collaborative care model for disease management. In rural areas, the shortage of specialized mental health resources may hamper management of depressed patients.

Objective: The aim was to test the feasibility, acceptability, and effectiveness of a remote collaborative care program for patients with depression living in rural areas.

Methods: In a nonrandomized, open-label (blinded outcome assessor), two-arm clinical trial, physicians from 15 rural community hospitals recruited 250 patients aged 18 to 70 years with a major depressive episode (DSM-IV criteria). Patients were assigned to the remote collaborative care program (n=111) or to usual care (n=139). The remote collaborative care program used Web-based shared clinical records between rural primary care teams and a specialized/centralized mental health team, telephone monitoring of patients, and remote supervision by psychiatrists through the Web-based shared clinical records and/or telephone. Depressive symptoms, health-related quality of life, service use, and patient satisfaction were measured 3 and 6 months after baseline assessment.

Results: Six-month follow-up assessments were completed by 84.4% (221/250) of patients. The remote collaborative care program achieved higher user satisfaction (odds ratio [OR] 1.94, 95% CI 1.25-3.00) and better treatment adherence rates (OR 1.81, 95% CI 1.02-3.19) at 6 months compared to usual care. There were no statically significant differences in depressive symptoms between the remote collaborative care program and usual care. Significant differences between groups in favor of remote collaborative care program were observed at 3 months for mental health-related quality of life (beta 3.11, 95% CI 0.19-6.02).

Conclusions: Higher rates of treatment adherence in the remote collaborative care program suggest that technology-assisted interventions may help rural primary care teams in the management of depressive patients. Future cost-effectiveness studies are needed.

Trial Registration: Clinicaltrials.gov NCT02200367; https://clinicaltrials.gov/ct2/show/NCT02200367 (Archived by WebCite at http://www.webcitation.org/6xtZ7OijZ)

(J Med Internet Res 2018;20(4):e158) doi:10.2196/jmir.8803

KEYWORDS

primary health care; depression; telemedicine; rural health care; medically underserved area

Introduction

Depression is a public health problem, a disabling condition that has devastating effects on people's life and generates high economic costs to society. According to the World Health Organization, depression is the leading cause of disability worldwide and is a major contributor to the overall global burden of disease [1]. Epidemiological studies have demonstrated that 5.5% of the Chilean adult population has suffered from a depressive episode in the last week and that major depression has a lifetime prevalence of 4.7% [2-4]. Studies focusing on primary care services in Santiago, the capital of Chile, have reported a depression prevalence of approximately 30% [5,6].

To reduce the burden of depression in Chile, the Program of Treatment for Depression in Primary Health Care (PTDPHC) was introduced in the early 2000s [7]. This program was later complemented with universal health coverage [8] and the dissemination of clinical practice guidelines as quality standards for primary care clinics [9].

In the PTDPHC, any primary care clinician may refer suspected cases of depression to an on-site physician who can diagnose and initiate treatment [7]. Severe cases are referred to specialized mental health services, and mild to moderate cases may receive a combination of antidepressants, psychosocial interventions, and monitoring visits in primary care, according to severity [7,9].

The PTDPHC was based on a randomized controlled trial in primary care that confirmed the effectiveness and cost-effectiveness of a stepped-care program for the management of depression in resource-limited settings [10,11]. In this trial, the provision of care was highly structured, team-driven, and nonspecialized using available resources in primary care [10].

After a decade of implementation, the PTDPHC has proven to be effective in decreasing depressive symptoms [7]. However, diagnosis inaccuracies, non-guideline-concordant management, and high treatment dropout rates have been identified as pitfalls that reduce PTDPHC performance [6,7]. Moreover, for severe cases, specialized mental health services are unevenly distributed throughout the country [7].

Collaborative care models, which involve primary care teams working in coordination with case managers and mental health specialists [12], have demonstrated to be effective and cost-effective in the treatment of depression [13], leading to implementation efforts to promote its adoption in routine practice [14], thereby being a feasible approach to improve PTDPHC performance.

Although the uneven distribution of specialized mental health services may preclude adoption of these models in rural primary care practices, the use of information and communications technology (ICT) allows mental health specialists to remotely assist primary care teams in underserved locations [15]. These technology-assisted interventions have had positive outcomes in the treatment of depression [16].

These new applications of the collaborative care model may help to reduce the treatment gap in countries with unequal distribution of specialized mental health services, such as Chile [17]. This study reports the feasibility, acceptability, and effectiveness of a remote collaborative care program for patients with depression living in rural areas of Chile.

Methods

Study Design

This was a nonrandomized, open-label, two-arm clinical trial with a blinded outcome assessor. Patients in the usual care group were recruited first; 3 months later, patients were enrolled for the intervention.

Recruitment

This study was carried out at 15 rural community hospitals in the Coquimbo, Bío Bío, and Los Lagos regions in Chile. These hospitals had internet access and a PTDPHC in operation.

Patients aged 18 to 70 years, who had received a new diagnosis of depression, were invited to participate in the study by their primary care physicians at the rural community hospitals and were subsequently interviewed via phone by a blinded research psychologist to assess eligibility. Participants were eligible if they met criteria for a current major depressive episode on the Mini-International Neuropsychiatric Interview (MINI), and were not being treated for depression (ie, patients were not currently attending the PTDPHC). The MINI is a structured psychiatric interview that allows clinicians to diagnose *DSM-IV* mental disorders, and it is available in Spanish [18,19]. Additionally, the MINI was used to assess suicide risk. For those patients with high suicide risk, their primary care physician was informed because physicians are responsible for referral of patients to specialized services.

Full ethical approval was granted by the Universidad de Chile Clinical Hospital Ethics Committee and by each of the Regional Health Services Ethics Committees. Informed consent was obtained after the nature and possible consequences of the study were explained.

Interventions

Before patients' recruitment, all primary care teams working at the participating rural community hospitals received 8 hours of training on guideline-concordant depression care based on the national clinical practice guidelines of the PTDPHC [9].

In accordance with the PTDPHC, the following treatment algorithms were used: (1) patients with mild depression received low-intensity psychosocial interventions (ie, guided self-help or physical activity); (2) patients with moderate depression or depression not resolving with initial treatment received selective serotonin reuptake inhibitors as first-line pharmacologic interventions, while gradually intensifying psychosocial

interventions and/or pharmacotherapy for those who did not respond; (3) patients with severe depression or who were not responding to previous therapy were offered high-intensity psychosocial interventions, antidepressants, and psychotherapy; and (4) patients with treatment-resistant depression, high suicide risk, psychosis, and/or bipolar disorder were referred to regional specialized mental health services.

Remote Collaborative Care Program

A remote collaborative care program is a complex intervention designed to remotely support rural primary care teams to treat depressed patients according to current national clinical practice guidelines for depression [9], maintaining medical decision making within local health services. Rural primary care teams—composed of physicians, psychologists, social workers, midwives, and nurses—provided guideline-concordant, face-to-face care to depressed patients according to the treatment algorithm described in the previous section.

These rural primary care teams were contacted directly via ICT with a centralized and specialized mental health team at the University of Chile Clinical Hospital in Santiago, Chile's capital city. This mental health team—provided by the study—was composed of six psychiatrists, who interacted with the rural primary care teams via Web-based shared clinical records and/or telephone. The rural primary care teams entered basic clinical data of the locally treated depressed patients in the Web-based shared clinical records, subsequently updating information about their patients' progress. The information uploaded to this system was reviewed by the mental health team once a week.

In order to complement these data, a call center was administered by a trained nonmedical health professional (midwife) at the university's facilities, who called the patients to gather information directly from them through a series of structured telephone interviews lasting approximately 30 minutes each. During these interviews, the nonmedical health professional monitored treatment compliance in general (ie, missing appointments, clinical progress, and treatment compliance and medication side effects, if prescribed). The telephone monitoring was carried out once a week during the first month and then every 2 weeks, and the information was available to the rural primary care teams and the mental health team through the Web-based shared clinical record system.

Finally, the mental health team reviewed the data gathered from both sources, once a week, providing remote assistance to the rural primary care teams by entering suggestions into the Web-based shared clinical records and, in special cases, by giving indications to rural primary care clinicians over the telephone.

Usual Care

Rural primary care teams in the usual care group were encouraged to follow national clinical practice guidelines of the PTDPHC [9], thus following the same treatment algorithm described in previously.

Outcomes Assessments

Baseline and follow-up assessments at 3 and 6 months after baseline evaluation were carried out via telephone by a research psychologist at Universidad de Chile Clinical Hospital in Santiago, Chile. The research assessor was blinded to intervention status.

Treatment adherence to antidepressants during the previous 3 months was estimated using the Simplified Medication Adherence Questionnaire [20], and user satisfaction was measured through a depression treatment satisfaction scale ranging from 1=very dissatisfied to 7=very satisfied that was dichotomized. The satisfaction scale has been used previously by members of the research team [10].

Depressive symptoms scores were assessed using the Beck Depression Inventory (BDI-I). The BDI-I is 21-item self-report depression symptom scale with scores ranging from zero to 63. A score of 10 to 19 indicates mild symptoms of depression, a score of 20 to 29 is considered moderate depression, and 30 or higher is considered severe depression. This instrument has good psychometric properties for assessing depressive symptoms [21], and has been previously used in Chile for the evaluation of the PTDPHC [7].

Finally, health-related quality of life was recorded by the 36-item Short Form Survey (SF-36). The SF-36 is a widely used self-report measure of generic health status, providing two summary scores (physical and mental components) ranging from zero to 100 (worse to best possible health status) [22]. The SF-36 has been validated in Chile [23].

Statistical Analysis

Data Analysis

The data input into the electronic platform was extracted to be processed using STATA version 14.0 (StataCorp, College Station, TX, USA). Baseline characteristics were compared between the treatment groups using the chi-square test or Fisher exact test for categorical data and Student *t* test for continuous variables. Mixed-effects logistic regression analyses were performed to test the association between intervention and treatment adherence or user satisfaction at each follow-up, with age, sex, and baseline BDI-I scores as covariates and random effects at the hospital level. The effectiveness of the intervention on depressive symptoms and health-related quality of life was determined using repeated-measures analyses with linear mixed models, with time and intervention group as the independent variables and with random effects at the patient and the hospital level. A *P* value of less than .05 was considered to be statistically significant.

Results

Recruitment and Follow-Up

A total of 409 adults were interviewed after local physicians considered they were depressed. After the blinded baseline interview, 159 patients were excluded because they did not meet criteria for depression according to the MINI or were already in treatment, and 250 patients were eligible. In all, 88.4% (221/250) of the sample were followed up at 3 months and 84.8% (212/250) at 6 months after baseline evaluation.

Sample Characteristics

There were no differences between groups on baseline assessment (Table 1). The majority of the sample were female (216/250, 86.4%), with a mean age of 41.3 (SD 12.6) years, and 47.6% (119/250) were homemakers. The baseline BDI-I score was mean 30.0 (SD 9.0), and 38.0% (95/250) of participants had a high suicide risk according to the MINI.

Treatment Adherence and User Satisfaction

The remote collaborative care program patients had higher treatment adherence rates than those in the usual care group at 3 months (73/104, 70.2% vs 73/119, 61.3%) and 6 months (60/98, 61.2% vs 55/114, 48.2%). A higher proportion of patients in the intervention group were “very satisfied and satisfied” (74/99, 74.7%) than in the usual care group (66/106, 62.3%) at 3-month follow-up, a trend that was maintained at the 6-month assessment (69/92, 75% vs 55/93, 59.1%). In the mixed-effects analyses, reported in Table 2, significant differences between groups in favor of the remote collaborative care program were observed at 6 months for treatment adherence and user satisfaction.

Table 1. Baseline characteristics of participants according to study group.

Variables	Total (N=250)	Usual care (n=139)	Remote collaborative care program (n=111)	*P* value
Sex (female), n (%)	216 (86.4)	117 (84.2)	99 (89.2)	.25[a]
Age, mean (SD)	41.3 (12.6)	41.8 (1.1)	40.6 (1.2)	.45[b]
Marital status, n (%)				**.08[a]**
Single	57 (22.8)	30 (21.6)	27 (24.3)	
Cohabiting	28 (11.2)	19 (13.7)	9 (8.1)	
Married	104 (41.6)	49 (35.3)	55 (49.6)	
Annulled/Divorced	48 (19.2)	33 (23.7)	15 (13.5)	
Widowed	13 (5.2)	8 (5.8)	5 (4.5)	
Education, n (%)				**.71[b]**
Illiterate	9 (3.6)	5 (3.6)	4 (3.6)	
Incomplete elementary	60 (24.0)	35 (25.2)	25 (22.5)	
Complete elementary	35 (14.0)	22 (5.8)	13 (11.7)	
Complete secondary	79 (31.6)	38 (27.3)	41 (36.9)	
Incomplete secondary	30 (12.0)	17 (12.2)	13 (11.7)	
Higher	37 (14.8)	22 (15.8)	15 (13.5)	
Occupation, n (%)				**.11[c]**
Housewife	119 (47.6)	56 (40.3)	63 (56.8)	
Student	10 (4.0)	6 (4.3)	4 (3.6)	
Worker	96 (38.4)	62 (44.6)	34 (30.6)	
Unemployed[d]	21 (8.4)	13 (9.4)	8 (7.21)	
Retired/pensioner	4 (1.6)	2 (1.4)	2 (1.8)	
Depressive symptoms (BDI-I)[e], mean (SD)	30.0 (9.0)	30.6 (9.2)	29.4 (8.8)	.29[b]
Prior depressive episode, n (%)	106 (42.4)	62 (44.6)	44 (39.6)	.45[a]
Suicide risk, n (%)				**.43[a]**
None	60 (24.0)	34 (24.5)	26 (23.4)	
Low	65 (26.0)	33 (23.7)	32 (28.8)	
Moderate	30 (12.0)	14 (10.0)	16 (14.4)	
High	95 (38.0)	58 (41.7)	37 (33.3)	

[a]Chi-square test.

[b]Student *t* test.

[c]Fisher exact test.

[d]Not currently working, but seeking work.

[e]BDI-I: Beck Depression Inventory

Table 2. Differences in treatment adherence and user satisfaction between the remote collaborative care program and usual care groups.

Variables	Remote collaborative care program, n/N (%)	Usual care, n/N (%)	Estimated intervention effect		*P* value	χ^2_4
			OR (95% CI)[a]	AOR (95% CI)[b]		
Treatment adherence						
3 months	73/104 (70.2)	73/119 (61.3)	1.48 (0.85-2.60)	1.53 (0.85-2.75)	.15	13.2
6 months	60/98 (61.2)	55/114 (48.2)	1.69 (0.98-2.93)	1.81 (1.02-3.19)	.04	12.7
User satisfaction						
3 months	74/99 (74.7)	66/106 (62.3)	1.79 (0.98-3.27)	1.83 (0.99-3.36)	.05	7.0
6 months	69/92 (75.0)	55/93 (59.1)	2.07 (1.11-3.88)	1.94 (1.25-3.00)	<.001	10.6

[a]Cluster-adjusted odds ratio.

[b]Odds ratio further adjusted by age, sex, and baseline BDI-I scores.

Table 3. Differences in depressive symptoms in the remote collaborative care program and usual care groups.

Variables	BDI-I[a] score, β (95% CI)	*P* value
Group		
Remote collaborative care program	-1.36 (-3.96 to 1.23)	.30
Time		
3 months	-10.16 (-12.14 to -8.18)	<.001
6 months	-14.00 (-16.00 to -12.01)	<.001
Group × time		
Remote collaborative care program × 3 months	-1.05 (-3.96 to 1.86)	.48
Remote collaborative care program × 6 months	-0.90 (-3.85 to 2.06)	.55

[a]BDI-I: Beck Depression Inventory

Table 4. Differences in health-related quality of life for the remote collaborative care and usual care groups.

Variables	SF-36[a] Mental component summary		SF-36 Physical component summary	
	β (95% CI)	*P* value	β (95% CI)	*P* value
Group				
Remote collaborative care program	1.12 (-1.09 to 3.33)	.32	0.56 (-1.41 to 2.54)	.58
Time				
3 months	7.61 (5.63 to 9.60)	<.001	–1.62 (-3.28 to 0.04)	.06
6 months	9.49 (7.49 to 11.49)	<.001	–3.19 (-4.87 to -1.51)	<.001
Group × time				
Remote collaborative care program × 3 months	3.11 (0.19 to 6.02)	.04	–1.51 (-3.95 to 0.93)	.23
Remote collaborative care program × 6 months	0.77 (-3.73 to 2.19)	.61	0.31 (-2.17 to 2.79)	.81

[a]SF-36: 36-item Short Form Survey

Depressive Symptoms

The intervention group had a mean decrease in BDI-I score from 29.4 (95% CI 27.7-31.0) to 14.6 (95% CI 12.1-17.0) compared with a decrease from 30.6 (95% CI 29.0-32.1) to 16.8 (95% CI 14.8-18.9) among the usual care group at 6 months. In the linear mixed-effects regression models, using BDI-score data at all time points, the remote collaborative care program had a 1.05 point greater decrease in mean BDI-I from baseline than the usual care group (95% CI -3.96 to 1.86, *P*=.48) at 3 months and a 0.90 point greater decrease from baseline at 6 months (95% CI -3.85 to 2.06, *P*=.55). These differences were not statistically significant (Table 3).

Health-Related Quality of Life

In the mixed-effects analyses, reported in Table 4, significant differences between groups in favor of remote collaborative care program were observed at 3 months for mental health component summary scores; however, at 6 months these differences were not statistically significant. There were no

clear trends over time for remote collaborative care program regarding the physical component summary scores.

Discussion

Principal Results

The remote collaborative care program, carried out in 15 community hospitals to support rural primary care teams in the treatment of depressed patients, was feasible and acceptable, achieving higher user satisfaction and better treatment adherence rates at 6 months as compared to usual care. In addition, although depressive symptoms at follow-up did not show significant differences between the remote collaborative care program and usual care, a trend was observed in favor of the intervention group. The remote collaborative care program had a specific effect on mental health-related quality of life at 3 months that disappeared at 6 months, and no differential effect was achieved on physical health-related quality of life at any time point.

Strengths and Weaknesses of the Study

The remote collaborative care program was an innovative and complex technology-assisted intervention to support rural primary care teams located in different parts of the country in the management of depressive patients.

The rural primary care teams had to face the challenge of treating depression without on-site psychiatrist. Thus, remote collaborative care program may provide timely and appropriate treatment recommendations from online psychiatrist to the local health providers.

Although no significant differences were observed in depressive symptoms between the patients treated in remote collaborative care program and usual care, patients in the remote collaborative care program group achieved higher rates of treatment adherence, suggesting that technology-assisted interventions, such as the one described in this paper, can bring additional benefits to the PTDPHC, afflicted by high rates of treatment dropout, which may hinder its effectiveness [6].

However, results must be viewed in the context of study limitations: it was not a randomized trial and rural primary care teams faced major time limitations and they were not paid for participating in the study.

Furthermore, collaborative care programs are complex because they include several components; therefore, it is necessary to identify the most active components of the programs in order to prioritize them in treatment [24]. In the case of remote collaborative care program, these components were the treatment provided by rural primary care teams (which could include pharmacotherapy and psychotherapy), Web-based supervision by a centralized and specialized mental health team (via Web-based shared clinical record system), and telephone monitoring by a nonmedical health professional. Unfortunately, the study design does not make it possible to determine the contribution of each component to the results obtained.

Comparison With Prior Work

Increased clinical attention and patient engagement, along with consultation for those patients not achieving improvement, have all been regarded as essential components for the implementation of effective collaborative care programs [25], and the remote collaborative care program integrated all these in a remote fashion with the assistance of ICT.

There is emerging evidence that some of the core elements of collaborative care programs for depression, such as those previously mentioned, can be delivered remotely, providing timely access to mental health care to vulnerable or underserved populations (eg, people living in rural areas). Studies conducted in the United Kingdom and the United States have found that remote collaborative care programs for depression (ie, those complex interventions in which a component is delivered through the use of ICT) are at least as effective as those collaborative care programs delivered face-to-face [26].

A previous study, carried out at primary care centers in Santiago, Chile, proved that a collaborative care program for depression, which included a pharmacological intervention with periodical telephone contact with lay health workers, improved depressive symptoms and health-related quality of life [27]. Although carried out by the same research team in the aforementioned trial, psychiatric consultation was provided face-to-face to physicians in urban—and more resourceful—practices.

Conclusions

Remote collaborative care programs may support rural primary care teams that do not have the possibility to collaborate with an on-site psychiatrist by providing an acceptable and highly satisfactory intervention for depressed adults and timely advice to primary care teams working in distant parts of a developing Latin American country. Future studies must evaluate treatment process outcomes in a more detailed manner, taking into account the acceptability of these interventions among teams, as well as their cost-effectiveness. Studies of this type should assess changes in users' symptomatology and functionality, and the direct costs of implementing these kinds of programs and the indirect costs, such as the variation in patient referrals from distant areas to specialized centers.

Acknowledgments

This study received support from the Innovation Fund for Competitiveness (FIC), part of the Ministry of Economy, Development, and Tourism, through the Millennium Scientific Initiative, Project IS130005, and the Project PII20150035 of the National Commission for Scientific and Technological Research (CONICYT). The authors wish to thank all those who participated in this study: municipal health departments, primary health care centers, and users.

This study was funded by the National Scientific and Technological Fund (FONDECYT), through project number 1100205. FONDECYT had no influence on the design of the study; it was not involved in the data collection, analysis, or interpretation processes nor did it participate in the preparation, review, or approval of the manuscript.

Conflicts of Interest

None declared.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. Washington, DC: American Psychiatric Association; 2013.
2. Vicente B, Rioseco P, Saldivia S, Kohn R, Torres S. Estudio chileno de prevalencia de patología psiquiátrica (DSM-III-R/CIDI) (ECPP). Rev Méd Chile 2002 May;130(5):527-536. [doi: 10.4067/S0034-98872002000500007]
3. Kohn R, Levav I, Almeida JM, Vicente B, Andrade L, Caraveo-Anduaga JJ, et al. Los trastornos mentales en América Latina y el Caribe: asunto prioritario para la salud pública. Rev Panam Salud Publica 2005 Nov;18(4-5):229-240. [doi: 10.1590/S1020-49892005000900002]
4. Araya R, Rojas G, Fritsch R, Acuña J, Lewis G. Common mental disorders in Santiago, Chile: prevalence and socio-demographic correlates. Br J Psychiatry 2001;178(03):228-233. [doi: 10.1192/bjp.178.3.228]
5. Rojas G, Araya R, Fritsch R. Mental health, psychosocial problems and primary care of health. Acta Psiquiatr Psicol Am Lat 2000;46(2):119-126.
6. Vöhringer PA, Jimenez MI, Igor MA, Fores GA, Correa MO, Sullivan MC, et al. Detecting mood disorder in resource-limited primary care settings: comparison of a self-administered screening tool to general practitioner assessment. J Med Screen 2013 Sep;20(3):118-124. [doi: 10.1177/0969141313503954]
7. Alvarado R, Rojas G, Minoletti A, Alvarado F, Domínguez C. Depression program in primary health care: the Chilean experience. Int J Ment Health 2012 Sep 19;41(1):38-47. [doi: 10.2753/IMH0020-7411410103]
8. Realizing Rights through Social Guarantees: An Analysis of New Approaches to Social Policy in Latin America and South Africa. Washington, DC: The International Bank for Reconstruction and Development / The World Bank; 2008 Feb. URL: http://siteresources.worldbank.org/EXTSOCIALDEVELOPMENT/Resources/244362-1164107274725/3182370-1164107324437/Realizing_Rights_through_Social_Guarantees-web1.pdf [WebCite Cache ID 6st4lMCt3]
9. Ministerio de Salud, Gobierno de Chile. Guía Clínica AUGE Depresión en personas de 15 años y más. Santiago: MINSAL; 2013.
10. Araya R, Rojas G, Fritsch R, Gaete J, Rojas M, Simon G, et al. Treating depression in primary care in low-income women in Santiago, Chile: a randomised controlled trial. Lancet 2003 Mar 22;361(9362):995-1000. [doi: 10.1016/S0140-6736(03)12825-5] [Medline: 12660056]
11. Araya R, Flynn T, Rojas G, Fritsch R, Simon G. Cost-effectiveness of a primary care treatment program for depression in low-income women in Santiago, Chile. Am J Psychiatry 2006 Aug;163(8):1379-1387. [doi: 10.1176/ajp.2006.163.8.1379] [Medline: 16877650]
12. Vanderlip E, Rundell J, Avery M, Alter C, Engel C, Fortney J, et al. Dissemination of Integrated Care Within Adult Primary Care Settings: The Collaborative Care Model. Washington, DC: American Psychiatric Association and Academy of Psychosomatic Medicine; 2016.
13. Woltmann E, Grogan-Kaylor A, Perron BE, Georges H, Kilbourne AM, Bauer MS. Comparative effectiveness of collaborative chronic care models for mental health conditions across primary, specialty, and behavioral health care settings: systematic review and meta-analysis. Am J Psychiatry 2012 Aug;169(8):790-804. [doi: 10.1176/appi.ajp.2012.11111616] [Medline: 22772364]
14. Solberg LI, Crain AL, Jaeckels N, Ohnsorg KA, Margolis KL, Beck A, et al. The DIAMOND initiative: implementing collaborative care for depression in 75 primary care clinics. Implement Sci 2013 Nov 16;8:135 [FREE Full text] [doi: 10.1186/1748-5908-8-135] [Medline: 24238225]
15. Hoeft TJ, Fortney JC, Patel V, Unützer J. Task-sharing approaches to improve mental health care in rural and other low-resource settings: a systematic review. J Rural Health 2018 Dec;34(1):48-62. [doi: 10.1111/jrh.12229] [Medline: 28084667]
16. Fortney JC, Pyne JM, Edlund MJ, Williams DK, Robinson DE, Mittal D, et al. A randomized trial of telemedicine-based collaborative care for depression. J Gen Intern Med 2007 Aug;22(8):1086-1093 [FREE Full text] [doi: 10.1007/s11606-007-0201-9] [Medline: 17492326]
17. World Health Organization, Ministry of Health of Chile. WHO-AIMS Report on Mental Health System in Chile. Santiago: World Health Organization and the Ministry of Health Chile; 2014.
18. Sheehan D, Lecrubier Y, Sheehan K, 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]

19. Bobes J, González M, Sáiz P, Bousoño M, Iglesias C, Bascarán M. La MINI (Mini International Neuropsychiatric Interview): una familia de entrevistas de ayuda diagnóstica en Psiquiatría y Atención Primaria. In: Gutiérrez M, Ezcurra J, González A, editors. Psiquiatría y otras especialidades médicas. Madrid: Grupo Aula Médica; 1998:401-419.
20. Knobel H, Alonso J, Casado JL, Collazos J, González J, Ruiz I, GEEMA Study Group. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. AIDS 2002 Mar 08;16(4):605-613. [Medline: 11873004]
21. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun 01;4(6):561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]
22. Ware J, Kosinski M, Keller S. SF-36 Physical and Mental Component Summary Scales: A User's Manual. Boston, MA: The Health Institute, New England Medical Center; 1993.
23. Olivares-Tiradotado P. Estado de salud de beneficiarios del sistema de salud de Chile: 2004-2005. Chile: MINSAL; 2006.
24. Coventry PA, Hudson JL, Kontopantelis E, Archer J, Richards DA, Gilbody S, et al. Characteristics of effective collaborative care for treatment of depression: a systematic review and meta-regression of 74 randomised controlled trials. PLoS One 2014 Sep;9(9):e108114 [FREE Full text] [doi: 10.1371/journal.pone.0108114] [Medline: 25264616]
25. Kroenke K, Unutzer J. Closing the false divide: sustainable approaches to integrating mental health services into primary care. J Gen Intern Med 2017 Apr;32(4):404-410 [FREE Full text] [doi: 10.1007/s11606-016-3967-9] [Medline: 28243873]
26. Fortney JC, Pyne JM, Mouden SB, Mittal D, Hudson TJ, Schroeder GW, et al. Practice-based versus telemedicine-based collaborative care for depression in rural federally qualified health centers: a pragmatic randomized comparative effectiveness trial. Am J Psychiatry 2013 Apr;170(4):414-425 [FREE Full text] [doi: 10.1176/appi.ajp.2012.12050696] [Medline: 23429924]
27. Fritsch R, Araya R, Solís J, Montt E, Pilowsky D, Rojas G. Un ensayo clínico aleatorizado de farmacoterapia con monitorización telefónica para mejorar el tratamiento de la depresión en la atención primaria en Santiago, Chile. Rev Méd Chile 2007 May;135(5):587-595. [doi: 10.4067/S0034-98872007000500006]

Abbreviations

BDI-I: Beck Depression Inventory
ICT: information and communications technology
MINI: Mini-International Neuropsychiatric Interview
PTDPHC: Program of Treatment for Depression in Primary Health Care
SF-36: 36-item Short Form Survey

Edited by G Eysenbach; submitted 21.08.17; peer-reviewed by C Miller, Z Cooper; comments to author 02.12.17; revised version received 16.02.18; accepted 18.02.18; published 30.04.18.

Please cite as:
Rojas G, Guajardo V, Martínez P, Castro A, Fritsch R, Moessner M, Bauer S
A Remote Collaborative Care Program for Patients with Depression Living in Rural Areas: Open-Label Trial
J Med Internet Res 2018;20(4):e158
URL: http://www.jmir.org/2018/4/e158/
doi:10.2196/jmir.8803
PMID:29712627

Original Paper

Stigma and Its Association With Glycemic Control and Hypoglycemia in Adolescents and Young Adults With Type 1 Diabetes: Cross-Sectional Study

Anne-Sophie Brazeau[1], PhD, RD; Meranda Nakhla[2], MD, FRCPC; Michael Wright[3], BSc; Mélanie Henderson[4], PhD, MD, FRCPC; Constadina Panagiotopoulos[5], MD, FRCPC; Daniele Pacaud[6], MD, FRCPC; Patricia Kearns[7], BFA; Elham Rahme[8], PhD; Deborah Da Costa[8], PhD; Kaberi Dasgupta[8], MSc, MD, FRCPC

[1]School of Human Nutrition, McGill University, Montreal, QC, Canada

[2]Department of Pediatrics, McGill University, Montreal, QC, Canada

[3]Department of Exercise Science, Concordia University, Montréal, QC, Canada

[4]Centre Hospitalier Universitaire Sainte-Justine, Department of Pediatrics, Université de Montréal, Montréal, QC, Canada

[5]BC Children's Hospital, University of British Columbia, Vancouver, BC, Canada

[6]Alberta Children's Hospital, University of Calgary, Calgary, QC, Canada

[7]Breast Cancer Action Quebec, Montréal, QC, Canada

[8]Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Department of Medicine, McGill University, Montreal, QC, Canada

Corresponding Author:
Kaberi Dasgupta, MSc, MD, FRCPC
Centre for Outcomes Research and Evaluation
Research Institute of the McGill University Health Centre
Department of Medicine, McGill University
5252 boul de Maisonneuve, Office 3E.09
Montreal, QC, H4A 3S5
Canada
Phone: 1 514 934 1934 ext 44715
Email: kaberi.dasgupta@mcgill.ca

Abstract

Background: Qualitative studies in type 1 diabetes indicate that visibility of diabetes supplies, self-care, and hypoglycemia symptoms are associated with stigma and suboptimal management. This may be particularly salient in youth who face concurrent challenges such as establishing autonomy and making vocational choices.

Objective: The aim of the study was to estimate stigma prevalence in youth (aged 14-24 years) with type 1 diabetes and its associations with glycemic control.

Methods: Participants, recruited largely through social media, were asked to complete a Web-based survey and to send via mail capillary blood samples for glycated hemoglobin (HbA_{1c}) measurement. The primary definition of stigma required endorsement of one or more of 3 stigma-specific items of the Barriers to Diabetes Adherence questionnaire. These addressed avoidance of diabetes management with friends present, difficulty telling others about diabetes diagnosis, and embarrassment in performing diabetes care with others present. Poor glycemic control was defined as HbA_{1c}>9% (ie, >75 mmol/mol; measured value when available, else self-report) and/or ≥1 severe hypoglycemic episode in the previous year (reported requiring assistance from someone else during the episode). Stigma prevalence was computed (95% CI), and associations with glycemic control were evaluated (multivariate logistic regression models).

Results: Among the 380 respondents, stigma prevalence was 65.5% (95% CI 60.7-70.3). Stigma was associated with a 2-fold higher odds of poor glycemic control overall (odds ratio [OR] 2.25, 95% CI 1.33-3.80; adjusted for age, sex, and type of treatment). There were specific associations with both HbA_{1c}>9% (75 mmol/mol; OR 3.05, 95% CI 1.36-6.86) and severe hypoglycemia in the previous year (OR 1.86, 95% CI 1.05-3.31).

Conclusions: There is a high prevalence of stigma in youth with type 1 diabetes that is associated with both elevated HbA_{1c} levels and severe hypoglycemia. Targeted strategies to address stigma are needed.

Trial Registration: ClinicalTrials.gov NCT02796248; http://clinicaltrials.gov/ct2/show/NCT02796248 (Archived by WebCite at http://www.webcitation.org/6yisxeV0B)

(J Med Internet Res 2018;20(4):e151) doi:10.2196/jmir.9432

KEYWORDS

type 1 diabetes; youth; young adult; social stigma; perception; glycated hemoglobin A1c

Introduction

Stigma related to chronic disease is a negative social judgment that leads to unwarranted rejection or exclusion. It is related to visible features of the disease or its management [1,2]. In conceptualizing stigma, it is important to consider these features, the sources of stigma (eg, individuals, groups, media), and the psychological mechanisms driving stigma such as fear, blame, or disgust [3]. Stigma is characterized by labeling, negative stereotyping, *us versus them* attitudes, and loss of status or discrimination [1]. It may be experienced or perceived, which, in turn, may engender self-stigmatization, an internalization, and acceptance of stigma. The harm that results may be psychological, social, behavioral, and medical. Chronic disease–related stigma has been studied in the context of mental illness, HIV/AIDS, and type 2 diabetes [4-6]. It has been less extensively studied in type 1 diabetes.

Type 1 diabetes is a chronic autoimmune disease with usual onset in childhood and youth. It is characterized by complex and noticeable self-management imperatives, including insulin injection or pump use, capillary blood glucose testing, and attention to meal timing, food choices, and physical activity levels. It also has strong potential for symptomatic hypoglycemia with confusion, distress, or loss of consciousness. The visibility of the equipment, blood testing, making adjustments to therapy, and hypoglycemic symptoms, if they occur, are the disease features that have the potential to lead to stigma [3,7,8]. Most studies examining stigma in type 1 diabetes have been qualitative evaluations that provide insight into the sources (eg, coworkers, family members, media) and characteristics (eg, name calling, rejection) of stigma [3,8,9]. These highlight the blame and discrimination experienced by individuals with type 1 diabetes, which may lead them to hide their condition.

Stigma may be particularly salient when combined with the challenges of adolescence and young adulthood (ie, youth). Many youth with diabetes struggle with self-esteem, body image, social role definition, and peer-related issues [10]. During adolescence, peer relationships and acceptance by friends are essential [11]. In an effort to avoid being seen as different by their peers, qualitative studies suggest that youth with type 1 diabetes may engage in passive coping strategies, such as avoidance of activities and nonadherence to treatment regimens [12-14]. These behaviors are not limited to adolescence but continue into early adulthood, a stage in life termed emerging adulthood, characterized by the challenges of establishing autonomy, personal identity, and making vocational and educational choices [15]. Emerging adults with diabetes must contend with complex developmental tasks while also dealing with their condition and its treatment [16]. There is a paucity of evidence addressing the prevalence of stigma and quantifying its consequences in youth with type 1 diabetes.

Two publications have examined stigma prevalence, one in a mixed population of people with type 1 and type 2 diabetes, in which 70% reported having experienced stigma [17] and a second that included both type 1 and type 2 diabetes patients but reported on each separately [18]. Youth were included in this latter study, but parents rather than patients completed questionnaires; 83% of parents believed that diabetes comes with social stigma. Although qualitative studies suggest that stigma is an important issue for people with type 1 diabetes, there have been no previous large-scale studies estimating stigma prevalence in youth through direct query, nor have associations with glycemic control been evaluated. To address these knowledge gaps, we conducted a cross-Canada study that incorporated social media–based recruitment, online questionnaires, and mailed-in capillary blood samples in youth (adolescents and emerging adults) with type 1 diabetes.

Methods

Overview

The study design and methods, described previously (Clinicaltrials.gov NCT02796248) [19], are briefly reviewed here. Procedures were approved by the Institutional Review Boards of the McGill University Health Centre, the Research Centre of the Centre Hospitalier Universitaire Sainte-Justine, the University of British Columbia, and the University of Calgary. Recruitment and data collection occurred between May 4, 2016, and January 4, 2017.

Questionnaire

The questionnaire incorporated existing instruments and new questions formulated by our team of researchers, patient representatives, and physicians. In a pilot study, high reliability was observed with intraclass coefficients >.95 for each scale [19]. We queried demographic and clinical information (age at diagnosis, insulin pump vs multiple daily injection, hypoglycemia frequency and severity, most recent glycated hemoglobin, HbA_{1c} value), incorporated the Barriers to Diabetes Adherence in Adolescence questionnaire (BDA; 21 items; maximum score of 5) [20], and included 12 closed-ended questions we developed (informed by Browne et al's diabetes-related stigma framework [3]), and open-ended questions (free text responses).

Stigma Definition

Stigma was assessed using the BDA stigma subscale, the only scale available to measure stigma in adolescents with type 1 diabetes. No specific cutoffs for this subscale have been established or validated. Our aim was to determine prevalence

rather than severity. Therefore, we defined stigma as an affirmative response to at least one of 3 key items on the BDA stigma subscale (score ≥2 on a 5-point Likert-type scale; alternate thresholds were also examined, see Multimedia Appendix 1). These (*I try not to deal with my diabetes in front of friends; I have a hard time telling people I have diabetes; I feel embarrassed taking care of my diabetes in front of other people*) were selected a priori by our team. These 3 items assess consequences of stigma.

Several other alternate definitions of stigma were evaluated such as providing a personal experience of feeling judged for having diabetes and combining a personal experience with endorsement of at least one of the 3 key BDA stigma subscale items.

Self-Efficacy and Well-Being

The *Self-Efficacy for Diabetes Self-Management* measure (SEDM; 10 items; maximum score of 10; higher score indicates greater self-efficacy) [21] and the *World Health Organization-5 Well-Being index* (WHO-5 Well-Being index; 5 items; maximum score of 100; higher score indicates greater sense of well-being) [22,23] were both incorporated in our questionnaire. These tools have been validated in adolescents [21,23].

Poor Glycemic Control

Poor glycemic control was defined as an HbA_{1c} level above 9% (75 mmol/mol) and/or at least 1 severe hypoglycemic episode in the last year, defined as requiring assistance from someone else during the episode. When only one type of HbA_{1c} measure was available (ie, self-reported vs direct measurement), this measure was used to classify into an HbA_{1c} category (ie, ≤9% vs >9%). When both types were available, the direct measurement was used. To assess agreement between types, a Pearson correlation was computed, and a Bland-Altman plot was generated.

Recruitment

Adolescents (aged 14-18 years) and emerging adults (aged 19-24 years) with type 1 diabetes were eligible for this study (ie, youth). A comprehensive prevalence survey would have captured all youth with type 1 diabetes across Canada or used these individuals as a sampling framework and subsampled among them. However, there is no diabetes registry that reliably identifies this group of individuals in Canada. We therefore opted to partner with diabetes-related organizations to reach out to this target population through social media. Diabetes Canada and several smaller diabetes and patient organizations (see Multimedia Appendix 1) partnered with us, tweeting members about the study and publicizing it on Facebook. Some patients were approached by their medical team (in person or by email) and provided with the study website address. The study focus was described as "living with type 1 diabetes." A purposive sample of respondents was recruited.

Data Collection

Participants registered on the secure study website and were then emailed a link to an online consent form and questionnaire. After survey completion, they were asked if they would agree to provide a blood sample for HbA_{1c} assessment. Those who consented were mailed a kit. They received a Can $10 gift card after survey completion and a second after mailing in a capillary blood sample. HbA_{1c} was measured on these samples (DTIL Laboratories, Inc, Thomasville, GA, USA).

Survey completion time was reviewed. Our pilot study in 30 participants from patients with whom our team had direct contact indicated a mean questionnaire completion time of 20:19 min (SD 8:52), ranging from 9:13 to 39:41 [8]; therefore, if participants completed the survey in <9 min and/or had little variation in answers (eg, selecting all 1 on a 1-10 Likert scale), we emailed a request to call us directly through a 1-800 number so that responses could be verified. In the absence of such verification, respondents were excluded.

Data Analysis

Means and SDs or proportions were used to report participant characteristics, as appropriate. Stigma prevalence was calculated with 95% CI overall, by sex, and separately among adolescents (ie, aged 14-18 years) and emerging adults (ie, aged 19-24 years). Logistic regression models were constructed to examine associations between stigma and poor glycemic control (primary outcome). Variables considered for inclusion were gender, age, diabetes duration, and insulin administration method. In an alternate model, SEDM and well-being were also included. Linear regression models were constructed to examine associations of stigma with SEDM and with well-being (secondary outcomes).

In an additional set of analyses, we computed the prevalence of stigma using alternate definitions (see Stigma definition) and assessed associations of these with poor glycemic control. We also evaluated associations of our main stigma definition with individual components of the poor glycemic control definition.

Results

Participants

Recruitment targets were achieved (384 respondents; 4 removed because of implausible answers or time to completion) with representation from all 10 Canadian provinces. Participants were largely of European origin (351/380, 92.4%) and English-speaking (315/380, 82.9%; see Table 1). Average age was 19.5 years (SD 3.3), and 46.8% (178/380) were aged between 14 and 18 years. In terms of gender identity, 257 were girls or women (67.6%), 118 were boys or men (31.1%), and 5 indicated being agender or gender fluid (1.3%). With respect to sexual orientation, the majority were heterosexual (302/380, 79.5%), 8.9% (34/380) were bisexual, 5.3% (20/380) were homosexual, 3.4% (13/380) did not know yet, and 2.6% (10/380) preferred not to respond.

Prevalence

The prevalence of stigma (Table 2) was 65.5% (95% CI 60.7%-70.3%) by our primary definition (ie, at least one of the 3 most stigma-relevant BDA questions). About two-thirds described a personal experience of stigma (63.4%; 95% CI 58.6%-68.3%). The prevalence of stigma by our primary definition was slightly higher among girls at 68.3% (95% CI 62.7%-74.0%) compared with 59.3% (95% CI 50.3%-68.3%)

among boys. Similarly, the proportion reporting stigma by our primary definition was slightly higher among young adults (aged 19-24 years) at 69.3% (95% CI 62.9%-75.7%) than among adolescents at 61.2% (95% CI 54.1%-68.5%).

Glycemic Control

An HbA_{1c} value was available for 312 out of 380 participants (82.1%). This included 112 with both mailed-in capillary blood samples for direct measurement and self-reported recent HbA_{1c}, 26 with the mailed-in sample only, and 174 with the self-reported HbA_{1c} only. Among those with both direct and reported HbA_{1c} measures (n=112), moderate agreement was observed (r=.41, 95% CI 0.25-0.56; Bland-Altman plot; see Multimedia Appendix 1). By our primary definition, poor glycemic control was observed in 36.9% (95% CI 31.5%-42.2%) of the 312 participants for whom HbA_{1c} was available; among these individuals, 17 out of 312 (5.4%) participants had both an HbA_{1c}>9% and at least one severe hypoglycemia in the past year, 64 out of 312 had experienced severe hypoglycemia without HbA_{1c}>9% (20.5%, 95% CI 16.0%-25.0%), and 34 out of 312 had an HbA_{1c}>9% (10.9%, 95% CI 7.4%-14.4%) without severe hypoglycemia.

Table 1. Respondent characteristics.

Characteristics	All	Adolescents[a]	Young adults[b]
Number of respondents	380	178	202
Gender, n (%)			
Male	118 (31.1)	74 (41.6)	44 (21.8)
Female	257 (67.6)	101 (56.7)	156 (77.2)
Other	5 (1.3)	3 (1.7)	2 (1.0)
Age, years, mean (SD)	19.5 (3.3)	16.3 (1.3)	22.2 (1.7)
Age at diagnosis, in years, mean (SD)	9.9 (5.3)	8.3 (4.2)	11.3 (5.8)
Diabetes duration, in years, mean (SD)	9.6 (5.4)	8.1 (4.5)	10.9 (5.8)
Currently in school, n (%)	271 (71.5)	155 (87.1)	116 (57.4)
Living with parents, n (%)	264 (69.5)	167 (93.8)	97 (48.0)
Insulin pump, n (%)	220 (57.9)	102 (57.3)	118 (58.4)
HbA_{1c}[c], mean (SD)	7.8 (1.7)	7.9 (1.4)	7.8 (1.9)
Self-reported hypoglycemic episodes per week, mean (SD)	3.2 (2.3)	3.5 (2.5)	3.0 (2.0)
Having experienced severe hypoglycemia in the last year, n (%)	106 (27.9)	48 (27.0)	58 (28.7)

[a]14-18 years.

[b]19-24 years.

[c]HbA_{1c} was available for 312 participants.

Table 2. Prevalence of stigma with 95% CIs. BDA: Barriers to Diabetes Adherence in Adolescence questionnaire

Definition	Prevalence (95% CI)				
	All	Girls[a]	Boys[a]	Adolescents	Young adults
Primary definition					
Endorsed ≥1 of the 3 most relevant items[b] of the BDA stigma subscale	65.5% (60.7-70.3)	68.6% (62.9-74.3)	59.3% (50.3-68.3)	61.2% (54.0-68.5)	69.3% (62.9-75.7)
Other definitions					
Provided an example	63.4% (58.6-68.3)	70.5% (64.9-76.1)	47.5% (38.3-56.6)	51.7% (44.3-59.1)	73.8% (67.6-79.9)
Endorsed ≥1 of the 3 most relevant items[b] of the BDA stigma subscale and provided an example	47.1% (42.1-52.2)	51.2% (45.1-57.2)	38.1% (29.2-47.0)	39.3% (32.1-46.6)	54.0% (47.0-60.9)

[a]According to their sex at birth.

[b]I try not to deal with my diabetes in front of friends. I have a hard time telling people I have diabetes. I feel embarrassed taking care of my diabetes in front of other people.

Associations of Stigma With Poor Glycemic Control

Among participants fulfilling our primary definition of stigma, the odds of poor glycemic control were more than twice as high as in the remaining respondents in both unadjusted (odds ratio [OR] 2.32, 95% CI 1.39-3.89) and adjusted models (OR 2.25, 95% CI 1.33-3.80; adjusted for age, sex, and type of treatment). The odds of HbA_{1c}>9% (>75 mmol/mol) was 3-fold greater (OR 3.05, 95% CI 1.36-6.86), and the odds of severe hypoglycemic episode in the past year was nearly 2-fold greater (OR 1.86, 95% CI 1.05-3.31) in those with versus without stigma (adjusted models; Table 3). Alternate stigma definitions demonstrated an approximately 2-fold higher odds of poor glycemic control.

In the alternate model including self-efficacy and well-being as independent variables, OR for poor glycemic control with versus without stigma was 1.82 (95% CI 1.06-3.13). A 1-point increase in the self-efficacy score (10-point scale) was associated with a 20% lower risk for poor glycemic control (OR 0.8, 95% CI 0.6-0.9); on average, the scores were 6.5 (SD 1.7) with stigma and 7.6 (SD 1.6) without using the primary stigma definition (see Multimedia Appendix 1). There was no association between well-being and poor glycemic control (OR 1.0, 95% CI 0.98-1.01).

Associations of Stigma (by the Primary Definition) With Perceived Well-Being and Self-Efficacy for Diabetes Self-Management

In an adjusted linear regression model (age, sex, type of treatment; Table 4), having stigma (by the primary definition) was associated with a 7.5-point lower score (95% CI −11.8 to −3.3) on the Well-Being index (range 0-100). The threshold for a clinically relevant change is considered to be 10 points [24].

In a separate model, having stigma was associated with a 0.9 (95% CI −1.3 to −0.6) lower SEDM scale score (range 1-10), which corresponds to an approximately 0.5 SD lower score (Table 5).

Table 3. Association between stigma and glycemic control, odds ratios with 95% CI. BDA: Barriers to Diabetes Adherence in Adolescence questionnaire; OR: odds ratio.

Stigma definition	Glycated hemoglobin, HbA_{1c}>9% (75 mmol/mol), OR (95% CI)		Self-reported ≥1 severe hypoglycemia in the previous year, OR (95% CI)		Poor glycemic control overall, OR (95% CI)	
	Univariate	Multivariate[a]	Univariate	Multivariate[a]	Univariate	Multivariate[a]
A: BDA 3 most relevant (at least one of the 3 items)	3.39 (1.53-7.51)	3.05 (1.36-6.86)	1.76 (1.00-3.09)	1.86 (1.05-3.31)	2.32 (1.39-3.89)	2.25 (1.33-3.80)
Item 1: I try not to deal with my diabetes in front of friends	2.58 (1.39-4.77)	2.62 (1.39-4.94)	1.67 (1.00-2.77)	1.74 (1.04-2.91)	2.02 (1.27-3.23)	2.03 (1.27-3.27)
Item 2: I have a hard time telling people that I have diabetes	1.32 (0.72-2.41)	1.11 (0.59-2.07)	1.64 (0.99-2.73)	1.77 (1.05-3.00)	1.56 (0.98-2.48)	1.48 (0.92-2.38)
Item 3: I feel embarrassed taking care of my diabetes in front of other people	2.34 (1.24-4.44)	2.03 (1.04-3.93)	1.52 (0.91-2.54)	1.63 (0.96-2.76)	1.62 (1.02-2.58)	1.54 (0.96-2.49)
B: Personal experience (open-ended question)	2.47 (1.18-5.14)	2.18 (1.03-4.65)	1.26 (0.73-2.16)	1.38 (0.78-2.43)	1.62 (0.99-2.68)	1.58 (0.94-2.65)
A and B: (BDA stigma 3 most relevant + personal experience)	2.69 (1.42-5.11)	2.44 (1.27-4.72)	1.66 (0.99-2.77)	1.79 (1.06-3.03)	2.06 (1.29-3.29)	2.02 (1.25-3.26)

[a]Multivariate: age, sex at birth, type of treatment (multiple daily injection or insulin pump).

Table 4. Linear regression model evaluating association between stigma and well-being, adjusted for age, sex, and use of insulin pump.

Variables	Change in well-being score (95% CI)
Stigma presence (primary definition)	−7.5 (−11.80 to -3.26)
Treatment (insulin pump)	3.7 (−0.39 to 7.80)
Age	−0.28 (−0.90 to 0.35)
Sex (female)	−3.2 (−7.68 to 1.28)

Table 5. Linear regression model evaluating association between stigma and self-efficacy, adjusted for age, sex, and use of insulin pump.

Variables	Change in self-efficacy score (95% CI for B)
Stigma presence (primary definition)	−0.90 (−1.25 to −0.55)
Treatment (insulin pump)	0.42 (0.09 to 0.76)
Age	−0.08 (−0.13 to −0.03)
Sex (female)	−0.61 (−0.97 to −0.24)

Discussion

Principal Findings

Among 380 youth with type 1 diabetes recruited through social media from across Canada, the prevalence of some degree of stigma was approximately 65% (ie, endorsement of at least one of 3 key items on the BDA stigma subscale). Youth with some degree of stigma were more likely to have poor glycemic control. They were twice as likely to have either an HbA_{1c} above 9% or one or more hypoglycemic events in the prior year. When these components of poor glycemic control were considered separately, youth with stigma were 3 times as likely to have an HbA_{1c} above 9%, and they were twice as likely to have had a hypoglycemic event in the prior year. Stigma was also associated with a lower sense of well-being and less self-efficacy for diabetes management. Our findings are a call to action to develop, test, and implement strategies to address stigma in youth with type 1 diabetes.

Previous qualitative studies of stigma in type 1 diabetes [3,8,25,26] provide important insights into the roots and experiences of stigma, but cannot capture the prevalence of the problem, in contrast to our study. To recruit the participants, we opted to partner with diabetes-related organizations to reach them through social media. In 2015, 100% of Canadian young adults had internet access [27]. Over 80% of Canadian youth report daily use of social networking sites [28]. The use of social media combined with an online questionnaire allowed us to meet recruitment targets at relatively low cost and to attract respondents throughout Canada, a country with an area of 10 million km². Our success with this approach is consistent with the findings from a systematic review evaluating social media–based recruitment of youth into health research studies [29]. However, there is a possibility that such a recruitment approach may also attract fictitious respondents or responses. To mitigate this possibility, study promotion occurred only through diabetes-specific organizations and incentives were not publicized. Furthermore, we examined all responses in terms of time to completion and variability in responses; the responses of 4 individuals appeared suspect in this regard and were excluded. We offered a response incentive to encourage survey completion, apparent when the survey was started. A meta-analysis reported that participants who start a survey are more likely to finish it (OR 1.27, 95% CI 1.12-1.44) if an incentive is offered [30].

The BDA questionnaire was specifically designed for adolescents and addresses consequences of stigma through one of its subscales [20]. For our main stigma definition, we made an a priori decision to use endorsement of one or more of the 3 BDA items that appeared to most directly query stigma. We also tested several alternative definitions, and the prevalence of stigma was consistently in the order of 60%. Almost half of respondents (47%) were captured by all definitions. We opted not to include the other 3 items of the BDA stigma subscale as we considered these to be reflective more of personal difficulties or challenges with diabetes management rather than necessarily being stigma-related. For example, “Restaurants are challenging for me” could be a consequence of not knowing the carbohydrate content of the different dishes offered. Other items were “Parties and social gatherings get in the way of taking care of my diabetes and I need to find a private place to take care of my diabetes.”

The two-thirds of participants whom we estimate to have some degree of stigma is similar to the proportion reported in a Swiss study, largely among adults (median age 67 years, ranging from 16 to 96 years) with type 1 or type 2 diabetes. Their definition was having been discriminated against because of their health condition [17]. Their estimate rose to 85% when stigma was defined as perceiving at least one stereotypical attribution [17]. In an adult population with type 1 diabetes, 74% of the respondents believed that diabetes comes with social stigma [18]. We observed a slightly higher prevalence of stigma, according to the primary definition, among young adults than among adolescents (69% vs 61%). The transition from pediatric to adult care occurs concurrently with the increase in stigma prevalence. Addressing stigma before or during this transition may be a means of countering the deterioration in glycemic control that typically occurs at this time.

Stigma was observed in a higher proportion of girls than boys (69% vs 59% by the primary definition). It has been reported that girls with type 1 diabetes feel more embarrassed by their disease [26] and have more diabetes-related distress [31] and higher rates of diabetes-related acute complications [32] and hospitalizations [33]. In contrast to prior studies, we queried not only biological sex but also gender identity; however, the low number of participants who considered themselves to be neither a boy nor a girl did not allow specifically estimation of stigma prevalence in this subgroup of individuals. In terms of sexual orientation, 20% of participants reported being homosexual, bisexual, or did not know yet; there were too few individuals in each category to draw conclusions on the impact of sexual orientation on stigma prevalence or associations with glycemic control. Examination of stigma prevalence and impact in these subpopulations of individuals with type 1 diabetes deserves dedicated study given that “being different” in terms of gender identity or sexual orientation may compound the experience of diabetes-related stigma [34]. In terms of ethnocultural representation, the sample was preponderantly of European origin but that is consistent with the epidemiology of type 1 diabetes [35], in contrast to type 2 diabetes. Studies on prevalence of stigma have addressed this issue mainly in (>90%) individuals of European origin [17,18]. Specific studies in individuals from other ethnic backgrounds are needed.

We evaluated the associations between stigma and glycemic control in youth. Importantly, our definition for poor control combined a high HbA_{1c} with severe hypoglycemia, defined as having at least one episode requiring assistance from someone else in the past year. Higher HbA1c levels are associated with greater risk for diabetes-related complications, with the merits of “tight” control demonstrated in the Diabetes Control and Complications Trial [36]. However, lower HbA_{1c} levels cannot be traded off for increased risk of severe hypoglycemia, with its attendant risk of death, loss of consciousness, and other injury [37,38]. Navigating the space between hypoglycemia and hyperglycemia is particularly difficult in type 1 diabetes, in

contrast to type 2 diabetes. We determined that stigma was associated with a 2.3-fold higher odds of poor glycemic control, defined as having an HbA_{1c} over 9% (75 mmol/mol) and/or having experienced a severe hypoglycemic episode in the previous year. Importantly, stigma was conclusively associated not only with this composite measure of poor control but also with its individual components. These associations were robust across various definitions of stigma evaluated (Table 3). Our study is the first to demonstrate a clear association between stigma and both hyperglycemia and hypoglycemia in youth with type 1 diabetes.

We demonstrated a negative association with SEDM, as captured by the SEDM measure. This is consistent with qualitative studies that have reported that patients neglect diabetes care to avoid stigma. For example, in public places, people will skip blood glucose testing or will delay insulin injections [12,39]. In our sample, stigma was associated with lower scores for every item on the self-efficacy scale used. Our survey did not directly capture self-care behavior, but lower self-efficacy has been shown to be associated with poorer diabetes self-management [40]. Indeed, in the logistic regression model evaluating SEDM alongside stigma in terms of associations with glycemic control, lower self-efficacy was associated with poor glycemic control.

It is important to emphasize that the importance of addressing stigma lies not only in optimizing glycemic control but also in enhancing overall well-being. Indeed, in people with diabetes, emotional well-being may be compromised by the burden of living with diabetes [41]. In this study, respondents with stigma reported a lower sense of well-being. In the Swiss study of adults with type 1 or type 2 diabetes previously discussed, respondents with stigma also reported higher levels of psychological distress [17]. Our identification of high stigma prevalence and a low sense of overall well-being in youth with type 1 diabetes is clearly important; addressing stigma may be one avenue to emotionally support youth with type 1 diabetes.

Limitations

Our study has some limitations. Given its cross-sectional nature, causality cannot be proven. Indeed, it is likely that there is a bidirectional effect of stigma and glycemic control such that stigma adversely affects glycemic control, and poor control contributes to stigma (eg, hypoglycemic episodes witnessed by others, high HbA_{1c} levels known to health care providers and family). We were able to estimate the prevalence of stigma in a large sample, but we cannot be certain that this sample is representative of all youth with type 1 diabetes. Those with stigma may have been more likely to participate. In comparison to prior studies, we queried gender identity and sexual orientation, but these subgroups of individuals were too small to study specifically. Not all participants provided a mailed-in capillary blood sample; however, there was a reasonable correlation between measured and reported HbA_{1c} in those who provided both. Some respondents may have provided fictitious answers; however, we endeavored to mitigate this through various approaches previously described.

Conclusions

Despite some limitations, our study provides important findings. Stigma is prevalent in youth with type 1 diabetes and is associated with lower diabetes-related self-management self-efficacy, high HbA_{1c} levels, severe hypoglycemia, and diminished sense of well-being. Our findings indicate that stigma can be captured through a few simple questions. These results should stimulate clinicians, friends, and family members to ask about stigma and work toward addressing it to help youth with type 1 diabetes avoid diabetes-related complications and lead happier and safer lives.

Acknowledgments

The authors thank Deborah Chan (research assistant) for her assistance with recruitment and study coordination and Cynthia Turcotte (Psychologist) for the development of the survey. This study was funded by a research grant from Diabetes Canada (Principal Investigator KD; Co-Principal Investigator ASB).

Conflicts of Interest

None declared.

Multimedia Appendix 1

STIGMA supplementary material.

[PDF File (Adobe PDF File), 117KB - jmir_v20i4e151_app1.pdf]

References

1. Link BG, Phelan JC. Conceptualizing stigma. Annu Rev Sociol 2001;27:363-385. [doi: 10.1146/annurev.soc.27.1.363]
2. Schabert J, Browne JL, Mosely K, Speight J. Social stigma in diabetes : a framework to understand a growing problem for an increasing epidemic. Patient 2013;6(1):1-10. [doi: 10.1007/s40271-012-0001-0] [Medline: 23322536]
3. Browne JL, Ventura A, Mosely K, Speight J. 'I'm not a druggie, I'm just a diabetic': a qualitative study of stigma from the perspective of adults with type 1 diabetes. Br Med J Open 2014;4(7):e005625 [FREE Full text] [doi: 10.1136/bmjopen-2014-005625] [Medline: 25056982]

4. Browne JL, Ventura A, Mosely K, Speight J. 'I call it the blame and shame disease': a qualitative study about perceptions of social stigma surrounding type 2 diabetes. Br Med J Open 2013 Nov 18;3(11):e003384 [FREE Full text] [doi: 10.1136/bmjopen-2013-003384] [Medline: 24247325]
5. Turan B, Budhwani H, Fazeli PL, Browning WR, Raper JL, Mugavero MJ, et al. How does stigma affect people living with HIV?The mediating roles of internalized and anticipated HIV stigma in the effects of perceived community stigma on health and psychosocial outcomes. AIDS Behav 2017 Jan;21(1):283-291 [FREE Full text] [doi: 10.1007/s10461-016-1451-5] [Medline: 27272742]
6. Corrigan PW, Watson AC. The paradox of self-stigma and mental illness. Clin Psychol Sci Pract 2002;9:35-53. [doi: 10.1093/clipsy.9.1.35]
7. Irani MD, Abdoli S, Bijan I, Parvizy S, Fatemi NS, Amini M. Strategies to overcome type 1 diabetes-related social stigma in the Iranian society. Iran J Nurs Midwifery Res 2014 Sep;19(5):456-463 [FREE Full text] [Medline: 25400672]
8. Abdoli S, Abazari P, Mardanian L. Exploring diabetes type 1-related stigma. Iran J Nurs Midwifery Res 2013 Jan;18(1):65-70 [FREE Full text] [Medline: 23983731]
9. Nishio I, Chujo M. Self-stigma of patients with type 1 diabetes and their coping strategies. Yonago Acta Med 2017 Sep;60(3):167-173 [FREE Full text] [Medline: 28959127]
10. International Diabetes Federation. Belgium; 2013. IDF diabetes atlas - 8th edition URL: http://www.diabetesatlas.org/ [accessed 2018-04-03] [WebCite Cache ID 6yOinZrdx]
11. Berntsson L, Berg M, Brydolf M, Hellström A. Adolescents' experiences of well-being when living with a long-term illness or disability. Scand J Caring Sci 2007 Dec;21(4):419-425. [doi: 10.1111/j.1471-6712.2006.00490.x] [Medline: 18036004]
12. Smith FJ, Taylor KM, Newbould J, Keady S. Medicines for chronic illness at school: experiences and concerns of young people and their parents. J Clin Pharm Ther 2008 Oct;33(5):537-544. [doi: 10.1111/j.1365-2710.2008.00944.x] [Medline: 18834369]
13. Yi-Frazier JP, Yaptangco M, Semana S, Buscaino E, Thompson V, Cochrane K, et al. The association of personal resilience with stress, coping, and diabetes outcomes in adolescents with type 1 diabetes: variable- and person-focused approaches. J Health Psychol 2015 Sep;20(9):1196-1206. [doi: 10.1177/1359105313509846] [Medline: 24271691]
14. Thomas AM, Peterson L, Goldstein D. Problem solving and diabetes regimen adherence by children and adolescents with IDDM in social pressure situations: a reflection of normal development. J Pediatr Psychol 1997 Aug;22(4):541-561 [FREE Full text] [Medline: 9302851]
15. Arnett JJ. Emerging adulthood. A theory of development from the late teens through the twenties. Am Psychol 2000 May;55(5):469-480. [Medline: 10842426]
16. Nylander C, Seidel C, Tindberg Y. The triply troubled teenager--chronic conditions associated with fewer protective factors and clustered risk behaviours. Acta Paediatr 2014 Feb;103(2):194-200. [doi: 10.1111/apa.12461] [Medline: 24117768]
17. Gredig D, Bartelsen-Raemy A. Diabetes-related stigma affects the quality of life of people living with diabetes mellitus in Switzerland: implications for healthcare providers. Health Soc Care Community 2017 Sep;25(5):1620-1633. [doi: 10.1111/hsc.12376] [Medline: 27489251]
18. Liu NF, Brown AS, Younge MF, Guzman SJ, Close KL, Wood R. Stigma in people with type 1 or type 2 diabetes. Clin Diabetes 2017 Jan;35(1):27-34 [FREE Full text] [doi: 10.2337/cd16-0020] [Medline: 28144043]
19. Brazeau AS, Nakhla M, Wright M, Panagiotopoulos C, Pacaud D, Henderson M, et al. Stigma and its impact on glucose control among youth with diabetes: protocol for a Canada-wide study. JMIR Res Protoc 2016 Dec 15;5(4):e242 [FREE Full text] [doi: 10.2196/resprot.6629] [Medline: 27979791]
20. Mulvaney SA, Hood KK, Schlundt DG, Osborn CY, Johnson KB, Rothman RL, et al. Development and initial validation of the barriers to diabetes adherence measure for adolescents. Diabetes Res Clin Pract 2011 Oct;94(1):77-83 [FREE Full text] [doi: 10.1016/j.diabres.2011.06.010] [Medline: 21737172]
21. Iannotti RJ, Schneider S, Nansel TR, Haynie DL, Plotnick LP, Clark LM, et al. Self-efficacy, outcome expectations, and diabetes self-management in adolescents with type 1 diabetes. J Dev Behav Pediatr 2006 Apr;27(2):98-105. [Medline: 16682872]
22. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
23. de Wit M, Pouwer F, Gemke RJ, Delemarre-van de Waal HA, Snoek FJ. Validation of the WHO-5 Well-Being Index in adolescents with type 1 diabetes. Diabetes Care 2007 Aug;30(8):2003-2006. [doi: 10.2337/dc07-0447] [Medline: 17475940]
24. Bech P, Lunde M, Bech-Andersen G, Lindberg L, Martiny K. Psychiatric outcome studies (POS): does treatment help the patients? A Popperian approach to research in clinical psychiatry. Nord J Psychiatry 2007;61 Suppl 46:4-34. [doi: 10.1080/08039480601151238] [Medline: 17365777]
25. Abdoli S, Hardy L, Hall J. The Complexities of “Struggling to Live Life”. Diabetes Educ 2017 Dec;43(2):206-215. [doi: 10.1177/0145721717697245] [Medline: 28340547]
26. Elissa K, Bratt EL, Axelsson AB, Khatib S, Sparud-Lundin C. Societal norms and conditions and their influence on daily life in children with type 1 diabetes in the West Bank in Palestine. J Pediatr Nurs 2017;33:16-22. [doi: 10.1016/j.pedn.2016.12.005] [Medline: 27979497]

27. Pew Research Center. 2016. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2018-03-27] [WebCite Cache ID 6yEZtkQ8R]
28. Sampasa-Kanyinga H, Chaput J, Hamilton HA. Use of social networking sites and perception and intentions regarding body weight among adolescents. Obes Sci Pract 2016 Mar;2(1):32-39 [FREE Full text] [doi: 10.1002/osp4.26] [Medline: 27812377]
29. Park BK, Calamaro C. A systematic review of social networking sites: innovative platforms for health research targeting adolescents and young adults. J Nurs Scholarsh 2013 Sep;45(3):256-264. [doi: 10.1111/jnu.12032] [Medline: 23676115]
30. Goritz AS. Incentives in web studies: methodological issues and a review. Int J Internet Sci 2006;1(1):58-70 [FREE Full text]
31. Lašaitė L, Dobrovolskienė R, Danytė E, Stankutė I, Ražanskaitė-Virbickienė D, Schwitzgebel V, et al. Diabetes distress in males and females with type 1 diabetes in adolescence and emerging adulthood. J Diabetes Complications 2016;30(8):1500-1505. [doi: 10.1016/j.jdiacomp.2016.08.013] [Medline: 27613444]
32. Rewers A, Chase HP, Mackenzie T, Walravens P, Roback M, Rewers M, et al. Predictors of acute complications in children with type 1 diabetes. J Am Med Assoc 2002 May 15;287(19):2511-2518. [Medline: 12020331]
33. Nakhla M, Daneman D, To T, Paradis G, Guttmann A. Transition to adult care for youths with diabetes mellitus: findings from a Universal Health Care System. Pediatrics 2009 Dec;124(6):e1134-e1141. [doi: 10.1542/peds.2009-0041] [Medline: 19933731]
34. Garnero TL. Providing culturally sensitive diabetes care and education for the lesbian, gay, bisexual, and transgender community. Diabetes Spectr 2010 Aug 19;23(3):178-182. [doi: 10.2337/diaspect.23.3.178]
35. Maahs DM, West NA, Lawrence JM, Mayer-Davis EJ. Epidemiology of type 1 diabetes. Endocrinol Metab Clin North Am 2010 Sep;39(3):481-497 [FREE Full text] [doi: 10.1016/j.ecl.2010.05.011] [Medline: 20723815]
36. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993;329(14):977-986. [doi: 10.1056/NEJM199309303291401]
37. Cryer PE. Hypoglycemia: still the limiting factor in the glycemic management of diabetes. Endocr Pract 2008 Sep;14(6):750-756. [doi: 10.4158/EP.14.6.750] [Medline: 18996798]
38. Gill GV, Woodward A, Casson IF, Weston PJ. Cardiac arrhythmia and nocturnal hypoglycaemia in type 1 diabetes--the 'dead in bed' syndrome revisited. Diabetologia 2009 Jan;52(1):42-45. [doi: 10.1007/s00125-008-1177-7] [Medline: 18972096]
39. Tak-Ying Shiu A, Kwan JJ, Wong RY. Social stigma as a barrier to diabetes self-management: implications for multi-level interventions. J Clin Nurs 2003 Jan;12(1):149-150. [Medline: 12519263]
40. Iannotti RJ, Schneider S, Nansel TR, Haynie DL, Plotnick LP, Clark LM, et al. Self-efficacy, outcome expectations, and diabetes self-management in adolescents with type 1 diabetes. J Dev Behav Pediatr 2006 Apr;27(2):98-105. [Medline: 16682872]
41. Rubin RR, Peyrot M. Quality of life and diabetes. Diabetes Metab Res Rev 1999;15(3):205-218. [Medline: 10441043]

Abbreviations

BDA: Barriers to Diabetes Adherence in Adolescence questionnaire
HbA_{1c}: glycated hemoglobin
OR: odds ratio
SEDM: self-efficacy for diabetes self-management

Edited by G Eysenbach; submitted 16.11.17; peer-reviewed by D Gredig, H Miller; comments to author 28.12.17; revised version received 01.02.18; accepted 20.02.18; published 20.04.18.

Please cite as:
Brazeau AS, Nakhla M, Wright M, Henderson M, Panagiotopoulos C, Pacaud D, Kearns P, Rahme E, Da Costa D, Dasgupta K
Stigma and Its Association With Glycemic Control and Hypoglycemia in Adolescents and Young Adults With Type 1 Diabetes: Cross-Sectional Study
J Med Internet Res 2018;20(4):e151
URL: http://www.jmir.org/2018/4/e151/
doi:10.2196/jmir.9432
PMID:29678801

Original Paper

Improving the Usefulness and Use of Patient Survey Programs: National Health Service Interview Study

Kelsey Flott[1], PhD; Ara Darzi[1], MD, FRS, FRCS, FACS, FRCP, FREng; Sarah Gancarczyk[2], MSc; Erik Mayer[1], PhD

[1]Centre for Health Policy, Imperial College London, London, United Kingdom

[2]Picker Institute Europe, Oxford, United Kingdom

Corresponding Author:
Kelsey Flott, PhD
Centre for Health Policy
Imperial College London
10th Floor QEQM
St Mary's Campus
London, W2 1NY
United Kingdom
Phone: 44 7909248515
Fax: 44 7909248515
Email: k.flott14@imperial.ac.uk

Abstract

Background: A growing body of evidence suggests a concerning lag between collection of patient experience data and its application in service improvement. This study aims to identify what health care staff perceive to be the barriers and facilitators to using patient-reported feedback and showcase successful examples of doing so.

Objective: This study aimed to apply a systems perspective to suggest policy improvements that could support efforts to use data on the frontlines.

Methods: Qualitative interviews were conducted in eight National Health Service provider locations in the United Kingdom, which were selected based on National Inpatient Survey scores. Eighteen patient-experience leads were interviewed about using patient-reported feedback with relevant staff. Interviews were transcribed and underwent thematic analysis. Staff-identified barriers and facilitators to using patient experience feedback were obtained.

Results: The most frequently cited barriers to using patient reported feedback pertained to interpreting results, understanding survey methodology, presentation of data in both national Care Quality Commission and contractor reports, inability to link data to other sources, and organizational structure. In terms of a wish list for improved practice, staff desired more intuitive survey methodologies, the ability to link patient experience data to other sources, and more examples of best practice in patient experience improvement. Three organizations also provided examples of how they successfully used feedback to improve care.

Conclusions: Staff feedback provides a roadmap for policy makers to reconsider how data is collected and whether or not the national regulations on surveys and patient experience data are meeting the quality improvement needs of local organizations.

(J Med Internet Res 2018;20(4):e141) doi:10.2196/jmir.8806

KEYWORDS

patient experience; surveys; patient data

Introduction

National Patient Feedback Surveys

Involving patients in their care has become a key feature of health care improvement policies across countries and health systems [1]. The value of patient experience has been recognized not only in terms of its centrality to respectful and conscientious care, but also its relationship to better clinical outcomes and pathway adherence [2]. Feedback on experience is now collected as a norm and regarded as a fundamental quality measure [3,4]. In the United Kingdom, the National Patient Survey Program (NPSP) was established in 2002 to systematically solicit feedback from patients across many different care settings at nearly all National Health Service (NHS) organizations on a range of experience metrics [5]. As is the case in many other countries' systems, the NPSP surveys are only one piece of the

feedback puzzle that providers can use for quality improvement. Staff must interpret and integrate feedback from a multitude of sources such as bespoke surveys, online platforms, social media outlets, audits, complaints, and, in the case of the NHS, the Friends and Family Test (FFT) [6,7].

The activity around patient-reported feedback is impressive, and in many ways indicates an actual shift towards a patient-centric paradigm of care. Policy documents have endorsed the use of patient-reported feedback and articulated its benefit across the health sector [8]. However, harnessing this feedback for improvement is still challenging; growing evidence suggests that feedback is still not used to drive improvement [6,9-14]. In the United Kingdom for example, improvements on patient experience metrics have mostly come in response to large-scale national campaigns, with only modest improvements and some declines witnessed in other areas [15]. The gap between feedback collection and use not only represents a costly misuse of resources, as national surveys cost upwards of a £640,000 per survey per year, but it also raises ethical concerns around not acting on critical patient information [16,17].

The value of data for organizational quality improvement is not commensurate to the volume of data the system supplies. This misalignment is a symptom of the tension between national agendas and local needs. It is, in many ways, the consequence of national survey programs being set up to satisfy a national agenda rather than being designed with respect to local circumstances.

The System

The underuse of data, and the substantial investment in it, calls into question the system around collecting, analyzing, and reporting patient experience feedback. The paramount actors in the system should arguably be the patients who report their experience and providers who use feedback for quality improvement. However, most of the decisions about how national patient experience feedback collections operate are made by a host of other actors.

In the NHS, the independent health care regulator known as the Care Quality Commission (CQC) is responsible for the content and roll-out of NPSP surveys. In addition, the Patient Survey Coordination Centre, housed within Picker Institute Europe, develops and deploys the surveys on behalf of the CQC. These two groups determine the questions included on surveys based on existing patient experience frameworks, with a focus on maintaining the ability to compare questions to those included on previous survey iterations [18]. Patients are consulted during the development and redevelopment of all NPSPs, but rarely invited to suggest entirely new concepts for the survey [19].

Regarding another nationally mandated feedback source, FFT, the system (namely NHS England and local commissioners) does allow for flexibility in questions, but it is more concerned with the volume of data the FFT can accumulate rather than the methodology by which it is collected. Organizations therefore have an incentive to boost response rates rather than include meaningful questions or make use of the data. The volume of data and its readily available nature should be a strength for most providers, especially when providers accompany their FFT collection with other rich free-text questions. However, this feedback has also become underutilized due to the existing necessity for most providers to sort through data manually or pay for external analysts. The real time vehicle has tremendous merit as an idea, but the system's execution of it hinders meaningful use.

External bodies also determine sampling procedures in both the NPSP and FFT. In the NPSP, NHS providers can conduct the survey either "in house" or with an approved contractor. Both operations involve sampling based on a 24-page sampling document provided by the Coordination Centre. When a survey is administered "in house," organizations complete all required tasks independently, which includes tasks such as the Disclosure and Barring Service checking each sampled patient to make sure they fit the criteria (ie, ensuring the patient is still alive), printing surveys, posting them, inputting data, and sending it to the Coordination Centre on time [20]. The sample for the National Inpatient Survey, the largest of the NPSP, has only recently moved from 850 to 1250; still a very low proportion of some hospitals' inpatients, but a much larger proportion of others. Nonetheless, sampling is administratively burdensome [20]. For this reason, the CQC maintains a list of approved survey contractors (another key player in the system) who can do this work for providers. Providers purchase a survey package from contractors determining the extent of analyses and data presentation to which the provider is entitled. Not all contractors provide the same service, offer the same analysis, or engage with providers on an equal basis.

The complexity of the system pertaining to patient experience feedback is not unique to the NHS, and it is important to note that the number of actors involved does not discredit the information that patients relay in their feedback. Rather, this complex system presents considerations that providers in all health systems need to account for when interpreting data; It raises questions around how the national system of patient feedback can supply local providers with better data that could be more meaningfully translated into quality improvement information.

Staff Perceptions

A King's Fund report explained that gleaning information from experience data requires the same analytical capability as interpreting clinical data; however, that capability is often unavailable [21]. Staff across health systems consider patient feedback to be valuable but have neither the time nor the expertise to use it [22]. Evidence from the field of Patient Reported Outcome Measures suggests a similar pattern: data goes unused when staff cannot make sense of the data or do not fully understand how it was collected [14].

In 2007 work was conducted to understand staff attitudes towards the NPSP and their ability to use its data in the NHS. Findings from this work explicitly revealed staff's concerns around using aggregate, organizational-level data to engage clinicians within specialties, and their difficulty navigating the statistical underpinnings of results [23]. Furthermore, this work put forth staff-driven recommendations for improvement, such as increased resources and organizational prioritization for patient experience. A full decade later, the lag in data use still

exists. Many organizations have made well-defined attempts to use patient experience feedback, especially from national surveys, but have found their efforts thwarted by a series of barriers [17].

Underuse of data is unacceptable from a quality assurance perspective, as the requirement to perform analyses without proper resources risks key details being missed. It is frustrating from clinical and operational perspectives, as time and money are being invested with little return of insights to improve care. Even the National NHS Staff Survey demonstrated that only 20% of staff strongly agree that their organization acts on patients' concerns [20]. It is demoralizing and dangerous from a patient perspective, as their input is going unheard and problems are persisting for others. Ultimately, it is ethically questionable, as patients have provided sensitive information but their feedback fails to drive change.

These concerns further expose the tensions between the data produced by the national system and the local needs within organizations. These factors compel inquiry into how data can be more usefully supplied to organizations so that it can serve as meaningful business intelligence for service improvement. It is crucial to understand how the national systems for patient feedback affect the use of data and how they can change to better support translation of feedback into insights for quality improvement.

Aims and Objectives

Using the NHS as a case study, the aim of this research was to determine how the national system related to patient surveys can be improved so that it supports the local needs of organizations in their endeavors to use patient experience feedback. The first objective is to identify a diverse range of health care professionals responsible for using patient experience feedback and interview them about their experiences using patient experience feedback, the barriers that still prohibit using feedback for improvement, and their ideas for improving the system. The final objective is to identify and showcase successful attempts to overcome barriers and use patient experience feedback for improvement.

Methods

Case Study Selection

This study used a qualitative case study design to gather input from a range of organizations. Organizations were selected based on 12 metrics within the National Inpatient Survey, as it is currently the largest and most robust source of patient experience feedback in the NHS. Three organizations were selected based on demonstrated improvements on the 12 key National Inpatient scores between 2010 and 2014 (the most recent data at the time of sampling), while three others were selected based on demonstrated declines on the 12 key scores during the same time. These were then referred to as the "increased" group and "decreased" group, respectively. A final three organizations whose scores remained consistent for the same years were also selected. Organizations then put forward relevant staff for interview.

The 12 National Inpatient Survey questions were identified as most important to patients through principle component analyses. These 12 were also deemed by the Picker Institute Europe to be good indicators of whether or not organizations exhibited meaningful shifts in experience (Textbox 1). It is important to note that the sampling strategy did not account for the baseline from which those scores changed; this ensured that any organization demonstrating improvement could be included regardless of initial high or low experience scores.

Organizations that demonstrated a significant increase or decrease on any of these questions were recorded. A list was then compiled of all the organizations that were recorded to see which three providers had the most increases and decreases. Another group of organizations was also identified that remained the most constant. With regard to the organizations with consistent scores, selection consideration was given for size (to have a range of small, medium, and large acute organizations) and geography (to have a distribution of rural, urban, southern, and northern acute organizations) in order to maintain a degree of diversity, as there were many organizations that demonstrated no changes on the 12 questions. Selected organizations were contacted, and 8 of the 9 that were sampled agreed to participate; the only one to not take part was one of the *consistent* organizations.

Interviews

Face-to-face, semi-structured interviews were conducted with as many staff as each organization elected to put forward. This was typically 1-3 staff per organization, and their most common job titles were Patient Experience Lead, Patient Experience Administrator, or Director/Deputy Director of Nursing. In total, 18 staff members were interviewed: seven from the *increased scores* group, seven from the *decreased scores* group, and four from the *consistent* group.

The topic guide covered questions such as staff responsibilities for using the patient-reported feedback, preferences for using it, current likes and dislikes regarding survey programs, and changes they would like to see made to it. Respondents were specifically asked about what changes they would like to see made to the system regarding patient experience feedback to facilitate better data use. Organizations in the *increased scores* group were asked to share their strategies for using patient reported feedback as a vehicle for shared learning. Staff from the *increased scores* group also submitted organizational information about how they had used patient-reported feedback in action planning and improvement.

Textbox 1. Questions used for grouping and sampling organizations.

Do you feel you got enough emotional support from hospital staff during your stay?
When you had important questions to ask a doctor, did you get the answers that you could understand?
Overall did you feel you were treated with respect and dignity while you were in the hospital?
Do you think the hospital staff did everything they could to control your pain?
Did you have confidence and trust in the doctors treating you?
Did you find someone on the hospital staff to talk to about your worries and fears?
Did a member of staff tell you about medication side effects to watch for when you went home?
Were you involved as much as you wanted to be in decisions about your care and treatment?
In your opinion, how clean was the hospital room or ward you were in?
Did doctors talk in front of you as if you weren't there?
In your opinion were there enough nurses on duty to care for you in hospital?
Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?

Thematic Analysis

Interviews were recorded and transcribed if the participants gave explicit permission in their consent form. Two interviewees consented to an interview but declined to be recorded. In these two cases, notes were taken by a team member and used in place of a full transcript.

Transcripts were then uploaded into the qualitative analysis software NVivo (QRS International Pty Ltd). A thematic analysis was conducted to demarcate different themes or topics within the transcripts. This study used thematic analysis to identify information relevant to the experience of using patient experience survey data to generate improvements.

The codes were developed a priori for the most part, as they were taken from the background literature regarding possible barriers and facilitators to data use. Some codes were identified a posteriori as they occurred unexpectedly but were important to answering the research question. Specifically, the coding looked for mention of themes relevant to answering the research question and then subthemes mentioned in relation to the primary themes. Sentiment was coded to capture how respondents felt about any particular theme, particularly whether staff referenced subthemes negatively (as a barrier to data use), positively (as a facilitator to data use), or as a desire for change in patient survey data (staff wish list).

Results

Key Themes

Four primary themes were identified with a range of subthemes relating to each of them. The subthemes were expressed with different sentiments, which fell into three distinct categories: negative (barriers to using data), positive (facilitators to using data), and desire for change (staff wish list). The themes, subthemes, and sentiments are mapped below.

While staff were specifically probed about their ability to use NPSP data, transcripts of the conversations naturally exposed the types of data staff found most useful. Transcripts also revealed the variation in sentiments towards themes and subthemes. For example, staff would reference a particular theme (ie, survey data) and subtheme (ie, the inability to link data) as a barrier, and then that same subtheme (ie, the ability to link data) as a facilitator. The sentiment behind each theme was coded to categories' subthemes. The four primary themes identified in interview transcripts related to survey methodologies, survey reports, survey data and organization, and staff factors that impact the ability to use patient experience data.

Survey Methodologies

While discussing how they used patient experience survey data, one of the most common topics that staff mentioned was the survey methodology used in NPSP surveys. It was clear from staff that difficulty interpreting results, and lack of clarity around the reasons for certain methods, created barriers to using the data. Staff were concerned that the methods not only led to confusing results but were also inappropriate given the size of their organizations, as illustrated by two quotations below. Staff mentioned that in order to facilitate data use, methods should be more intuitive to staff who do not have survey training and should also include the scope for larger sample sizes. Some staff also expressed a desire for more real-time and qualitative methods to accompany NPSP results.

> *The other problem with the national survey is the way that they actually design it; the CQC part of it makes it really difficult.*
>
> *So in a Trust that sees 1.6 million patients a year, although the majority are out-patients, 850 is a tiny sample. I know it's increased this year but it's still quite small*
>
> *[The free text is] so much better because what it does is it elicits the things which matter to people, not what we think matter to them.*

Survey Reports

The second primary theme related to survey reports and how data is presented back to staff. In terms of data from the NPSP surveys, reports from the CQC and survey contractors are the principle source of patient survey data provided to NHS

organizations. Staff felt that these reports often caused confusion. Despite many staff referencing these reports as barriers, other staff members gave examples of contractor reports facilitating the use of feedback. This finding related mostly to contractors' ability to provide a report with more personalized information for each organization than that available in CQC reports. Staff also cited contractors' reports being accompanied by workshops to explain the results. In terms of how staff discussed this themes in relation to their wish list of changes, the main desire expressed was for enhanced opportunity to share success stories, rather than simply receiving benchmarking tables.

> *So we got amber on every single question. Every single question we got the same as everybody else which just happens to be the same score that Morecombe Bay got who are in special measures...*
>
> *So both having those stories and the information but also make the workshops not just around the outcome and the talks but actually the best practice workshops, maybe on a regular basis, so someone from Newcastle getting up and presenting to all the other Trusts who want to be there about discharge, next it will be Birmingham about food or whatever it might be but Picker being almost a coordinating body for that because that's the vehicle with which it's been done. Something like that would be good.*

Survey Data

The third primary theme identified related to the actual survey data that staff received from NPSP surveys. Most of the conversation about this topic related to the ability to link data to other quality indicators. Staff found NPSP data difficult to work with because it could not be compared at a granular level to other data sources and left them with an inexact picture of how patient experience data fits in with other organizational data. Another prominent subtheme revolved around the inability to glean which NPSP data points matter most to patients. Staff expressed an interest in more explanation of the data, support to analyze it, and better indication of what was most important from a patient perspective.

> *This could be related to any survey, but the idea of linking results at the patient level would help the patient know they were listened to see feedback on incident reporting to support the need for response to feedback.*
>
> *What I would want is that to be linked in with complaints, so I'd love to have some kind of dashboardy thing that pulls all that stuff together.*
>
> *In terms of understanding the data, I think when they come and do workshops with us or present the data we need—that's very helpful but I need—we need them, in there, telling the story of how they collected the data and how it's reliable.*
>
> *For example, the question about the call button may mean different things to different patients, and they need to know what to improve.*

Organization and Staff Factors

The final theme identified related to the factors outside of survey programs that impacted how staff could use NPSP data. The subthemes related to aspects of organizational structure, the extent of training staff had in using survey data, and the priority given to patient experience within organizations. Some staff members mentioned that there was sometimes lack of clarity about whose responsibility it was to use patient experience data, and more frequently, the people in charge of using data did not feel sufficiently trained to do so.

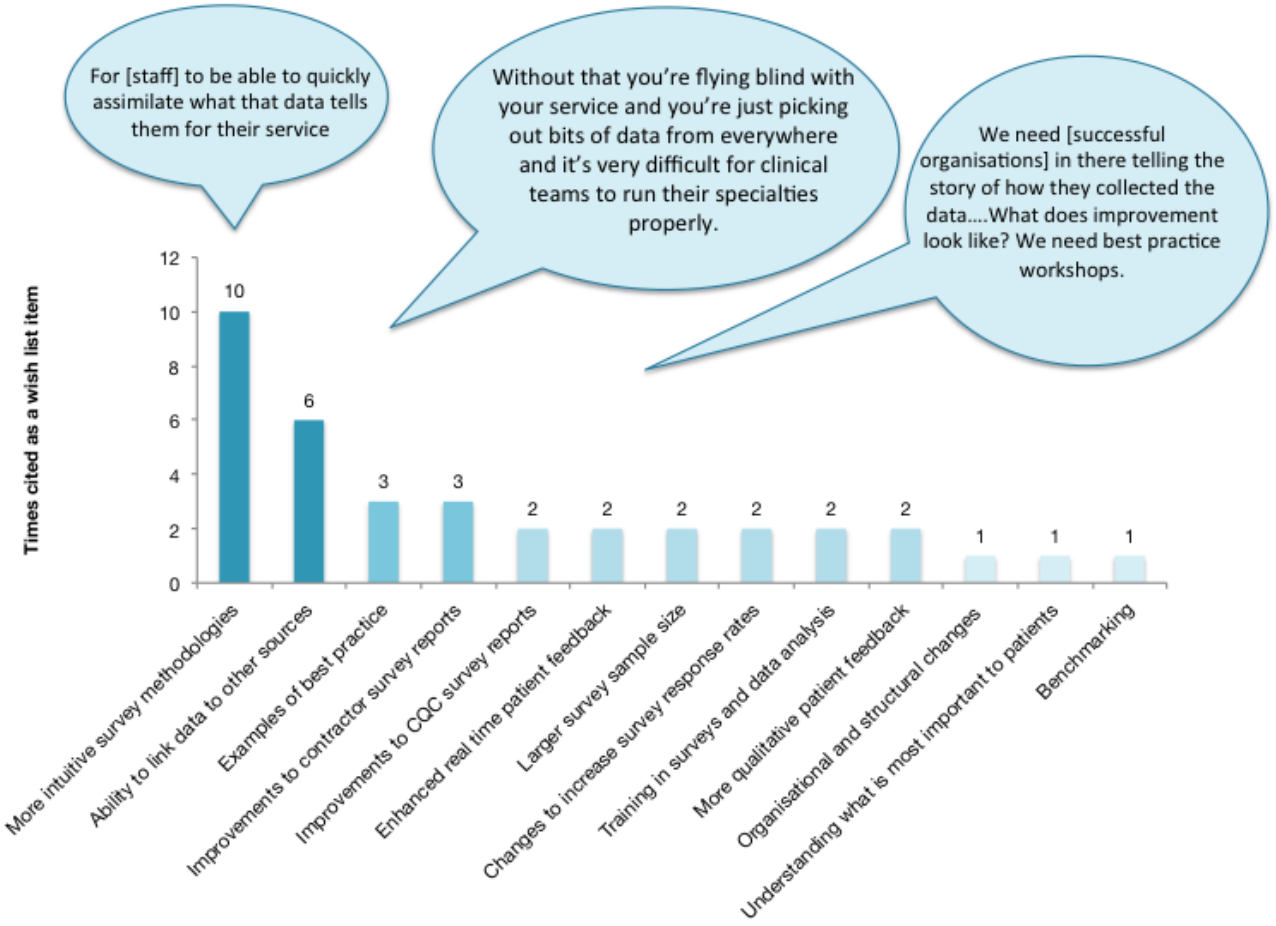
Very few staff members were concerned about the priority given to patient experience in their organization; however, some did cite it as a key factor in being able to pursue improvements. This idea led to many staff members desiring more information about what other organizations had done to achieve success in patient experience.

> *And I think sharing that nationally, because I want to know what other people are doing, because even if it's things that we're doing well but we could do better, I don't want to reinvent the wheel.*
>
> *What I'd be interested in, is sharing best practice and stories from others.*

Figure 1 depicts what staff specifically said they would put on their wish list of changes in order to improve how patient-reported feedback is collected, analyzed, and presented. Examples of best practice was its own theme, but the idea of learning from others came through in a few of the themes. Staff articulated a desire for survey methodologies and resulting data to be presented in a more intuitive way that resonates with their day-to-day practice.

Finally, Textbox 2 lists the examples of how organizations in the *improved scores* category overcame identified barriers and used patient-reported feedback effectively to improve their service and their survey scores. This list reveals the general technique used and Multimedia Appendix 1 provides a full description of that process. These themes are mapped in Figure 2.

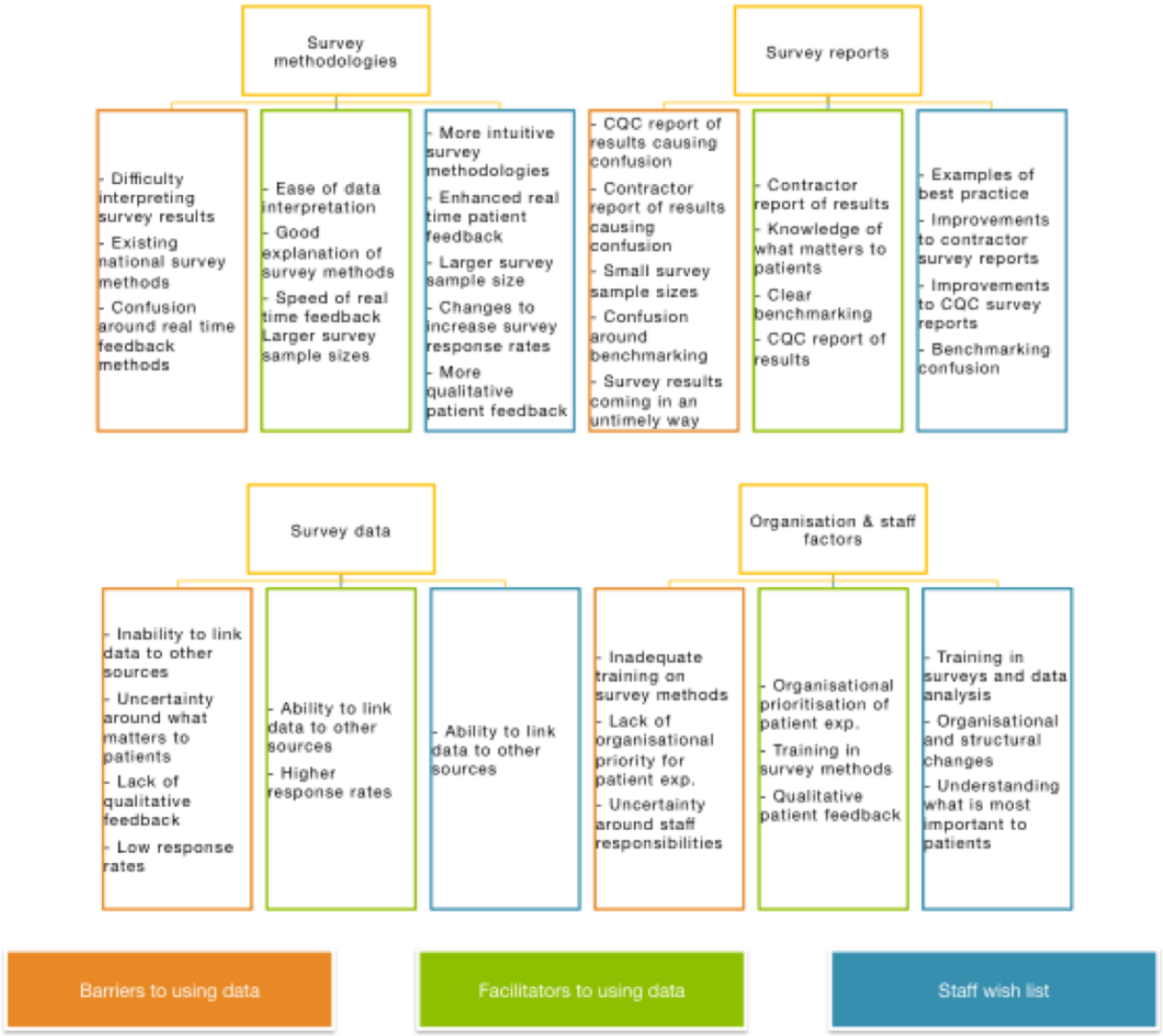
Figure 1. Staff wish list.



Textbox 2. Examples of improvements techniques.

Data triangulation even when data cannot be directly linked
Emotional intelligence training
Identifying communication breakdowns
Bespoke survey methodologies across services
Following-up with people after they give feedback
Provision of better information about patient pathway
Values-based improvement
Competition to drive innovations
Comfort packs on the wards
Including patient experience in staff inductions
Giving staff the positive feedback from patients
Identifying priority questions on surveys

Figure 2. Map of themes, subthemes, and sentiment.



Discussion

Overview of Results

In total, 18 staff members who work with patient experience feedback were identified through the sampling process based on National Inpatient Survey scores; this group was deliberately diverse, representing three organizations that had demonstrated improvements, decreased performance, and consistencies in important National Inpatients Survey questions. Respondents were also from a range of small, medium, large, and teaching hospitals from across England.

The most frequently cited barriers to using patient-reported feedback had to do with interpreting results (14 mentions), understanding survey methodology (14 mentions), presentation of data in both national CQC (13 mentions) and contractor reports (12 mentions), inability to link data to other sources (7 mentions), and organizational structure (7 mentions). The most frequently cited facilitators were: ability to link data (9 mentions), ease of survey interpretation, and clarity around methodologies (7 mentions). In terms of a wish list for improved practice, staff desired more intuitive survey methodologies (10 mentions), ability to link patient experience data to other sources (6 mentions), and more examples of best practice in patient experience improvement (3 mentions).

Staff feedback varied slightly when segmented by organizational group. Organizations whose scores had decreased cited training, organizational structure, and interpreting results as barriers more often than other organizations. Those whose scores had increased focused more on difficulty understanding survey methods and confusion around CQC reports, but cited knowledge of what is most important to patients as a key facilitator. The difference in focus could relate to other aspects of organizational health. For instance, organizations with declines who cited structural issues might have obstructions to the use of patient feedback that are not necessarily a product of the system regarding patient experience feedback.

Interpretation of Results

One of the most interesting findings emerging from the thematic analysis relates to what was *not* discussed. Even when specifically probed, staff virtually never cited finances or lack

of senior-level interest as a major barrier to using patient-reported feedback. While some staff members did say they would like things like "free extra analysis," and many cited paid-for survey contractors as providing the most useful analytic tools and workshops, staff did not feel that lack of funding was a chronic barrier to using data. Furthermore, very few staff criticized organizational leadership when citing the barriers to using data. These findings demonstrate enormous progress in the field of patient experience, as only a decade ago such feedback was highly underprioritized by staff and within budgets [23].

In terms of wish list items, the idea that received the most consensus by far, and was articulated in a variety of different themes, was the idea of sharing best practice to help other organizations emulate the successful use of data. This finding was a compelling plea for collaboration, rather than competition, to improve experiences for patients. Ideally some of the examples of improvement provided in this research will resonate with staff and provide a first step towards such shared learning. Furthermore, organizations wanted to see a variety of changes to survey methods (ie, shortening the questionnaire, offering it in different modalities, larger sampling), the ability to link data to other sources of information, clarity about how to interpret results, and better reports of results both from their contractors and the CQC. Less frequently mentioned but also on the list, organizations wanted help to improve response rates as well as changes to the analysis methods so that they include qualitative sentiment analysis.

The results also brought forward ideas for improvement strategies, such as mapping organizational values to questions, triangulating data from multiple sources to identify trends even when data is not directly linkable, sharing feedback with staff, and using "improvement maps" to gauge which questions are most important to patients (currently only provided by hired survey contractors). The underlying theme of the improvements was that chasing individual questions was not as fruitful as rectifying the root causes that are behind negative scores. This approach included a focus on integrating survey findings into conversations involving operational development to stimulate better patient experiences. More successful organizations found certain survey questions symptomatic of larger organizational health issues and recognized that improvement was going to take a more concerted effort than a single focus on one particular question. These approaches emphasized working on organizational values and staff experience rather than targeting a specific question. Furthermore, these examples support the idea that provision of clear information and supplies to make ward life more enjoyable can improve experiences without making drastic changes to care delivery. Finally, there was support for involving staff in the process of learning from feedback, both in giving them positive feedback from patients and working with them to design experience feedback collections specific to their patients.

Limitations

The research approach identified what providers are struggling with when it comes to using patient-reported feedback. Although the case studies were relatively diverse, it would have been beneficial to gather feedback from a wider range of respondents. Furthermore, some interviewees declined to be recorded, meaning that their files were based on notes, which contained less precise information for coding. Finally, some of the results regarding successful approaches to data use cannot be taken at face value, because things like improvement maps (where certain survey questions are mapped to an overall question) are conceptually questionable in terms of providing accurate levels or importance [6].

Implications for Health Policy

Through this research, health care staff have provided a blueprint for optimizing national systems related to patient experience feedback including how it is collected, analyzed, and presented. In order for patient-reported feedback to be an effective improvement tool, and avoid the ethical grey zone around soliciting patient input and not acting on it, feedback programs need to make efforts to facilitate data comprehension and use.

Staff have offered a considerable amount of insight into how best to improve the system regarding patient experience feedback, such that it generates useful intelligence for organizational improvement. Some of the staff's suggestions could be seen as simple adjustments to existing surveys, such as larger sample sizes, reports that are more appropriately pitched to the audience, and revisions in survey methodologies such that they make sense to service providers. Facilitating local data use also requires the system regarding patient surveys to provide relevant data breakdowns and intuitive reports and presentations. This is true in the NHS, but also in international contexts. Not only are staff eager to have ward- and service-level data, they need survey results to explain which aspects of experience are most important to various patient groups. This approach likely requires soliciting and relaying a different kind of data entirely. Different types of data have different utilities for staff, but the feedback of staff in this study present a desire to further explore near-real-time feedback (that does not risk confidentiality) and extraction from unstructured data; more appropriately called *patient stories*.

These more difficult challenges are perhaps the most important. The idea of linking feedback to other information represents staff's inclination to move towards more holistic quality improvement rather than continue to analyze and respond to a wide range of disparate, uninterpretable data. Enabling wish list items like this would require a paradigm shift in patient experience feedback collection.

The paradigm regarding patient experience feedback is heavily rooted in large national initiatives, the NPSP and the FFT, both of which are accompanied by a sluggish bureaucracy and political concerns. It is likely that these initiatives are neither capturing, nor producing, what is most useful to the organizations trying to use patient feedback to improve care. Listening to what staff said in this interview study should ignite a change in thinking and compel the actors within the system to collect clear, linkable, digitally mature, and timely information. Furthermore, truly understanding what matters to patients (another wish list item) requires a different level of engagement with patients beyond testing surveys and asking people what they expect from their care pathway.

These ideas for change do not suggest abolishing national survey initiatives. Currently, these initiatives still hold the only academically robust source of patient experience feedback and are likely to play a role for a long time. Rather, these ideas call for improvement in the investment in feedback collection; they demand modernizing feedback collection and revamping it to be flexible to patient priorities, which is reflective of the whole patient population and more useful to the frontline. If the system is open to new approaches, these changes will help transform unused data into business intelligence insights for clinical informatics.

Conclusion

Experience has joined effectiveness and safety to form the quality pyramid that has been accepted by policy makers, providers, and patients. Patient-reported feedback programs are now a staple of developed health care systems; however, they have not yet achieved their full potential as a conduit for patient needs and preferences to enter into quality improvement strategies. This research demonstrates which barriers lie behind the problem. More importantly it illuminates what staff want and need from the system related to patient experience feedback, in order to put the data to use.

The focus on enhanced data presentation came through very strongly, as did the desire for patient-reported feedback to be explained in a way that is meaningful not only to analysts, but also to frontline staff. It is also clear that more needs to be done to enable data linkage so that staff can explore problems within specialties and across datasets. Finally, the most powerful finding of this study was one for a shared network of success and shared learning. The examples in this research makes inroads towards such shared learning and will hopefully be the beginning of a growing repository of successful approaches to using patient feedback that can help the system adapt to changing local needs.

Acknowledgments

The authors acknowledge the contributions of the eight NHS organizations that agreed to participate in this work and extend thanks to the 18 participants who offered their time and shared their insights. The research was supported by the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre and NIHR Imperial Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of improvement.

[PDF File (Adobe PDF File), 64KB - jmir_v20i4e141_app1.pdf]

References

1. National Health Service. London; 2014 Oct. Five Year Forward View URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [accessed 2018-03-12] [WebCite Cache ID 6xrPz1CGn]
2. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013 Jan 03;3(1):e001570 [FREE Full text] [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
3. Garratt A, Solheim E, Danielsen K. National and cross-national surveys of patient experiences: a structured review [Internet]. NIPH Methods Resources 2008:1 Oslo, Norway: Knowledge Centre for the Health Services at The Norwegian Institute of Public Health (NIPH); 2008 Feb. Report from Norwegian Knowledge Centre for the Health Services (NOKC) No. 07-2008. [Medline: 29320000]
4. Darzi A. High quality care for all: NHS next stage review final report. London: Department of Health; 2008.
5. Care Quality Commission. 2016. Inpatient questionnaire (scored questionnaire) URL: http://www.nhssurveys.org/Filestore/Inpatient_2016/IP16_Scored_Questionnaire_v1.pdf [accessed 2018-03-12] [WebCite Cache ID 6xrOqcGsR]
6. Flott K, Graham C, Darzi A, Mayer E. Can we use patient-reported feedback to drive change? The challenges of using patient-reported feedback and how they might be addressed. BMJ Qual Saf 2016 Jun 20;26(6):502-507. [doi: 10.1136/bmjqs-2016-005223]
7. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
8. Department of Health. London; 2009. Understanding what matters: a guide to using patient feedback to transform services URL: http://www.nhssurveys.org/Filestore/documents/DH_Understanding_what_matters.pdf [accessed 2018-03-12] [WebCite Cache ID 6xrOv7jZs]

9. Coulter A, Locock L, Ziebland S, Calabrese J. Collecting data on patient experience is not enough: they must be used to improve care. BMJ 2014 Mar 26;348(mar26 1):g2225-g2225. [doi: 10.1136/bmj.g2225]
10. DeCourcy A, West E, Barron D. The National Adult Inpatient Survey conducted in the English National Health Service from 2002 to 2009: how have the data been used and what do we know as a result? BMC Health Serv Res 2012 Mar 21;12:71 [FREE Full text] [doi: 10.1186/1472-6963-12-71] [Medline: 22436670]
11. Wensing M, Vingerhoets E, Grol R. Feedback based on patient evaluations: a tool for quality improvement? Patient Educ Couns 2003 Oct;51(2):149-153. [Medline: 14572944]
12. Reeves R, West E. Changes in inpatients' experiences of hospital care in England over a 12-year period: a secondary analysis of national survey data. J Health Serv Res Policy 2015 Jul;20(3):131-137. [doi: 10.1177/1355819614564256] [Medline: 25534393]
13. CQC Patient Survey Team. Care Quality Commission. 2016 Jan. NHS patient survey programme: survey output user report URL: http://www.cqc.org.uk/sites/default/files/20160127_survey_output_user_report.pdf [accessed 2018-03-12] [WebCite Cache ID 6xrP4KCPW]
14. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf 2014 Jun;23(6):508-518. [doi: 10.1136/bmjqs-2013-002524] [Medline: 24505110]
15. Raleigh V, Graham C, Thompson J, Sizmur S, Jabbal J, Coulter A. The King's Fund. London, UK: Picker Institute Europe; 2005. Patients' experience of using hospital services: An analysis of trends in inpatient surveys in NHS acute trusts in England, 2005–13 URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Patients-experience-Kings-Fund-Dec-2015.pdf [accessed 2018-03-12] [WebCite Cache ID 6xrPFY9ND]
16. United Kingdom Parliament. House of Commons Proceeding. 2007 Feb 20. Hansard written answers (pt 0022) URL: https://publications.parliament.uk/pa/cm200607/cmhansrd/cm070220/text/70220w0022.htm [accessed 2018-03-12] [WebCite Cache ID 6xrPMANze]
17. InHealth Associates & Membership Engagement Services. 2015. Making sense and making use of patient experience data URL: https://www.membra.co.uk/case-studies/making-sense-and-making-use-patient-experience-data [accessed 2018-03-12] [WebCite Cache ID 6xrPR7wWq]
18. Department of Health. 2011. NHS patient experience framework URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215159/dh_132788.pdf [accessed 2018-03-12] [WebCite Cache ID 6xrPSlgLX]
19. Care Quality Commission. 2016 Jan. NHS patient survey programme engagement strategy URL: http://www.cqc.org.uk/sites/default/files/20160125%20User%20Engagement%20Strategy.pdf [accessed 2018-03-12] [WebCite Cache ID 6xrPUneOk]
20. Care Quality Commission.: Picker Institute Europe; 2016. Sampling handbook for the NHS adult inpatient survey 2016: in-house trusts URL: http://www.nhssurveys.org/Filestore/Inpatient_2016/IP16_Sampling_Handbook_InHouse_v3.pdf [accessed 2018-03-12] [WebCite Cache ID 6xrPYqer5]
21. Robert G, Brearley S. What matters to patients? Developing the evidence base for measuring and improving patient experience. London, UK: The King's Fund; 2011. URL: http://www.wales.nhs.uk/sites3/documents/420/Final%20Project%20Report%20pdf%20doc%20january%202012%20(2).pdf [accessed 2018-03-12] [WebCite Cache ID 6xrPbdqeU]
22. Picker Institute Europe. 2016. Focused on staff experience national NHS staff survey co-ordination centre URL: http://www.nhsstaffsurveys.com/Page/1056/Home/NHS-Staff-Survey-2016/ [accessed 2018-03-12] [WebCite Cache ID 6xrPetkT8]
23. Reeves R, Seccombe I. Do patient surveys work? The influence of a national survey programme on local quality-improvement initiatives. Qual Saf Health Care 2008 Dec;17(6):437-441 [FREE Full text] [doi: 10.1136/qshc.2007.022749] [Medline: 19064659]

Abbreviations

CQC: Care Quality Commission
FFT: Friends and Family Test
NHS: National Health Service
NIHR: National Institute for Health Research
NPSP: National Patient Survey Program

Edited by G Eysenbach; submitted 22.08.17; peer-reviewed by S Kula, R Pankomera; comments to author 20.09.17; revised version received 26.09.17; accepted 27.09.17; published 24.04.18.

Please cite as:
Flott K, Darzi A, Gancarczyk S, Mayer E
Improving the Usefulness and Use of Patient Survey Programs: National Health Service Interview Study
J Med Internet Res 2018;20(4):e141
URL: http://www.jmir.org/2018/4/e141/
doi:10.2196/jmir.8806
PMID:29691207

Original Paper

Analysis of College Students' Personal Health Information Activities: Online Survey

Sujin Kim[1], PhD; Donghee Sinn[2], PhD; Sue Yeon Syn[3], PhD

[1]Division of Biomedical Informatics, Department of Internal Medicine, University of Kentucky, Lexington, KY, United States

[2]Department of Information Science, University at Albany – State University of New York, Albany, NY, United States

[3]Department of Library and Information Science, The Catholic University of America, Washington, DC, United States

Corresponding Author:
Sujin Kim, PhD
Division of Biomedical Informatics
Department of Internal Medicine
University of Kentucky
208H, MDS Building
725 Rose Street
Lexington, KY
United States
Phone: 1 859 218 0110
Fax: 1 859 257 0483
Email: skim3@uky.edu

Abstract

Background: With abundant personal health information at hand, individuals are faced with a critical challenge in evaluating the informational value of health care records to keep useful information and discard that which is determined useless. Young, healthy college students who were previously dependents of adult parents or caregivers are less likely to be concerned with disease management. Personal health information management (PHIM) is a special case of personal information management (PIM) that is associated with multiple interactions among varying stakeholders and systems. However, there has been limited evidence to understand informational or behavioral underpinning of the college students' PHIM activities, which can influence their health in general throughout their lifetime.

Objective: This study aimed to investigate demographic and academic profiles of college students with relevance to PHIM activities. Next, we sought to construct major PHIM-related activity components and perceptions among college students. Finally, we sought to discover major factors predicting core PHIM activities among college students we sampled.

Methods: A Web survey was administered to collect responses about PHIM behaviors and perceptions among college students from the University of Kentucky from January through March 2017. A total of 1408 college students were included in the analysis. PHIM perceptions, demographics, and academic variations were used as independent variables to predict diverse PHIM activities using a principal component analysis (PCA) and hierarchical regression analyses (SPSS v.24, IBM Corp, Armonk, NY, USA).

Results: Majority of the participants were female (956/1408, 67.90%), and the age distribution of this population included an adequate representation of college students of all ages. The most preferred health information resources were family (612/1408, 43.47%), health care professionals (366/1408, 26.00%), friends (27/1408, 1.91%), and the internet (157/1408, 11.15%). Organizational or curatorial activities such as Arranging, Labeling, Categorizing, and Discarding were rated low (average=3.21, average=3.02, average=2.52, and average=2.42, respectively). The PCA results suggested 3 components from perception factors labeled as follows: Assistance (alpha=.85), Awareness (alpha=.716), and Difficulty (alpha=.558). Overall, the Demographics and Academics variables were not significant in predicting dependent variables such as Labeling, Categorizing, Health Education Materials, and Discarding, whereas they were significant for other outcome variables such as Sharing, Collecting, Knowing, Insurance Information, Using, and Owning.

Conclusions: College years are a significant time for students to learn decision-making skills for maintaining information, a key aspect of health records, as well as for educators to provide appropriate educational and decision aids in the environment of learning as independent adults. Our study will contribute to better understand knowledge about specific skills and perceptions for college students' practice of effective PHIM throughout their lives.

(J Med Internet Res 2018;20(4):e132) doi:10.2196/jmir.9391

KEYWORDS

health records, personal; health information management; student health services

Introduction

Background

With abundant personal health information at hand, individuals are faced with a critical challenge in evaluating the informational value of the health care records to keep useful information and discard that which is determined useless. College students, in particular, are confronted with a similar issue; however, their situation is quite different from that of the senior population. Young, healthy college students who were previously dependents of adult parents or caregivers are less likely to be concerned with disease management. As such, their lack of interest in health care [1] leads to further disinterest in personal health document management. Personal health information management (PHIM) is a special case of personal information management (PIM) that is associated with multiple interactions among varying users (eg, patients, providers, insurance companies), complex health information and systems (eg, labs, medications, insurance), and advanced health information technology tools (eg, personal health records, PHRs; personal health devices) [2-4]. In PHIM research, little is known about college students' information management activities in the context of health. Thus, this study investigates the demographic and academic profiles of college students with regard to diverse PHIM activities. Additionally, this study aims to discover the major determinants of key information management activities among college students for health information. This study reviews existing literature about diverse PHIM activities and document types and college students' health information–seeking with relevance to their PIM behaviors.

Personal Health Information Management Activities and Document Types

What individuals do with their personal health documents has been studied to understand diverse information management activities, document types, and related personal behaviors. As a health focus of PIM, core tasks of PIM or PHIM activities include "the search, retrieval, and re-finding of previously encountered information from both personal and shared space" [4,5]. Among these activities, individuals develop and use their own strategies to manage and organize their personal records. However, it is not clear if the strategies are effective or efficient. In the PIM context, researchers have observed that "the individual characteristic of being orderly has a positive bearing at a later point in time when the individual needs to find this information" [3,6-8]. Furthermore, successful PIM retrieval is dependent on the "prior processes used to organize relevant information and the extent to which those processes were appropriately planned" [3].

There is no comprehensive understanding of sources or document types contained in PHR systems. However, some PHIM studies reported specific or situational aspects of PHIM sources and document types. The most important PHR sources are health care providers who are broadly responsible for delivery and administration of health care. This group generates diverse types of health records (or documents) at clinical encounters such as care notes (eg, discharge summary, physical exam), therapeutic notes (eg, operative notes, treatment regimes, procedure information, surgeries), imaging or lab results (eg, x-rays, pathology, cytology), or administrative or legal or financial information (eg, appointment schedule, medical bills and receipts, birth certificate or death certificate, date of birth) [3]. Health care insurers were also reported as the relevant PHR source representing any health insurance program that "helps pay for medical expenses; whether through privately purchased insurance, social insurance or a social welfare program funded by the government" [3]. This information is then accessible for further investigation at times of inquiry or need.

Nowadays, some tethered PHRs can selectively or potentially include calendar or diary entries, daily planners, medications and tools, reference material, referrals, poison control, cancer surveys, over-the-counter medications, exercise and diet or self-care logs, home-monitored data (eg, blood pressure, glucose, peak flow), logs of symptoms, or pedometer data [9-11]. Additionally, individuals' social networks such as family, friends, and other informal human sources were reported as relevant PHR sources. In recent times, online support communities of people with similar diseases, such as PatientsLikeMe, also constitute relevant PHR sources [12]. Traditional public health sources such as the mass media, public health departments, and libraries still play important roles as PHR sources through health websites, printed health publications, public library classes, etc.

Personal Information Management Perceptions

Although the previous studies have not focused on college population for their health documents, they have identified some interesting findings regarding factors influencing PIM activities [13,14]. Technological solutions or individual knowledge about diverse PIM tools and methods were found to be associated with individuals' success at PIM management [14], especially in digital environments [13,14], as individual users often have limited knowledge about appropriate technical tools or techniques for management and preservation [15]. As PIM technology evolves, diverse PIM activities happen in digital forms, and personal data stores could be at risk in terms of digital service providers' policies and standards [16]. Williams et al reported that technologically perceptive interviewees were diligent with aspects concerning back-ups and mindful about the risk of loss, which was also confirmed in the research of Sinn et al [17,18]. Still, how technology influences young college students in the digital generation in terms of PHIM remains unknown.

The difficulties in PIM activities were investigated, and the 2 most critical challenges that were identified were curatorial and organizational activities. Bruce and colleagues (2005) found

that individuals have difficulty in determining the future value of digital content [19]. Marshall similarly described curatorial decisions as a "cognitively demanding exercise" [13]. Individuals' management methods for information are much more diverse in personal settings than those in organizational settings. People often allow their information to accumulate without taking action to organize it. Actions for decluttering or organizing personal information often happens with trigger events, such as moving offices, buying new devices, and reaching the limits of space capacity [17]. Even in cases where individuals preserve their content, their organizational activities for long-term use seem unlikely to meet a required level, and they especially lack "creating appropriate metadata, and migrating materials to maintainable formats" in a proper and secure data management system [16,20].

To achieve a satisfying level of information retrieval for individuals' needs, some types of assistance, whether technological or professional, might be useful. The patterns of individuals' preservation seem haphazard or premature, such as simple replications or keeping everything including old computers [13,15,17,21-23]. Obviously, those patterns are neither sustainable nor efficient. Many researchers argue that professional intervention in PIM would benefit them greatly to preserve important personal information as well as to preserve cultural heritage from which personal histories could be found [14,17,18]. However, the era of professional support or technological assistance in PIM is still in its infancy, with only limited technological support available mainly for the aging population [24,25].

The sense of ownership or home-grown organization was one of the ways to observe the characteristics of a personal archive [26]. The same applies to the online environment. For example, users perceive the Cloud as a "storage box" on the internet, not going much beyond the concept of ownership [27]. In addition, individuals strongly felt that they should be able to preserve even their own social media data [28]. Hence, the sense of possession or ownership may influence PIM activities.

Another factor was awareness of the importance of personal information. When someone thinks that his or her personal information may be important in a different context (eg, financial, academic, personal history), then he or she may make more of an effort to preserve that information. Personal information builds personal life history, documents important occasions for achievements or memorials, and presents identity construction evidence [26,29]. Although proven to be associated with PIM activity [18], the awareness, however, has not been tested for any specific context, such as health information, college students, or other PIM activities.

College Students' Health Information–Seeking Behaviors

College students enter a critical transition and begin to become independent and responsible for their own health during college years. As they are away from their parents, college students must acquire their health records, such as immunization records, drug test results, or vaccination records, and present them whenever asked for academic admission or employment. Moreover, college students are thought to be a vulnerable population in that they are exposed to pandemic outbreaks such as meningococcal disease and influenza [30-32]. However, they often exhibit lack of interest in either disease management or health information management. Most importantly, this age group is the least insured in the United States [32], in spite of the fact that they are exposed to risky behaviors, such as the highest rates of motor vehicle injury and death, homicide, mental health problems, sexually transmitted infections, and substance abuse [31,32]. In addition, these young adults do not normally seek assistance with finding or maintaining their PHRs until an illness or accident occurs [30].

Studies have reported that college students are using online resources for health information due to their easy access, although the students do not consider them to be credible [33-36]. Given that health and medical information requires professional knowledge to interpret and manage [35,36], this situation could lead to critical health decisions. In this sense, the fact that personal health record keeping has not been a part of college education in a conventional academic setting is problematic. Particularly for health matters, having unknown digital records that hold important personal information may mean being at an increased risk of chronic conditions and their associated complications for many more years, thus making college students an important population in need of immediate health promotion and intervention.

With relevance to health information seeking and sharing activities, Syn and Kim (2016) reported that both contextual and user variations were influencing factors [37]. Ivanitskaya and colleagues reported that "most students (89%) understood that a one-keyword search is likely to return too many documents," and that "few students were able to narrow a search," showing search inefficacy among college students [38]. They also reported that "students' self-perceptions of skill tended to increase with increasing level of education" [38]. Notably, as part of Project Information Literacy, Head and Eisenberg (2009) reported that college students in their survey "used course readings and Google first for course-related research and Google and Wikipedia for everyday life research" [39]. As such, there has been limited evidence to understand informational or behavioral underpinning of the college students' PHIM activities, which can influence their health, in general, throughout their lifetime. This knowledge can help students practice effective PHIM throughout their lives.

Therefore, the aim of this study was to investigate perceived behaviors of college students by asking questions that focused on various information management–related activities through an online survey. Three research questions were tested within our samples. The first research question investigated demographic and academic profiles of college students with relevance to PHIM activities. The second research question sought to construct major PHIM-related activity components and perceptions among college students. The third research question sought to discover the major factors predicting core PHIM activities among college students that we sampled.

Methods

Survey Sample

The target sample was 28,254 students who were included on the University of Kentucky student mailing list (as of January 2017). We excluded those who signed off from the University mailing list according to the Family Educational Rights and Privacy Act. Our online survey responses were collected from March through May 2017. The overall response rate was 9.12% (2578/28,254), and the study included only responses with a survey completion rate greater than 90% (1408/2578, 54.61%). The participants who included their emails participated in a drawing for compensation. The study has been approved by the University of Kentucky institutional review board.

Nonresponse Analysis and Common Method Bias

Low response rate for Web surveys among college students is not a surprising phenomenon. As reported in the recent National Survey of Student Engagement, the response rate ranged from 5% to 77%, with an average of 29% [40]. Our data show a high dropout rate of 44.57% (1149/2578) where the remaining responses were completely missing. Due to low response rate (ie, 9.12%), nonresponse analysis recommended by Babbie (1990) was performed by comparing the initial 30% and the last 30% responses (considered as a proxy for nonresponses) [41]. To compare the 2 groups, we performed the analysis of variance test, which indicates no statistically significant differences between the 2 groups of respondents for the independent and dependent variables. For instance, the demographics variables entered in hierarchical regression analyses, age ($F_{1,784}$=2.532, P=.11 gender ($F_{1,785}$=0.588, P=.44), ethnicity ($F_{1,784}$=0.849, P=.36), and relationship ($F_{1,788}$=0.247, P=.62). Remaining variables entered in the regression analyses were found to be insignificant between the 2 groups. Therefore, the nonresponse bias is considered to be minimal in this study.

Additionally, Harman single-factor test based on confirmatory factor analysis was performed to avoid the common method bias [42]. This study employed the online survey method to measure college students' information management behaviors and perceptions with relevance to personal health record management within the same survey respondents. Therefore, the common method bias issue can be introduced by measuring both independent and dependent variables that were collected from the same survey respondents. Harman single-factor test shows that the largest variance for one factor (ie, age) is 12.92%, which is less than the acceptable value of 50% [43]. Therefore, the common method bias is also considered to be marginal in this study.

Measures

An aggregate sum of 84 PHIM-related activities measuring from 1 (strongly disagree) to 5 (strongly agree) on the Likert scale for each question was used as a dependent variable, namely, Overall PHIM Activity. These 84 questions used as a PHIM activity measure were based on literature [3-8] in our reference. In addition to the overall PHIM score, we formed 11 additional dependent variables using a principal component analysis (PCA) using SPSS v.24. The PCA allows us to convert possibly correlated variables into principle components by the strengths of possible variances, so that we could create principle PHIM activities and record type constructs. The surveyed items and accompanying results are reported in Tables 1-3. From the PCA analysis, the 11 PHIM constructs formed include Labeling, Sharing, Categorizing, Collecting, Health Education Materials, Understanding, Discarding, Insurance Information, Organizing, Using, and Owning. Reliability scores of Cronbach alpha for these 11 PHIM components range from .803 to .969, indicating high internal consistency in PHIM measures (Table 4). For predictor variables, 16 survey questions were analyzed to extract major PIM perception components using the PCA technique. We included only highly reliable components among our predictors for a series of regression analyses. The PIM perceptions used as predictors are labeled as Assistance, Awareness, and Difficulty (Table 3). Additionally, demographics (age, gender, ethnicity), health concerns (number of clinic visits, preferred health information resources), and academic characteristics (status, relationship, grade point average, number of courses taken) were entered in hierarchical regression analyses to predict major predictors for the PHIM activities extracted. To control demographics or academic variances, we recoded some variables into a binary comparison (Female: 1, others: 0; White: 1, others: 0; Undergraduate: 1, others: 0; Single: 1, others: 0) in our hierarchical regression analyses. The preferred health information sources were measured from 1 (representing least preferred) to 9 (most preferred).

Results

We first performed a descriptive analysis to characterize our student sample, which was followed by PCA analyses. On the basis of the PCA results, a series of hierarchical regression analyses were performed to assess which variables were statistically meaningful to predict diverse PHIM activities.

Research Question 1: Sample Characteristics

The first research question sought to profile demographic characteristics of the survey participants (N=1408; Table 1). The majority of participants were female (956/1408, 67.90%), and the age distribution of this population included adequate representation of college students of all ages, except for adults aged 66 years or above (n=8). The participants were oversampled in female population in comparison with University of Kentucky's current student demographics (N=16,628; 54.10%). This sample lacked racial and ethnic diversity in that 72.70% (1023/1408) were white, with African Americans representing the next most sampled population (5.90%, 83/1408). For academic status, 852 students (852/1408, 60.50%) were undergraduate, and the rest represented graduate or certification program students. Half of the students resided in off-campus housing (824/1408, 58.50%), and 48.90% (689/1408) reported being in a relationship. High grade averages were reported, with A (800/1408, 56.80%), B (413/1408, 29.30%), C (87/1408, 6.20%), D (9/1408, 0.60%), and F (9/1408, 0.60%), Participants reported that they predominantly use parent-provided health insurance (787/1408, 55.90%), the student health plan through the University (206/1408, 14.60%), and employment-based insurance (177/1408, 12.60%).

Table 1. Sample description—demographics, academics, health, and information resources (N=1408).

Variables	Statistics
Age, mean (SD)	24.65 (7.10)
Gender, n (%)	
Male	354 (25.14)
Female	956 (67.90)
No response	98 (7.00)
Ethnicity, n (%)	
White, not Hispanic	1023 (72.70)
Black, not Hispanic	83 (5.90)
Hispanic or Latino	53 (3.80)
Asian or Pacific Islander	106 (7.50)
Native American or Alaskan Native	3 (0.20)
Other	44 (3.10)
No response	96 (6.80)
Academic status, n (%)	
1st year undergraduate	237 (16.80)
2nd year undergraduate	197 (14.00)
3rd year undergraduate	189 (13.40)
4th year undergraduate	173 (12.30)
5th year or more undergraduate	56 (4.00)
Mater's program	154 (10.90)
Doctoral program	267 (19.00)
Certification program	5 (0.40)
Other: please specify	36 (2.60)
No response	94 (6.70)
International students, n (%)	
Yes	86 (6.10)
No	1227 (87.10)
No response	95 (6.70)
Housing, n (%)	
Campus residence hall	315 (22.40)
Fraternity or sorority house	30 (2.10)
Other university housing	36 (2.60)
Off-campus housing	824 (58.50)
Parent or guardian's home	71 (5.00)
Other: please specify	38 (2.70)
No response	94 (6.70)
Relationship, n (%)	
Single	689 (48.90)
Married/domestic partner	188 (13.40)
Engaged/committed dating relationship	403 (28.60)
Separated	8 (0.60)
Divorced	13 (0.90)

Variables	Statistics
Widowed	2 (0.10)
Other: please specify	13 (0.90)
No response	92 (6.50)
Tuition support, n (%)	
Parents	521 (37.00)
Student loans	509 (36.20)
Self	441 (31.30)
Your employer	94 (6.70)
Scholarships (eg, teaching/research assistantship)	683 (48.50)
Grade point average, n (%)	
A	800 (56.80)
B	413 (29.30)
C	87 (6.20)
D/F	9 (0.60)
No response	99 (7.00)
Health insurance, n (%)	
Parent health insurance	787 (55.90)
Employment-based insurance	177 (12.60)
Student health plan through universities	206 (14.60)
Subsidized Obamacare coverage	17 (1.20)
Catastrophic coverage	2 (0.10)
Medicaid	68 (4.80)
Not insured	23 (1.60)
Other: please specify	24 (1.70)
No response	104 (7.40)
Health information sources sought first, n (%)	
Professionals (eg, doctors, nurses, etc.)	366 (26.00)
Family	612 (43.50)
Friends	27 (1.90)
Colleagues (eg, other patients)	6 (0.40)
Internet	157 (11.20)
Social media	2 (0.10)
Mass media	1 (0.10)
Government agencies	3 (0.20)
Libraries	3 (0.20)
Other	16 (1.10)
Number of courses taken, mean (SD)	27.08 (22.64)
Number of clinic visit, mean (SD)	4.32 (6.33)
Number of digital devices owned, mean (SD)	3.18 (1.93)
Number of mobile phones owned, mean (SD)	3.57 (11.94)

Table 2. Personal health information management (PHIM) activities by document types.

PHIM activities by document types	Immunization records, mean (SD)	Family medical history, mean (SD)	Emergency information, mean (SD)	Surgery, mean (SD)	Drugs, mean (SD)	Insurance information, mean (SD)	Health education materials, mean (SD)	Average activities
I already have a collection of_____	3.46 (1.28)	2.99 (1.23)	3.89 (1.12)	3.42 (1.26)	3.71 (1.19)	4.15 (0.95)	2.95 (1.30)	3.51
I have a habit of collecting _____ whenever providing for my health	3.27 (1.32)	3.04 (1.25)	3.56 (1.24)	3.28 (1.29)	3.52 (1.26)	3.81 (1.18)	2.91 (1.31)	3.34
I know which of _____ are needed for my doctor's visits	3.55 (1.28)	3.65 (1.21)	3.95 (1.10)	3.71 (1.21)	3.95 (1.11)	4.09 (1.04)	3.29 (1.31)	3.74
I discard _____ when they are no longer needed	2.2 (1.10)	2.2 (1.08)	2.21 (1.12)	2.21 (1.11)	2.53 (1.26)	2.49 (1.28)	3.08 (1.38)	2.42
I have my own method to manage and organize	3.24 (1.30)	3.14 (1.29)	3.41 (1.26)	3.23 (1.29)	3.4 (1.26)	3.6 (1.21)	3.01 (1.30)	3.29
I categorize _____ on a regular basis	2.47 (1.21)	2.41 (1.17)	2.58 (1.24)	2.45 (1.19)	2.66 (1.26)	2.72 (1.28)	2.36 (1.16)	2.52
I arrange _____ effectively so that I can find it easily for my doctor's appointment	3.15 (1.35)	3.03 (1.31)	3.35 (1.32)	3.09 (1.32)	3.37 (1.31)	3.66 (1.24)	2.79 (1.32)	3.21
I label _____ in a meaningful way so I can find it easily for later use for my doctor's appointment	3.02 (1.37)	2.9 (1.33)	3.12 (1.36)	2.97 (1.34)	3.16 (1.35)	3.3 (1.36)	2.7 (1.30)	3.02
Usually, I try to personally own a copy of _____ in my possession	3.33 (1.36)	3.02 (1.32)	3.47 (1.32)	3.09 (1.33)	3.47 (1.31)	3.96 (1.15)	2.75 (1.33)	3.30
I can easily find _____ in an efficient manner	3.39 (1.31)	3.28 (1.29)	3.76 (1.20)	3.37 (1.29)	3.69 (1.21)	3.98 (1.08)	3.07 (1.34)	3.51
I can easily share my_____ records, when needed	3.36 (1.31)	3.28 (1.27)	3.74 (1.21)	3.42 (1.27)	3.67 (1.23)	3.91 (1.10)	3.01 (1.32)	3.48
I use _____ when I discuss my health matters with a health professional	3.24 (1.27)	3.59 (1.18)	3.69 (1.16)	3.57 (1.21)	3.89 (1.10)	3.86 (1.11)	2.94 (1.30)	3.39
Average by data types	3.14	3.04	3.39	3.15	3.42	3.63	2.91	

The most preferred health information resources were as follows: family (43.50%, 612/1408), health care professionals (366/1408, 26.00%), friends (27/1408, 1.90%), and the internet (157/1408, 11.20%). Compared with other studies, this sample prefers depending more on family for health information sources than health care professionals [1,33]. Although this is not a direct comparison, the average number of clinic visits in this sample was 4.32 times more than those in the past year, implying a relatively healthy population compared with the national average of 12.9 visits in 2001 and 11.6 visits in 2010 among people aged between 18 and 64 years who reported fair or poor health [44].

Among the 12 PHIM activities, the participants reported that "I know which of (document types) are needed for my doctor's visits" (average=3.74) ranked the highest (Table 2). Organizational or curatorial activities such as Arranging, Labeling, Categorizing, and Discarding were rated low (average=3.21, average=3.02, average=2.52, and average=2.42, respectively). For the record types–related questions, we found that Insurance information was the PHIM data type that was most actively managed (average=3.63), whereas Health Education Materials and Family Medical Histories were the least favorably pursued PHIM data types (average=2.91 and 3.04, respectively).

Research Question 2: Major Personal Health Information Management Constructs

The second research question sought to identify principle factors for PHIM activities using PCA analyses. In addition to the demographic information, we included PIM perception factors as predictors. The PCA results suggested 5 components from perception factors, 2 of which were eliminated due to low reliability scores, resulting in 3 components labeled as follows: Assistance (alpha=.85), Awareness (alpha=.716), and Difficulty (alpha=.558). Table 3 reports further details on PCA results performed on PIM perceptions.

Table 3. Primary factors of personal information management (PIM) perceptions.

Components	Factor 1	Factor 2	Factor 3
Factor 1: Assistance			
If I have professional assistance, I think I will be able to manage my personal records better	0.812		
I would like professional advice about managing personal records	0.844		
Training would be useful to manage my personal records better	0.848		
I would like to have technology assistance to manage my personal records	0.738		
Factor 2: Awareness			
It is important to keep my personal records for future use		0.812	
It is critical to collect my academic records for my future career		0.764	
It is essential to store my health records to better manage my health		0.707	
Factor 3: Difficulty			
It takes considerable time to look through my personal records to determine what to keep and what to delete			0.830
I find it difficult to know how I should organize my personal records			0.659
Mean (SD)	3.300 (1.051)	4.220 (0.782)	3.370 (1.093)
Cronbach alpha	.85	.716	.558
Eigenvalue	3.607	2.882	1.003
Percentage of variance explained	22.542	18.013	6.271

Table 4. Summary of PCA results by component for primary factors in personal health information management (PHIM) activities for record types.

PCA Result Summary by Components	1	2	3	4	5	6	7	8	9	10	11
Cronbach alpha	.969	.933	.971	.924	.895	.944	.925	.803	.964	.922	.933
Mean	3.052	3.550	2.500	3.335	2.962	3.795	2.414	3.933	3.250	3.692	3.252
SD	1.347	1.244	1.205	1.270	1.305	1.166	1.179	1.113	1.289	1.162	1.329
Eigenvalue	40.867	6.606	4.502	2.901	2.415	2.102	2.050	1.727	1.655	1.508	1.420
Percentage of variance explained	48.651	7.865	5.359	3.454	2.875	2.503	2.440	2.056	1.970	1.795	1.690

On the basis of the responses to 84 PHIM questions, we performed a PCA analysis to form statistically meaningful constructs for use as dependent variables in the hierarchical regression analyses. As a result, the model yielded 11 distinct factors that represent 11 PHIM activities (Multimedia Appendix 1 and Table 4). The factors accounted for about 78.9% of the variance. The scores for the scales were summed and divided by the number of items in the scale to produce variables ranging from 1 to 5, with smaller values indicating lower levels of agreement. The reliability of the 11 factors was also assessed to measure strengths of the scales. The 11 factors were subsequently labeled as follows: Labeling (alpha=.969), Sharing (alpha=.933), Categorizing (alpha=.971), Collecting (alpha=.924), Health Education Materials (alpha=.895), Knowing (alpha=.944), Discarding (alpha=.925), Insurance Information (alpha=.803), Organizing (alpha=.964), Using (alpha=.922), and Owning (alpha=.933). Multimedia Appendix 1 reports the full PCA result.

Research Question 3: Predicting Primary Factors for Personal Health Information Management Activities

The last research question sought to discover which independent variables are affecting factors to the major PHIM activities constructed from the PCA. The relationship between possible factors from the college students' characteristics and the 12 PHIM activity constructs is the focus of the third research question. A series of 12 dependent variables (overall PHIM activity + 11 PHIM activity constructs) were tested with regression analyses using 4 groups of factor variables, including Demographics, Academics, Health Resources, and PIM Perceptions. For the hierarchical regression analyses, Demographics variables were entered in the first block, Academics variables—GPA, number of courses taken, academic status—were entered in the second block, Health and Information Resources variables—number of clinic visits and the 5 health information sources including professionals, family, friends, the internet, and mass media—were entered in the third block, and 3 PIM Perceptions variables of assistance, awareness, and difficulty were entered in the fourth block. Multimedia Appendix 2 shows an aggregate result of the hierarchical regression analyses between the 4 independent variables

(predictors) and the 12 dependent variables (PHIM activities), and Table 5 shows the hierarchical regression analysis predicting overall personal health information management (PHIM) activity.

Overall, Health and Information Resources and Perceptions significantly increased the explanatory power of the regression model. More specifically, the Demographics and Academics variables were not significant in predicting the dependent variables such as Labeling, Categorizing, Health Education Materials, and Discarding, whereas they were significant for other outcome variables. Among Demographics variables, gender (coded female=1) significantly explained the outcome variables of Sharing, Collecting, Knowing, Insurance Information, Using, and Owning. Although the overall Academics variables significantly explain some activities, none of the individual Academics variables are significant for each dependent construct. For Health and Information Resources variables, the number of clinic visits is significant in Sharing and Using variables. Some of preferred health information resources such as professionals and friends are found to be significant factors in some PHIM activity variables. The internet and mass media did not significantly predict most of the PHIM variables. Among the 3 PIM Perceptions variables, Awareness is the most significant of the 12 outcome variables, whereas the Difficulty variable is not significant in health education–related and discarding activity. Interestingly, the Assistance variable is found to be significant in Labeling, Sharing, and Organizing variables.

A 4-stage hierarchical multiple regression was conducted with the overall PHIM activity construct as a dependent variable for predictor variables used in the below analysis.

Table 5. Hierarchical regression analysis predicting overall personal health information management (PHIM) activity.

Dependent predictors	Overall personal health information management (PHIM) activity							
	R	R^2	ΔR^2	ΔF	Degrees of freedom	B	*t*	*P* value
Demographics	.139	.019	.019	5.272	4,1069		5.272	<.001
Age						0.532	1.360	.17
Gender[a] (Female)						13.392	2.885	.004[b]
Ethnicity[c] (White)						4.745	0.945	.35
Relationship[d] (single)						−5.940	−1.393	.16
Academics	.148	.022	.003	0.979	7,1066		3.432	<.001
Grade point average						3.594	1.082	.28
Number of courses taken						0.028	0.312	.76
Academic[e] status						7.863	1.406	.16
Health and information resources[f]	.262	.069	.047	8.872	13,1060		6.025	<.001
Number of clinic visit						0.626	1.964	.05[b]
Professional						5.271	3.467	<.001
Family						0.755	0.444	.66
Colleague						3.993	2.614	.009[b]
Internet						−1.931	−1.235	.22
Mass media						−2.496	−1.314	.19
Personal information management (PIM) perceptions[g]	.379	.144	.075	30.874	16,1057		11.098	<.001
Assistance						−1.445	−2.215	.03[b]
Awareness						9.506	8.589	<.001
Difficulty						−4.216	−3.431	<.001

[a]Gender: Dummy-coded with Female=1 and Male=0.

[b]*P*<.05.

[c]Ethnicity: Dummy-coded with White=1 and Others=0.

[d]Relationship: Dummy-coded with Single=1 and Others=0.

[e]Status: Dummy-coded with Undergraduates=1 and Others=0.

[f]Preferred Health Information Resources: 1 (least preferred) to 9 (Most preferred).

[g]PIM Perceptions: five-point response scale from Strongly Disagree (1) to Strongly Agree (5).

The hierarchical multiple regression revealed that at stage one, Demographics variability as a group contributed significantly to the regression model, $F_{4,1069}$=5.272, P<.001, and accounted for 1.9% of the variation in overall PHIM variables. However, when considering the individual variables, only gender variation indicates significant contribution, whereas the other Demographic variables are not significant. Introducing the Academics variables at the next stage does not explain any additional variation in overall PHIM activity, but this change in R^2 is significant statistically at $F_{7,1066}$=3.432, P<.001. Adding Health and Information Resources variables to the regression model explains an additional 4.7% of the variation in overall PHIM activity, and this change in R^2 is significant by $F_{13,1060}$=6.025, P<.001. At the final stage, the addition of PIM Perceptions to the regression model explains an additional 7.5% of the variation in overall PHIM activity, and this change in R^2 is also significant by $F_{16,1057}$=11.098, P<.001. When all the 4 blocks of independent variables were included in stage 4 of the regression model, Assistance, Awareness, and Difficulty perceptions were found to be significant. Female, Professional, and Mass Media Resources variables were also found to be significant in predicting overall PHIM activities. The most important predictor of overall PHIM activity is the PIM Perceptions predictor, which uniquely explains 7.5% of the variation in overall PHIM activity. Taken together, all the 5 independent variables accounted for 14.4% of the variance in overall PHIM activity.

Discussion

Overall Characteristics

This study examined college students' behaviors and perceptions about information management of their health-related information. Demographically, our sample is young, female, single, and white dominant, and this study sample is similar to that of the University of Kentucky student body, except for oversampled female students. These young adults reported high GPAs and generally seek health information from family and health professionals. Their total number of annual clinic visits indicates a relatively healthy population compared with the national average. The most active PHIM behaviors found were to know the value of health records, and collect and easily find the records. Their least active PHIM behaviors were to discard health information when no longer needed, and to categorize and label them for proactively organizing them. They highly value insurance and drug information but consider health education materials and family medical history less important for PHIM activities.

Descriptive Characteristics

The descriptive analysis suggests that our study participants are heavily dependent on parents in terms of their financial support, including tuition payments and health insurance. It is not surprising, then, that the majority of college students are transitioning to build financial as well as medical independence from their parents. Their financial responsibilities speak not only to monetary dependency but also to information dependency on their parents or family members. In terms of information sources they use, we anticipate that authentic information sources such as clinicians or other health care providers would be sought by the students. However, majority of the participants report that they sought health information through family members first. This result is not consistent with other studies reporting that health care providers are initially or most frequently sought out as a health information source [1,25,36], and it is tied to our previous discussion about college students' transitioning phase of information dependency.

Affecting Factors

Our PCA analyses suggested 3 perception constructs, namely, Assistance, Awareness, and Difficulty. Among these, Assistance represents any professional or technological help in managing various PHIM activities, which have been discussed in previous research. Conversely, this study found that neither Technology nor Ownership questions were formed as significant factors, so we did not include them in our regression analyses. Surprisingly, we found that PIM Perceptions overall were a predictor construct, as well as the individual variables of Assistance, Awareness, and Difficulty. Additionally, Awareness and Difficulty were formed as statistically viable constructs for further analyses. Awareness indicates how college students perceive the importance of information management in their personal records and was shown to be highly perceived in this group. Thus, we entered this variable in our regression analyses.

As dependent variables, we formed 11 constructs out of 84 questions. Interestingly, the constructs formed such as Collecting, Organizing, or Using were consistent with previous PHIM literature [3,25]. However, some of the major PHIM activities such as Retrieving, which were heavily studied in information retrieval or seeking studies, were not statistically significant in internal consistency to form a construct and thus not included in our analyses. Although the descriptive values were not high, some variables, such as Labeling and Categorizing, presented good patterns as components. Curatorial activities such as Categorizing, Labeling, or Organizing were also important constructs in our PCA result. Among health information types, Health Education Materials and Insurance Information were used as outcome constructs in this analysis, even though their descriptive values were low. This result calls for further study to evaluate critical values of information contained in these underutilized resources.

The most interesting result was based on a series of 12 hierarchical regression analyses, as these provided an opportunity to examine the aforementioned predictors and their influence on diverse PHIM activities. With these regression results, we could identify possible factors for further analysis to better prepare in PHIM. For instance, many studies have examined the challenges that individuals experience during digital archiving practices [14,17,19,20,23,28,45]. As we stated, our goal was to identify a profile for active PHIM performers and to discover possible improvements based on the current state of PHIM practices and perceptions among college students. The overall results in Table 5 confirm that Demographics (Gender), Academics in general, Health Information Sources (number of clinic visits, professionals and colleagues as Information Sources), and PIM Perceptions (Assistance,

Awareness, and Difficulty) have meaningful influences on PHIM variables. Therefore, we can profile active PHIM performers as female students with good academic standing, who visit clinics frequently and seek information from professionals and colleagues, who are aware of the importance of personal history, who acknowledge the importance of professional assistance, and who understand the difficulties of information management activities.

Females were the dominant population of our study group, and other PIM or information-seeking studies have already reported that females are active information seekers who were found to be better organizers than their counterparts [1,24]. In the second block, we found significant association between overall academic variances and PHIM activities. Academic status—undergraduate versus graduate—was found to be an insignificant predictor of certain PHIM activities such as Collecting or Sharing. In the third block, the number of clinic visits was an important predictor for Sharing and Using, and some health information sources such as professional and colleagues were found to be significant predictors. These 2 sources were found to be highly ranked health information sources and significant predictors of some curatorial PHIM activities including Labeling, Sharing, Categorizing, and Collecting. Such curatorial activities for deciding what to keep and what to delete were reported as the least performed PIM activity for a number of reasons [19]. For instance, whether a record has specific values for future use was difficult to judge for those who have no experience in PIM practices. Even if individuals have PIM experiences, anticipating changing PHIM status is a moving target that any human could adjust depending on unpredictable health conditions. This so-called "post-value recall" is not known until individual situations come into play. In other words, the values of the archived information cannot be perfectly predicted for later use when it is needed [46]. Thus, people often keep more than they need, and they do not expect to use all the information they have archived. At the same time, they still look for assistance with archiving or discarding decisions for better PHIM practices.

Finally, the findings suggest that 3 PIM Perception constructs entered in the last block are significantly associated with PHIM activity predictions. Accordingly, it appears that Awareness and Difficulty predict all PHIM activities, whereas Assistance only predicts the curatorial activities such as Labeling or Categorizing. On the basis of the findings, college students' PHIM activities are influenced by their perceived behaviors, such as PIM awareness and difficulty. Relatedly, the Difficulty construct indicates the areas that information professionals, such as librarians and archivists, have traditionally addressed to assist their users. These professionals are trained to extract major informational values by anticipating future uses while discarding useless information from their public collections. Although archivists work for personal collections in their archives, their value-judgment is based on social, historical, and cultural schema. Although PHIM collections are archived for events of personal interest, information professionals' skills and expertise can provide assistance in developing best PHIM practices, which could be taught to college students through training.

Contribution

This paper contributed to understanding college students' information management activities and tested multiple factors to predict major PHIM activities. In a large-scale survey, data collected from a young adult population with regards to personal record management is a novel finding in PIM studies. In particular, healthy young individuals in a college setting have been neglected in disease-based patient education as well as in health information–seeking studies. We believe that our findings provide critical values of information management skills among college students to promote self-help care management. Predictors that were tested range from demographic variation to academic measures to general PIM perceptions applied in a health information management setting. The findings about diverse PHIM activity constructs could be used as PHIM outcome measures for future validation, as in a training outcome measure for specific PHIM activity. We believe that further studies about cost-benefit analyses on diverse training methodologies on individual PHIM outcomes would even be beneficial to improve PHIM training design and evaluation. We also suggest validating diverse PHIM activity constructs with relevance to clinical or wellness outcomes.

Limitations

This study presents a number of limitations. First, we used large-scale survey data locally collected from one state-owned university. Therefore, our sample might be biased when generalizing to other college settings. The demographics considered in this study represent a female-dominant sample. Although other demographic or academic variables, except for gender, were found to be insignificant predictors, it is advisable to use a more diverse ethnic group in future studies. Additionally, technological variations did not form any significant factor in our analyses. However, our results could change if we collect data years after PHRs or other patient portal services are introduced to the arena of college health. As of today, our participants were not fully aware of PHRs or other patient portal services that could be utilized for managing PHRs. Therefore, further postimplementation studies with PHR users are needed for validating some of our findings.

Conclusions

In conclusion, PHIM among college students is a neglected topic of health information seeking, health services, or health IT research and tool development. Due to the relatively healthy nature of young adults in a rich educational environment, their information management skills could be improved drastically once we pay close attention to individual PHIM skill training and development. The findings of this study indicate that the awareness of PHIM's importance is in place, but the reality of weak skills, such as curatorial activity and least utilized records such as patient education materials, should be acknowledged and remedied while in college or by hospitals' health services. Most particularly, a special focus should be given to train college students about how to assess informational values of personal records and their efficient organization by utilizing various health information technologies. Public collections management and strategies devised for public use in libraries or archives may not be suitable for personal archival or health record collection

management. The dynamic nature of informational values or individual levels of health information literacy will come to the forefront in a personalized PHIM education setting. With the advent of PHRs or any similar patient portal services by large academic hospitals or individual providers, college students will be faced with overloaded health records that have never been addressed in their education [47]. For preventive care, numerous studies have confirmed that informed health consumers exhibit better self-care than those that are not informed. As college students' independence starts from their college years, their health record management is soon to be declared free from parental control. Understanding information flow and values of their personal records is important in understanding college students' health conditions, and relevant PHIM educational endeavors will definitely boost these efforts.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Primary factors in personal health information management (PHIM) activities for record types.

[PDF File (Adobe PDF File), 77KB - jmir_v20i4e132_app1.pdf]

Multimedia Appendix 2

Aggregate result of hierarchical regression analyses for all the 12 personal health information management (PHIM) constructs.

[PDF File (Adobe PDF File), 54KB - jmir_v20i4e132_app2.pdf]

References

1. Kim S. An exploratory study of inactive health information seekers. Int J Med Inform 2015 Feb;84(2):119-133 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.003] [Medline: 25453277]
2. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
3. Agarwal R, Khuntia J. Rockville, MD: Agency for Healthcare Research and Quality, AHRQ Publication No. 09-0075-EF Personal Health Information and the Design of Consumer Health Information Technology: Background Report URL: https://healthit.ahrq.gov/sites/default/files/docs/citation/09-0075-EF.pdf [accessed 2018-01-07] [WebCite Cache ID 6wIVXybxL]
4. Jones W, Bruce H. 2005 Jan 27. A Report on the NSF-Sponsored Workshop on Personal Information Management, Seattle, WA URL: http://pim.ischool.washington.edu/final%20PIM%20report.pdf [accessed 2018-01-07] [WebCite Cache ID 6wIViYun4]
5. Dumais S, Cutrell E, Cadiz JJ, Jancke G, Sarin R, Robbins DC. Stuff I've seen: a system for personal information retrieval and re-use. Presented at: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval; July 28-August 01, 2003; Toronto, Canada. [doi: 10.1145/860435.860451]
6. Malone TW. How do people organize their desks?: Implications for the design of office information systems. ACM Trans Inf Syst 1983;1(1):99-112. [doi: 10.1145/357423.357430]
7. Mander R, Salomon G, Wong YY. A "pile" metaphor for supporting casual organization of information. Presented at: CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 03-07, 1992; Monterey, California, USA p. 627-634.
8. Whittaker S, Hirschberg J. The character, value,management of personal paper archives. ACM Trans Comput Hum Interact 2001;8(2):150-170.
9. Civan A, Skeels MM, Stolyar A, Pratt W. Personal health information management: consumers' perspectives. AMIA Annu Symp Proc 2006:156-160 [FREE Full text] [Medline: 17238322]
10. 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]
11. Zayas-Cabán T, Dixon BE. Considerations for the design of safe and effective consumer health IT applications in the home. Qual Saf Health Care 2010 Oct;19 Suppl 3:i61-i67. [doi: 10.1136/qshc.2010.041897] [Medline: 20959321]
12. Patients Like Me. 2018. URL: https://www.patientslikeme.com/?format=html [accessed 2018-01-07] [WebCite Cache ID 6wIVm7hEY]
13. Marshall CC. 2011. Challenges and opportunities for personal digital archiving URL: http://www.csdl.tamu.edu/~marshall/I-Digital-Marshall.pdf [accessed 2018-03-23] [WebCite Cache ID 6y91fdoM2]
14. Sinn D, Kim S, Syn SY. Personal digital archiving: influencing factors and challenges to practices. Library Hi Tech 2017;35(2):222-239. [doi: 10.1108/LHT-09-2016-0103]

15. Strodl S, Motlik F, Stadler K, Rauber A. Personal & SOHO archiving. Presented at: Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries; June 16-20, 2008; Pittsburgh, PA, USA p. 115-123.
16. Van Kleek M, OHara K. The future of social is personal: the potential of the personal data store. In: Miorandi D, Maltese V, Rovatsos M, Nijholt A, Stewart J, editors. Social Collective Intelligence. Cham, Switzerland: Springer International Publishing; 2014:125-158.
17. Williams P, Rowlands I, Dean K, John JL. Ariadne. 2008. Digital lives: Report of interviews with the creators of personal digital collections URL: http://www.ariadne.ac.uk/issue55/williams-et-al [accessed 2018-03-23] [WebCite Cache ID 6y927u1TJ]
18. Sinn D, Syn S, Kim S. Personal records on the web: Who's in charge of archiving, Hotmail or archivists? Library & Information Science Research 2011;33(4):320-330.
19. Bruce H, Jones W, Dumais S. Information Behaviour That Keeps Found Things Found. Inf Res Int Electron J 2004;1:1-23 [FREE Full text]
20. Marshall CC, Bly S, Brun-Cottan F. The long term fate of our digital belongings: Toward a service model for personal archives. Presented at: Proceedings of IS&T Archiving 2006; May 23-26, 2006; Ottawa, Canada p. 25-30 URL: http://www.csdl.tamu.edu/~marshall/marshall-personal-archiving.pdf
21. Copeland AJ. Analysis of public library users' digital preservation practices. J Assoc Inf Sci Technol 2011;62(7):1288-1300. [doi: 10.1002/asi.21553]
22. John J, Rowlands I, Williams P, Dean K. British Library. 2010 Mar 03. Digital lives: personal digital archives for the 21st century: an initial synthesis URL: http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf [accessed 2018-03-06] [WebCite Cache ID 6xid9Eqnt]
23. Marshall C, Shipman F. Social media ownership: using twitter as a window onto current attitudesbeliefs. 2011 May 7 Presented at: The SIGCHI Conference on Human Factors in Computing Systems, ACM; 2011; Vancouver, BC, Canada p. 1081-1090.
24. Kim S, Real K. A profile of inactive information seekers on influenza prevention: a survey of health care workers in Central Kentucky. Health Info Libr J 2016 Sep;33(3):222-238. [doi: 10.1111/hir.12132] [Medline: 26725746]
25. Kim S, Abner E. Predictors affecting personal health information management skills. Inform Health Soc Care 2016;41(3):211-229 [FREE Full text] [doi: 10.3109/17538157.2015.1008490] [Medline: 26268728]
26. Kaye J, Vertesi J, Avery S, Dafoe A, David S, Onaga L, et al. To haveto holdxploring the personal archive. 2011 Mar 7 Presented at: The SIGCHI conference on Human Factors in computing systems, ACM; 2006; Vancouver, BC, Canada p. 275-284.
27. Odom W, Sellen A, Harper R, Thereska E. Lost in translation: understanding the possession of digital things in the cloud. Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 05-10, 2012; Austin, TX, USA p. 781-790 URL: http://willodom.com/publications/paper1202-odom.pdf [doi: 10.1145/2207676.2207789]
28. Marshall CC. Rethinking personal digital archiving, Part 1: Four challenges from the field. D-Lib Mag 2008;14(3/4):2 [FREE Full text]
29. Sinn D, Syn SY. Personal documentation on a social network site: Facebook, a collection of moments from your life? Arch Sci 2014;14(2):95-124.
30. Mulye TP, Park MJ, Nelson CD, Adams SH, Irwin CE, Brindis CD. Trends in adolescent and young adult health in the United States. J Adolesc Health 2009 Jul;45(1):8-24. [doi: 10.1016/j.jadohealth.2009.03.013] [Medline: 19541245]
31. Park MJ, Paul Mulye T, Adams SH, Brindis CD, Irwin Jr CE. The health status of young adults in the United States. J Adolesc Health 2006 Sep;39(3):305-317. [doi: 10.1016/j.jadohealth.2006.04.017] [Medline: 16919791]
32. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J, Campbell MK. Physical activity in young adults: a signal detection analysis of Health Information National Trends Survey (HINTS) 2007 data. J Health Commun 2015;20(2):134-146 [FREE Full text] [doi: 10.1080/10810730.2014.917745] [Medline: 25375396]
33. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
34. Kim H, Park SY, Bozeman I. Online health information search and evaluation: observations and semi-structured interviews with college students and maternal health experts. Health Info Libr J 2011 Sep;28(3):188-199 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00948.x] [Medline: 21831218]
35. Kwan MY, Arbour-Nicitopoulos KP, Lowe D, Taman S, Faulkner GE. Student reception, sources, and believability of health-related information. J Am Coll Health 2010;58(6):555-562. [doi: 10.1080/07448481003705925] [Medline: 20452932]
36. Vader AM, Walters ST, Roudsari B, Nguyen N. Where do college students get health information? Believability and use of health information sources. Health Promot Pract 2011 Sep;12(5):713-722. [doi: 10.1177/1524839910369995] [Medline: 21282492]
37. Syn SY, Kim SU. College students' health information activities on Facebook: investigating the impacts of health topic sensitivity, information sources, and demographics. J Health Commun 2016 Jul;21(7):743-754. [doi: 10.1080/10810730.2016.1157652] [Medline: 27220029]

38. Ivanitskaya L, O'Boyle I, Casey AM. Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). J Med Internet Res 2006;8(2):e6 [FREE Full text] [doi: 10.2196/jmir.8.2.e6] [Medline: 16867969]
39. Head AJ, Eisenberg MB. 2018. Lessons learned: How college students seek information in the digital age URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2281478 [accessed 2018-01-07] [WebCite Cache ID 6wIVxVh8g]
40. National Survey of Student Engagement. 2018. NSSE Response Rate FAQ URL: http://nsse.indiana.edu/pdf/Resp_Rate_FAQ.pdf [accessed 2018-01-09] [WebCite Cache ID 6wLZss7al]
41. Babbie ER. Survey Research Methods. Belmont, CA: Wadsworth Publishing; 1990.
42. Craighead CW, Ketchen Jr DJ, Dunn KS, Hult GT. Addressing common method variance: guidelines for survey research on information technology, operations, and supply chain management. IEEE Trans Eng Manag 2011;58(3):578-588. [doi: 10.1109/TEM.2011.2136437]
43. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
44. O?Hara B, Caswell K. Health status, health insurance, and medical services utilization: 2010. Curr Pop Rep 2012;2010:70-133 [FREE Full text]
45. Beagrie N. Plenty of room at the bottom? Personal digital libraries and collections. Dlib Mag 2005;11(6):00-00 [FREE Full text] [doi: 10.1045/june2005-beagrie]
46. Czerwinski M, Gage DW, Gemmell J, Marshall CC, Pérez-Quiñones MA, Skeels MM, et al. Digital memories in an era of ubiquitous computing abundant storage. Commun ACM 2006;49(1):44-50. [doi: 10.1145/1107458.1107489]
47. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015 Jun 04;17(6):e137 [FREE Full text] [doi: 10.2196/jmir.4381] [Medline: 26043709]

Abbreviations

PCA: principal component analysis
PHIM: personal health information management
PHRs: personal health records
PIM: personal information management

Edited by G Eysenbach; submitted 10.11.17; peer-reviewed by M Abdelhamid, I Adeleke; comments to author 07.12.17; revised version received 22.01.18; accepted 23.01.18; published 20.04.18.

Please cite as:
Kim S, Sinn D, Syn SY
Analysis of College Students' Personal Health Information Activities: Online Survey
J Med Internet Res 2018;20(4):e132
URL: http://www.jmir.org/2018/4/e132/
doi:10.2196/jmir.9391
PMID:29678798

Original Paper

Usage of a Digital Health Workplace Intervention Based on Socioeconomic Environment and Race: Retrospective Secondary Cross-Sectional Study

Conor Senecal[1], MD; R Jay Widmer[1], MD, PhD; Kent Bailey[1], PhD; Lilach O Lerman[1], MD, PhD; Amir Lerman[1], MD

Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Amir Lerman, MD
Mayo Clinic
200 1st St SW
Rochester, MN, 55905
United States
Phone: 1 507 255 1622
Email: lerman.amir@mayo.edu

Abstract

Background: Digital health tools have been associated with improvement of cardiovascular disease (CVD) risk factors and outcomes; however, the differential use of these technologies among various ethnic and economic classes is not well known.

Objective: To identify the effect of socioeconomic environment on usage of a digital health intervention.

Methods: A retrospective secondary cross-sectional analysis of a workplace digital health tool use, in association with a change in intermediate markers of CVD, was undertaken over the course of one year in 26,188 participants in a work health program across 81 organizations in 42 American states between 2011 and 2014. Baseline demographic data for participants included age, sex, race, home zip code, weight, height, blood pressure, glucose, lipids, and hemoglobin A_{1c}. Follow-up data was then obtained in 90-day increments for up to one year. Using publicly available data from the American Community Survey, we obtained the median income for each zip code as a marker for socioeconomic status via median household income. Digital health intervention usage was analyzed based on socioeconomic status as well as age, gender, and race.

Results: The cohort was found to represent a wide sample of socioeconomic environments from a median income of US $11,000 to $171,000. As a whole, doubling of income was associated with 7.6% increase in log-in frequency. However, there were marked differences between races. Black participants showed a 40.5% increase and Hispanic participants showed a 57.8% increase in use with a doubling of income, compared to 3% for Caucasian participants.

Conclusions: The current study demonstrated that socioeconomic data confirms no relevant relationship between socioeconomic environment and digital health intervention usage for Caucasian users. However, a strong relationship is present for black and Hispanic users. Thus, socioeconomic environment plays a prominent role only in minority groups that represent a high-risk group for CVD. This finding identifies a need for digital health apps that are effective in these high-risk groups.

(J Med Internet Res 2018;20(4):e145) doi:10.2196/jmir.8819

KEYWORDS

race and ethnicity; socioeconomic position; computers; health services research; health disparities

Introduction

Cardiovascular disease (CVD) remains the most prominent cause of morbidity and mortality in the world [1]. A disproportionate amount of disease burden affects people from specific ethnic groups and lower socioeconomic backgrounds [2,3]. These disparities permeate all aspects of cardiovascular care, including prevention, diagnosis, treatment, and outcomes [4,5]. Prevention of CVD through risk factor reduction is an effective and valuable strategy for reducing the majority of disease burden [6,7]. In recent years, wide arrays of digital interventions have been created to aid in this effort [8]. These tools, widely described as digital health, have taken a variety of different forms including mobile apps, texting, desktop-based apps, wearables, digital workplace interventions, and virtual

reality. The perceived advantages of digital health tools include increased access, streamlined communication, lower patient costs, and an increase in efficiency and value of health care [9]. Many of these tools have yet to be reliably realized and proven through randomized controlled trials, but the current summation of the evidence points toward benefit in reducing cardiovascular risk [8] and potentially CVD outcomes [10].

For all the perceived benefits of digital health, potential limitations remain. Even with the belief that health care will be improved with these tools, there remain concerns that disadvantaged populations such as those of low socioeconomic status, minorities, and the elderly may obtain fewer benefits. Previous work has shown that people with lower socioeconomic status are less likely to proactively seek online health information, use online health care communication, or participate in user-generated online health activities [11,12]. In one large health system, online patient portal usage was found to be markedly decreased in blacks, Hispanics, patients of low socioeconomic status, and the elderly; differences that largely stem from reduced access in these groups [12]. However, patient education, digital literacy, and other factors are possible contributors as well [13]. This trend is especially concerning given the increased prevalence of cardiovascular risk factors in these populations [3,14,15]. The increased CVD risk coupled with the perceived lack of benefit from digital health tools has been termed the "digital divide." This divide raises the question of whether these tools provide similar benefit in terms of CVD risk reduction to all users, and if further investigation into groups based on both socioeconomic background and ethnic group may provide more insight.

Previous work from our group examined the usage of a digital health intervention (DHI) as part of a workplace health program (WHP) to determine association with cardiovascular risk factors [16,17]. These previous data provided evidence that DHI usage and benefit, particularly with regard to weight loss over one year, was dependent upon race. In an effort to further explore the digital divide, we utilized publicly available data from the American Community Survey (ACS) cross-referenced with the WHP in a retrospective secondary cross-sectional analysis to determine the relationship between income and DHI use through the lens of race, gender, and age. We hypothesized that DHI use would increase with median income and that the relationship would vary among racial and gender subgroups.

Methods

Employee Recruitment and Study Parameters

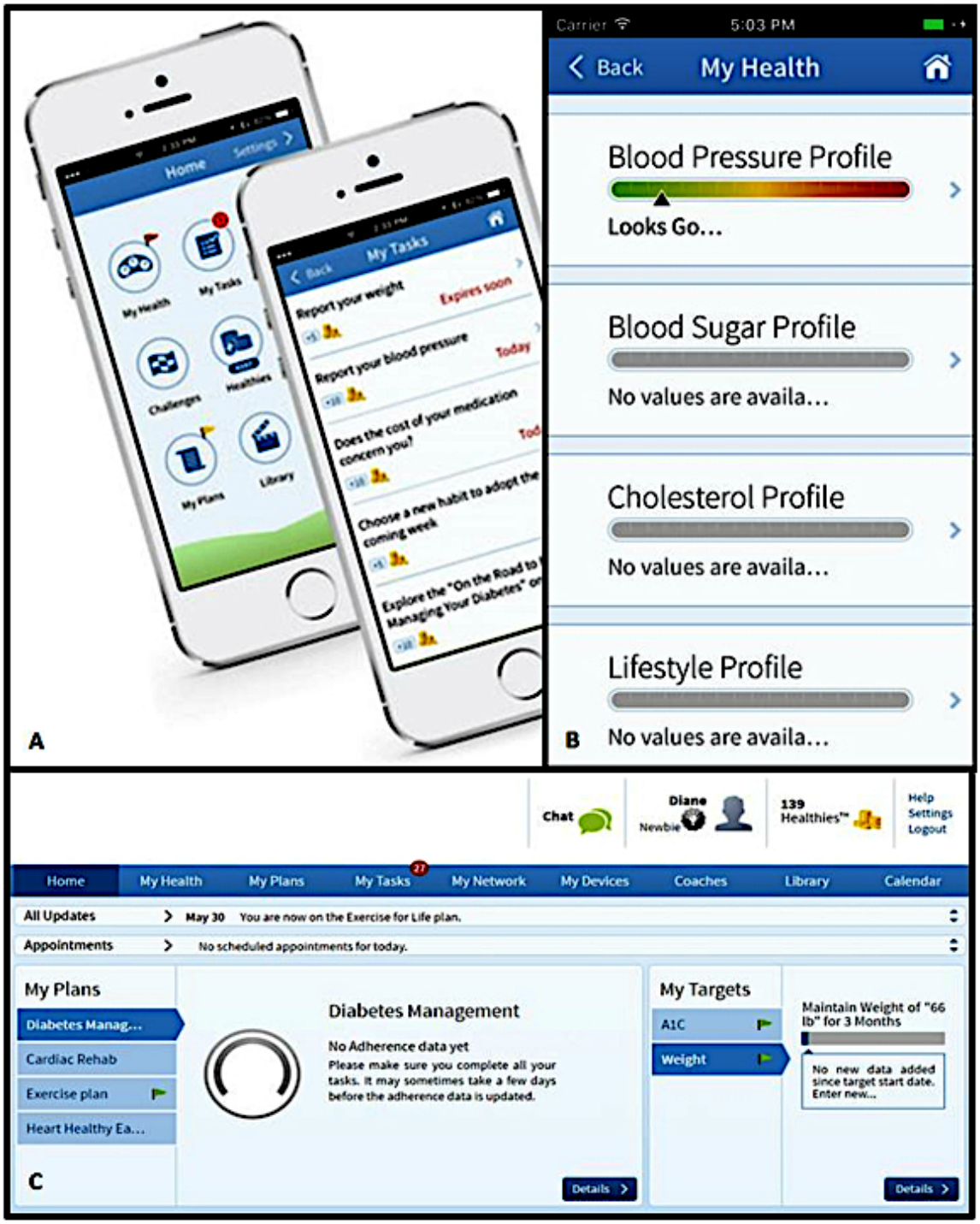
As described previously [16,17], between 2011 and 2014 CareHere, LLC (Nashville, TN) created and implemented an incentive plan for employees to improve health across 81 employers in 42 of the United States, encompassing 30,974 employees from a variety of ethnic backgrounds and in a variety of occupations in governmental, white collar, and blue-collar settings. CareHere, LLC's onsite clinic vendor managed the individual programs and tracked results both manually and with the Online CareHere Connect Personal Health Assistant designed and produced by Healarium, Inc (Dallas, TX). All employees enrolled in the employer-sponsored health insurance program were offered the opportunity to complete the biometric screening. The DHI software implementations were branded to each employer but were similar in that they covered basics of CVD prevention that were conceived and designed by CareHere. Delivery methods and interventions did not vary by employer. Employees were given the choice to use the available digital health component of the program upon the initial intake but were not consented at the time of entry, as de-identified data were to be used in the analyses. Race was included as a self-identified attribute at the initial visit. The multivariate analysis was limited to the three predominant racial groups (black, Caucasian, and Hispanic), while other racial groups were excluded due to a small sample size, with the largest identifiable group of others being Asian with approximately 200 individuals. The study and consent process were approved by the Institutional Review Board of Mayo Clinic. No extramural funding was used to support this work. The authors are solely responsible for the design and conduct of this study, all study analyses, and drafting and editing of the manuscript.

Standard laboratory blood tests including fasting lipid panels and serum glucose values were assessed at baseline and every 90 days for one year for those participating, and were paid for through their WHP and/or insurance. The patients' primary health care providers assessed blood pressure, height, weight, and administered the health behavior questionnaires at baseline and every 90 days in a standard fashion. Employees were asked to follow-up with a health care provider every 90 days for the duration of their WHP (at least one year) to review results and see if all health benchmarks set by the employer were met.

Digital Health Intervention

The DHI studied here has been described previously [16,17]; the log-in screen is visualized in Figure 1. In brief, it is a platform that is accessible online through a desktop and smartphone-based portal which provides educational materials and concrete health improvement tasks, while the app collects data on the number log-ins to the app via mobile or desktop platforms. Upon enrollment, participants upload baseline data and additional data is compiled as they progress through the program. The platform provides individualized care plans based on medical comorbidities such as obesity, hyperlipidemia, and diabetes, including health status information tasks, targets, and plans that encourage the adoption and maintenance of a healthier lifestyle for improved wellness without physician intervention. Communication occurs directly in the app on a smartphone or desktop as well as through email and short message service (SMS) text messaging. These means of communication can be tailored to user preference in the app's settings. The number of instances an individual logged into the app throughout the one year studied (log-in frequency) was recorded; this did not include email or SMS messages sent from the app to the user. There were no requirements for log-in frequency. The software directly communicates with an electronic health record that captures International Classification of Diseases codes, medical diagnoses, medical information, demographic information, and lab and vital sign values.

Figure 1. Digital health app mobile and desktop. A) Initial menu and tasks screen of the mobile app; B) My Health visual scales screen within mobile app; C) Desktop version of the digital health intervention. Displayed is the log-in screen that greeted participants following log-in.



Socioeconomic Data

The ACS is an ongoing statistical survey conducted by the United States Census Bureau [18] that has been used extensively in clinical research in a similar context [19,20]. The ACS is sent to approximately 3.5 million Americans annually and collects information regarding ancestry, educational attainment, income, language proficiency, migration, disability, employment, and housing characteristics. The ACS is the largest survey administered by the Census Bureau aside from the decennial census. Aggregate data from the ACS is publicly available through American Fact Finder [18]. A query was placed to determine 5-year median income information by zip code from 2008-2013. These zip codes were cross-referenced with the residential zip code for the study participants in order to obtain the residential median income for each participant.

Statistical Analyses

Statistical analyses were performed by an independent statistician (KB). Baseline characteristics were summarized by

frequency percent or by mean and standard deviation and were compared between ethnic groups by analysis of variance (ANOVA) tests or by chi-square tests. Log-in frequency was analyzed in two ways. First, simple and multiple linear regression was used to predict log (1+ "number of log-ins") as a function of demographic variables: age, sex, ethnicity, and income. A model with only main effects, as well as a second model including all 2-way interactions, was fit. The second method, for verification only, involved a Poisson regression model with "log-ins" as the count variable to be predicted, with a logarithmic link function. The same models with main effects only, and including 2-way interactions, were used.

Results

Baseline demographics of participants from Caucasian, black, Hispanic, and other groups are shown in Table 1. Participants were predominantly Caucasian (22,278/30,953, 71.97%) and represented 42 separate states. Actual differences between the groups in baseline characteristics were small, but statistically significant differences were seen in all categories except gender and low-density lipoprotein cholesterol measurement. The cohort was found to represent a wide range of socioeconomic backgrounds from a median income of US $11,000 to $171,000 from 2530 separate residential zip codes. The lowest median income of US $11,021 represented an area of Chattanooga, TN and the highest represented Southlake, TX with a median income of US $170,975.

In an additive model, all four demographic variables yielded significant independent effects on log-in frequency; this is displayed in Figure 2. Racial groups showed wide variation in average annual log-in frequency. With the Caucasian majority used as a reference, there was an increase in utilization of 4.68% (P<.001) in Hispanics, and a decrease of 27.17% (P<.001) in blacks. Increasing age was positively associated with DHI log-in frequency, with a 10-year age increase associated with an average of 1.54% (P=.003) increase in log-ins. However, this result differed significantly by racial group. Overall, female participants had a 32.59% higher log-in frequency compared to their male counterparts. The interaction of gender and race showed that female gender was consistently associated with increased usage among all racial groups; however, there was considerable variability. Black women showed the strongest gender association with a female to male ratio showing a 18.94% increase and Hispanic women showing a 3.04% increase (both P<.001) compared to the female to male ratio in Caucasians.

Figure 3 displays the relationship between log-in frequency and median income of associated zip code. Caucasian participants showed a very modest nonsignificant (3.03%) increase in usage with a doubling of median income. Black participants showed a much larger increase in utilization related to a doubling of median income (40.48%; P<.001). Similarly, Hispanic participants also saw a larger increase of monthly log-ins per doubling of median income of 57.8% (P<.001). Differences between each group were statistically significant. Median income by associated zip code had separate effects in male and female groups, with females showing a 14.73% (P<.001) smaller effect of a doubling of income compared to males.

Table 1. Baseline characteristics of workplace health program participants. Cardiovascular risk factors were those collected at initial enrollment visit. Continuous variables are shown with standard deviation. *P* values were derived with ANOVAs to compare means for continuous variables. Chi-square tests were used to compare proportions for categorical variables.

Parameter	Caucasian	Black	Hispanic	Other	*P* value
Participants (n)	22,278	2698	1212	4765	
Age, years (SD)	48.6 (11.3)	48.2 (11.0)	45.0 (11.0)	46.2 (13.4)	<.001
Sex, male (%)	42	44	41	43	.11
Weight, pounds (SD)	197 (50)	214 (51)	193 (57)	190.8 (50)	<.001
Waist circumference, inches (SD)	37.3 (7.0)	39.3 (7.4)	37.1 (6.5)	36.7 (7.1)	<.001
Body mass index, kg/m^2 (SD)[a]	30.6 (7.2)	33.3 (7.7)	31.6 (6.9)	29.8 (6.9)	<.001
Systolic blood pressure, mmHg (SD)	123 (14)	127 (15)	122 (14)	122 (15)	<.001
Diastolic blood pressure, mmHg (SD)	77 (9)	80 (10)	78 (10)	77.4 (10)	<.001
Triglycerides, mg/dL (SD)	140 (100)	104 (67)	155 (106)	135.8 (84)	<.001
Low-density lipoprotein cholesterol, mg/dL (SD)	111 (32)	112 (33)	111 (32)	111 (33)	.64
High-density lipoprotein cholesterol, mg/dL (SD)	52.1 (15.1)	53.6 (14.1)	49.8 (13.8)	51.8 (15.2)	<.001

[a]Windsored at body mass index: 70

Figure 2. Relative log-in frequency based on demographic variables. The effect of individual demographic variables on log-in frequency to the digital health application. All values are reported as relative changes in log-in frequency, with age compared to 10 years younger, females being compared to males, and with Hispanic and black groups being compared with Caucasians.

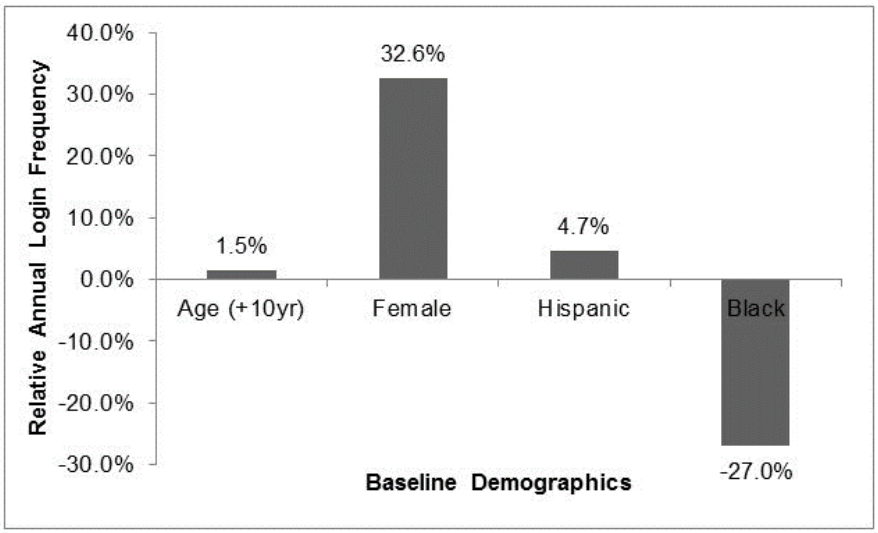
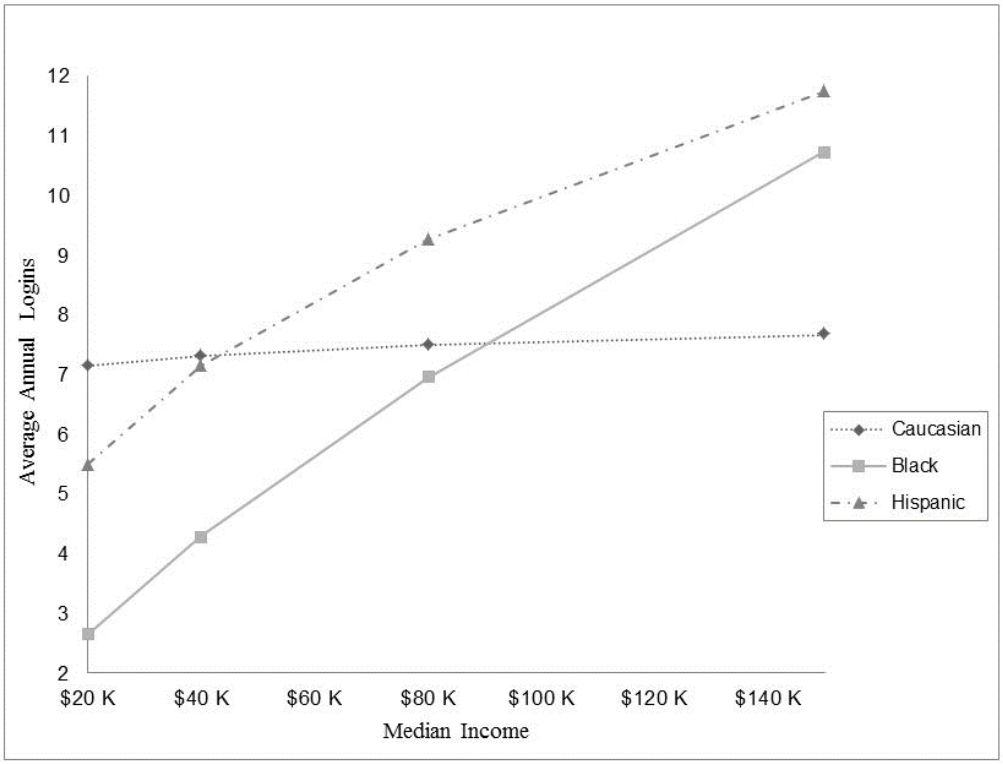


Figure 3. Log-in frequency as determined by combined effect of race and socioeconomic environment. The relationship of monthly log-in frequency by median income of associated zip code separated by racial group. The Y axis displays annual log-ins to the digital health interventionand the X axis displays median annual income of the associated zip code. Differential effects of race on income were statistically significant (*P*<.001).



Discussion

This retrospective secondary cross-sectional analysis demonstrated that socioeconomic status, as derived from the zip code of residential addresses, had a significant effect on digital health usage in this large cohort of working adults. Among Caucasians, the frequency of DHI log-ins showed a small nonsignificant increase as socioeconomic status increased. In comparison, Hispanic participants saw a>50% increase in log-in frequency with a doubling of income and black populations displayed a>40% increase. The current study may have significant implications for future design and implementation of digital health and personalized health care delivery.

Previous studies have identified decreased digital health engagement in people of low socioeconomic status [11,21,22] and in racial minorities [11,23]. Our study reaffirms these findings in individuals of both low socioeconomic status and minority racial backgrounds. However, our study is the first to identify a differential relationship between increasing

socioeconomic status and utilization according to race; it is also the first to identify minorities of high socioeconomic status to be more frequent digital health users than their Caucasian counterparts.

These results expand upon our previous DHI findings and demonstrate that baseline racial and socioeconomic characteristics were predictive of DHI log-in frequency [16,17]. Furthermore, female sex was strongly associated with a significant increase in log-in frequency. This correlation had been identified in previous digital health usage studies that addressed online health information seeking and health care communication [24]. Kontos et al examined the Health Information National Trends Survey and found that women were more likely to use health care and user-generated content domains [11]. The mechanism behind these findings may be increased engagement in online health care activities as well as increased use of social media by women as a whole [25,26]. In contrast to previously published work [11,23], but in line with our previous work in this cohort, age in our study was associated with increased DHI across ages from 23 to 88 years. While the increased usage was modest at 1.54% for every decade of age increase, it was notable that the association was positive. This result may be because our study focused on a working age population, which is a group that is more likely to be digitally literate. Individuals beyond working age may not show this increased usage due to decreased access or diminished digital literacy [26,27]. The disparity in log-in frequency by racial groups is especially notable for the highest log-in frequency being among Hispanic participants. Given that increasing age and racial minorities are known associations with worsening cardiovascular risk factors, these new findings may have implications on the use of digital health-based interventions moving forward. For example, future apps may need to be more individually tailored based on underlying user demographics.

This study adds to a growing body of evidence that socioeconomic status affects the utilization of digital health tools. It is notable that the populations in this study were working adults: not included were jobless, disabled, or retired individuals; populations that are likely to have lower incomes and more medical comorbidities. Even with this healthier and more affluent portion of the population, socioeconomic status appears to play a significant role in the utilization of this digital health tool. More work is needed to assess a possible mechanism for this finding, however digital health literacy is a possible consideration. Previous research has documented low digital health literacy in these groups [13,26,27]. The fact that our study took place as part of a WHP also raises the question of whether this relationship had an effect on DHI usage. Although all Health Insurance Portability and Accountability Act requirements were followed, participants could have perceived poorer health as a threat to job security. Future research into user satisfaction and concerns is necessary to further elucidate whether this factor generated an important effect on the results seen in the study.

To our knowledge, no previous study has directly investigated the differing effectiveness of digital health tools based on race and income simultaneously. While more mechanistic studies are needed to elucidate a cause-effect relationship, this study clearly adds to the growing body of evidence describing a digital divide within low-income minorities. The increased usage among minorities of a more affluent socioeconomic background raises several pressing questions about the future of digital health tools. Further research may be able to identify the roots of this gap and find ways to apply these findings to minority groups of lower socioeconomic backgrounds. While true that the design and usability aspects of DHIs should be focused on real-world usability studies with socioeconomic variables as secondary considerations, these data underscore the importance of inclusion in DHI design and implementation.

This study relies on data from the ACS, which provides median income by zip code for individuals in the study, that has been used in a similar fashion previously [19,20]. These values approximate the individuals' socioeconomic environment but are an imperfect correlate for incomes. The use of ACS data assumes an identical median income for all participants with the same residential zip code. A better measure of income may be through participant survey; however, this was not conducted at the time of the study. Follow-up for this study was limited to one year and the durability of these findings over a longer duration are unknown. Future research to investigate causal mechanisms for these observations is necessary to realize equitable gains from DHIs in the workplace and digital health as a whole.

In conclusion, the combination of socioeconomic background and race has a dynamic effect on digital health tool usage in a working adult population, with black and Hispanic populations showing a positive association with increased affluence. Future research may be directed at leveraging high-use affluent minorities to extrapolate strategies to bridge the digital divide in low-income minorities.

Conflicts of Interest

None declared.

References

1. Mozaffarian D, Benjamin E, Go A, Arnett D, Blaha M, Cushman M, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation 2015 Jan 27;131(4):e29-322 [FREE Full text] [doi: 10.1161/CIR.0000000000000152] [Medline: 25520374]
2. Kaplan GJ, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. Circulation 1993 Oct;88(4 Pt 1):1973-1998. [Medline: 8403348]

3. Odutayo A, Gill P, Shepherd S, Akingbade A, Hopewell S, Tennankore K, et al. Income disparities in absolute cardiovascular risk and cardiovascular risk factors in the United States, 1999-2014. JAMA Cardiol 2017 Jul 01;2(7):782-790. [doi: 10.1001/jamacardio.2017.1658] [Medline: 28593301]
4. Bonow R, Grant AA, Jacobs AK. The cardiovascular state of the union: confronting healthcare disparities. Circulation 2005 Mar 15;111(10):1205-1207 [FREE Full text] [doi: 10.1161/01.CIR.0000160705.97642.92] [Medline: 15769758]
5. Graham G. Disparities in cardiovascular disease risk in the United States. CCR 2015 May 14;11(3):238-245. [doi: 10.2174/1573403X11666141122220003] [Medline: 25418513]
6. Stampfer M, Hu F, Manson J, Rimm EW, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. N Engl J Med 2000 Jul 06;343(1):16-22. [doi: 10.1056/NEJM200007063430103] [Medline: 10882764]
7. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364(9438):937-952. [doi: 10.1016/S0140-6736(04)17018-9] [Medline: 15364185]
8. Burke L, Ma J, Azar K, Bennett G, Peterson E, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213 [FREE Full text] [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
9. Chow C, Ariyarathna N, Islam S, Thiagalingam AJ, Redfern J. mHealth in cardiovascular health care. Heart Lung Circ 2016 Aug;25(8):802-807. [doi: 10.1016/j.hlc.2016.04.009] [Medline: 27262389]
10. Widmer R, Collins N, Collins C, West C, Lerman LA, 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]
11. Kontos E, Blake K, Chou WA, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
12. Graetz I, Gordon N, Fung V, Hamity CM, Reed ME. The digital divide and patient portals: Internet access explained differences in patient portal use for secure messaging by age, race, and income. Med Care 2016 Dec;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
13. Neter EE, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
14. Ogden CL, Lamb MM, Carroll MD, Flegal KM. Obesity and socioeconomic status in adults: United States, 2005-2008. NCHS Data Brief 2010 Dec(50):1-8 [FREE Full text] [Medline: 21211165]
15. Kurian AK, Cardarelli KM. Racial and ethnic differences in cardiovascular disease risk factors: a systematic review. Ethn Dis 2007;17(1):143-152. [Medline: 17274224]
16. Widmer R, Allison T, Keane B, Dallas A, Bailey K, Lerman LA, et al. Workplace digital health is associated with improved cardiovascular risk factors in a frequency-dependent fashion: a large prospective observational cohort study. PLoS One 2016;11(4):e0152657 [FREE Full text] [doi: 10.1371/journal.pone.0152657] [Medline: 27092940]
17. Widmer R, Allison T, Keane B, Dallas A, Lerman LA, Lerman A. Using an online, personalized program reduces cardiovascular risk factor profiles in a motivated, adherent population of participants. Am Heart J 2014 Jan;167(1):93-100. [doi: 10.1016/j.ahj.2013.09.019] [Medline: 24332147]
18. United States Census Bureau. American FactFinder. 2013. American community survey 5-year estimates 2008-2013 URL: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml [accessed 2018-03-27] [WebCite Cache ID 6yElksCnr]
19. Miller A, Simon D, Roe M, Kontos M, Diercks D, Amsterdam ED, et al. Comparison of delay times from symptom onset to medical contact in blacks versus whites with acute myocardial infarction. Am J Cardiol 2017 Apr 15;119(8):1127-1134. [doi: 10.1016/j.amjcard.2016.12.021] [Medline: 28237284]
20. Clerkin K, Garan A, Wayda B, Givens R, Yuzefpolskaya M, Nakagawa S, et al. Impact of socioeconomic status on patients supported with a left ventricular assist device: an analysis of the UNOS database (United Network for Organ Sharing). Circ Heart Fail 2016 Oct;9(10):9-10 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.116.003215] [Medline: 27758810]
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 Dec 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
22. Jensen J, King A, Davis LL, Guntzviller LM. Utilization of Internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. J Aging Health 2010 Sep;22(6):804-826. [doi: 10.1177/0898264310366161] [Medline: 20495159]
23. Gordon NM, Hornbrook MC. Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]

24. Nikoloudakis I, Vandelanotte C, Rebar A, Schoeppe S, Alley S, Duncan MC, et al. Examining the correlates of online health information-seeking behavior among men compared with women. Am J Mens Health 2016 May 18:1557988316650625. [doi: 10.1177/1557988316650625] [Medline: 27193765]
25. Duggan M. Pew Research Center. 2016. Social media update 2016 URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [accessed 2018-03-27] [WebCite Cache ID 6yEmEzske]
26. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige SJ, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
27. Levy H, Janke AK, Langa KM. Health literacy and the digital divide among older Americans. J Gen Intern Med 2015 Mar;30(3):284-289 [FREE Full text] [doi: 10.1007/s11606-014-3069-5] [Medline: 25387437]
28. Gell N, Rosenberg D, Demiris G, LaCroix AK, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421 [FREE Full text] [doi: 10.1093/geront/gnt166] [Medline: 24379019]
29. Choi ND, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
30. Fox S. Pew Research Center. 2013. Health fact sheet: highlights of the Pew Internet Project's research related to health and health care URL: http://www.pewinternet.org/fact-sheet/ [accessed 2017-08-22] [WebCite Cache ID 6sufDC0yM]
31. Jargowsky P. The Century Foundation. 2015. The architecture of segregation: civil unrest, the concentration of poverty, and public policy URL: https://s3-us-west-2.amazonaws.com/production.tcf.org/app/uploads/2015/08/07182514/Jargowsky_ArchitectureofSegregation-11.pdf [accessed 2018-04-04] [WebCite Cache ID 6yRCKARAn]
32. Pew Research Center. 2017. Internet/broadband fact sheet 2017 URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2017-08-20] [WebCite Cache ID 6suf15KWx]

Abbreviations

ACS: American Community Survey
CVD: cardiovascular disease
DHI: digital health intervention
SMS: short message service
WHP: workplace health program

Edited by G Eysenbach; submitted 28.08.17; peer-reviewed by S Abdulrahman, P Jimenez, S Pezaro; comments to author 02.12.17; revised version received 23.01.18; accepted 24.02.18; published 23.04.18.

Please cite as:
Senecal C, Widmer RJ, Bailey K, Lerman LO, Lerman A
Usage of a Digital Health Workplace Intervention Based on Socioeconomic Environment and Race: Retrospective Secondary Cross-Sectional Study
J Med Internet Res 2018;20(4):e145
URL: http://www.jmir.org/2018/4/e145/
doi:10.2196/jmir.8819
PMID:29685862

Original Paper

Real-World Implementation of Video Outpatient Consultations at Macro, Meso, and Micro Levels: Mixed-Method Study

Trisha Greenhalgh[1*], FMedSci; Sara Shaw[1*], PhD; Joseph Wherton[1], PhD; Shanti Vijayaraghavan[2], MD; Joanne Morris[2], PhD; Satya Bhattacharya[2], PhD; Philippa Hanson[2], MD; Desirée Campbell-Richards[2], SRN; Seendy Ramoutar[2], SRN; Anna Collard[2], PhD; Isabel Hodkinson[3], MRCGP

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]Barts Health NHS Trust, London, United Kingdom

[3]Tower Hamlets Clinical Commissioning Group, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Trisha Greenhalgh, FMedSci
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865289293
Email: trish.greenhalgh@phc.ox.ac.uk

Abstract

Background: There is much interest in virtual consultations using video technology. Randomized controlled trials have shown video consultations to be acceptable, safe, and effective in selected conditions and circumstances. However, this model has rarely been mainstreamed and sustained in real-world settings.

Objective: The study sought to (1) define good practice and inform implementation of video outpatient consultations and (2) generate transferable knowledge about challenges to scaling up and routinizing this service model.

Methods: A multilevel, mixed-method study of Skype video consultations (micro level) was embedded in an organizational case study (meso level), taking account of national context and wider influences (macro level). The study followed the introduction of video outpatient consultations in three clinical services (diabetes, diabetes antenatal, and cancer surgery) in a National Health Service trust (covering three hospitals) in London, United Kingdom. Data sources included 36 national-level stakeholders (exploratory and semistructured interviews), longitudinal organizational ethnography (300 hours of observations; 24 staff interviews), 30 videotaped remote consultations, 17 audiotaped face-to-face consultations, and national and local documents. Qualitative data, analyzed using sociotechnical change theories, addressed staff and patient experience and organizational and system drivers. Quantitative data, analyzed via descriptive statistics, included uptake of video consultations by staff and patients and microcategorization of different kinds of talk (using the Roter interaction analysis system).

Results: When clinical, technical, and practical preconditions were met, video consultations appeared safe and were popular with some patients and staff. Compared with face-to-face consultations for similar conditions, video consultations were very slightly shorter, patients did slightly more talking, and both parties sometimes needed to make explicit things that typically remained implicit in a traditional encounter. Video consultations appeared to work better when the clinician and patient already knew and trusted each other. Some clinicians used Skype adaptively to respond to patient requests for ad hoc encounters in a way that appeared to strengthen supported self-management. The reality of establishing video outpatient services in a busy and financially stretched acute hospital setting proved more complex and time-consuming than originally anticipated. By the end of this study, between 2% and 22% of consultations were being undertaken remotely by participating clinicians. In the remainder, clinicians chose not to participate, or video consultations were considered impractical, technically unachievable, or clinically inadvisable. Technical challenges were typically minor but potentially prohibitive.

Conclusions: Video outpatient consultations appear safe, effective, and convenient for patients in situations where participating clinicians judge them clinically appropriate, but such situations are a fraction of the overall clinic workload. As with other technological innovations, some clinicians will adopt readily, whereas others will need incentives and support. There are complex

challenges to embedding video consultation services within routine practice in organizations that are hesitant to change, especially in times of austerity.

(J Med Internet Res 2018;20(4):e150) doi:10.2196/jmir.9897

KEYWORDS

remote consultations; diabetes mellitus; ethnography; interviews; organizational case studies; health systems

Introduction

Background

Outpatient services, particularly for people with long-term conditions, have changed little in recent decades. Yet, in many countries, population demographics, disease epidemiology, and care priorities have changed a great deal, with, for example, an aging population, rising rates of chronic illness and multimorbidity, and an increasing emphasis on multidisciplinary team care and supported self-management. Outpatient nonattendance rates in some patient groups, especially the disadvantaged and those with multiple and complex needs, are high and may be associated with poor disease control, increased use of emergency services, and high costs [1,2]. Patient-borne costs (in terms of time and travel) of attending outpatient appointments are high, especially for tertiary care [3-5].

There is a strong policy push in the United Kingdom [6-9] and elsewhere [10-12] to harness the potential of digital technologies to improve care models and redesign care pathways in a way that improves the accessibility and efficiency of services and maximizes the potential for patient self-management. The UK's National Information Board recently argued that a different kind of health service is needed, in which the traditional outpatient consultation will become increasingly obsolete [8]. Remote video consulting using Skype or FaceTime is one potential solution.

Published research on video outpatient consultations has been summarized in several recent narrative reviews [13-15]. Randomized controlled trials (RCTs) have shown such consultations to be acceptable, safe, and effective (and, when measured, to reduce patient-borne costs) in patients deemed clinically eligible in a range of conditions, including adult and teenage diabetes [16-18], chronic kidney disease [19], chronic obstructive pulmonary disease [20,21], mental health conditions [22,23], chronic pain [24-26], support after premature birth [27], support of patients in care homes [28], postoperative follow-up for orthopedics [29-32], plastic surgery [33], and prostate cancer [34].

However, all these trials were small; publication bias cannot be excluded, uptake rates were not always reported, and the cost of establishing and maintaining the remote service was rarely measured. Although no specific safety issues or critical events were reported in any of the above studies, exclusion or withdrawal rates in some studies were high [18,23,25].

Notwithstanding the positive findings of randomized trials, audits of actual practice suggest that video outpatient consultations, in common with other forms of telehealth, account for only a tiny fraction of encounters in any specialty [35,36]. Nonadoption, abandonment, and failure of scale-up, spread, and sustainability (the NASSS framework) are the norm when technology-supported service models are introduced in real-world settings for multiple and complex reasons [37-39]. In the words of one critic, the benefits of Skype- and FaceTime-supported outpatient services demonstrated in proof-of-concept studies and experimental trials should be weighed against contextual realities, including "the vagaries of technology, negative views among clinicians, poor uptake by providers, and legal, ethical and administrative barriers" [40].

Some authors have queried whether video consultations might be less reliable [41-43], less safe [44], and less cost-effective than traditional encounters [40]. The online environment is known to produce subtle alterations in the dynamics of human interaction, with a potential risk that clinical clues will be missed or the clinician-patient dynamic altered adversely [45]. The introduction of video outpatient consultations also brings operational and cultural challenges, including the need to develop new ways of organizing clinical and administrative work and train and support both staff and patients in technology use [37,46].

In short, there is a growing mismatch between the positive evidence base emerging from experimental trials and the variable (mostly negative) experiences of teams who try to introduce video outpatient services in the real world [35,36]. The discussion sections of randomized trial reports often suggest that further research is needed into implementation challenges—especially national policy and economic context, the practicalities and costs of organizational change, and the fine-grained detail of how video consultations unfold (in particular, how "quality" and "safety" are achieved and assured) [17,25]. No previous studies have addressed all these issues and their interdependencies rigorously and prospectively. Yet, such research is critical to enrich our understanding of video consulting and to inform and support the development and scaling up of such services. For this reason, we undertook an in-depth, real-world case study of an attempt to introduce video outpatient consultations across several clinical services (in three different hospitals) in a London-based acute trust.

Aim, Objectives, and Research Questions

Our aims were (1) to define good practice and inform its implementation in relation to video outpatient consultations via Skype and similar media and (2) to generate transferable new knowledge about challenges to scaling up and routinizing this service model in health care organizations.

Specifically, our objectives were as follows:

1. At macro level, to build relationships with key national stakeholders, identify from their perspective how to

overcome policy and legal barriers, and create a receptive context for our findings.
2. At meso level, to illuminate and explore the sociotechnical system that supports the video consultation at organizational level and identify how organizations can best support the introduction and sustainability of this service model where appropriate.
3. At micro level, to study the clinician-patient interaction in a maximum variety sample of video consultations and a comparator sample of face-to-face consultations, exploring both effective and less effective communication.

Our research questions were as follows:

1. Macro level: what is the national-level context for the introduction of video outpatient consultations in the United Kingdom, and what measures might incentivize and make such consultations easier to achieve?
2. Meso level: how can a video consultation service best be established, routinized, and sustained?
3. Micro level: what defines "quality" in a video consultation, and what are the barriers to achieving this?

Methods

Outline

We have published a detailed study protocol previously [14]; this section provides a summary and refinement of our sampling and analytic approach.

Study Design

The multilevel, mixed-method study of video outpatient consultations in three hospital departments (diabetes, antenatal diabetes, and cancer surgery: micro level), embedded in an organizational case study of the introduction and rollout of this new service model (meso level), taking account of the evolving national context (macro level), was used.

Setting

The research took place from 2015 to 2017 in Barts Health, a large, multisite acute trust in a socioeconomically deprived and multiethnic borough in inner-city London, United Kingdom. The organization was under pressure to deliver services more cost-effectively while responding to rising need and demand; outpatient clinics were crowded, and travel to and between its multiple sites was difficult, time-consuming, and (for patients on low incomes) costly.

Clinicians in the diabetes service had been working for several years to establish a remote video option for outpatient consultations to improve accessibility and address high "did not attend" rates. In an early pilot study, we found high acceptance rates by staff and patients for the video option, a significant reduction in "did not attend" rates and small efficiency savings [47,48]. We subsequently demonstrated greater engagement, improved self-management, and better glycemic control in patients with challenging social circumstances and a history of defaulting from appointments who were offered the option of video consultations (see Multimedia Appendix 1 for our 2014 Diabetes Review, Engagement and Management via Skype [DREAMS] report).

This study occurred at a time when the trust was seeking to learn from the diabetes pilot and make video outpatient consultations part of business as usual whenever clinically appropriate. We worked mainly with three services on separate sites in east London: diabetes (adult and young adult), which had been piloting the virtual consultation model; antenatal diabetes, which sought to use such consultations to support close (sometimes daily) monitoring of diabetes in pregnancy; and hepatobiliary and pancreatic cancer, a tertiary care service to which patients were sometimes referred from beyond London, requiring a round trip of several hours. By the end of the study period, other specialties at Barts Health (including neurology, rheumatology, hematology, and endocrinology) had begun to introduce video outpatient consultations.

Project Management and Governance

The study was delivered by a core working group (TG, SV, JW, JM, and SS), supported by a 6-monthly independent steering group and a patient advisory group (see below). The steering group had a lay chair and cross-sector stakeholder representation, including patients, National Health Service (NHS) stakeholders, and national-level decision makers (details in Multimedia Appendix 2). The study received ethics approval from City Road and Hampstead NHS Research Ethics Committee on December 9, 2014 (reference 14/LO/1883) and subsequent amendments.

Participants and Data Sources

We collected data over a 28-month period from 36 national-level stakeholders; 24 staff at Barts Health (9 clinicians, 5 managers, 3 technical support staff, 7 administrative support staff); and a total of 50 patients. Data sources are summarized in Table 1 and described in more detail below.

Sampling and Data Collection: Macro Level

We began with individuals charged with delivering information technology (IT) strategy in NHS England, as well as those leading on patient participation. Alongside review of policy documents (from 2000), we used snowball sampling (asking each interviewee to nominate a colleague) to build up a picture of the national context. We invited a maximum variety sample of 39 stakeholders from across government (eg, NHS England, Care Quality Commission, and NHS Improvement), professional organizations (eg, Royal College of Physicians and Medical Protection Society), patient groups (eg, National Voices), industry (eg, Microsoft), and charitable and third sector organizations (eg, Health Foundation), of whom 36 agreed to talk informally with the study team (3 were uncontactable). We then undertook audiotaped, semistructured interviews with a purposive sample of 12 of these stakeholders, ensuring variation in the different institutions, groups, and perspectives represented. Stakeholder details and interview guides are available in Multimedia Appendix 2.

Table 1. Overview of multilevel data collection and analysis in Virtual Online Consultations: Advantages and Limitations (VOCAL) study.

Data source	Type and nature of data	First-order interpretation	Higher order categories
Macro-level study of the wider context for introducing video consulting	Accounts of national-level stakeholders (36 informal and 12 formal semistructured interviews); 50 national-level documents from 2000 onwards (including policies, guidance, and national-level announcements)	Historical and policy drivers for the move to video consultations; system-level blocks	External social structures such as political, regulatory and economic context; background and context to multilevel analysis
Meso-level study of organizational change	Accounts of 24 staff involved in delivering video consultations; approximately 300 hours of observations across 3 clinics; 16 documents (eg, operating procedures and meeting minutes) and researcher field notes about people and technologies delivering video consultations; diagrams and accounts of how people, technologies, and clinical work relate and interact	Key interactions and interdependencies; key organizational routines and how these are changing over time	External social structures (such as professional standards and definitions of excellence, symbolic meaning of illness); internal social structures (what actors "know" and how they interpret the strategic terrain, such as "scripts" held by patients and staff about how they should behave and how they change over time); assumptions built into the technology about, for example, capability of users, how people interact, privacy and consent, the nature of clinical work and routines and how all these interact
Micro-level study of virtual consultations	Video-recording and screen capture (at patient end and clinician end) of 30 virtual consultations (18 diabetes, 12 cancer); field notes from before or after the consultation at patient and clinician end	What is said and done in (video and face-to-face) consultations; unfolding interaction and strategies for communication; how technology shapes and constrains (video and face-to-face) consultations; how participants felt	External social structures (such as professional standards and definitions of excellence, symbolic meaning of illness); internal social structures (what actors "know" and how they interpret the strategic terrain, such as "scripts" held by patients and staff about how they should behave and how they change over time); assumptions built into the technology about, for example, capability of users, how people interact, privacy and consent, the nature of clinical work and routines and how all these interact
Micro-level study of matched face to face consultations	Video-recording of 17 face-to-face consultations (12 diabetes, 5 cancer); field notes from before or after the consultation	What is said and done in (video and face-to-face) consultations; unfolding interaction and strategies for communication; how technology shapes and constrains (video and face-to-face) consultations; how participants felt	External social structures (such as professional standards and definitions of excellence, symbolic meaning of illness); internal social structures (what actors "know" and how they interpret the strategic terrain, such as "scripts" held by patients and staff about how they should behave and how they change over time); assumptions built into the technology about, for example, capability of users, how people interact, privacy and consent, the nature of clinical work and routines and how all these interact
Descriptive and demographic data in the video consultation service	Number of patients offered video consultation option and proportion who accept and persist with it; start and finish time; DNA rate for video and face-to-face options; unscheduled encounters (eg, urgent care) for index condition	Acceptability/popularity of the service; demographic data (eg, uptake by age or ethnicity); failed encounter rate; risk of missing serious problems; consultation length	Background and context to multilevel analysis

Sampling and Data Collection: Meso Level

The goal of sampling was to map the people, interactions, and organizational routines that support the virtual consultation with a view to building a rich "ecological" picture of the sociotechnical microsystem [49] (and its wider embedding in the organization) needed to make this service model work as business as usual. We began from participating clinics, mapped the individuals and technologies involved there, and then moved outwards to include, for example, finance and clinical informatics departments.

In total, we conducted over 300 hours of observation of consultations and the clinical and administrative work supporting them, combined with semistructured or naturalistic interviews (asking people "on the job" what they are doing and why they are doing it, as people often find it easier to talk about the detail of their job while they are actually doing it [50]) with 24 staff. We also collected 16 local documents (business plans, informal guides, emails, and minutes of meetings) and descriptive statistics from each clinic.

Sampling and Data Collection: Micro Level

We used audio, video, and screen capture to produce rich multimodal data on a total of 30 virtual consultations and made audio recordings of 17 face-to-face recordings matched for clinical condition (in which the clinician stated the encounter could have happened virtually). Details of these consultations are shown in Table 1. We sought maximum variation in age; ethnicity; and clinical, social, and personal circumstances. It was a precondition of ethical approval that clinicians were able to exercise judgment about which patients to invite to join the study.

Specific exclusion criteria were as follows: no 3G access at home, lack of familiarity (by patient or carer) with the relevant

technology, clinical inappropriateness (eg, need for direct physical examination), inability to give informed consent, and comorbidity preventing participation (eg, severe visual impairment). Clinic populations included a high proportion of limited English speakers. In the young adult diabetes clinic, bilingual health advocates were available and trained in the use of remote consulting, so limited English was not an exclusion criterion there. In the antenatal diabetes and cancer clinics, a remote interpreting service was not available, but patients comfortable with a family member interpreter were included.

Our micro-level dataset consisted of video recordings of consultations incorporating two video streams: one from the clinic and one from the remote site (typically the patient's home). We recorded consultations using a small digital camcorder with wide-angle lens and remote control (Sony Handycam DCR-SR72), mounted on a mini tripod. We used a commercially available screen capture software tool (ACA Systems) run directly from an encrypted Universal Serial Bus (USB) stick to record screen images showing on each party's computer screen as a video file. A researcher started and stopped the recordings but left the room during the consultation. When the patient used a mobile, tablet, or Mac computer (which could not run the encrypted USB device), the researcher positioned a second digital camera to capture the screen. In 12 cases, the consultation was recorded at the clinic end but not at the patient end for logistical or patient preference reasons.

Each end of the consultation resulted in 2 digital files, one screen capture and one video. We synchronized these into one file using video editing software (Sony Movie Studio)—allowing us to play the video of the computer screen exactly in parallel with a video of the patient looking at the screen and to view the patient and clinician "ends" in parallel. Face-to-face consultations were audiorecorded using a digital dictaphone. Further details of the informed consent process are given in Multimedia Appendix 2.

The micro-level dataset also included contemporaneous field notes from patients' homes (eg, material circumstances) and the clinic (eg, physical surroundings and use of paper or electronic records). We also collected demographic data (age, gender, and ethnicity) on patients using the Skype option for remote consulting and (for comparison) on the clinic population as a whole over a 12-month period at each clinic setting.

Theoretical Framework

In our original study protocol [14], we drew on a technology-enriched version of Giddens' structuration theory [14], which proposes a dynamic and reciprocal link between (1) the external social environment, (2) human interpretations and action, and (3) technologies; it considers how the relationship between these evolves over time, each shaping the others. For example, the theory explores how human action reproduces and reinforces social norms; how societal expectations (including professional norms and codes of practice) influence the "scripts" built into technologies; how technologies, through their functionality and affordances, make some actions possible but others impossible; and how laws, regulatory restrictions, policy priorities, and professional codes of conduct mean that even when an action is technically and physically possible, it is not in reality an option [51].

The health care setting is heavily institutionalized (ie, our behavior is influenced primarily by expectations of how we *should* or *must* behave in this setting rather than simply by economic or personal concerns, such as maximizing efficiency or pleasure). In such circumstances, behavior is often ritualized (ie, we know and play out the roles expected of us as doctors, nurses, patients, and so on). A key question driving our data collection and analysis was how would the technological and material aspects of the remote consultation shape, enable, and constrain the playing out of these institutionalized roles and behaviors.

As the study unfolded, we enriched and extended this initial theoretical framework with additional material on, for example, clinical aspects of the illness or condition, the kind of knowledge that is foregrounded (or backgrounded) by the technology, and commercial and regulatory considerations. The resulting theoretical framework (NASSS) has recently been published [37].

Action Research

As described in detail previously [14], our study was informed by the principles of action research [52], defined as "a mutual learning process within which people work together to discover what the issues are, why they exist, and how they might be addressed" [53]. Data collected by and with the research team were fed back formatively to inform development of the service (for instance, where appropriate, we sought to support plans for rollout of virtual consultations across the hospital). In the early stages of the study, we held two formative learning workshops to feed back our findings. As the study progressed, we were welcomed into the trust's existing governance structures (including an outpatient strategy group set up to drive the rollout of virtual consultations and the local information governance department to develop standards and guidance for such consultations) so bespoke feedback meetings became unnecessary. We also fed back emerging findings periodically to national-level decision makers (for instance, relating to national payment systems) both via bespoke meetings and also because a national policy maker with responsibility for NHS IT was on the VOCAL steering group and another worked closely with us on a related project.

Data Analysis: Macro Level

Interviews with national stakeholders were initially analyzed thematically to provide context for the statements, actions, and interpretations made by organizational actors at local level. We also used interpretive policy analysis [54,55] to identify the key "storylines" [56] shaping policy and debate around remote consultations and to surface inconsistencies and ambiguities between local and national perspectives.

Data Analysis: Meso Level

Our approach to mapping the sociotechnical health care ecosystem [49] provided detailed data about the logistical and technical barriers involved in introducing and running remote consultation services (in diabetes and cancer clinics, as well as

the wider hospital such as IT and information governance departments) and workarounds to overcome them. This included data about issues related to technology, clinic management, administrative processes, patient enrollment, and the exercise of clinical judgment. We used diagrams and narrative as synthesizing devices to draw together a visual representation and linked verbal account of these human and technical interactions and interdependencies.

We also drew on the notion of "organizational routines" [57,58] defined as "recognizable, repetitive patterns of interdependent action carried out by multiple actors" [59]. Routines are how organizational life is patterned, and hence, studying them provides important insights into how innovations such as remote consultations may (or may not) be assimilated in health care and how that assimilation changes over time. We identified the work required (at clinic, departmental, and executive levels) to "routinize" aspects of the virtual consultation service; examined the dynamics within and across different routines; and analyzed the convergence between stated (or proxy) routines, clinician and staff understandings about how to enact it (ostensive routine), and the range of ways in which it is then carried out (performative routine). This allowed us to reveal the tension between current business as usual and the new ways of working implied by a video consultation model.

Data Analysis: Micro Level

Our initial analysis of micro-level data involved repeated viewing of selected virtual consultations and discussion in our interdisciplinary team (including sociology, linguistics, human computer interaction, and medicine), alongside review of interview data with patients and clinicians. This led us to identify a number of questions relating, for instance, to the ways in which the context of the consultation (often involving patients in their home setting and clinician at the clinic) shaped communication; whether the usual format of the medical consultation (opening, history taking, examination, diagnosis, and review) might shift when conducted remotely; how talk about technology might reorient patient-clinician interaction; and how sensitive topics (such as breaking bad news) might play out differently.

On the basis of these early emerging themes, we sought a methodology to add depth and detail to our findings and identified the Roter interaction analysis system (RIAS), a widely used method for coding medical dialogue [60]. Broadly derived from social exchange theories related to interpersonal influence, problem-solving, and reciprocity [61-63], RIAS offers a validated coding system [60], allowing researchers to systematically quantify the occurrence of different types of talk that occur during medical encounters that reflects accepted patient and provider roles and obligations in a "meeting between experts" [60]. It has been used extensively to analyze face-to-face consultations but rarely in remote settings. We identified one paper (a conference proceeding) that explored the theoretical potential of RIAS in technology-mediated consultations [64], 3 small empirical studies in different clinical conditions [65-67], and a validation study of new RIAS codes for technology-related talk [68].

Roter's original taxonomy distinguishes three main categories of talk: "task-focused" (eg, application of medical expertise to diagnose and manage disease), "socioemotional" (eg, greeting, building rapport, and showing concern), and "process" (eg, inviting the patient to sit down). In this study, we also used a fourth category: "technology talk" (eg, that the picture is fuzzy), initially introduced by other researchers [66,68] and adapted and extended by our own team. Table A1 in Multimedia Appendix 2 shows the high-level clusters and more detailed categories used in RIAS with examples drawn from our data.

Following a 3-day training course delivered by the team that originally developed RIAS, we applied the RIAS coding scheme to our micro-level data, addressing five questions about the differences between virtual and face-to-face consultations for the same clinical condition:

1. Are remote consultations shorter and more "to the point" than face-to-face ones?
2. How do they differ in the different kinds of (nontechnology-related) talk that occurs?
3. What kind of technology-related talk occurs?
4. What kinds of breaches (misunderstandings, "repairs," and so on) of talk occur in virtual consultations, when do such breaches occur, to what extent do they matter, and how might they be reduced?
5. How do interruptions (in the patient's home and/or in the clinician's office) affect the flow of talk in the virtual consultation?

The assumptions for normal distribution of the data were not accepted (Shapiro-Wilk normality test was significant at $P<.05$). Therefore, Mann-Whitney U tests (nonparametric) were used to compare interactions for virtual and face-to-face consultations.

Patient Involvement Statement

The initial impetus for introducing virtual outpatient consultations in the diabetes service was from service users (many from deprived socioeconomic backgrounds and/or minority ethnic groups) who found it difficult to attend face-to-face appointments. As noted above, the VOCAL steering group had a lay chair and patient representation. In addition, we sought ongoing patient feedback on both the research process and the developing video outpatient services at Barts Health via a dedicated patient advisory group, with 12 patients (and one spouse) representing a wide range of ages and ethnic groups and facilitated by an anthropologist with close knowledge of the local community. This group met three times throughout the study, supplemented with ad hoc contact with individual members (eg, to invite input on lay summaries). In one of the meetings, the patient advisory group was shown (with the written consent of participants) two video clip recordings of virtual consultations as the basis for discussion. In addition, some preexisting lay groups (one antenatal group consisting of 9 mothers and one spouse, a peer support group for gestational diabetes, and a young people's peer support group for diabetes) and a cancer support charity were consulted to extend the range of patient and lay input.

Results

Macro-Level Findings

The external context for technological innovation in the UK public sector is currently extremely challenging. There is a strong policy push to develop UK's digital economy [69-73], digital government [74-77], and digital health [7,8,70,78-85]. We found no national policy specifically relating to the introduction of virtual consultations. As one senior decision maker said, "There are pockets of success, and there are certain vanguards exploring it, there's bits and bobs. But there's not actively a digital fund for telehealth." Rather, the policy of using technologies to support new service models appears to be folded into other programs such as the NHS Five-Year Forward View (2015-20) [2] and the General Practice Forward View (2016-21) [86]. The former includes the NHS "vanguards" to test 50 local innovative service models supported by a dedicated tranche of innovation funding [87]. An independent review in 2016 of IT in the NHS called for "new national funding to help trusts go digital and achieve maximum benefit from digitisation" [7] and led to the appointment of 12 NHS hospital trusts as "digital exemplars" with generous additional funding for introducing new technologies [88].

Notwithstanding these initiatives, constraints imposed by financial austerity—spending plans, for example, indicate a decreasing share of gross domestic product being devoted to the NHS from around 7.3% in 2016 and 2017 to 6.9% by 2022 and 2023 [89]—have meant that there has been a little slack in supporting the piloting, organizational learning, and extensive groundwork that is often needed to routinize new technologies or practices. Low growth in NHS budgets [81] combined with sustained increases in demand are taking their toll on providers [90-92].

National-level stakeholders and documents depicted technological innovation as imminently poised to deliver efficiency savings in the NHS "at scale" and "at pace" [81,93,94], thereby helping solve the growing challenge of rising demand and falling budgets and also producing "a virtuous circle of economic growth for the UK" [8]. The industry sector (whether global companies seeking long-term strategic partnerships or small- and medium-sized enterprises offering "niche" products) is depicted in these documents as the innovator and producer of "transformative technologies" [81], with the implication that the technologies, once produced, will *drive* transformation of the NHS. The potential time lag between adoption of technology and realization of productivity gains (if any) was rarely acknowledged either by interviewees or in policy documents.

Few interviewees mentioned low digital literacy and digital access. In the United Kingdom, 9% of all citizens (and a disproportionate number of the poor, sick, and elderly) have never used the Internet [95]. The UK government has an active digital inclusion strategy [70,71,96], which appears to be predicated on a "deficit" model (ie, it assumes that with support and training, everyone will be able and willing to connect digitally to public services). An alternative perspective, taken by patient and advocacy organizations, is that digital exclusion has complex underlying causes linked to the social determinants of health (eg, poverty, social exclusion, language, and literacy issues) and that "training" alone will not overcome these [97].

There was broad consensus among national-level stakeholders that the current evidence base for technology-supported new service models is weak. However, there were striking sectoral differences in what kind of evidence stakeholders believed was needed. Industry interviewees expressed confidence in the "fail early, fail often" approach of iterating software design to optimize the use of a technology in a particular setting. Interviewees from regulatory bodies and professional organizations appeared keen on qualitative evidence (they wanted to know more about, for instance, *what makes a good quality remote consultation*). In contrast, clinicians and policy makers placed high value on RCT evidence generated elsewhere but assumed to be transferable to the current setting [2].

The introduction of virtual consultations was viewed by industry informants as a particularly difficult and risky aspect of NHS IT development because it required major changes to service models and institutional embedding. Suppliers lamented slow time frames, "decision making by committee," "so much duplication," "consultation about everything," and the "need for everyone to go out and evaluate every single thing."

The prevailing emphasis by NHS England and local providers on one-off procurement contracts for particular technologies contrasts with the strategic desire of many industry stakeholders to develop mature partnerships in which industry commits to supporting an evolving service via an evolving package of technology and support. Industry informants were therefore, perhaps reluctantly, prioritizing business initiatives that rested on "off-the-shelf" technologies that could be bought, installed, and made to work with minimal ongoing work to embed, routinize, and sustain them. They saw greater potential, for example, in what one industry executive called the "wellness and wearables market"—technologies that were more or less freestanding and could often be promoted direct to consumers.

A key issue repeatedly raised by interviewees but rarely evident in published documents was how reimbursement for virtual consultations would be implemented. As one senior decision maker in NHS England told us, "we have a drug tariff that does prescriptions very well, but we don't have anything for digital."

Our analysis highlighted several strands of work being undertaken by national-level bodies (NHS Improvement and NHS England) to review payment and pricing. The development of a national Innovation and Technology Tariff for England has gone some way to addressing this by removing the need for multiple local price negotiations (instead assuring reimbursement when an approved innovation is used) but is currently limited to 6 specific medical devices or apps and does not include virtual consultation technologies [98]. The need for individual provider organizations to negotiate payment for virtual consultations with local clinical commissioning groups (even as a temporary "pass through" solution) is likely to be a significant barrier to national rollout and might also compromise the long-term sustainability of existing virtual consultation services. The proposed shift away from activity-based funding to capitated payments in the NHS could potentially overcome

this problem but is likely to be several years in development [2,99,100].

Meso-Level Findings

The action research element of the study meant that we experienced first-hand the reality of setting up and delivering the virtual consultation service in a busy public-sector hospital trust. Despite the fact that Barts Health had been an early adopter of virtual consultations and a willing partner in the research proposal, implementation proved far more complex and difficult than anticipated. One key barrier to progress was the extreme pressure on human and financial resources. Clinicians, managers, and technical staff in every department were under pressure; key posts were unfilled or frozen, and clinics were very heavily booked.

Early success of video consultations in the diabetes service had been achieved partly through workarounds (eg, installation and support of Skype on a small number of machines as a "favor"). Mainstreaming this same service as business as usual across the trust—a shift that required many processes and activities to be formalized and managed centrally—was initially strongly resisted by the IT department because it threatened to "open the floodgates" in an overstretched department and risked impacting the network bandwidth. Yet, support from the IT department (to set up machines, troubleshoot, provide on-the-job training, and configure electronic booking systems to log "video" appointments) was mission-critical, as was the development and implementation of an information governance protocol to ensure compliance with legal and regulatory standards around privacy and data protection. The latter proved extremely time-consuming even with support from the research team who worked with key local staff to develop standard operating procedures and help navigate these through the trust's approval mechanisms. The resulting document (see Multimedia Appendix 2) was reviewed and endorsed by the UK Information Governance Alliance (IGA), who used it as the starting point for developing national policy guidelines on the use of Skype and FaceTime across the NHS. We also developed a patient information leaflet with input from our patient advisory group and a guidance document for staff on setting up Skype services, which was used to support rollout across the trust (see Multimedia Appendix 3).

Although some clinicians embraced the new technology with enthusiasm, many others were unwilling to try it (claiming to be "too busy"). Those who did join the study talked positively in interviews about the high quality of most video consultations and believed (correctly as it turned out) that video consultations were generally shorter than the equivalent face-to-face encounter. However, they also commented that they had had to take on a number of new roles and practices. These included triage (judging a patient's suitability for virtual consultation), finding appropriate physical space for "private" Skype interactions outside clinic hours, troubleshooting IT, patient setup (ensuring the technology worked and supporting patients with its use), and medical documentation (adjusting to the different ways in which electronic and paper documents and other artefacts were used in consultations).

Virtual consultations appeared to work best for long-term conditions in which the clinician and patient had a preexisting relationship with a high degree of mutual trust and "common ground" when interdepartmental coordination over clinical care was not required, when the need for close physical examination could be excluded in advance, when there was a clear relative advantage to virtual consulting (eg, a history of defaulting from face-to-face appointments or clinical or practical barriers to the patient traveling), when both parties were confident and competent with technical issues, and/or when there was a pressing clinical need to have repeated contacts with the patient. In the (relatively rare) circumstances in which key clinical, technical, and practical preconditions were met, video consultations appeared to be safe and popular with both patients and staff (patients were particularly positive about convenience and not having to take time off work).

The many fine-grained differences between the clinic routine for a face-to-face consultation (see example Figure 1) and a video one (Figure 2) illustrate why the process of "embedding" generated both new work for immediate staff and also knock-on effects elsewhere in the organization. Embedding work broadly related to four key processes or subroutines: generating data or information (highlighted gray in Figures 1 and 2), enabling access to data or information (highlighted yellow), facilitating patient access through the clinic (highlighted blue), and tracking the patient through the clinic and care pathway (highlighted green). Crucially, each of these processes was supported by technical and physical artefacts, the movement of artefacts across (virtual and physical) spaces, and the input of multiple clinical and nonclinical actors.

The physical presence of the patient within the clinic setting was fundamental to existing ways of identifying, scheduling, conducting, rebooking, and monitoring patient appointments. For instance, in Figure 1, the physical presence of the patient at reception prompted "check in" and generation of the clinic outcome form: this enabled the nurse assistant to identify the patient, conduct relevant tests, and record the results on the form, which in turn enabled the remainder of the consultation to take place. Embedding virtual consultations in the work of the clinic involved significant reworking of those processes in ways that took account of the "virtual" presence of the patient—as illustrated in Figures 1 and 2 and the additional examples in figures A1 to A4 in Multimedia Appendix 2. The extent to which face-to-face consultation routines needed to be reoriented, and how this reorientation was managed by staff, varied across the three clinic settings, depending on the people, technologies, material artefacts, physical and spatial arrangements, clinical pathways, and assessment procedures already in place.

Another key challenge to the introduction of video outpatient consultations was the reconfiguration of the electronic patient record system to allow the booking of video appointments on the clinic appointment schedules. Each consultant had a "profile" on the electronic record through which appointments were booked by the administration teams. The available appointment types and time slots that could be booked were configured according to their existing clinic schedule. The reconfiguration of the booking schedule to enable the video appointment option

was managed by the trust's information and communication technology (ICT) support department, which developed protocols for the types of appointments that could be integrated into staff profiles and specifications for the payment tariff allocated to these appointments.

Figure 1. Routine for face-to-face consultation in diabetes adult or young adult clinic.

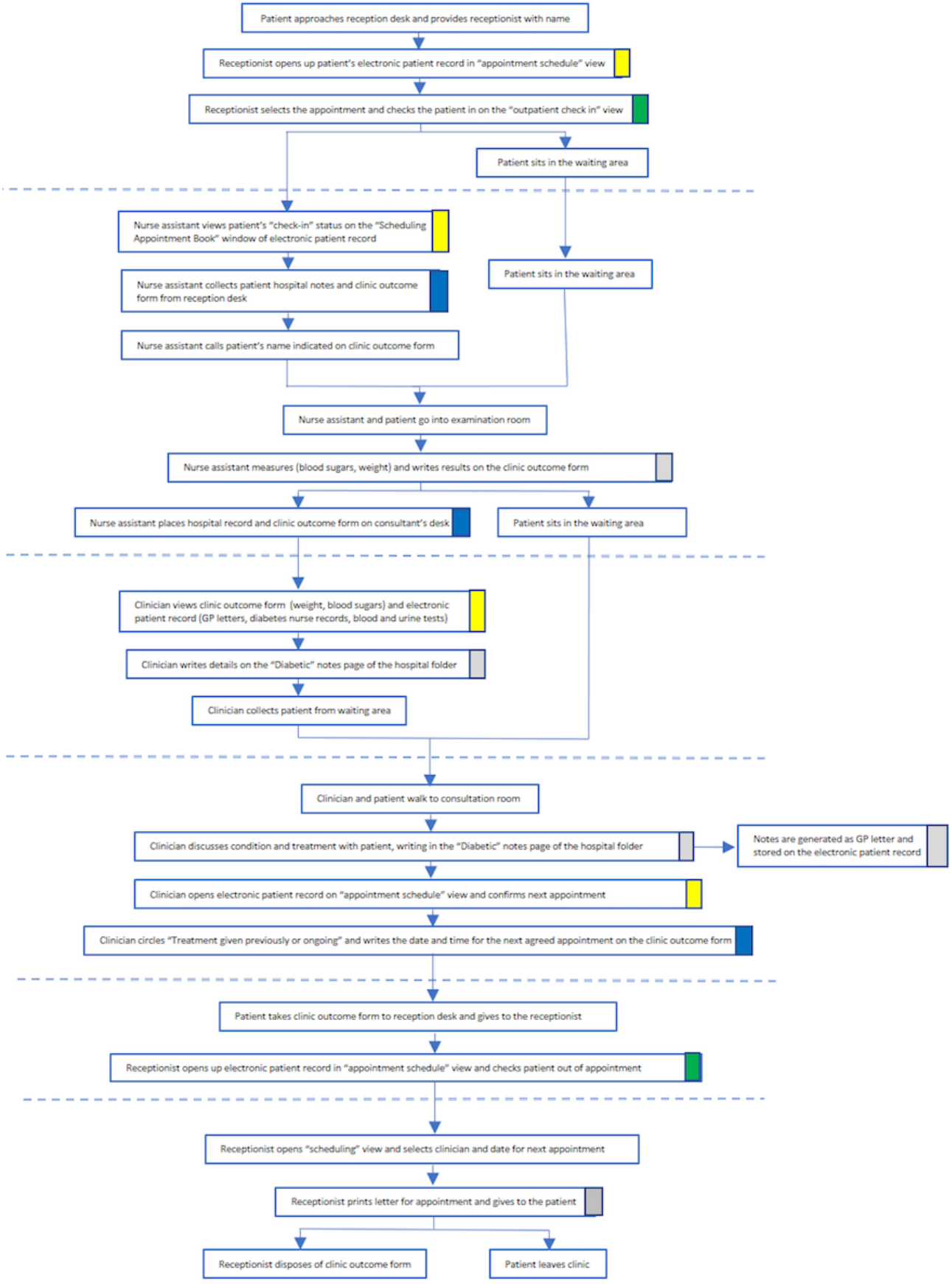
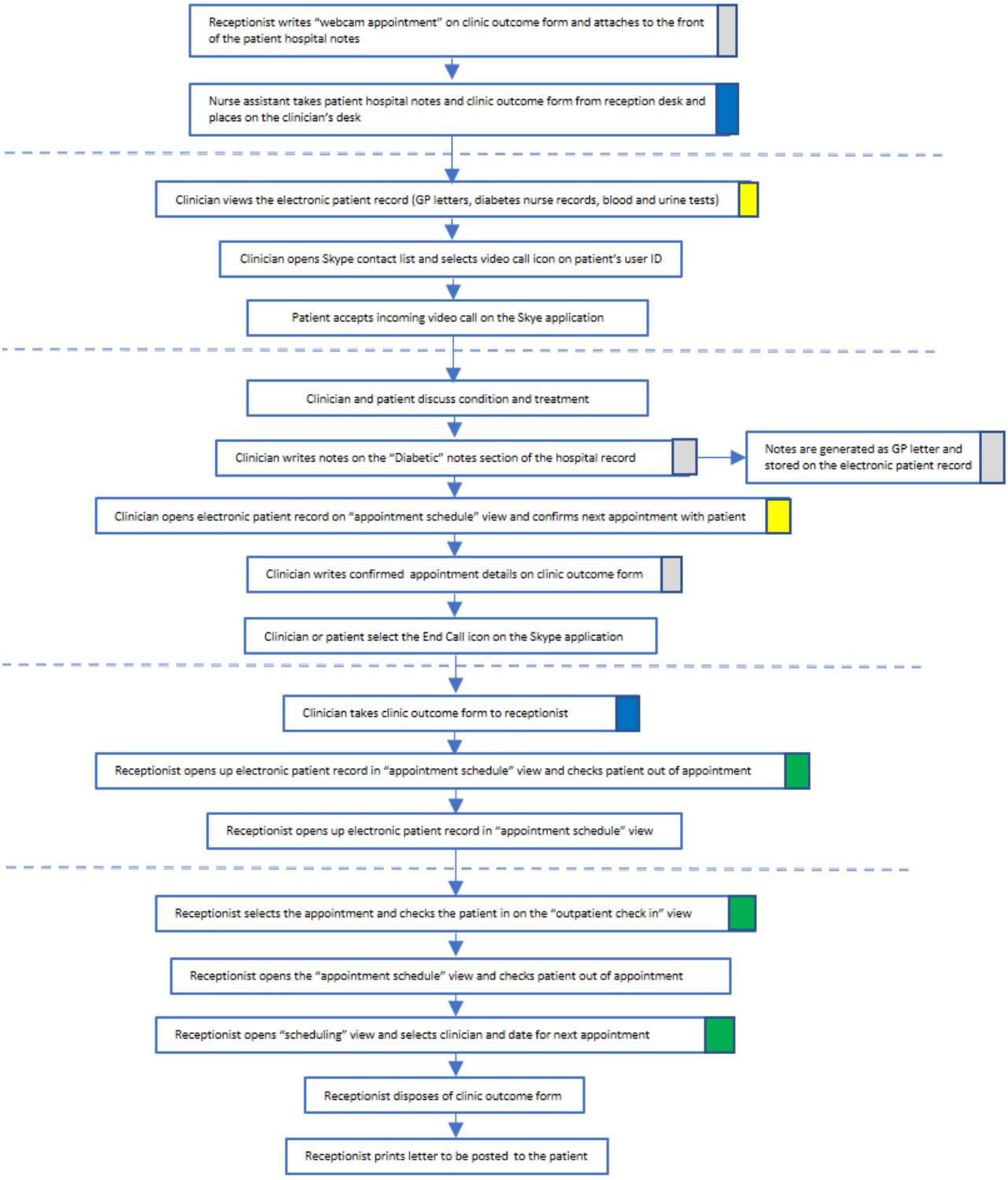


Figure 2. Routine for a virtual consultation in the diabetes adult or young adult clinic.



The introduction of the video consultation option in any clinical service involved a lengthy process of form filling, enquiries (email or phone), and discussions and agreements among senior service managers.

By the end of the study, the video option had to some extent become business as usual in the diabetes adult and young adult clinic (see Figures 1 and 2) but had evolved in a different way from the original plan. Although only 3.6% of prebooked outpatient appointments for the consultant diabetologist were formally coded on the electronic record as having occurred via Skype, the actual proportion was much higher because not all remote consultations were coded as such. This was partly because a new electronic record system (without a code for remote consultations) was introduced as part of the service part-way through the study and partly because of ad hoc consultations, discussed below.

A significant use of video consultation by Skype was for supplementary clinician-initiated and/or spontaneous

patient-initiated encounters (eg, as an ad hoc measure for keeping in close touch with patients who were undergoing a temporary period of instability or heightened need). If ad hoc encounters are added to the denominator, approximately 11% of all outpatient consultations in the diabetes clinic (and 22% of the lead diabetologist's) were undertaken by video. Clinicians liked the ease with which vulnerable and "hard to reach" patients could send a message via Skype requesting a virtual encounter, allowing prompt clinical input that clinicians believed improved patients' confidence in self-management, and which in some cases may have averted a serious complication or hospital admission. It was not possible to produce a reliable estimate of the extent to which these ad hoc encounters replaced rather than added to the clinic's conventional workload or prevented a hospital admission.

In the antenatal diabetes clinic (whose detailed routines are represented diagrammatically in figures A1 and A2 in Multimedia Appendix 2), only one clinician ever used the Skype service (for 2% of her encounters), and it was abandoned after a pilot period. In this (extremely busy) clinic, virtual consultations aligned poorly with a context involving multidisciplinary teams (patients were typically consulting multiple clinicians across departments) with a relatively short-term but high-risk condition (gestational or preexisting diabetes in pregnancy) and in the absence of integrated medical records (paper medical notes being held by the patient and so not physically present at the clinician end).

In the hepatobiliary and pancreatic cancer clinic (a tertiary referral service, figures A3 and A4 in Multimedia Appendix 2), virtual consultations were popular and generally unproblematic for follow-up after cancer surgery (a time when it was neither convenient nor clinically recommended for patients to make a long journey to the clinic). Clinicians reported that the dynamic of consultations was more relaxed (eg, being introduced to family members and pets), and some patients said they preferred to receive bad news (eg, signs of recurrence) in the comfort of their home without the ordeal of a long journey home afterwards. The proportion of all cancer follow-up consultations undertaken via video link rose from 7% to 20% during the course of the study.

There was no significant difference in demographic characteristics (age, gender, or ethnicity) between patients using the Skype option and the overall demographic of the patient population in the antenatal diabetes and cancer surgery settings. For adult and young adult consultant-led appointments in diabetes, there was no significant difference in terms of gender and ethnicity. However, there was a significant difference in age profile, with underrepresentation of patients older than 55 years taking up the Skype option (χ^2_3[N=307]=11.7, P=.01). Older patients opting for video consultations usually had a technology-savvy younger relative who offered to help.

In all virtual consultation services, there were multiple technical issues to be addressed. These were often easily resolvable, but not all patients (or staff) were sufficiently skilled or confident to undertake the necessary "troubleshooting" to achieve and maintain the video connection or resolve audio quality problems.

The VOCAL study came at a time when senior management at the trust were turning their attention to rollout of virtual consultations beyond a handful of clinics—a factor that was crucial to establishing an outpatient project strategy group (chaired by the chief medical officer) focused on supporting local rollout. That work needed to accommodate competing policy priorities locally and nationally and work with national policy makers, regulators, and industry partners (including the IGA, NHS England, Clinical Commissioning Groups, and Microsoft UK) to find workable ways forward through close dialogue and practical problem-solving.

Micro-Level Findings

Our micro-level dataset is summarized in Table 2. Key findings are described below.

The opening sequence of a video consultation was very different from that of a face-to-face one. In the former, there was invariably a "technical set-up" phase before clinical talk began. Clinicians sometimes conducted test calls so the patient could familiarize themselves with the Skype technology and/or check that the video and audio worked before the consultation. Clinicians often talked patients through minor technical problems while consulting. Examples of video and face-to-face opening sequences are shown in Textboxes 1 and 2, respectively.

Technical problems included lack of sound, poor sound quality, loss of picture, and patient failing to activate video. All were relatively minor and resolved satisfactorily, sometimes through "workarounds." For example, poor audio quality in two consultations required patient and clinician to communicate via telephone, muting the sounds while simultaneously running the video display; the problem was later found to be caused by a letter dictation device plugged into the clinician's computer. Technical issues sometimes led to conversational breaches, typically characterized by concern and/or humor, but the flow of conversation was usually quickly restored (see Textbox 3 for illustration).

Face-to-face consultations were characterized by shared physical space (eg, across the corner of a desk); patients and clinicians typically engaged together with numerous physical artefacts (paper notes, diary, smartphone, insulin pen or pump, scraps of paper, and sticky notes) as consultations unfolded. In video consultations, both parties had to compensate for lack of shared space and artefacts (eg, by holding a page up to the screen or reading aloud from a set of home blood glucose readings).

Apart from technical issues and differences linked to physical layout, video and face-to-face consultations within any specialty were strikingly similar. The content and flow of most video consultations in our dataset appeared to be high quality, though a small fraction appeared awkward and disjointed, with parties frequently misunderstanding or talking over each other and/or needing to seek clarification. Further analysis of these "awkward" consultations is ongoing, but no major safety concerns were identified.

Table 2. Overview of consultations in our micro-level dataset.

Clinic	Total recorded	Male or female	Age in years, range (median)	Ethnicity (n)
Diabetes (video)	12	5 male and 7 female	21-50 (23)	White British (5); White other (2); Black Caribbean (1); Asian Bangladeshi (1); Asian Indian (3)
Diabetes (face-to-face)	6	3 male and 3 female	21-58 (26)	White British (2); Black Caribbean (1); Asian Bangladeshi (2); Asian other (1)
Antenatal diabetes (video)	6	6 female	30-37 (34)	White British (1); Asian Bangladeshi (1); Asian other (3); Black Caribbean (1)
Antenatal diabetes (face-to-face)	6	6 female	26-36 (33)	White British (0); Asian Bangladeshi (3); Asian other (1); Asian Indian (1); Black Caribbean (1)
Cancer (video)	12	4 male and 8 female	55-85 (74)	White British (9); White other (1); Asian Indian (1); Black Caribbean (1)
Cancer (face-to-face)	5	3 male and 2 female	45-75 (69)	White British (2); Asian other (1); Black Caribbean (2)

Textbox 1. Example of opening exchange of a virtual consultation for antenatal diabetes. (xx)=length of pause in seconds.

Connection established and video display appears on both patient and clinician screen...

Patient: Ah!

Doctor: Ah hello! (0.53)

Patient: Can't hear anything. (0.5) Hold on. (1.26) Uh. (2.26) One minute, can't hear you. One minute, can't hear you.

Doctor: Are you alright, can you hear me now? (0.04)

Researcher: Can you hear us?

Doctor: I can hear you.

Patient: Is it this one? (0.11) No, no. (1.29) Volume, this one.

Doctor: Hello?

Patient: There it is. Hold on. (0.47) OK, can you hear me?

Doctor: I can hear you, can you hear me.

Patient: Ah, brilliant, yeah.

Doctor: We're on! Great! How are you?

Patient: I'm fine. Um. (0.27) OK. Um. (0.27) OK.

Doctor: Great. Alright. Now, just looking at what I wrote down at our last meeting, we'd started you on some insulin.

Patient: Yep. (0.04)

Doctor: How's that been going?

Textbox 2. Example of opening exchange of a face-to-face antenatal diabetes consultation. (xx)=length of pause in seconds.

Clinician brings patient from the waiting room to the consultation room, and reads through the patient's maternity folder...

Clinician: Right, so we met last time, we've met a few times.

Patient: Mhm.

Clinician: So, you've had a scan today.

Patient: Yes.

Clinician: How was the scan?

Patient: The scan was good!

Clinician: Was it?

Patient: Yeah.

Clinician: Brilliant, and you've seen the baby doctors, what did they say, were they happy?

Patient: Yes, they're happy, everything is OK, nice growing.

Clinician: Fantastic!

Patient: And they're preparing for my caesarean. (2.10)

Clinician: So, C section booked for the sixteenth of June!

Textbox 3. Conversational breach related to reduced video quality during cancer surgery follow-up appointment. (()) refers to unintelligible speech.

Clinician: Sorry—your your uh, the picture has frozen.

Patient: Right (( ))

Clinician: We can hear you very well, but the—

Patient: I can see you moving, (( )) that's fine.

Clinician: Yeah but (( )) your picture has frozen. But a uh—at a very happy expression so we don't mind.

Patient: Yes [Laugh]

Clinician: [Laugh] Um. (0.39) So we will see you again, or touch touch ba-+base—oh yeah you are moving again, now...

Nurse: That's better.

Patient: Right.

Clinician: We–we'll make contact again in November or December, after you've had another computed tomography (CT) scan and another set of blood tests.

Video consultations presented new possibilities for interruption. This included disruptions related to the technology (eg, loss of sound and incoming call on the mobile device being used for the consultation), as well as nontechnological interruptions in the domestic environment (eg, family members entering the room). In all cases in our dataset, flow of the consultation resumed readily after such interruptions.

Findings from our RIAS analysis are summarized in Table 3. Consultation length (defined by RIAS as frequency of utterances) was, overall, 13.34% (584/4379) shorter than comparable face-to-face encounters in all three clinical services studied, even taking account of the small amount of initial "technical talk" to establish the connection, constituting 4.43% (168/3795) of all talk during virtual consultations. However, these differences in length were not statistically significant (U=121.5, P=.43). The extent to which the clinician did more talking ("dominance") and exerted more control ("directedness") was similar in both video and face-to-face consultations in each specialty (though it varied across specialties, perhaps reflecting differences in clinical scope and/or clinicians' consulting styles). The one statistically significant difference in clinician dominance was in the diabetes antenatal setting, in which consultations were slightly more clinician-dominated during remote (median=1.2; interquartile range [IQR]=0.3) than face-to-face (median=1.7, IQR=0.5) consultations (U=3.5, P=.02). This was probably explained by patients, at the clinician's request, reading out home blood glucose readings and insulin doses in the video consultations.

A more fine-grained analysis of the different types of talk, which we will present in a separate publication (Wherton et al, in preparation) likewise confirmed only small and mostly nonstatistically significant differences in categories such as "verbal attentiveness," "making requests," "giving information," and "counseling" (see full list of categories in Multimedia Appendix 2); significant differences were again explained by the material circumstances of the consultation.

None of the other differences between video and face-to-face consultations in the above table were statistically significant.

Table 3. Median and interquartile ranges (IQR) for clinician and patient talk in virtual and face-to-face consultations, based on Roter interaction analysis system.

Clusters of talk	Consultations, median (IQR)					
	Video			Face-to-face		
	Clinician	Patient	Total	Clinician	Patient	Total
Diabetes (adult or young adult)						
Socioemotional	72 (26.5)	55 (34.0)	120 (51.5)	54 (30.0)	88 (71.8)	117 (71.0)
Task-focused	82 (38.8)	82 (49.8)	170 (53.5)	122 (24.5)	74 (34.3)	206 (4.5)
Process oriented	31 (21.5)	3 (4.0)	35 (21.5)	29 (8.3)	7 (9.5)	35 (11.8)
Technology-related	1 (6.0)	1 (2.8)	2 (8.8)			
Total number of utterances	181 (42.3)	143 (84.3)	337 (112.5)	204 (38.8)	173 (82.5)	366 (93.8)
Clinician dominance			1.3 (0.6)			1.3 (0.7)
Clinician directedness			0.7 (0.5)			0.5 (0.4)
Antenatal diabetes						
Socioemotional	35 (44.5)	38 (39.0)	74 (80.8)	43 (24.8)	36 (26.0)	83 (38.0)
Task-focused	37 (27.5)	29 (19.0)	66 (48.5)	42 (22.3)	23 (24.2)	73 (30.2)
Process oriented	6 (8.0)	2 (2.5)	8 (10.3)	11 (12.8)	1 (2.8)	14 (13.3)
Technology-related	5 (6.3)	3 (3.5)	7 (9.0)			
Total number of utterances	89 (66.0)	77 (59.5)	167 (125.5)	103 (38.8)	69 (51.3)	168 (76.3)
Clinician dominance			1.2 (0.3)			1.6 (0.5)[a]
Clinician directedness			0.8 (1.3)			0.8 (0.9)
Hepatobiliary cancer surgery						
Socioemotional	23 (46.5)	35 (34.5)	77 (35.0)	31 (39.0)	49 (38.0)	71 (72.5)
Task-focused	42 (40.5)	33 (26.5)	73 (39.0)	70 (38.5)	35 (44.5)	114 (63.0)
Process oriented	9 (14.5)	5 (6.5)	15 (20.5)	19 (16.5)	4 (5.5)	23 (21.0)
Technology-related	8 (8.5)	4 (13.5)	12 (22.0)			
Total number of utterances	108 (148.5)	84 (21.0)	192 (69.5)	137 (62.5)	72 (57.5)	217 (142.5)
Clinician dominance[b]			1.3 (1.8)			1.4 (0.5)
Clinician directedness[c]			1.0 (1.6)			0.9 (2.5)

[a]Statistically significant difference between video and face-to-face at $P<.01$ level (Mann-Whitney U test).

[b]Clinician dominance=ratio of clinician talk to patient talk (a figure above 1.0 means clinician talks more).

[c]Clinician directedness=ratio of clinician to patient control over consultation (higher number ≥ clinician has more control).

The RIAS analysis did not include the “ad hoc” consultations that occurred in the diabetes clinic (in which, eg, patients sought an immediate, and often very quick, Skype encounter with a clinician known to them to sort out a problem with insulin dosage).

Discussion

Statement of Principal Findings

This study has confirmed findings from randomized trials that when clinical, technical, and practical preconditions are met, video consultations are safe, effective, and popular with participating patients and staff. In most cases, video consultations consisted of similar types of talk, in similar proportions, to comparable face-to-face consultations, and differences between different clinical specialties were more striking than those because of the technology. By the end of this study, between 2% and 22% of all consultations were being undertaken via video link by participating clinicians. In the remainder, the video option was considered impractical, technically unachievable, or clinically inadvisable for the patient. Technical challenges were typically minor but potentially prohibitive.

Although these findings confirm that video consultations may have an important place in transforming care models, some staff members chose not to participate, and patients for whom video consultations were deemed appropriate represented a fraction of the overall clinic workload in all specialties studied.

Notwithstanding policy interest in digital solutions, the reality of establishing video outpatient services in a busy and financially stretched acute trust proved far more complex and

time-consuming than anticipated—mainly due to lack of "organizational slack" [101], disruption of traditional clinic routines, and real and perceived information governance challenges.

Although national policy makers viewed video consultations as a driver of change (supporting new, more efficient service models), industry informants viewed this option as low priority because of anticipated (and experienced) challenges of working with the NHS on projects that required complex organizational, policy, and regulatory changes.

These findings can be theorized using our recently published NASSS framework that was developed to explain why, despite significant investment and high expectations, five problems persist: digital technologies are either *not adopted* or soon *abandoned* by professionals and/or their patients and clients or else the technology-supported service succeeds as a small-scale demonstration project but fails to *scale up* locally, *spread* to other comparable settings, or be *sustained* over time. The NASSS framework analyses these problems in terms of seven interacting domains: the condition, the technology, the value proposition, individual adopters (staff and patients), the organization, the external (eg, regulatory and policy) context, and emergence over time [37]. Each domain can be simple (few components, predictable—as in making a sandwich), complicated (multiple components but still largely predictable—as in building a rocket), or complex (dynamic, composed of multiple interacting elements, and unpredictable—as in raising a child).

To the extent that VOCAL was successful in establishing a video consultation service, this was explained by the various NASSS domains: straightforward, predictable, and low-risk clinical conditions; simple and dependable technology that was fit-for-purpose; clear benefits for both the technology supplier and the patient; acceptance of the technology by staff (who considered that the technology supported and extended their professional role) and patients (who were able and willing to develop new skills and ways of engaging); alignment with existing—or emerging—organizational routines; and a strong policy push. To the extent that efforts to introduce video consultations were *unsuccessful*, this can be explained by complexity and unpredictability in the clinical condition; lack of dependability and fitness-for-purpose of the technology; lack of acceptance by staff or patients; limited organizational slack, lack of shared vision, and/or clashes with long-held and difficult-to-change routines; and tricky regulatory or policy issues (eg, national concerns about information governance and lack of a national tariff for virtual consultations).

Our study has also illustrated, through detailed multilevel analysis, the interdependence of the different domains in the NASSS framework. For example, our national-level interviews identified a reluctance among major technology vendors in the United Kingdom (not just Microsoft) to make major investments in partnerships with the NHS. This meant that, at the time of writing, the technology being used was an off-the-shelf product that had not been specially adapted for use in video consultations and that this technology was not high priority for support from the local IT department. This partly explains why significant clinician time (and an extension of the clinician role) was needed to complete such tasks as new appointment booking and management of a virtual waiting room. Our findings suggest that proactive codesign between technology suppliers, participating health care organizations, and national policy makers could potentially produce three things: video software that is more fit for purpose, organizational routines that are better aligned to support video consulting, and better incentives for major suppliers to work in a collaborative and ongoing way with health care providers.

Although our study was not designed to generate a simple or universal "checklist" for implementing video consultations, it is worth reproducing here the five "key recommendations for practice" aimed at clinicians and managers that were coproduced through action research in this study:

- Introduce the service slowly and incrementally with direct involvement of the team to ensure compatibility between the technology and existing practice
- Allow plenty of time for discussion with staff and patients about how it affects the service
- Work in collaboration with your ICT department and technical support teams to establish roles and processes to assist use of the technology
- Use with an understanding of the patients' lives and how the technology relates to the management of their health condition
- Support flexible use, allowing scope to fit the service around the needs of the patient

Strengths and Weaknesses of the Study

To our knowledge, this is the first research study in any clinical field to have taken an in-depth, mixed-methods, and multilevel approach to the study of video outpatient consultations. We succeeded in our goal of collecting rich qualitative data that exposed the "messy reality" of establishing a virtual consultation service and illuminated the pros and cons of using this medium for clinical interaction in different settings. Using action research, we were able to inform and facilitate the work of embedding the new service model and gain detailed insight into organizational complexities and how these changed over time. Working with front-line clinical and technical services, we have developed significant expertise, standard operating procedures, information governance and technical guidance documents, and protocols for setting up and running video outpatient clinics. National policy makers have been engaged from the outset, and the study has attracted interest from other hospitals. A rollout phase continues within the trust, and further work is also ongoing to extend the model to other NHS organizations across the United Kingdom.

The main limitation of the study is its focus on a single (albeit large and multisite) health care organization. Barts Health was even more financially stretched than most acute trusts in the United Kingdom; hospitals under less tight financial and staffing pressures may have found the implementation work easier. In addition, the sample size for the detailed analysis of virtual consultations was relatively small.

Comparison With Other Studies

Almost all previous research on video consultations in health care has either addressed the technical detail of the remote connection or undertaken an RCT of virtual vs face-to-face consultations [13-18,20-34]. Such studies lend support to the conclusion that in selected patients, video consultations are noninferior to face-to-face ones—but (often by their own admission) they leave unanswered the question of how to establish the service as a real-world option and/or move from a small-scale research or demonstration project to sustainable business as usual. Our finding that, in contrast with concerns raised by previous authors, the technical quality of Skype interactions via available broadband in London was almost always adequate affirms a recent study by others of 4G mobile technology [43].

Some critics will view it as a limitation that we did not emulate the experimental methodology of previous studies. This was deliberate. Our findings have confirmed that virtual consultations cannot be treated like a drug or even as a complex behavioral intervention to be tested "on" patients. Rather, they are the result of a hugely complex sociotechnical system in which "successful" virtual consulting is contingent on multiple factors at multiple levels. If we appear to have produced ambiguous findings, this is perhaps because ambiguity and tension are *inherent* to complex sociotechnical systems. To questions such as "do virtual consultations work?," "are virtual consultations safe?," and "are virtual consultations cost effective?," we suggest the answer will always be "it depends."

Meaning of the Study

In the context of a strong policy push to develop digital alternatives to the traditional consultation, delivering video outpatient services at scale is likely to be far from straightforward, as rollout in any locality will be influenced by (among other things) prevailing organizational culture, financial and human resources and priorities for allocating these, existing organizational and technical infrastructure, the nature and causes of professional resistance, information governance challenges, and the logistics of payment. Video consultations, although safe and effective for selected patients, fundamentally change the nature of outpatient care and require clinician buy-in (which may or may not be forthcoming). Industry, although not opposed to the idea of developing software to support video consultations, appears to view organizationally embedded technical solutions as relatively low priority.

The finding that efforts to implement a video consultation service met with multiple challenges in relation to workability and integration aligns with numerous previous studies of other forms of remote care (see in particular Finch and May's work on telehealth, which formed the empirical basis for May's normalization process theory [38,39]). Indeed, these difficulties may even have worsened in recent years as clinical work has become more protocolized and financial pressures more severe.

Unanswered Questions and Future Research

This study has, in some way, revealed and explored the challenges to establishing video outpatient consultations as a real-world service. Overcoming those challenges will not be easy, but further in-depth case studies in both comparable and contrasting settings are likely to enrich our understanding. As more health care organizations make the strategic decision to introduce video consultation services, research could explore the collaboration and mutual learning that occurs between them and test approaches to supporting that interorganizational interaction.

Our macro-level interviews identified a consistent finding from industry informants that the NHS is currently a uniquely difficult setting in which to attempt to introduce technologies that imply major changes in service models. Industry's preferred model—of long-term partnerships (for technologies plus service support to embed them) rather than one-off procurement contracts—should be introduced in test sites and carefully researched using longitudinal ethnography. The research agenda here is methodological as well empirical; it is founded (we believe) on the notion that technologies and services are continually evolving and mutually shaping; they cannot be fixed in time nor (therefore) be adequately tested using traditional randomized trial designs.

One of the most interesting findings of this study was that the technology provided opportunities for clinicians and patients to use the technology adaptively and differently, allowing new modes of consulting to evolve (eg, patient-initiated contacts direct to the clinician via Skype messaging, which appeared to help supported self-adjustment of insulin dosage in diabetes). Further qualitative research could pursue the consequences of such adaptive usage.

Conclusions

This study has applied a sociological lens (specifically, an empirically oriented adaptation of Giddens' structuration theory), as well as the recently-published NASSS framework to a real-world empirical study of video outpatient consultations across three contrasting clinical specialties.

We found that these consultations appear safe, effective, and convenient for patients in situations where participating clinicians judge them clinically appropriate; however, such patients are a fraction of the overall clinic workload. As with other technological innovations, some clinicians will adopt video consultations readily, whereas others will need incentives and support. There are complex challenges to embedding video consultation services within routine practice in health care organizations that are hesitant to change, especially at a time of austerity.

Acknowledgments

The study was funded by NIHR Health Services and Delivery Research programme (HS&DR-13/59/26); TG and SS's input was also part-funded by the NIHR Biomedical Research Centre, Oxford, grant BRC-1215-20008 to the Oxford University Hospitals

NHS Foundation Trust and the University of Oxford. The authors would like to thank Barts Health NHS Trust and all of the staff, patients, and carers who have given up their time to support and/or participate in the VOCAL study. Without them the research study would not have been possible. They also thank the members of their external steering group, especially Mrs Jenni Bowley who chaired the group, and the members of our Patient Advisory Group. The authors also thank Charles Gutteridge (Chief Clinical Information Officer at Barts Health) and Emma Byrne (previously at Queen Mary University of London) who helped to shape the initial proposal and the early stages of the project and Susan Larson from RIASWorks for providing training in the RIAS approach to data analysis. Several peer reviewers provided helpful comments that improved the paper.

Authors' Contributions

SV had the original idea for introducing virtual consultations at Barts Health and established the service as a pilot in 2010. TG was the chief investigator of the VOCAL study and (as such) its guarantor. She was involved in all aspects of study design, data collection and analysis, and writing; she drafted the first version of the paper. SS and JW are senior academics; they were involved in all aspects of the research. SS led on the macro element of the study and produced the full project report for NIHR. JW led on ethical applications and on the meso and micro elements of the study. At Barts Health, SB, SV, and PH are all consultants; DCR and SV are nurses; and JM is a senior project manager; all were involved in refining the study design, facilitating access to the study sites, data collection, and patient engagement. AC, a community anthropologist, led the patient advisory group. IH, a general practitioner, was the lead contact with local clinical commissioning groups. All authors contributed to dissemination, report writing, and publications, and all have seen and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Final report of Health Foundation DREAMS study (2014): Diabetes Review, Engagement and Management via Skype.

[PDF File (Adobe PDF File), 539KB - jmir_v20i4e150_app1.pdf]

Multimedia Appendix 2

Supplementary material.

[PDF File (Adobe PDF File), 368KB - jmir_v20i4e150_app2.pdf]

Multimedia Appendix 3

Skype guidance for clinicians and managers.

[PDF File (Adobe PDF File), 233KB - jmir_v20i4e150_app3.pdf]

References

1. QualityWatch. Quality at a cost: QualityWatch annual statement 2016 URL: http://www.qualitywatch.org.uk/sites/files/qualitywatch/field/field_document/QW%20annual%20statement%202016%20%28final%29%20WEB.pdf [accessed 2018-03-31] [WebCite Cache ID 6yK81OWsA]
2. NHS England. Five Year Forward View URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [accessed 2018-03-31] [WebCite Cache ID 6yK87xXgW]
3. Gerlach UA, Vrakas G, Holdaway L, O'Connor M, Macedo R, Reddy S, et al. Skype clinics after intestinal transplantation - follow-up beyond post codes. Clin Transplant 2016 Jul;30(7):760-766. [doi: 10.1111/ctr.12745] [Medline: 27140671]
4. Campbell-Richards D. Exploring diabetes non-attendance: an Inner London perspective. J Diabetes Nurs 2016;20(2):73-78.
5. de Oliviera C, Bremner KE, Ni A, Alibhai SM, Laporte A, Krahn MD. Patient time and out-of-pocket costs for long-term prostate cancer survivors in Ontario, Canada. J Cancer Surviv 2014 Mar;8(1):9-20. [doi: 10.1007/s11764-013-0305-7] [Medline: 23975612]
6. The King's Fund. A Digital NHS? An introduction to the digital agenda and plans for implementation URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/A_digital_NHS_Kings_Fund_Sep_2016.pdf [accessed 2018-03-31] [WebCite Cache ID 6yK8T4oMe]
7. UK Government. Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2018-03-31] [WebCite Cache ID 6yK8eXeFK]
8. National Information Board. 2014. Personalised Health and Care 2020: Using Data and Technology to Transform Outcomes for Patients and Citizens URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf [accessed 2018-03-31] [WebCite Cache ID 6yK8mdQV8]

9. Department of Health, UK Government. 2012. Digital strategy: Leading the culture change in health and care URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213222/final-report1.pdf [accessed 2018-03-31] [WebCite Cache ID 6yK8siiku]
10. DigitalHealth. Safe, seamless and secure: evolving health and care to meet the needs of modern Australia URL: https://www.digitalhealth.gov.au/about-the-agency/publications/australias-national-digital-health-strategy/ADHA-strategy-doc-(2ndAug).pdf [accessed 2018-03-31] [WebCite Cache ID 6yK92u7vL]
11. Canada Health Infoway. 2013. Opportunities for Action: A Pan-Canadian Digital Health Strategic Plan URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/i-infoway-i-corporate/vision/1658-opportunities-for-action-a-pan-canadian-digital-health-strategic-plan [accessed 2018-03-31] [WebCite Cache ID 6yK9E3mks]
12. European Commission. 2012. Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions: eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/com_2012_736_en.pdf [accessed 2018-03-31] [WebCite Cache ID 6yK9PbXVK]
13. Armfield NR, Bradford M, Bradford NK. The clinical use of Skype--For which patients, with which problems and in which settings? A snapshot review of the literature. Int J Med Inform 2015 Oct;84(10):737-742. [doi: 10.1016/j.ijmedinf.2015.06.006] [Medline: 26183642]
14. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016 Jan 29;6(1):e009388. [doi: 10.1136/bmjopen-2015-009388]
15. Gentles SJ, Lokker C, McKibbon KA. Health information technology to facilitate communication involving health care providers, caregivers, and pediatric patients: a scoping review. J Med Internet Res 2010;12(2):e22. [doi: 10.2196/jmir.1390]
16. Freeman KA, Duke DC, Harris MA. Behavioral health care for adolescents with poorly controlled diabetes via Skype: does working alliance remain intact? J Diabetes Sci Technol 2013 May 1;7(3):727-735 [FREE Full text] [doi: 10.1177/193229681300700318] [Medline: 23759406]
17. Hansen CR, Perrild H, Koefoed BG, Zander M. Video consultations as add-on to standard care among patients with type 2 diabetes not responding to standard regimens: a randomized controlled trial. Eur J Endocrinol 2017 Jun;176(6):727-736. [doi: 10.1530/EJE-16-0811] [Medline: 28325823]
18. Harris MA, Freeman KA, Duke DC. Seeing is believing: using Skype to improve diabetes outcomes in youth. Diabetes Care 2015;38(8):1427-1434. [doi: 10.2337/dc14-2469]
19. Katz IJ, Pirabhahar S, Williamson P, Raghunath V, Brennan F, O'Sullivan A, et al. iConnect CKD - Virtual Medical Consulting: a web-based chronic kidney disease, hypertension and diabetes integrated care program. Nephrology (Carlton) 2017 May 4:-. [doi: 10.1111/nep.13070]
20. Nield M, Hoo GW. Real-time telehealth for COPD self-management using Skype™. COPD 2012 Dec;9(6):611-619. [doi: 10.3109/15412555.2012.708067] [Medline: 22946768]
21. Mark D, Ikehara C, Matsuura C, Hara K, Li D. Validating the impact of teaching pursed-lips breathing with Skype: a pilot study. J Hosp Palliat Nurs 2013;15(8):424-432. [doi: 10.1097/NJH.0000000000000015]
22. Choi NG, Hegel MT, Marti N, Marinucci ML, Sirrianni L, Bruce ML. Telehealth problem-solving therapy for depressed low-income homebound older adults. Am J Geriatr Psychiatry 2014 Mar;22(3):263-271 [FREE Full text] [doi: 10.1097/JGP.0b013e318266b356] [Medline: 23567376]
23. Chavooshi B, Mohammadkhani P, Dolatshahi B. A randomized double-blind controlled trial comparing Davanloo intensive short-term dynamic psychotherapy as internet-delivered vs treatment as usual for medically unexplained pain: a 6-month pilot study. Psychosomatics 2016;57(3):292-300. [doi: 10.1016/j.psym.2016.01.001] [Medline: 27017450]
24. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 4;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
25. Somers TJ, Kelleher SA, Westbrook KW, Kimmick GG, Shelby RA, Abernethy AP, et al. A small randomized controlled pilot trial comparing mobile and traditional pain coping skills training protocols for cancer patients with pain. Pain Res Treat 2016;2016:2473629 [FREE Full text] [doi: 10.1155/2016/2473629] [Medline: 27891252]
26. Müller KI, Alstadhaug KB, Bekkelund SI. Acceptability, feasibility, and cost of telemedicine for nonacute headaches: a randomized study comparing video and traditional consultations. J Med Internet Res 2016 May 30;18(5):e140 [FREE Full text] [doi: 10.2196/jmir.5221] [Medline: 27241876]
27. Gund A, Sjöqvist BA, Wigert H, Hentz E, Lindecrantz K, Bry K. A randomized controlled study about the use of eHealth in the home health care of premature infants. BMC Med Inform Decis Mak 2013 Feb 9;13:22 [FREE Full text] [doi: 10.1186/1472-6947-13-22] [Medline: 23394465]
28. Orlandoni P, Jukic PN, Spazzafumo L, Venturini C, Cola C, Sparvoli D, et al. Utility of video consultation to improve the outcomes of home enteral nutrition in a population of frail older patients. Geriatr Gerontol Int 2016 Jun;16(6):762-767. [doi: 10.1111/ggi.12551] [Medline: 26171754]
29. Sharareh B, Schwarzkopf R. Effectiveness of telemedical applications in postoperative follow-up after total joint arthroplasty. J Arthroplasty 2014 May;29(5):918-22.e1. [doi: 10.1016/j.arth.2013.09.019]

30. Marsh J, Hoch JS, Bryant D, MacDonald SJ, Naudie D, McCalden R, et al. Economic evaluation of web-based compared with in-person follow-up after total joint arthroplasty. J Bone Joint Surg Am 2014 Nov 19;96(22):1910-1916. [doi: 10.2106/JBJS.M.01558] [Medline: 25410510]
31. Marsh JD, Bryant DM, MacDonald SJ, Naudie DD, McCalden RW, Howard JL, et al. Feasibility, effectiveness and costs associated with a web-based follow-up assessment following total joint arthroplasty. J Arthroplasty 2014 Sep;29(9):1723-1728. [doi: 10.1016/j.arth.2014.04.003] [Medline: 24881023]
32. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Quality of care for remote orthopaedic consultations using telemedicine: a randomised controlled trial. BMC Health Serv Res 2016;16:483 [FREE Full text] [doi: 10.1186/s12913-016-1717-7] [Medline: 27608768]
33. Westra I, Niessen FB. Implementing real-time video consultation in plastic surgery. Aesthetic Plast Surg 2015 Oct;39(5):783-790. [doi: 10.1007/s00266-015-0526-4] [Medline: 26169952]
34. Viers BR, Lightner DJ, Rivera ME, Tollefson MK, Boorjian SA, Karnes RJ, et al. Efficiency, satisfaction, and costs for remote video visits following radical prostatectomy: a randomized controlled trial. Eur Urol 2015;68(4):729-735. [doi: 10.1016/j.eururo.2015.04.002]
35. Zanaboni P, Wootton R. Adoption of routine telemedicine in Norwegian hospitals: progress over 5 years. BMC Health Serv Res 2016 Dec 20;16:496 [FREE Full text] [doi: 10.1186/s12913-016-1743-5] [Medline: 27644324]
36. Adler-Milstein J, Kvedar J, Bates DW. Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption. Health Aff (Millwood) 2014 Feb;33(2):207-215. [doi: 10.1377/hlthaff.2013.1054] [Medline: 24493762]
37. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
38. Finch T, May C, Mair F, Mort M, Gask L. Integrating service development with evaluation in telehealthcare: an ethnographic study. BMJ 2003;327:1205. [doi: 10.1136/bmj.327.7425.1205]
39. May C, Mort M, Williams T, Mair F, Gask L. Health technology assessment in its local contexts: studies of telehealthcare. Soc Sci Med 2003 Aug;57(4):697-710. [Medline: 12821017]
40. Chakrabarti S. Usefulness of telepsychiatry: a critical evaluation of videoconferencing-based approaches. World J Psychiatry 2015;5(3):286-304. [doi: 10.5498/wjp.v5.i3.286]
41. Krout RE, Baker F, Muhlberger R. Designing, piloting, and evaluating an on-line collaborative songwriting environment and protocol using Skype telecommunication technology: perceptions of music therapy student participants. Music Ther Perspect 2010;28(1):79-85. [doi: 10.1093/mtp/28.1.79]
42. Baker DC, Bufka LF. Preparing for the telehealth world: navigating legal, regulatory, reimbursement, and ethical issues in an electronic age. Prof Psychol Res Pr 2011;42(6):405-411. [doi: 10.1037/a0025037]
43. Caffery LJ, Smith AC. Investigating the quality of video consultations performed using fourth generation (4G) mobile telecommunications. J Telemed Telecare 2015 Mar 11;6:348-354. [doi: 10.1177/1357633X15577311] [Medline: 25766856]
44. van Eck CF. Web-based follow-up after total joint arthroplasty proves to be cost-effective, but is it safe? commentary on an article by Jacquelyn Marsh, PhD, et al.: "Economic evaluation of web-based compared with in-person follow-up after total joint arthroplasty". J Bone Joint Surg Am 2014;96(22):e192. [doi: 10.2106/JBJS.N.00829]
45. Arminen I, Licoppe C, Spagnolli A. Respecifying mediated interaction. Res Lang Soc Interact 2016;49(4):290-309. [doi: 10.1080/08351813.2016.1234614]
46. Kahn JM. Virtual visits--confronting the challenges of telemedicine. N Engl J Med 2015 Apr 30;372(18):1684-1685. [doi: 10.1056/NEJMp1500533] [Medline: 25923547]
47. Newham University NHS Trust. London: Health Foundation; 2012. Diabetes Appointments via Webcam in Newham: Final Report for Shine URL: http://www.health.org.uk/sites/health/files/Shine2011_DAWN_poster.pdf [accessed 2018-04-01] [WebCite Cache ID 6yMu8Cbju]
48. Vijayaraghavan S, O'Shea T, Campbell-Richards D, Sudra R, Morris J, Byrne E, et al. DAWN: Diabetes Appointments via Webcam in Newham. Br J Diabetes Vasc Dis 2015;15(3):123-126. [doi: 10.15277/bjdvd.2015.026]
49. Brown MJ, Shaw NT, Mador RL. Mapping the sociotechnical healthcare ecosystem: expanding the horizons of sociotechnical inquiry. AMIA Annu Symp Proc 2008:1233-1235. [Medline: 18998775]
50. Barley SR, Kunda G. Bringing work back in. Organ Sci 2001;12(1):76-95. [doi: 10.1287/orsc.12.1.76.10122]
51. Greenhalgh T, Stones R. Theorising big IT programmes in healthcare: strong structuration theory meets actor-network theory. Soc Sci Med 2010 May;70(9):1285-1294. [doi: 10.1016/j.socscimed.2009.12.034] [Medline: 20185218]
52. Waterman H, Tillen D, Dickson R, de Koning K. Action research: a systematic review and guidance for assessment. Health Technol Assess 2001;5(23):iii-157 [FREE Full text] [Medline: 11785749]
53. Bate P. Synthesizing research and practice: using the action research approach in health care settings. Soc Policy Adm 2000;34(4):478-493. [doi: 10.1111/1467-9515.00205]
54. Wagenaar H. Meaning in Action: Interpretation and Dialogue in Policy Analysis. New York: M. E. Sharpe; 2011.
55. Yanow D. How Does a Policy Mean? Interpreting Policy and Organisational Actions. Washington, DC: Georgetown University Press; 1996.

56. Hajer MA. Doing discourse analysis: coalitions, practices, meaning. In: Van der Brink M, Metze T, editors. Words Matter in Policy and Planning: Discourse Theory and Method in the Social Sciences. Utrecht: Netherlands Graduate School of Urban and Regional Research; 2006:65-74.
57. Feldman MS. A performative perspective on stability and change in organizational routines. Ind Corp Change 2003;12(4):727-752. [doi: 10.1093/icc/12.4.727]
58. Feldman MS, Pentland BT. Reconceptualizing organizational routines as a source of flexibility and change. Adm Sci Q 2003;48(1):94-118. [doi: 10.2307/3556620]
59. Feldman MS, Pentland BT, D'Adderio L, Lazaric N. Beyond routines as things: introduction to the special issue on routine dynamics. Organ Sci 2016;27(3):505-513. [doi: 10.1287/orsc.2016.1070]
60. Roter D, Larson S. The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions. Patient Educ Couns 2002 Apr;46(4):243-251. [Medline: 11932123]
61. Roter D, Hall JA. Doctors Talking With Patients/Patients Talking With Doctors: Improving Communication in Medical Visits. Westport, CT: Auburn House; 1992.
62. Bales RF. Interaction Process Analysis. Cambridge: Addison-Wesley; 1950.
63. Emerson R. Social exchange theory. Annu Rev Sociol 1976;2:335-362 [FREE Full text] [doi: 10.1146/annurev.so.02.080176.002003]
64. Cloud-Buckner J, Gallimore JJ, Scott R. Theoretical discussion of CAMOCS and RIAS communication models to support healthcare information technology development. Proc Hum Factors Ergon Soc Annu Meet 2010;54(5):517-521. [doi: 10.1177/154193121005400505]
65. Wakefield BJ, Bylund CL, Holman JE, Ray A, Scherubel M, Kienzle MG, et al. Nurse and patient communication profiles in a home-based telehealth intervention for heart failure management. Patient Educ Couns 2008 May;71(2):285-292. [doi: 10.1016/j.pec.2008.01.006] [Medline: 18337049]
66. Nelson E, Spaulding R. Adapting the Roter interaction analysis system for telemedicine: lessons from four specialty clinics. J Telemed Telecare 2005;11 Suppl 1:105-107. [doi: 10.1258/1357633054462010] [Medline: 16036014]
67. Agha Z, Roter DL, Schapira RM. An evaluation of patient-physician communication style during telemedicine consultations. J Med Internet Res 2009 Sep 30;11(3):e36 [FREE Full text] [doi: 10.2196/jmir.1193] [Medline: 19793720]
68. Miller EA. The continuing need to investigate the nature and content of teleconsultation communication using interaction analysis techniques. J Telemed Telecare 2011;17(2):55-64. [doi: 10.1258/jtt.2010.100203] [Medline: 21097559]
69. UK Government. Information Economy Strategy URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206944/13-901-information-economy-strategy.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKGPoTgr]
70. UK Government. 2014. Government Digital Inclusion Strategy URL: https://www.gov.uk/government/publications/government-digital-inclusion-strategy [accessed 2018-03-31] [WebCite Cache ID 6yKGZPtf4]
71. UK Government. 2014. UK Digital Inclusion Charter URL: https://www.gov.uk/government/publications/government-digital-inclusion-strategy/uk-digital-inclusion-charter [accessed 2018-03-31] [WebCite Cache ID 6yKGg6whV]
72. UK Government. 2016. Digital Economy Bill Overview URL: https://www.gov.uk/government/publications/digital-economy-bill-overview [accessed 2018-03-31] [WebCite Cache ID 6yKGy2xn6]
73. UK Government. 2016. Government's first Director General for Digital and Media to help turn Britain into world-leading digital economy URL: https://www.gov.uk/government/news/governments-first-director-general-for-digital-and-media-to-help-turn-britain-into-world-leading-digital-economy [accessed 2018-03-31] [WebCite Cache ID 6yKH6IPQW]
74. Fox M. London: Cabinet Office, Efficiency and Reform Group, and Government Digital Service; 2010. Directgov 2010 and beyond: revolution not evolution URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60993/Martha_20Lane_20Fox_s_20letter_20to_20Francis_20Maude_2014th_20Oct_202010.pdf [accessed 2018-04-01] [WebCite Cache ID 6yMvMo3QO]
75. UK Government. 2012. Digital Landscape Research URL: https://www.gov.uk/government/publications/digital-landscape-research [accessed 2018-03-31] [WebCite Cache ID 6yKHHAjBL]
76. UK Government. 2011. Government Digital Service opens for business URL: https://www.gov.uk/government/news/government-digital-service-opens-for-business [accessed 2018-03-31] [WebCite Cache ID 6yKHMTncC]
77. UK Government. Government Transformation Strategy 2017 to 2020 URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/590199/Government_Transformation_Strategy.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKHTTNR9]
78. House of Commons Health Committee. London: The Stationery Office; 2005. The Use of New Medical Technologies within the NHS URL: https://publications.parliament.uk/pa/cm200405/cmselect/cmhealth/398/398i.pdf [accessed 2018-04-01] [WebCite Cache ID 6yMvS7EGZ]
79. UK Government. 2008. High Quality Care for All URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKHgurG2]
80. UK Government. 2012. Civil Service Reform Plan URL: https://www.gov.uk/government/publications/civil-service-reform-plan [accessed 2018-03-31] [WebCite Cache ID 6yKHmK4RX]

81. UK Government. 2016. Accelerated Access Review: Final Report. Review of innovative medicines and medical technologies URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/565072/AAR_final.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKHug1Cv]
82. UK Government. 2015. Digital Health in the UK: An industry study for the Office of Life Sciences URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/461479/BIS-15-544-digital-health-in-the-uk-an-industry-study-for-the-Office-of-Life-Sciences.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKI6Hvv4]
83. UK Government. 2016. Office for Life Sciences: a How To guide. A guide to navigating the innovation pathway in England URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/525787/AAR_how_to_guides.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKIDYZdB]
84. NHS England. London: NHS England; 2016. NHS England announces first medtech devices and apps to join fast track payment scheme for innovation URL: https://www.england.nhs.uk/2016/11/innov-tech-tariff/ [accessed 2018-04-01] [WebCite Cache ID 6yMvZwKNF]
85. NHS England. Securing Excellence in GP IT Services: 2016-18 Operating Model URL: https://www.england.nhs.uk/digitaltechnology/wp-content/uploads/sites/31/2015/04/gp-it-operating-model-16-18.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKIREyhB]
86. NHS England. 2016. General Practice Forward View URL: https://www.england.nhs.uk/wp-content/uploads/2016/04/gpfv.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKIdRwTb]
87. NHS England. 2016. New Care Models: Vanguards - developing a blueprint for the future of NHS and care services URL: https://www.england.nhs.uk/wp-content/uploads/2015/11/new_care_models.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKJOEP1h]
88. NHS England. 2016. 12 NHS Hospital trusts to trail blaze NHS's Digital Revolution URL: https://www.england.nhs.uk/2016/09/digital-revolution/ [accessed 2018-03-31] [WebCite Cache ID 6yKJUGpc8]
89. Nuffield Trust. 2017. NHS funding choices and the 2017 General Election URL: https://www.nuffieldtrust.org.uk/research/nhs-funding-choices-and-the-2017-general-election [accessed 2018-03-31] [WebCite Cache ID 6yKJjLySi]
90. The King's Fund. 2017. Sicker patients the main reason for A&E winter pressures URL: https://qmr.kingsfund.org.uk/2017/22/ [accessed 2018-03-31] [WebCite Cache ID 6yKK4Wgr1]
91. The King's Fund. 2017. Half of NHS areas planning to cancel or delay spending because of financial pressures URL: https://www.kingsfund.org.uk/press/press-releases/half-nhs-areas-planning-cancel-or-delay-spending-because-financial-pressures [accessed 2018-03-31] [WebCite Cache ID 6yKKIHflb]
92. Appleby J. 2015. NHS Spending: squeezed as never before URL: https://www.kingsfund.org.uk/blog/2015/10/nhs-spending-squeezed-never [accessed 2018-03-31] [WebCite Cache ID 6yKKWW4XD]
93. NHS England. Innovation into Action: Supporting delivery of the NHS Five Year Forward View URL: https://www.england.nhs.uk/wp-content/uploads/2015/10/nhs-inovation-into-action.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKKg9C5E]
94. National Archives. 2011. Innovation, health and wealth: Accelerating adoption and diffusion in the NHS URL: http://webarchive.nationalarchives.gov.uk/20130107070708/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_134597.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKKnyHOi]
95. ONS. Internet users in the UK: 2017 URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2017 [accessed 2018-03-20] [WebCite Cache ID 6y3XhVJIN]
96. UK Government. Digital inclusion URL: https://www.gov.uk/performance/digital-inclusion [accessed 2018-03-20] [WebCite Cache ID 6y3XocjPl]
97. Tinder Foundation. 2016. Health & Digital: Reducing Inequalities, Improving Society. An evaluation of the Widening Digital Participation programme URL: http://nhs.tinderfoundation.org/wp-content/uploads/2016/07/Improving_Digital_Health_Skills_Report_2016.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKL9ghtG]
98. NHS England. The NHS England Innovation and Technology Tariff 2017 to 2019 Technical notes URL: https://www.england.nhs.uk/wp-content/uploads/2017/05/innovation-tech-tariff-technical-notes.pdf [accessed 2018-03-20] [WebCite Cache ID 6y3Xvb6Zd]
99. UK Government. 2014/15 National Tariff Payment System URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/300547/2014-15_National_Tariff_Payment_System_-Revised_26_Feb_14.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKLJvxvP]
100. Charlesworth A, Hawkins L, Marshall L. 2014. NHS payment reform: lessons from the past and directions for the future URL: https://www.nuffieldtrust.org.uk/files/2017-01/2014-payment-reform-policy-response-web-final.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKLPIlx7]
101. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]

Abbreviations

ICT: information and communication technology
IGA: Information Governance Alliance
IT: information technology
IQR: interquartile range
NASSS: nonadoption, abandonment, and failure of scale-up, spread, and sustainability
NHS: National Health Service
RCT: randomized controlled trial
RIAS: Roter interaction analysis system
USB: Universal Serial Bus

Edited by G Eysenbach; submitted 20.01.18; peer-reviewed by C May, S Hinder, T Schopf, H Yu; comments to author 08.02.18; revised version received 19.02.18; accepted 20.02.18; published 17.04.18.

Please cite as:
Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, Campbell-Richards D, Ramoutar S, Collard A, Hodkinson I
Real-World Implementation of Video Outpatient Consultations at Macro, Meso, and Micro Levels: Mixed-Method Study
J Med Internet Res 2018;20(4):e150
URL: http://www.jmir.org/2018/4/e150/
doi:10.2196/jmir.9897
PMID:29625956

Original Paper

Risk Assessment for Parents Who Suspect Their Child Has Autism Spectrum Disorder: Machine Learning Approach

Ayelet Ben-Sasson[1], ScD; Diana L Robins[2], PhD; Elad Yom-Tov[3], PhD

[1]Department of Occupational Therapy, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel

[2]AJ Drexel Autism Institute, Drexel University, Philadelphia, PA, United States

[3]Microsoft Research, Herzelia, Israel

Corresponding Author:
Ayelet Ben-Sasson, ScD
Department of Occupational Therapy, Faculty of Social Welfare and Health Sciences
University of Haifa
Aba Khoushy Ave 199
Haifa, 3498838
Israel
Phone: 972 4828388
Email: asasson@univ.haifa.ac.il

Abstract

Background: Parents are likely to seek Web-based communities to verify their suspicions of autism spectrum disorder markers in their child. Automated tools support human decisions in many domains and could therefore potentially support concerned parents.

Objective: The objective of this study was to test the feasibility of assessing autism spectrum disorder risk in parental concerns from Web-based sources, using automated text analysis tools and minimal standard questioning.

Methods: Participants were 115 parents with concerns regarding their child's social-communication development. Children were 16- to 30-months old, and 57.4% (66/115) had a family history of autism spectrum disorder. Parents reported their concerns online, and completed an autism spectrum disorder-specific screener, the Modified Checklist for Autism in Toddlers-Revised, with Follow-up (M-CHAT-R/F), and a broad developmental screener, the Ages and Stages Questionnaire (ASQ). An algorithm predicted autism spectrum disorder risk using a combination of the parent's text and a single screening question, selected by the algorithm to enhance prediction accuracy.

Results: Screening measures identified 58% (67/115) to 88% (101/115) of children at risk for autism spectrum disorder. Children with a family history of autism spectrum disorder were 3 times more likely to show autism spectrum disorder risk on screening measures. The prediction of a child's risk on the ASQ or M-CHAT-R was significantly more accurate when predicted from text combined with an M-CHAT-R question selected (automatically) than from the text alone. The frequently automatically selected M-CHAT-R questions that predicted risk were: following a point, make-believe play, and concern about deafness.

Conclusions: The internet can be harnessed to prescreen for autism spectrum disorder using parental concerns by administering a few standardized screening questions to augment this process.

(J Med Internet Res 2018;20(4):e134) doi:10.2196/jmir.9496

KEYWORDS

autistic disorder; early diagnosis; screening; parents; child; expression of concern; technology; machine learning

Introduction

Challenges of Early Autism Spectrum Disorder Screening

Many parents of children with autism spectrum disorder (ASD) report their concerns when their child is 12- to 19-month old [1,2], yet the average age of ASD diagnosis is above 3 years [1]. Evidence shows that 5 to 6 months, on average, pass from when the parent of a child with ASD becomes concerned until they approach a professional. Moreover, over 32 months pass until a diagnosis is made [3]. This delay wastes a critical time period, during which targeted intensive intervention is most efficient [4]; the delay may also increase parent stress.

Prompting parents to approach a health care provider may lead to earlier intervention.

Early identification of ASD in young children is often delayed due to a mix of child and health care factors. At the child level, the varying patterns of onset and diversity in the early markers presented [5,6] can impede early identification. At the health care level, the lack of knowledge and expertise in ASD of many primary care providers and the lack of infrastructure for handling the increase in referrals further hinder early screening [7,8]. This study examined the utility of automated strategies for verifying initial parental concern in Web-based sources. In the long run, such a system can be programmed to prompt parents to seek the advice of a professional, when their concerns indicate risk.

Parental Concerns Predict Autism Spectrum Disorder Status

Parents of children with ASD are often concerned about their child's development long before they seek professionals and receive a diagnosis [1]. Research has indicated that the type and number of early parental concerns predict ASD [2,3,6] particularly after 12 months of age [6,9]. Early concerns specific to ASD features were more prevalent among parents who already had a child with ASD [6,9] and arose earlier than in families without a previously diagnosed child with ASD [10]. The type of early concerns differentiated children later diagnosed with ASD from those with other developmental disorders [2]. The most common type of concerns are related to speech and communication [2,10]. In another study, the type and number of parental concerns correlated with Modified Checklist for Autism in Toddlers (M-CHAT) scores [11]. At the same time, for about 30% of preschool children, concerns regarding ASD did not necessarily predict an ASD diagnosis [12]; it is therefore worthwhile investigating whether specific follow-up questions would facilitate ASD prediction.

The literature reviewed points to the importance of eliciting, attending, and trusting early parental concerns in the early identification of ASD. The clinical validity of early parental concerns has been recognized by the American Academy of Pediatrics (AAP) [13] in its guidelines for the early detection of ASD in primary care. The guidelines list parental concerns related to ASD as an indicator of higher risk, and state that they should trigger ASD-specific screening at any age.

The Internet as an Opportunity for Prescreening

Although many parents of pediatric patients use the internet to research medical information and relieve anxiety, they are likely to find incorrect/inappropriate advice [14], and only 21% share this information with their health care provider [15]. As opposed to searching, parents are turning to Web communities more and more to ask about their child's development or health [15-17]. Parents' high involvement and frequent access to informational exchanges on the internet offer a unique public health opportunity for promoting early screening. Furthermore, spontaneous open-ended parental concerns regarding ASD enable researchers to document an unbiased view of markers outside the scope of the standardized screening tools [1]. Advances in natural language processing and machine learning (ML) technologies offer tools for predicting ASD risk by classifying text from the growing collection of parental concerns that have been documented in Web-based sources. The long-term goal of our work is the design of an automated support system evaluating the degree of risk based on parents' concerns related to ASD markers. Such a system could prompt parents to act early or alternatively relieve their anxiety.

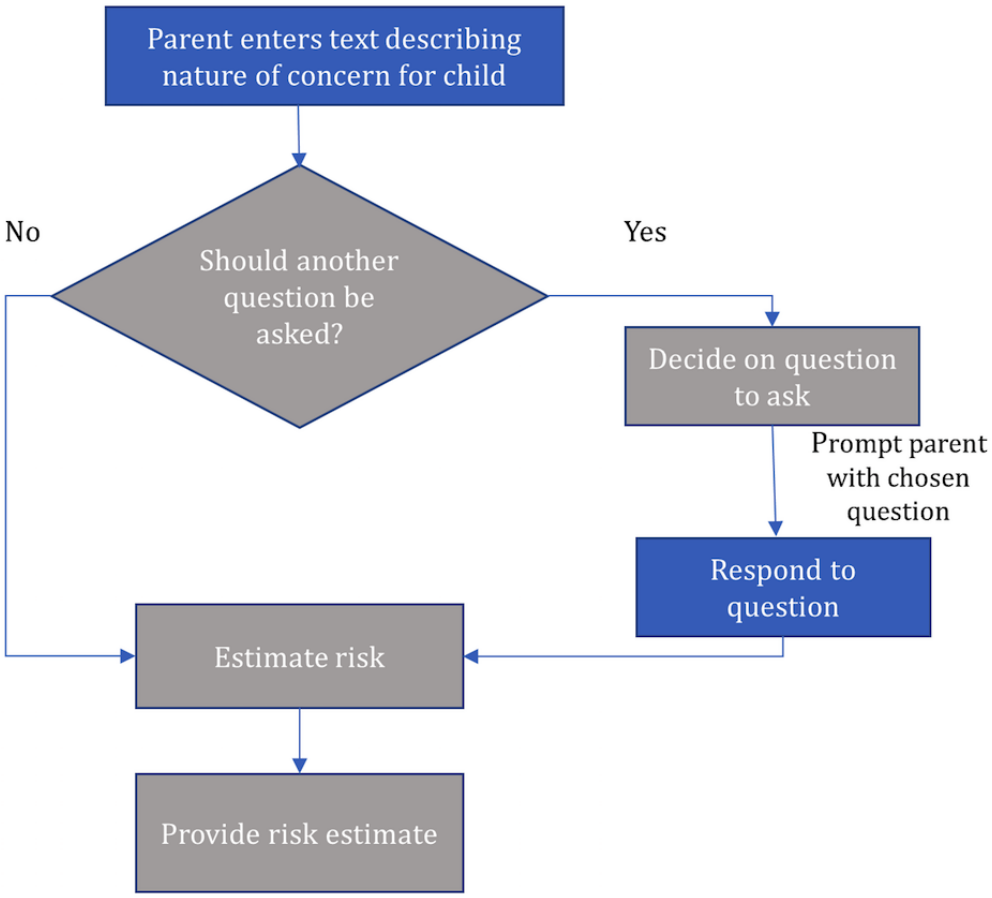
Evidence from other health conditions supports the effect of automated prompting on patients' health related actions such as seeking formal screening or a provider. A review of studies examining the effect of text messaging interventions on screening for cancer showed 4% to 63% higher screening rates for those who received text messages versus controls [18]. In the mental health domain, 38% of participants using a depression-screening app reported that they consulted with their health care professional about the app screening result [19]. A scoping review of studies of the impact of health information technologies for pediatric care showed, in diabetes type I studies and asthma studies, benefits for behavior change and patients' and caregivers' use of resources [20]. These diverse studies all point to the clinical utility of using automated systems to prompt health-related actions, in our case seeking formal ASD screening.

Machine Learning Prediction From Web-Based Querying

This study builds upon research of Ben-Sasson and Yom-Tov [21], which analyzed existing queries on Yahoo Answers where parents expressed concerns regarding ASD markers in their child. The study showed that (1) parents associate a rich array of behaviors with ASD; (2) to accurately predict ASD risk status, signs mentioned in the text must be classified into symptom categories; and (3) the accuracy of automated prediction differs based on the combination of categories reported. Using ML tools, which automatically learn to distinguish between labeled examples, enabled the researchers to classify children's risk based on expert categorization of texts into specific concern categories, though with moderate success. Area under the receiver operating curve (AUC) using the text combined with the classification of signs was 0.84 compared with 0.54 for text alone. Understanding that parental concerns alone are not sufficient for automated prediction using ML, we decided to test a prediction model that combines text analysis with structured ASD screening questions.

There is evidence that ML can use free-form text to elicit information about psychopathology markers; for example, ML has been used in the past to identify signs of postpartum depression in new mothers' social media posts. They found an average AUC of 0.82 for predicting behavioral change of mothers between prenatal and postpartum periods [22]. There is a need for an ML-based tool that, when receiving a combination of symptoms, can assess the minimum amount of specific information needed to provide an accurate risk prediction.

Figure 1. Flowchart for the proposed method of screening of parental concerns.



One of the challenges ML tools face when predicting risk from descriptions of parental concerns is their variability and individual nature, which is difficult to capture from the text of these descriptions alone. In a recent study [2], over 80 combinations of different types of signs were mentioned in parental concerns. At the same time, screening all children for a broad range of signs would entail a lengthy and costly screening process, posing a particular problem for pediatricians [7]. There is therefore a need for a mixed screening approach of user- and expert-driven report.

ML tools offer an efficient way to draw generalizations from available data. We hypothesize that these tools will be able to draw generalizations from the individual nature of parental concerns to predict ASD risk [23], so long as parents can be asked for additional information to complement their specific narrative. In this study, we wished to test whether, based on the text submitted by the parent, together with a minimal set of well-established ASD signs (M-CHAT-R questions) we could predict ASD risk. ASD risk was defined in this study by a gold standard global developmental screening tool and ASD-specific screening tools.

This study aims to (1) evaluate the ASQ and M-CHAT-R risk status of children with parental concerns in the social-communication domain, and whether this status correlates with a family history of ASD and expert assessment of ASD risk; (2) examine whether an automatically selected single M-CHAT-R question improves the accuracy of automated prediction of ASD risk based on parental concerns in Web-based sources; and (3) identify which M-CHAT-R questions contribute the most to the automated prediction of ASD risk. See Figure 1 for a description of the screening stages put forward in this study.

Methods

Participants

Inclusion criteria were as follows: (1) child's age between 16 and 30 months; and (2) parent had a concern regarding their social-communication development. The final sample (n=132) included 25 parents recruited via social media and 107 recruited via Instant.ly, a service for Web-based surveys and recruitment of relevant survey participants. A total of 6 participants entered gibberish or nonsense; they were excluded from the analyses. Of all children, 7 were already diagnosed with ASD, and 4 parents did not write about their concerns but rather a different topic; they too were excluded from the analyses. The remaining 115 participants' children had a mean age of 25.06 months (SD 4.46, range 16-30), of whom 58.3% (67/115) were boys. Of the participants, 57.4% (66/115) reported having a family member with ASD; these participants will be referred to as the ASD family group. The ASD family members included 37 siblings, 21 cousins, 7 aunts or uncles, and 1 grandson.

Measures

We used M-CHAT-R/F [24]. The M-CHAT-R is a short gold-standard ASD-specific screening tool, in which parents of children aged 16 to 30 months rate their children's early ASD signs. The M-CHAT-R includes 20 binary (Yes/No) questions regarding signs from the social-communication and repetitive behavior domains. Scoring of the questionnaire is based on the number of answers reporting an at-risk symptom (potential range 0-20 months); the child may be classified as low risk (0-2), medium risk (3-7), or high risk (8-20). In this study, when a parent's answer to an item indicated risk for ASD, the structured follow-up question for that item was immediately administered. This administration method is different from the method used by others [25] who administered the initial 20 items first, later followed by the relevant follow-up items. For this study, the follow-up was programmed for administration interspersed with initial items, without first tallying a total M-CHAT-R risk status. Children whose final M-CHAT-R/F score is 2 or higher are at risk for ASD and warrant referral for diagnostic evaluation and early intervention; children whose final score is 0 to 1 are at low risk and are not in need of referral unless parents or professionals had ASD concerns.

The M-CHAT-R/F was previously validated in a sample of 16,071 low-risk toddlers [25] and showed adequate internal consistency in the 2-stage screening process (Cronbach alpha=.79). Using the threshold of 3 initial risk items to trigger follow-up, and 2 or more risk items after follow-up, they found a sensitivity of 0.85 and specificity of 0.99. Although the positive predictive value for ASD was modest (0.48), 83% (184/221) of the children who screened positive were diagnosed with a developmental disorder, and nearly 95% (209/221) demonstrated developmental delays or concerns.

Additionally, we used the Ages and Stages Questionnaire-3 (ASQ) [26]. The ASQ is a short gold-standard broad developmental screening tool, in which parents rate their child's current skills and development. Parents answer 30 questions, 6 for each of 5 developmental domains: gross motor, fine motor, personal-social, problem solving, and communication. There are 21 versions of the ASQ for different age groups, from birth to 6 years. Items are rated by parents as "not yet" (0), "sometimes" (5), or "yes" (10). Each developmental domain yields a summary score and a risk score relative to age cutoffs. If the score is below 2 SDs from the norm, the child is considered in a need of referral. A score between 1 to 2 SDs indicates the need for monitoring in that domain. ASQ global developmental risk is defined as a score below 2 SDs in at least one domain.

Procedures

Following ethics committee approval, participants were recruited through social media and Instant.ly. Instant.ly is a Web-based platform for conducting structured surveys, where participants are recruited based on researchers' criteria by the company using financial incentives. All participants completed the survey on the Instant.ly survey platform after confirming that their child is between 16 to 30 months of age and that they have a concern related to their child's social-communication development. Following this initial screening, parents were asked to describe, in their own words, their concerns regarding their child's social-communication development. Parents answered a minimal set of background questions including age and gender of the child and whether they had a family member diagnosed with ASD. Then they answered the M-CHAT-R/F questions. Finally, parents completed the ASQ version that was appropriate for their child's age. Parents were compensated US $35 for their participation.

To further validate the child's risk status, we sought clinical judgment regarding the ASD risk of the child based on the parent's narrative. A total of 3 clinicians with clinical expertise in early ASD screening rated the degree of risk of developing ASD based on reading the parents' reported concerns, from 1 (no risk) to 4 (high risk). Raters were blind to the M-CHAT-R/F and ASQ scores. Intraclass correlation coefficient for rating ASD risk was 0.76 (95% CI 0.67-0.83).

Data Analysis

Spearman correlations were used for correlations between standardized screening tests and with clinician ASD risk rating. Chi-square tests were conducted to compare risk status on the 3 standardized measures between the groups with a family history of ASD compared with the group without a family history of ASD.

We used the initial M-CHAT-R risk status as well as the final M-CHAT-R/F risk status to ensure that we were covering 2 scenarios. One is a parent who wants to have his/her child screened but will not see another provider for that procedure, in which case the M-CHAT-R/F is the important score. The other is a parent who wants to know whether his/her child is at risk so that they can consult with a clinician for further screening, in which case the M-CHAT-R is the relevant score.

Machine Learning Automatic Assessment of Autism Spectrum Disorder Risk From Text

Our primary task was to predict the risk for ASD, either solely using the text provided by the parent or by augmenting the text with the parent's response to a single question. Specifically, the algorithm for predicting risk proceeded as follows: first, the algorithm used the text provided by parents to decide which question to ask, if any. Once a response was given to the chosen question, the algorithm predicted risk.

The question to add was selected from the 20 M-CHAT-R questions. The algorithm selected which question to add (if any) based on an estimate of whether the prediction from text combined with a question is better than one based on text alone. Specifically, if we knew the risk score to a given question and were asked to choose between using the output of a text-based risk predictor and that of a text-based risk predictor augmented with an answer to the *j*-th question, we would choose the most extreme of the 2 predictor outputs. That is, if the true risk is high, we would prefer the predictor output which is higher of the two, and vice versa. In practice, we do not know the actual risk score or the answer to the *j*-th question for a test question. Therefore, we trained a second predictor to estimate whether the output of a risk predictor with an added question would give a more extreme answer (in the right direction) than that of a risk predictor without an added question. The second predictor was

trained when the actual risk scores are known (in the training set), given the following independent variables: (1) output of the risk predictor using only text, (2) output of the risk predictor using text and a negative answer to the *j*-th question, (3) output of the risk predictor using text and a positive answer to the *j*-th question, and (4) variance of the 3 abovementioned predictors at the node where the test example is labeled. This procedure is repeated for all 20 augmenting questions. In the third stage, if the outputs of all selection decisions are below a particular threshold (set empirically to 0.5), we chose not to add a question to the text. If any are above that threshold, we add the question with the highest decision score.

We examined 6 possible risk outcomes: (1) ASQ, (2) M-CHAT-R, (3) M-CHAT-R/F, (4) ASQ Personal-Social, (5) ASQ Communication, and (6) Meeting ASQ Personal-Social or ASQ Communication risk. The narratives of parental concerns contained between 2 and 253 words (mean word count 72, SD 37). We represented the text of questions using single words, word pairs, and triplets, and lexical affinities [27], excluding stopwords [28], and keeping terms that appeared in 10 or more questions. The reported performance of the predictor was estimated using leave-one-out [23]. We trained the algorithm on all but one example and predicted the risk for that example, repeating the process for all examples. The algorithm used for predicting risk (using text or using text with an additional question) was a regression tree [23].

The performance of the predictors is reported using the AUC. AUC is a commonly used measure of classifier accuracy, calculated from a graph of the true-positive rate versus the false-positive rate, for different thresholds of the classifier. A perfect classifier would have an AUC of 1, whereas a random decision would achieve an AUC of 0.5.

Results

Risk Status Across Standardized Screening Measures

Table 1 presents the number of children who were at risk according to each of the screening measures; 79.1% (91/115) show a global developmental risk on the ASQ, and 73.9% (85/115) an ASD-specific risk on the M-CHAT-R/F. All those who were at high risk on the M-CHAT-R showed a confirmed risk on the M-CHAT-R/F. Due to the interspersed administration of the Follow-Up, fourteen children with low risk on the M-CHAT-R completed Follow-Up items. Of all the children, 14.3% (2/13) of the children at low risk and 58% (25/43) of those at medium risk on M-CHAT-R ended up categorized as at risk on the two-stage M-CHAT-R/F.

Spearman correlation coefficients between M-CHAT-R and M-CHAT-R/F total score and ASQ domain scores were significant and high (r_s=−0.58 to −0.65, P_s<.001). Of those at global risk according to the ASQ, 94.5% (85/91) were also at medium/high risk according to the M-CHAT-R and 82.4% (75/91) according to the M-CHAT-R/F. Only 17.6% (16/91) of those at risk based on the ASQ were not at risk on the M-CHAT-R/F, whereas 41.7% (10/42) of those who were not at risk based on the ASQ were at risk according to the M-CHAT-R/F.

Looking at ASD-related ASQ domains, 97% (65/67) with a refer score on the Personal-Social ASQ domain were at medium/high risk according to the M-CHAT-R, and 86.6% (58/67) were also at risk according to the M-CHAT-R/F. Similarly, 95.5% (64/67) with a refer score on the Communication ASQ domain were at medium/high risk on the M-CHAT-R, and 88.1% (59/67) were also at risk on to the M-CHAT-R/F.

Significant differences in odds of risk between those with versus without a family history of ASD (see Table 1) provided further validation of these screening scores. Chi-square tests indicated that those with a history of ASD in their family had significantly higher probability of risk on the M-CHAT-R and M-CHAT-R/F, as well as ASQ Personal-Social scores. Those with a history of ASD in family had nearly a 4 times higher likelihood of scoring at risk on the M-CHAT-R/F compared with those without a family history. This subgroup had a 2.63 times higher likelihood for a need for referral on the Personal-Social ASQ subscale.

Table 1. Modified Checklist for Autism in Toddlers-Revised (M-CHAT-R) and Ages and Stages Questionnaire (ASQ) risk rates, n (%). M-CHAT-R/F: Modified Checklist for Autism in Toddlers-Revised, with Follow-up.

Screening measures	Full sample n=115, n (%)	No ASD in family n=49, n (%)	ASD in family n=66, n (%)	Odds ratio (95% CI)	Chi-square (*df*)	*P* value
M-CHAT-R risk						
M-CHAT-R medium + high risk	101 (87.8)	39 (75)	62 (93)	3.97 (1.17-13.55)	5.4 (1)	.02
M-CHAT-R/F Risk (count fail ≥2)	85 (73.9)	29 (59)	56 (84)	3.86 (1.60-9.32)	9.6 (1)	.002
ASQ Factors						
ASQ risk	91 (79.1)	36 (73)	55 (83)	1.81 (0.73-4.47)	1.7 (1)	.20
Communication	67 (58.3)	25 (51)	42 (63)	1.68 (0.79-3.56)	1.8 (1)	.18
Personal-Social	85 (74.0)	22 (44)	45 (68)	2.63 (1.22-5.65)	6.3 (1)	.01
Gross motor	57 (49.6)	22 (44)	35 (53)	1.39 (.66-2.91)	0.7 (1)	.39
Fine motor	51 (44.3)	17 (34)	34 (51)	2.00 (0.93-4.28)	3.2 (1)	.07
Problem solving	68 (59.1)	29 (59)	39 (59)	1.00 (0.47-2.11)	0 (1)	.99

Table 2. Spearman correlations between screening scores and expert rating. *P* values were adjusted for multiple correlations. Ages and Stages Questionnaire (ASQ) correlations are negative as ASQ scores denote competencies whereas expert ratings reflect risk level. ASD: autism spectrum disorder; M-CHAT-R: Modified Checklist for Autism in Toddlers-Revised; M-CHAT-R/F: Modified Checklist for Autism in Toddlers-Revised, with Follow-up.

Screening measures	Expert rating of ASD risk
M-CHAT-R	0.43[a]
M-CHAT-R/F	0.36[a]
ASQ Communication	−0.21
ASQ Personal-Social	−0.26[a]
ASQ Gross Motor	−0.26[a]
ASQ Fine Motor	−0.21
ASQ Problem Solving	−0.29[a]

[a]*P*<.007.

Table 3. Area under the receiver operating curve (AUC) values of risk prediction. ASQ: Ages and Stages Questionnaire; M-CHAT-R: Modified Checklist for Autism in Toddlers-Revised; M-CHAT-R/F: Modified Checklist for Autism in Toddlers-Revised, with Follow-up.

Screening measures	AUC text only	AUC text + question selected
M-CHAT-R risk		
M-CHAT R/F	0.39	0.85
M-CHAT-R	0.54	0.88
ASQ Factors		
ASQ Risk	0.55	0.85
Communication	0.60	0.74
Personal-Social	0.36	0.80
Communication or Personal-Social	0.49	0.79

Expert Rating of Child Risk Based on Parental Concerns

Experts rated a total of 33.9% (39/115) children, as at high (3/4) ASD risk, based on the parents' text. M-CHAT-R, M-CHAT-R/F, and ASQ-Risk scores were all significantly associated with the experts' rating of ASD risk (see Table 2).

Machine Learning Automatic Assessment of Autism Spectrum Disorder Risk From Parental Concerns

Table 3 and Multimedia Appendix 1 show the proportion of children predicted as "at risk" for each of the risk measures, the AUC for risk prediction using the text of questions alone, and the AUC when a question is added. The accuracy of the predictor that selects between text-based and text-augmented predictor (the second stage predictor) had an AUC greater than 0.74 for all risk measures (10-fold cross-validation; AUCs ranged between 0.74 and 0.88 across outcome measures). Notably, M-CHAT-R initial score had the largest AUC from text and questions selected and the highest prediction from text alone (34%).

Adding a question significantly improves the accuracy of risk prediction (Friedman test, *P*=.02). The table also presents 11 single M-CHAT-R questions that when asked contributed to the prediction of risk for 5% or more of the children in at least one of the risk measures (in addition to the text). The selection of question is automatically guided by the text. Note that for the ASQ risk prediction 17% of the questions required no additional questions for estimating the children's risk status.

Discussion

Principal Findings

This study examines the potential for providing automated Web-based support for parents expressing concerns regarding ASD signs in their toddler. The unique findings from this study demonstrate the potential for accurately predicting a child's ASD risk or global development risk status from a parent's report of concerns in Web-based forums. This prediction was made possible by supplementing open-ended concern reports with 1 question from the M-CHAT-R. Designing such a system encounters the challenges of predicting from the diverse free texts entered by parents, as well as the need to be brief. Parents are using Web-based platforms to seek advice regarding their child's health and development [21,29]. Web-based screening for ASD is important as it may reduce the time gap between parents' first worry and approaching a professional [3] and consequently enable the family to seek services earlier.

Automated Support for Parents in Web-Based Forums

Findings from this study indicate that by applying ML tools, parental concerns in Web-based platofrms can be utilized for ASD prescreening. Adding a single question to the coded free response of the parents yielded AUCs ranging between 0.74 and 0.88, depending on the risk outcome measure predicted (ie, M-CHAT-R(/F), ASQ total, or specific domains). The M-CHAT-R AUCs in our study were similar to the AUC values reported previously in a study using ML tools for predicting ASD risk from parental queries combined with an ontology of the signs [21] and to the AUC in a postpartum depression screening study using ML tools [22]. The text alone predicted up to 34% of those at risk, depending on the risk outcome score predicted. These findings are consistent with research indicating that parental concerns alone cannot fully predict ASD diagnostic status [30,31]; rather, parents benefit from targeted follow-up questions to increase accuracy of prediction. Consistent with this finding, the AAP guides health care providers to follow-up on parental concerns with ASD-specific screening. The innovative ML methodology enables the selection of a single screening question, tailored based on the content of the concern, to significantly improve accuracy.

Autism Spectrum Disorder Screening Questions That Complement Parental Concerns

Of all questions, 3 M-CHAT-R questions showed consistent contribution when combined with text to predict all 6 screening outcome measures. These items measured following a parent's pointed finger (16%-35%), appearing deaf (9%-22%), and lack of make-believe play (6%-21%; see Table 3). Interestingly, following a pointed finger has been reported as a key question in several ASD screening studies as has lack of make-believe play [25,32] and appearing deaf [26]. However, these questions did not differentiate ASD-related risk from global developmental risk. We hypothesize that these questions inquire about behaviors that parents are less aware of and may require knowledge regarding child development to notice them. This is supported by their absence/low appearance in parental first concerns related to ASD [1,2]. However, these items were not the best added item for all parents; the diversity of ASD-related concerns reported by parents requires an individualized follow-up approach, as demonstrated by this study.

A challenge in ASD screening is differentiating global development risk from an ASD-specific risk. There is evidence that most of the false positives in ASD screening studies have a different type of developmental disorder [25,33]. The challenge of differential screening is highlighted in the significant correlations between M-CHAT-R and ASQ scores and the finding that same 3 M-CHAT-R questions contributed to ASD-specific and nonspecific outcome scores. This is in line with studies showing high correspondence between the ASQ global developmental risk and an ASD diagnosis [34]. At the same time, there was a subgroup of children who only showed risk of ASD; 41.7% (48/115) of the sample who were at risk according to the M-CHAT-R/F were not at risk according to the ASQ, whereas most of those who were at risk according to the ASQ were also at risk according to the M-CHAT-R/F. Questions that specifically predicted ASD-related risk (ie, M-CHAT-R/F, ASQ Personal-Social, ASQ Communication) and not global risk status on the ASQ were pointing to show, interest in others, showing, and responding to name. Although they added to the prediction of risk in a smaller percentage of cases, they seem to be more specific to ASD risk status. These items were among the best 7 M-CHAT-R questions for predicting ASD diagnosis [24]. As certain ASD-related questions (finger movements and getting your attention to watch him/her) improved the prediction of ASQ global risk status but did not improve the prediction of ASD-related risk status, an expert evaluation is necessary for differential diagnosis.

Screening Risk Status for Children Whose Parents Are Concerned About Them

In absence of diagnostic outcomes, we relied upon standardized screening scores as outcome measures. Of the children in our sample, 87.8% (101/115) were at medium or high risk according to M-CHAT-R, 73.9% (85/115) according to M-CHAT-R/F, 79.1% (91/115) according to the ASQ total, and 58%-74% (20/115 to 97/115) according to the Personal-Social and/or Communication ASQ subscales. Participants in the study had concerns that likely motivated their participation. One would expect such a high positive risk rate when screening a sample of children for whom there is already concern in the social-communication domain [2,9]. Children in our study with versus without a family history of ASD had a much higher likelihood of presenting risk on ASD-specific screening and social-communication developmental screening, consistent with previous research [35,36]. It is not surprising that in a sample of parents concerned about their child's social-communication development, there is a high rate of family history of ASD, which justifiably raises parents' concerns given the high recurrence rates of ASD [6] and parents' awareness of ASD symptom manifestation [9]. When interpreting the results of this study we must consider the potential impact of including concerns of parents with a family history of ASD. First, evidence indicates that siblings of children with ASD are not necessarily representative of the general ASD population [35]. Second, the language of parents with a family history of ASD may be enriched with terms that are more specific to ASD than in the general public leading to higher accuracy prediction from their text alone.

It is noteworthy that in standard use of the M-CHAT-R, some of the children in the M-CHAT-R medium-risk range eventually drop into the low-risk range after follow-up questions are administered, as opposed to the high-risk children, who are referred for evaluation without follow-up. This is consistent with prior studies [25] and suggests that in some cases, parents need additional support from the structured follow-up items to describe their child's behavior. Hence, the M-CHAT-R/F results are more likely to predict typical versus delayed development; however, it is common to report both initial M-CHAT-R and M-CHAT-R/F outcomes.

Limitations

Future studies should randomize the order of the M-CHAT-R and ASQ completion as the fixed order may have led to elevated ASQ scores. In addition, the interspersed administration of the M-CHAT-R/F within the M-CHAT-R was not the standardized

method for administering this tool. However, this approach identified 2 children who would have been missed if the M-CHAT-R/F had been administered only to those who screened positive on the initial M-CHAT-R questionnaire. The interspersed procedure may have led to elevated risk rates on the M-CHAT-R as our risk rates were higher than those reported in another study of infant siblings of children with ASD [36]. At the same time in addition to the family history of ASD, parents in our study had specific social-communication concerns suggesting that elevated risk rates were not solely associated with measurement. In the absence of diagnostic outcomes, we cannot determine who true false-negative children were. This study requires replication with diagnostic outcomes, which would increase the clinical validity of the automated prediction and follow-up procedure. Sociodemographic features of the sample were unknown, hence further research is needed to identify potential biases in such data. In addition, to generalize results to patients in general primary care settings, this method needs to be tested with a low-risk representative sample.

Conclusions

This study aimed to test the possibility of automatically estimating a child's risk for ASD based on his/her parent's description of concerns. Utilizing ML methods, we showed satisfying performance of prediction models, relying on the parent's text and a particular M-CHAT-R question that complemented that unique text. The selection of one tailored question promoted accurate screening. Such a screening method that alerts parents can reduce the gap between first concern and approaching a health care provider. This study offers an opportunity to capitalize on digital health methods for designing a real-time support system that would ask users screening questions based on their concerns.

Acknowledgments

The authors received financial support for the research of this paper from the Caesarea Rothschild Institute.

Conflicts of Interest

ABS and EYT have no conflicting interests. The second author of the paper, DLR, is a co-owner of M-CHAT, LLC, which licenses the Modified Checklist for Autism in Toddlers and related materials to commercial entities. No royalties were received that relate to any data presented in this manuscript.

Multimedia Appendix 1

Percentage of contribution of each question to different risk measures.

[PDF File (Adobe PDF File), 341KB - jmir_v20i4e134_app1.pdf]

References

1. Guinchat V, Chamak B, Bonniau B, Bodeau N, Perisse D, Cohen D, et al. Very early signs of autism reported by parents include many concerns not specific to autism criteria. Res Autism Spectr Disord 2012 Apr;6(2):589-601 [FREE Full text] [doi: 10.1016/j.rasd.2011.10.005]
2. Richards M, Mossey J, Robins D. Parents' concerns as they relate to their child's development and later diagnosis of autism spectrum disorder. J Dev Behav Pediatr 2016 Sep;37(7):532-540 [FREE Full text] [doi: 10.1097/DBP.0000000000000339] [Medline: 27541581]
3. Siklos S, Kerns K. Assessing the diagnostic experiences of a small sample of parents of children with autism spectrum disorders. Res Dev Disabil 2007;28(1):9-22. [doi: 10.1016/j.ridd.2005.09.003] [Medline: 16442261]
4. Klintwall L, Eldevik S, Eikeseth S. Narrowing the gap: effects of intervention on developmental trajectories in autism. Autism 2015 Jan;19(1):53-63. [doi: 10.1177/1362361313510067] [Medline: 24212258]
5. Ozonoff S, Williams BR, Landa R. Parental report of the early development of children with regressive autism: the delays-plus-regression phenotype. Autism 2005 Dec;9(5):461-486. [doi: 10.1177/1362361305057880] [Medline: 16287700]
6. Ozonoff S, Young GS, Steinfeld MB, Hill MM, Cook I, Hutman T, et al. How early do parent concerns predict later autism diagnosis? J Dev Behav Pediatr 2009 Oct;30(5):367-375 [FREE Full text] [Medline: 19827218]
7. Crais E, McComish C, Humphreys B, Watson LR, Baranek GT, Reznick JS, et al. Pediatric healthcare professionals' views on autism spectrum disorder screening at 12-18 months. J Autism Dev Disord 2014 Sep;44(9):2311-2328. [doi: 10.1007/s10803-014-2101-2] [Medline: 24700359]
8. Dobrez D, Sasso AL, Holl J, Shalowitz M, Leon S, Budetti P. Estimating the cost of developmental and behavioral screening of preschool children in general pediatric practice. Pediatrics 2001 Oct;108(4):913-922. [Medline: 11581444]
9. Sacrey L, Zwaigenbaum L, Bryson S, Brian J, Smith I, Roberts W, et al. Can parents' concerns predict autism spectrum disorder? A prospective study of high-risk siblings from 6 to 36 months of age. J Am Acad Child Adolesc Psychiatry 2015 Jun;54(6):470-478. [doi: 10.1016/j.jaac.2015.03.014] [Medline: 26004662]
10. Herlihy L, Knoch K, Vibert B, Fein D. Parents' first concerns about toddlers with autism spectrum disorder: effect of sibling status. Autism 2015 Jan;19(1):20-28 [FREE Full text] [doi: 10.1177/1362361313509731] [Medline: 24216070]

11. Glascoe F, MacLean W, Stone W. The importance of parents' concerns about their child's behavior. Clin Pediatr (Phila) 1991 Jan;30(1):8-11; discussion 12. [doi: 10.1177/000992289103000101] [Medline: 1995206]
12. Lo B, Klopper F, Barnes E, Williams K. Agreement between concern about autism spectrum disorder at the time of referral and diagnosis, and factors associated with agreement. J Paediatr Child Health 2017 Aug;53(8):742-748. [doi: 10.1111/jpc.13511] [Medline: 28374573]
13. Johnson C, Myers S, American Academy of Pediatrics Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. Pediatrics 2007 Nov;120(5):1183-1215. [doi: 10.1542/peds.2007-2361] [Medline: 17967920]
14. 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]
15. Wainstein B, Sterling-Levis K, Baker S, 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]
16. Harvey S, Memon A, Khan R, Yasin F. Parent's use of the Internet in the search for healthcare information and subsequent impact on the doctor-patient relationship. Ir J Med Sci 2017 Nov;186(4):821-826. [doi: 10.1007/s11845-017-1555-6] [Medline: 28130665]
17. Khoo K, Bolt P, Babl F, Jury S, Goldman R. Health information seeking by parents in the Internet age. J Paediatr Child Health 2008;44(7-8):419-423. [doi: 10.1111/j.1440-1754.2008.01322.x] [Medline: 18564080]
18. Uy C, Lopez J, Trinh-Shevrin C, Kwon S, Sherman S, Liang P. Text messaging interventions on cancer screening rates: a systematic review. J Med Internet Res 2017 Aug 24;19(8):e296 [FREE Full text] [doi: 10.2196/jmir.7893] [Medline: 28838885]
19. BinDhim N, Alanazi E, Aljadhey H, Basyouni M, Kowalski S, Pont L, et al. Does a mobile phone depression-screening app motivate mobile phone users with high depressive symptoms to seek a health care professional's help? J Med Internet Res 2016 Dec 27;18(6):e156 [FREE Full text] [doi: 10.2196/jmir.5726] [Medline: 27349441]
20. Gentles S, Lokker C, McKibbon K. Health information technology to facilitate communication involving health care providers, caregivers, and pediatric patients: a scoping review. J Med Internet Res 2010 Jun 18;12(2):e22 [FREE Full text] [doi: 10.2196/jmir.1390] [Medline: 20562092]
21. Ben-Sasson A, Yom-Tov E. Online concerns of parents suspecting autism spectrum disorder in their child: content analysis of signs and automated prediction of risk. J Med Internet Res 2016 Nov 22;18(11):e300 [FREE Full text] [doi: 10.2196/jmir.5439] [Medline: 27876688]
22. De Choudry M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. Palo Alto, California: AAAI press Presented at: Proceedings of the 7th International Conference on Weblogs and Social Media; 8-10 July, 2013; Boston, Massachusetts p. 128-137 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6124/6351
23. Duda RO, Hart PE, Stork DG. Pattern Classification (2nd ed.). New York: Wiley-Interscience; 2001.
24. Robins D, Fein D, Barton M. Autism Speaks. 2009. Modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F) URL: https://www.autismspeaks.org/sites/default/files/docs/sciencedocs/m-chat/m-chat-r_f.pdf?v=1 [WebCite Cache ID 6vKmaN7WV]
25. Robins DL, Casagrande K, Barton M, Chen CA, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). Pediatrics 2013 Dec 23;133(1):37-45 [FREE Full text] [doi: 10.1542/peds.2013-1813] [Medline: 24366990]
26. Squires J, Twombly E, Bricker D, Potter L. Ages and stages questionnaires user's guide (3rd ed.). Baltimore, MD: Paul H. Brookes; 2009.
27. Carmel D, Farchi E, Petruschka Y, Soffer A. Automatic query refinement using lexical affinities with maximal information gain. New York: ACM Presented at: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval; 11-15 August, 2002; Tampere, Finland. [doi: 10.1145/564376.564377]
28. Wilbur WK, Sirotkin K. The automatic identification of stop words. J Inf Sci 2016 Jul;18(1):45-55. [doi: 10.1177/016555159201800106]
29. 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):1-12 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
30. Glascoe FP, Macias MM, Wegner LM, Robertshaw NS. Can a broadband developmental-behavioral screening test identify children likely to have autism spectrum disorder? Clin Pediatr (Phila) 2007 Nov;46(9):801-805. [doi: 10.1177/0009922807303928] [Medline: 17641122]
31. Veness C, Prior M, Bavin E, Eadie P, Cini E, Reilly S. Early indicators of autism spectrum disorders at 12 and 24 months of age: a prospective, longitudinal comparative study. Autism 2012 Mar;16(2):163-177. [doi: 10.1177/1362361311399936] [Medline: 21733958]
32. Baron-Cohen S, Wheelwright S, Cox A, Baird G, Charman T, Swettenham J, et al. Early identification of autism by the CHecklist for Autism in Toddlers (CHAT). J R Soc Med 2000 Oct;93(10):521-525 [FREE Full text] [doi: 10.1177/014107680009301007] [Medline: 11064690]

33. Sturner R, Howard B, Bergmann P, Morrel T, Andon L, Marks D, et al. Autism screening with online decision support by primary care pediatricians aided by M-CHAT/F. Pediatrics 2016 Sep;138(3):1-11 [FREE Full text] [doi: 10.1542/peds.2015-3036] [Medline: 27542847]
34. Hardy S, Haisley L, Manning C, Fein D. Can screening with the Ages and Stages Questionnaire detect autism? J Dev Behav Pediatr 2015 Sep;36(7):536-543 [FREE Full text] [doi: 10.1097/DBP.0000000000000201] [Medline: 26348972]
35. Pandey J, Verbalis A, Robins DL, Boorstein H, Klin AM, Babitz T, et al. Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. Autism 2008 Sep;12(5):513-535. [doi: 10.1177/1362361308094503] [Medline: 18805945]
36. Weitlauf AS, Vehorn AC, Stone WL, Fein D, Warren ZE. Using the M-CHAT-R/F to identify developmental concerns in a high-risk 18-month-old sibling sample. J Dev Behav Pediatr 2015 Sep;36(7):497-502 [FREE Full text] [doi: 10.1097/DBP.0000000000000194] [Medline: 26171765]

Abbreviations

AAP: American Academy of Pediatrics
ASD: Autism Spectrum Disorder
ASQ: Ages and Stages Questionnaire
AUC: area under the receiver operating curve
M-CHAT: Modified Checklist for Autism in Toddlers
M-CHAT-R/F: Modified Checklist for Autism in Toddlers-Revised, with Follow-upML: machine learning

Edited by G Eysenbach; submitted 30.11.17; peer-reviewed by MN Sadat, E McGhee Hassrick; comments to author 18.01.18; revised version received 04.02.18; accepted 16.02.18; published 24.04.18.

Please cite as:
Ben-Sasson A, Robins DL, Yom-Tov E
Risk Assessment for Parents Who Suspect Their Child Has Autism Spectrum Disorder: Machine Learning Approach
J Med Internet Res 2018;20(4):e134
URL: http://www.jmir.org/2018/4/e134/
doi:10.2196/jmir.9496
PMID:29691210

Original Paper

A Comparison and Calibration of a Wrist-Worn Blood Pressure Monitor for Patient Management: Assessing the Reliability of Innovative Blood Pressure Devices

Sarah Melville[1,2], BSc (Hons), CRA; Robert Teskey[2,3], MD, FRCP(C); Shona Philip[1,2], MD, MSc; Jeremy A Simpson[4], PhD; Sohrab Lutchmedial[2,3], MD, FRCP(C); Keith R Brunt[1,2,3], PhD

[1]Department of Pharmacology, Dalhousie Medicine New Brunswick, Saint John, NB, Canada

[2]Cardiovascular Research New Brunswick, New Brunswick Heart Centre, Saint John, NB, Canada

[3]Department of Cardiology, Saint John Regional Hospital, Horizon Health Network, Saint John, NB, Canada

[4]Department of Human Health Nutritional Sciences, University of Guelph, Guelph, ON, Canada

Corresponding Author:
Keith R Brunt, PhD
Department of Pharmacology
Dalhousie Medicine New Brunswick
P.O. Box 5050
100 Tucker Park Road
Saint John, NB, E2L 4L5
Canada
Phone: 1 506 636 6974
Fax: 1 506 636 6001
Email: keith.brunt@dal.ca

Abstract

Background: Clinical guidelines recommend monitoring of blood pressure at home using an automatic blood pressure device for the management of hypertension. Devices are not often calibrated against direct blood pressure measures, leaving health care providers and patients with less reliable information than is possible with current technology. Rigorous assessments of medical devices are necessary for establishing clinical utility.

Objective: The purpose of our study was 2-fold: (1) to assess the validity and perform iterative calibration of indirect blood pressure measurements by a noninvasive wrist cuff blood pressure device in direct comparison with simultaneously recorded peripheral and central intra-arterial blood pressure measurements and (2) to assess the validity of the measurements thereafter of the noninvasive wrist cuff blood pressure device in comparison with measurements by a noninvasive upper arm blood pressure device to the Canadian hypertension guidelines.

Methods: The cloud-based blood pressure algorithms for an oscillometric wrist cuff device were iteratively calibrated to direct pressure measures in 20 consented patient participants. We then assessed measurement validity of the device, using Bland-Altman analysis during routine cardiovascular catheterization.

Results: The precalibrated absolute mean difference between direct intra-arterial to wrist cuff pressure measurements were 10.8 (SD 9.7) for systolic and 16.1 (SD 6.3) for diastolic. The postcalibrated absolute mean difference was 7.2 (SD 5.1) for systolic and 4.3 (SD 3.3) for diastolic pressures. This is an improvement in accuracy of 33% systolic and 73% diastolic with a 48% reduction in the variability for both measures. Furthermore, the wrist cuff device demonstrated similar sensitivity in measuring high blood pressure compared with the direct intra-arterial method. The device, when calibrated to direct aortic pressures, demonstrated the potential to reduce a treatment gap in high blood pressure measurements.

Conclusions: The systolic pressure measurements of the wrist cuff have been iteratively calibrated using gold standard central (ascending aortic) pressure. This improves the accuracy of the indirect measures and potentially reduces the treatment gap. Devices that undergo auscultatory (indirect) calibration for licensing can be greatly improved by additional iterative calibration via intra-arterial (direct) measures of blood pressure. Further clinical trials with repeated use of the device over time are needed to assess the reliability of the device in accordance with current and evolving guidelines for informed decision making in the management of hypertension.

Trial Registration: ClinicalTrials.gov NCT03015363; https://clinicaltrials.gov/ct2/show/NCT03015363 (Archived by WebCite at http://www.webcitation.org/6xPZgseYS)

(J Med Internet Res 2018;20(4):e111) doi:10.2196/jmir.8009

KEYWORDS

patient self-management; diastolic hypertension; telemonitoring; vital signs; smartphone applications

Introduction

Hypertension is a global health problem affecting over a billion people [1]. If uncontrolled, it is a major risk factor for stroke, myocardial infarction, and kidney failure; it remains the leading cause of cardiovascular morbidity and mortality [2,3]. Diagnosis and management of hypertension relies intensively on the indirect measurement of blood pressure in outpatient settings.

Manual auscultatory and automated oscillometry (ie, clinical and patient blood pressure monitoring methods) are common means for diagnosing hypertension and guiding appropriate medical management. Current guidelines (Canadian Hypertension Education Program, CHEP; American Society of Hypertension; International Society of Hypertension; European Society of Hypertension, ESH) recommend confirmation of hypertension (ie, blood pressure ≥135/85) with ambulatory or home blood pressure monitoring [4-6] . In recent years, there has been a dramatic growth in automated devices and increased use of mobile health apps [7]. However, there is limited information about their accuracy or precision, which creates a risk for inappropriate therapy or a treatment gap [8,9]. Clinically, the term *treatment gap* refers to hypertensive patients who are left untreated because of underestimated noninvasive blood pressure readings, which could be a health risk for the patient.

Automated clinical oscillometric or consumer-level devices have been generally compared with manual auscultatory measurements [10-12], while studies using invasive blood pressure measurements for validation or calibration are increasing [13-15]. Regulatory agencies (eg, Health Canada; Food and Drug Administration, FDA) license devices without mandatory independent third-party, peer-reviewed assessment of the validity of measurements or calibration standards. The principal directive of regulatory agencies for these sphygmomanometer devices (Health Canada and FDA Class II) is to ensure physical safety and personal data security [16,17] rather than guarantee accuracy and precision for clinical diagnostic purposes. The minimal requirements prescribed by Health Canada are not the minimal requirements of a clinician.

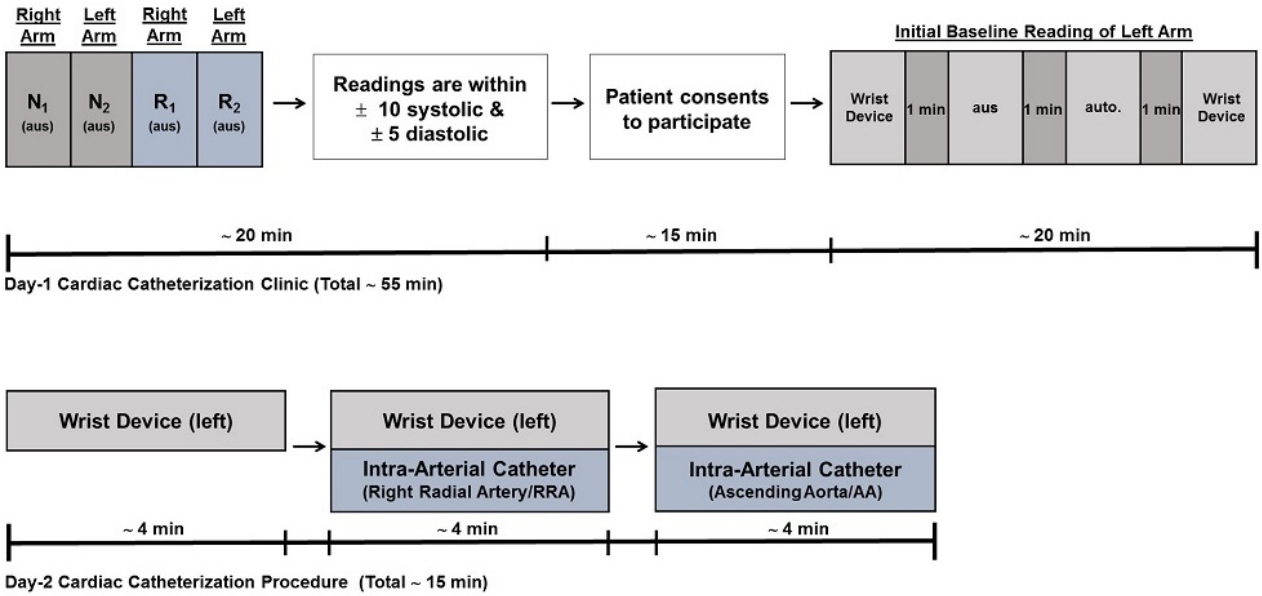
Increasingly, automated devices are being assessed against direct intra-arterial standards for clinical certainty, yet, most still do not report the uncertainty of the single blood pressure measurement. Comparing direct and indirect blood pressure measures simultaneously ensures that intraphysiological variability can be accounted for in the measures. The difference between 2 methods can then be validated with multiple measures within a patient population. There is a growing need for more accurate devices to measure blood pressure [18] to better diagnose and manage hypertension according to clinical guidelines. This could be accomplished by calibration to simultaneous direct measures in addition to auscultatory calibration using indirect measurements to reduce the treatment gap in hypertension. Generally, consumer devices, clinical devices, and the true invasive blood pressure measures in healthy and hypertensive patients should be in agreement with each other. Our objective was to assess the validity of indirect measures of blood pressure by a wrist-worn blood pressure device in direct comparison with simultaneously recorded gold standard intra-arterial blood pressure measures for the purpose of iterative device calibration.

Methods

Recruitment and Screening

The clinical protocol was approved by the Horizon Health Network Research Ethics Board, and the study is registered with the National Institute of Health Clinical Trials Registry database (NCT 03015363). The patient participant inclusion/exclusion criteria were: (1) a referral from the patient's attending cardiologist to undergo a first-time nonemergent diagnostic cardiac catheterization procedure for clinically valid indications; (2) participants aged ≥19 years; (3) wrist circumference should be in the range of 13.5 to 23 cm; (3) participant should be willing to voluntarily sign the study-specific informed consent form; and (4) participant should have no previous percutaneous coronary intervention, coronary artery bypass graft, abdominal aortic aneurysm, peripheral vascular disease, aortic stenosis, arrhythmia, tremors (before or during procedure), or carotid bruits. In accordance with the Association for the Advancement of Medical Instrumentation (AAMI) standards for clinical investigation using reference-based invasive blood pressure monitoring [19], we undertook a 2-day protocol (Figure 1).

Figure 1. Procedural overview: timelines for patient recruitment and data collection. Day 1-Timeline intake of patients for inclusion screening followed by consent and initial data collection. N and R are 2 trained investigators measuring blood pressure twice using the auscultatory method (aus); initial baseline readings used an automatic upper arm cuff (auto.). Day 2-Timeline for simultaneous invasive and noninvasive blood pressure measurements.



Blood Pressure Measurements

Logistically, the wrist cuff cannot be applied to the same arm that is being cannulated for radial pressure. Thus, patients were screened for bilateral upper arm auscultatory blood pressure equality (day 1; within ±10 mm Hg systolic and ±5 mm Hg diastolic pressure; see Multimedia Appendix 1) by 2 separate blinded readings performed using an upper arm aneroid sphygmomanometer (Welch Allyn Canada Ltd, Mississauga, ON) by 2 health professionals (Figure 1). This ensured that all patients did not have undiagnosed peripheral vascular disease to cause arm inequality in blood pressure. All eligible participating patients voluntarily provided documented informed consent. After obtaining the consents, 4 blood pressure readings were taken approximately 1 min apart to establish a baseline (Figure 1); 3 different devices were used for general observational comparison, according to standard procedures or manufacturers' instructions: (1) an oscillometric wrist cuff device (Cloud Diagnostics Inc, PULSEWAVE Health Monitor, Kitchener ON), (2) an upper arm oscillometric device (Welch Allyn Canada Ltd, Mississauga, ON), (3) and an upper arm aneroid sphygmomanometer (Welch Allyn Canada Ltd, Mississauga, ON). The pulse pressure waveform recorded by the wrist cuff device was digitized at 100 Hz and stored securely on a cloud-based server for subsequent analysis.

On the day of the cardiac catheterization procedure (Day 2), patients were given 1mg of midazolam and 50 µg of fentanyl for presedation ; blood pressure readings were taken as indicated in Figure 1 with the patient in the supine position. A total of 10 readings were taken with 2 measurements for each method (Figure 1); the first set of 2 wrist cuff measures were taken for patient conditioning and were not subsequently used for comparison in this study. The second set of 2 wrist cuff measures were taken while simultaneously recording 2 intra-arterial pressures at the right radial artery. The third set of 2 wrist cuff measures were taken while simultaneously recording 2 intra-arterial pressures at the ascending aorta. Each measurement was treated independently throughout this study (N=160 intra-arterial measurements in total for recalibration; 20 participants). Digital records of previously monitored data in a clinical setting using indirect, double-observer auscultatory measurements were secured for the purpose of clinical comparative analysis (N=375 measurements; 97 participants).

Intra-arterial pressure was measured with a fluid-filled 5 or 6 French gauge catheter (Cordis AVANTI+) attached to a pressure transducer (NAMIC, Navilyst Medical, Marlborough, MA, USA), which is consistent with the recent recommendation by the ARTERY Society task force [20]. The pressure transducer was zeroed before the start of the catheterization procedure. Intra-arterial pressure was recorded with hemodynamic software (MAC-LAB, General Electric Company). To avoid verapamil or heparin-induced changes in vascular tone, both drugs were administered as per standard of care procedure after pressure readings were complete.

Calculated device values of systolic and diastolic pressures were initially derived from the Cloud Diagnostics Inc application reports (original auscultatory calibration). The intra-arterial dataset was split into a training set and a testing set using the jackknife technique, as originally described [21]. The jackknife technique is a power data analysis tool suitable for small original data samples. In a dataset of N readings, the jackknife iterative processing can be described as the systematic resampling of a single reading from the entire dataset to be used as the testing set and the rest N−1 readings are used as the training set. This is repeated N times and during the ith iteration, the ith reading is chosen as the testing set and the rest N-1 readings are chosen as the training set. During each iteration the training set is used to obtain calibration coefficients, which are then used on the testing set. After making algorithm adjustment, new systolic and diastolic pressures were obtained directly from the engineers at Cloud Diagnostics Inc.

Statistical Analysis and Data Reporting

Data analysis was performed via Bland-Altman (Tukey mean difference) plot analysis [22,23] to assess the accuracy of the wrist cuff relative to intra-arterial measures. Systolic and diastolic validity was set a priori as any 2 measures being within 10 and 5 mm Hg, respectively (based on a normally distributed relevant clinical population) [24,25] (See Multimedia Appendix 2). The absolute mean difference and SD of all Bland-Altman plots are reported as the bias [20,26,27] (Also, see Multimedia Appendix 3). A one-tailed Fisher exact test was used to determine the sensitivity of the device in assessing measures of hypertension versus normotension.

Results

Blood Pressure Device Calibration

Here, we used intra-arterial blood pressure measurements in comparison with the noninvasive wrist cuff method (Figure 2).

A total of 74 potentially eligible patients were screened, of which 37 consented, 3 withdrew, and 14 failed to meet the criteria during the (day 2) procedure. A summary of the patients' characteristics that completed all aspects of this study is presented in Table 1.

The mean direct right radial arterial systolic pressure was 145.7 (SD 20.2) mm Hg, and the mean direct ascending aorta systolic pressure was 133.4 (SD 22.0) mm Hg. This illustrates the real physiological difference of 14.4 (SD 10.3), with $P<.001$, as a result of pressure augmentation [28]. There was no physiological difference in direct right radial arterial diastolic pressure and direct ascending aorta diastolic pressure (66.2 [SD 9.1] mm Hg vs 67.4 [SD 8.7] mm Hg), respectively; absolute mean difference of 5.1 (SD 3.6), $P=.23$).

Initially, the absolute mean difference of the wrist cuff compared with direct systolic measures using the original auscultatory calibration settings was 10.8 (SD 9.7), with $P<.001$, while the absolute mean difference of the diastolic measures was 16.1 (SD 6.3), with $P<.001$ (See Multimedia Appendix 4). Next, we adjusted the algorithm using the intra-arterial blood pressure datasets. First, we applied an iterative calibration of the wrist cuff to radial artery pressures, and the absolute mean difference of the systolic and diastolic measures was 7.9 (SD 6.6), with $P=.87$ and 4.3 (SD 3.3), with $P>.99$, respectively (Figure 3). However, we noted a negative slope trend line that may suggest an attenuated pressure when intraradial systolic pressure is greater than 150 mm Hg. Then, we calibrated the wrist cuff to ascending aortic pressures, and the absolute mean difference of the systolic and diastolic measures was 7.2 (SD 5.1), with $P=.97$ and 4.3 (SD 3.3), with $P=.98$, respectively (Figure 2), with a near 0 systolic pressure trend.

To further assess the value of the central algorithm independently of our calibration dataset, we sourced an arms-length dataset of double-observer auscultatory blood pressure measures from Cloud Diagnostics Inc (375 measurements; 97 participants). The average absolute mean difference of the initial algorithm markedly improved with the central pressure calibrated algorithm in both the systolic and diastolic pressures (7.5 [SD 7.3] vs 6.1 [SD 4.7] and 18.0 [SD 7.6] vs 9.8 [SD 6.0], respectively; see Multimedia Appendix 5). This is an improvement in accuracy of 20% with 38% less variability for systolic measures and 46% more accuracy with 19% less variability for diastolic measures using an independent dataset. Therefore, central pressure calibration improved the accuracy and reliability of the wrist-cuff device comparisons to direct pressures and upper arm cuff measures, which is relevant to clinical practice guidelines.

Blood Pressure Variability Assessment

To demonstrate measurement variability, a representative illustration of instantaneous intra-arterial pressures that were recorded every 10 s over a period of 4 min is shown in Figure 4. Measurements varied by approximately 20 and 10 mm Hg for systolic and diastolic pressures, respectively.

Subsequently, all invasive measures having SD 1 were plotted (Figure 5). Direct pressure analysis (both radial and aortic) illustrates pressure augmentation of peripheral blood pressure (Figure 5). Several measures have an SD that crosses the clinical threshold for diagnosing hypertension based on the CHEP guidelines [4] or the Systolic Blood Pressure Intervention Trial (SPRINT) [29]. Similar results were obtained by plotting the indirect measures (Figure 5). Taken together, in the absence of intraphysiological variability reporting (Figure 5), the wrist cuff produces diagnostically similar data to direct arterial pressure with a lower treatment gap risk than a noncalibrated indirect measure, such as the upper arm measures used in this study. Approximately 75% (15/20) of the patients in this study were already diagnosed with hypertension. According to direct intra-aortic and indirect wrist cuff measures, approximately 40% (Figure 5) were above the blood pressure target provided by the CHEP guidelines. When measures were plotted as being either hypertension positive or negative, the centrally-calibrated wrist cuff measures were on par with direct aortic measures using either CHEP- [4] or SPRINT-based [29] thresholds (Figure 5), whereas there was a significant difference between direct aortic pressure measures and calibrated wrist cuff measures compared with the upper arm cuff ($P=.04$; see Figure 5).

Figure 2. Physiological tracings. Representative hemodynamic tracing from the intra-arterial pressure catheter with electrocardiogram (ECG). Representative pulse waveform report from the wrist-cuff blood pressure device.

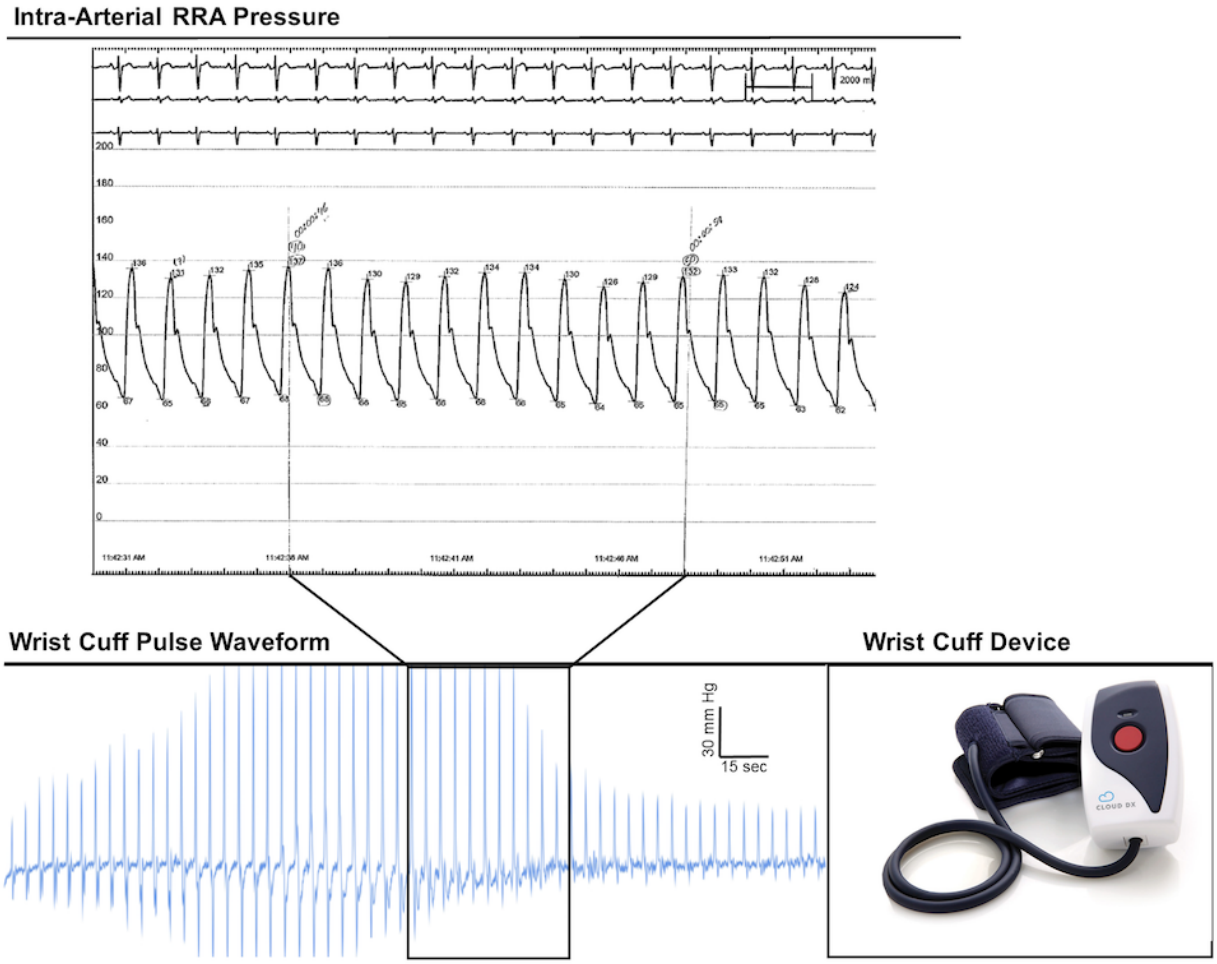


Table 1. Patient participant characteristics.

Characteristics	Patient participants (N=20)
Sex (male/female)	15/5
Age in years, mean (SD); range	62.0 (SD 9.0); 43-77
BMI[a], mean (SD); range	30.6 (SD 5.7); 21-45.4
Central aortic pressure (mm Hg)	
Systolic, mean (SD); range	133.4 (SD 22.0); 96.7-179.3
Diastolic, mean (SD); range	67.4 (SD 8.7); 50.6-85.1
Peripheral arterial pressure (mm Hg)	
Systolic, mean (SD); range	145.7 (SD 20.2); 113.0-184.8
Diastolic, mean (SD); range	66.2 (SD 9.1); 49.4-84.8
Wrist circumference in cm (left), range	15.5-21.5
Smoking, % (Y/N)	15 (3/17)
Diabetes, % (Y/N)	35 (7/13)
Statin, % (Y/N)	70 (14/6)
Hypertension, % (Y/N)	75 (15/5)

[a]BMI: body mass index.

Figure 3. Direct intra-arterial blood pressure agreement with indirect wrist cuff measures. Bland-Altman plot analyses of pressure measurement agreement with: systolic and diastolic blood pressure wrist cuff measurements after peripheral algorithm (PA) adjustment and direct right radial artery (RRA) blood pressure measurements (N=20; mean=80), and systolic and diastolic blood pressure wrist cuff measurements after central algorithm (CA) adjustment and direct ascending aorta (AA) blood pressure measurements (N=20; mean=80).

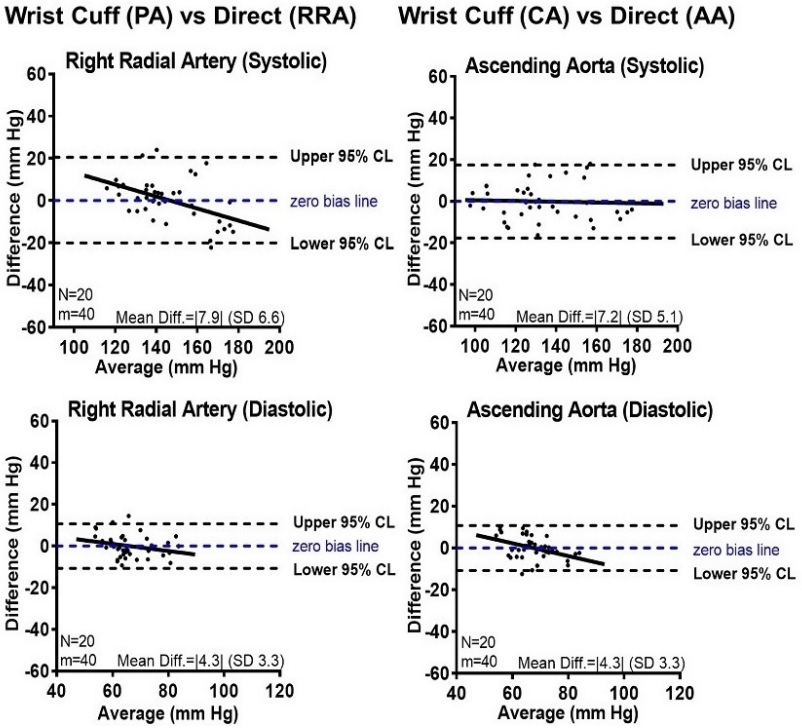


Figure 4. Simultaneous recordings of direct and indirect blood pressures. Representative illustration of instantaneous intra-arterial blood pressures every 10 s over 4 min (blue=systolic; red=diastolic) along with 2 simultaneous readings of wrist cuff blood pressure (green=systolic; purple=diastolic) with SD. Wrist cuff blood pressure is acquired over approximately 1 min (gray boxes).

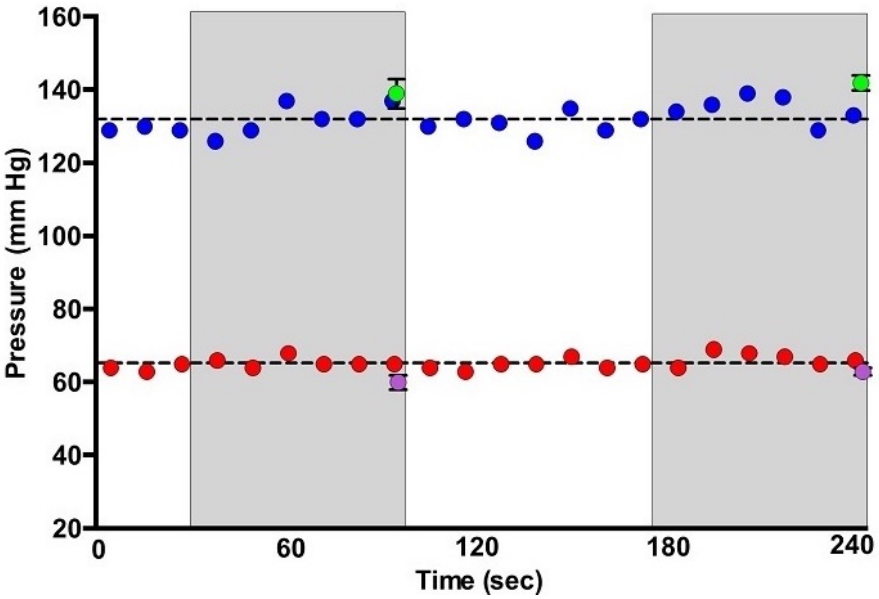
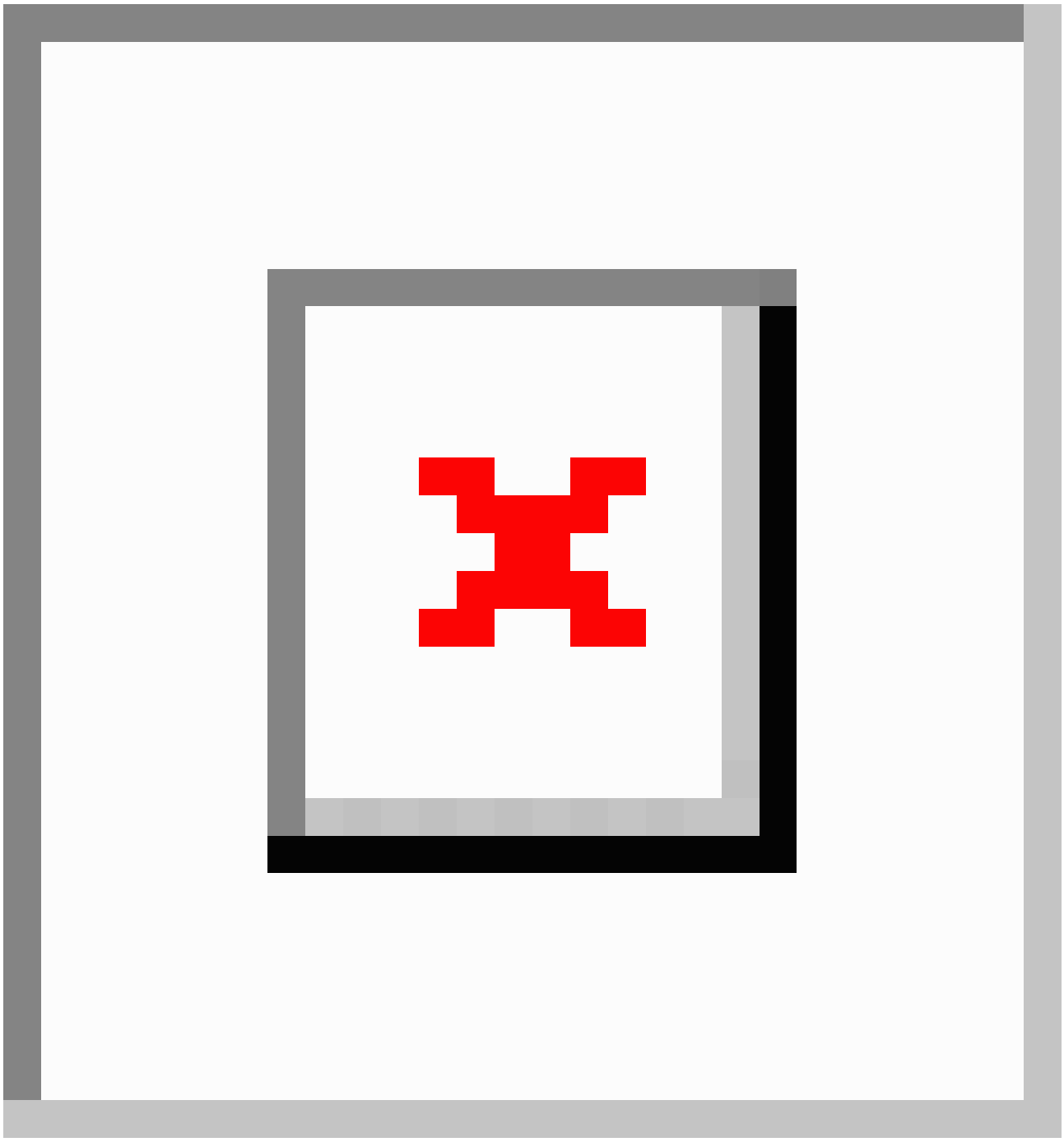


Figure 5. Relevant threshold comparison of peripheral and central pressures, directly and indirectly. Direct intra-arterial systolic pressure measurements from the right radial artery (magenta) and ascending aorta (green). Indirect wrist cuff systolic pressure measurements using the peripheral algorithm (black) and central algorithm (blue); upper arm measurements included with no possibility of a measure of uncertainty. All course of mean systolic measures (without SD) with guideline and trial target thresholds. Note: gray box shows upper arm measures at risk of treatment gap based on Canadian Hypertension Education Program (CHEP) and Systolic Blood Pressure Intervention Trial (SPRINT) thresholds.



Discussion

Principal Findings

In this study, we assessed the validity of indirect measures of a novel wrist cuff blood pressure device in direct comparison with simultaneously recorded gold standard intra-arterial pressures. This study shows that a calibrated wrist cuff blood pressure device is in agreement with the precision and accuracy of intra-arterial pressure measurements in an in-patient setting. To our knowledge, this study is the first to test and incorporate iterative calibration of a commercially available blood pressure device and also the first to show measurement uncertainty in the device output. This study shows that an automated upper arm device when compared with other indirect methods was less accurate than the gold standard calibration used here; this could reduce the risk of a treatment gap. A calibrated wrist cuff device, if used properly, is capable of producing measures that are clinically accurate for the management of systolic hypertension in accordance with the guidelines for home blood pressure monitoring. In our center, we identified a clinical need to use wrist cuff devices for patients in intensive care (eg,

minimize sleep disturbance) or with physical limitations to upper arm cuff use (eg, frailty, obesity). To achieve broad clinical reliability, additional use and monitoring of the device, such as in clinical trials, is warranted.

Clinical Relevance

This wrist cuff device offers several advantages in comparison with many upper arm devices for the measurement and medical management of hypertension. A common complaint among patients and front-line health care providers, especially when performing repetitive blood pressure monitoring, is that they experience pain with upper arm compression [30]. Frail patients are more likely to become physically intolerant to multiple daily measures with an arm cuff. Furthermore, these devices, when activated in hospitals at night, disrupt essential circadian sleep quality [31]. Obese, senior, and frail patients most often experience cuff malposition issues, either from obese, conical upper arms or from mobility or dexterity issues for self-fastening and positioning [30]. Patients with conical or obese arms and the frail elderly are key target demographics in need of hypertension identification and management.

Individual blood pressure measures from this wrist cuff device report the degree of uncertainty in each measure. While several mean blood pressure measurements were below the threshold for hypertension based on the CHEP guidelines [4] or the SPRINT trial target [29], they were within 1 SD of the threshold. This poses the question as to whether these patients would benefit from more aggressive therapeutic intervention. Future work should assess the frequency by which this occurs in an outpatient setting using a longitudinal study.

Calibration Issues With Indirect Oscillometric Devices: Peer Review, Interdependence on Therapeutics, and the Level of Uncertainty

The variability between radial and aortic pressures is a function of differences in compliance, vasoactivity, and pressure augmentation, which become more variable with aging and disease [25,28,32]. Thus, noninvasive blood pressure can be calibrated to either radial artery or aortic pressure. Specifically, a universal protocol that recommends intra-arterial pressure as a reference standard to validate noninvasive blood pressure is currently being developed via collaboration of the AAMI and the ESH [20]. While this wrist cuff blood pressure device was calibrated to aortic pressures [20,33-37] in this study, we also presented wrist cuff blood pressure data calibrated to either radial artery or aortic pressures for scientific interest. This natural intraphysiological variability from the mean (Figure 4, dashed line) is derived from variability of neuro-endocrine stimulation, arterial tone, heart rhythm, cardiac output, and respiration. Thus, the correct and accurate way to calculate blood pressure should include intrinsic measures of variance [38] (measure of uncertainty or SD). This was achieved for the first time with a blood pressure device that reported an SD for each measure.

Blood pressure devices are almost exclusively calibrated to manual auscultatory measurements [10-12,17]. In addition to the calibration inconsistency among blood pressure devices as previously described [18], unreliable measurements among automated blood pressure devices are common [39,40]. Most protocols for device comparison are focused on variability within a population rather than variability of blood pressure measures within an individual. Most automated devices currently do not report the uncertainty of the blood pressure measure. We also observed that upper arm measures were not in agreement with aortic pressures and could produce a treatment gap of approximately 20%. The need to align blood pressure measures with the medical management of hypertension is of paramount importance [41]. Additionally, many (medical or life) insurers have a vested financial interest in the reliability of blood pressure as an index of health in determining premiums and eligibility. Fundamentally, the management of hypertension and blood pressure devices are interdependent, yet a barrier to achieving optimal disease management is, at least in part, related to the current lack of information about their precision, accuracy, and level of uncertainty. Indeed, systematic review of noninvasive blood pressure devices reportedly meeting the engineering standards of either the AAMI, ESH, or British Hypertension Society protocols [19,42,43] are inconsistently adhered to and are not always in agreement [18]. Often, incongruent variables are reported in the Bland-Altman analyses that invariably lead to a mean difference of 0 with increasing n-values, which is statistically unacceptable [27], whereas clauses in the protocol can allow for removal of potentially relevant measures (ie, 12/8 rule) [44]. Furthermore, engineering and clinical standards are not comparable—the former is concerned with device reproducibility, while the latter is concerned with interpatient and intrapatient variability. Regulatory agencies (eg, Health Canada, FDA) that license these devices are responsible only for aspects relating to product safety and the comparability to other market devices and not for determining the validity of the measurements, [16,17]. More efforts are required to advance device quality and functionality using a patient-centered approach to accommodate the interdependencies between blood pressure devices and (clinical trial-approved) therapeutics for the management of hypertension [8,9]. A universal protocol that is clinically practical and can consistently determine device validity, including when challenged by a direct pressure analysis, is currently being developed and estimated to be released in 2018 [20].

Study Limitations and the Impact of Precise and Accurate Blood Pressure Measures on Current Clinical Guidelines

Further calibration studies using this device should include special populations such as pediatric, obese, or frail patients, or those with an underlying arrhythmia or peripheral vascular disease. Also, patients with peripheral movement artifacts (eg, tremors) represent an incremental challenge to wrist cuff devices that would be less pronounced for upper arm cuff devices. Finally, more work is required to determine whether diastolic pressure and diagnoses using indirect diastolic pressure measures are required to be accurate to direct central pressures.

To diagnose hypertension according to guidelines, we must be confident with device measures [45,46]. It is encouraging that large clinical trials have demonstrated improved outcomes in populations achieving even 1 mm Hg reductions in blood

pressure, yet devices have no provision for this single digit resolution for an individual patient. Guidelines tend to utilize 5 mm Hg increments and we are not aware of any device that has this resolution without applying variability filtering [44]. Indeed, intrapatient variability of blood pressure within a short period of time can vary by increments of >10 mm Hg, notwithstanding intraday variability of 25 mm Hg or more. This not only necessitates serial averaging of blood pressure measures but also compounds the standard error in the mean of measures. A greater awareness of uncertainty could help to establish criteria of acceptability in a measure. Now, the question, given the recent results of the SPRINT trial [29], whether the lowering of the blood pressure target was achieved by having patients at 120 mm Hg or the consequence of greater certainty in patient catchment below the existing 135 mm Hg threshold will have to be faced. Future clinical trials should report the devices used for blood pressure analysis and their level of uncertainty. This will provide an opportunity to factor in-device uncertainty as guidelines are further refined.

Conclusions

The wrist cuff calibrated here to the gold standard—the central (ascending aortic) pressure—is an accurate device that can be used in accordance with guidelines for informed decision making in the management of systolic hypertension.

Acknowledgments

The following funding sources were pertinent to this study: New Brunswick Health Research Foundation and Health+Life Sciences New Brunswick Inc (KRB and SL) and The Natural Sciences and Engineering Research Council of Canada (KRB and JAS). JAS is a New Investigator with the Heart and Stroke Foundation of Canada. Devices and technical support were provided in-kind by Cloud Diagnostics Inc. Cloud Diagnostics Inc contracted Health+Life Sciences New Brunswick Inc to fund, at arms-length, this study with partnership funding provided by the New Brunswick Health Research Foundation. The authors are grateful to Dr Fong, Dr Hubacek, Dr Barry, and Dr Paddock for their support in recruitment and data collection and Alexandra Yip and Dr Stoica for helpful discussions. We are also appreciative of the patients, nurses, and nurse associates of The New Brunswick Heart Centre.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Assessment of bilateral blood pressure equality.

[JPG File, 930KB - jmir_v20i4e111_app1.jpg]

Multimedia Appendix 2

Normality analysis of acquired BP.

[JPG File, 804KB - jmir_v20i4e111_app2.jpg]

Multimedia Appendix 3

Correlational, actual indirect-direct error trend, paired simultaneous, and normalized % ratio difference comparative analysis of measured direct and indirect data BP.

[JPG File, 1MB - jmir_v20i4e111_app3.jpg]

Multimedia Appendix 4

Device initial algorithm vs intra-arterial pressure measures.

[JPG File, 302KB - jmir_v20i4e111_app4.jpg]

Multimedia Appendix 5

Device initial and central algorithms vs auscultatory pressure measures.

[JPG File, 326KB - jmir_v20i4e111_app5.jpg]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. Lancet 2017 Dec 7;389(10064):37-55 [FREE Full text] [doi: 10.1016/S0140-6736(16)31919-5] [Medline: 27863813]

2. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
3. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet 2005;365(9455):217-223. [doi: 10.1016/S0140-6736(05)17741-1] [Medline: 15652604]
4. Leung AA, Daskalopoulou SS, Dasgupta K, McBrien K, Butalia S, Zarnke KB, Hypertension Canada. Hypertension Canada's 2017 guidelines for diagnosis, risk assessment, prevention, and treatment of hypertension in adults. Can J Cardiol 2017 May;33(5):557-576. [doi: 10.1016/j.cjca.2017.03.005] [Medline: 28449828]
5. Weber MA, Schiffrin EL, White WB, Mann S, Lindholm LH, Kenerson JG, et al. Clinical practice guidelines for the management of hypertension in the community a statement by the American Society of Hypertension and the International Society of Hypertension. J Hypertens 2014 Jan;32(1):3-15. [doi: 10.1097/HJH.0000000000000065] [Medline: 24270181]
6. Authors/Task Force Members, Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Additional Contributor: Simone Binno (Italy), Document Reviewers, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts): Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur J Prev Cardiol 2016 Jul;23(11):NP1-N96. [doi: 10.1177/2047487316653709] [Medline: 27353126]
7. Dimitrov DV. Medical internet of things and big data in healthcare. Healthc Inform Res 2016 Jul;22(3):156-163 [FREE Full text] [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
8. Campbell NR, Gelfer M, Stergiou GS, Alpert BS, Myers MG, Rakotz MK, et al. A call to regulate manufacture and marketing of blood pressure devices and cuffs: a position statement from the World Hypertension League, International Society of Hypertension and Supporting Hypertension Organizations. J Clin Hypertens (Greenwich) 2016 May;18(5):378-380 [FREE Full text] [doi: 10.1111/jch.12782] [Medline: 26852890]
9. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. N Engl J Med 2014 Jul 24;371(4):372-379. [doi: 10.1056/NEJMhle1403384] [Medline: 25054722]
10. Deutsch C, Krüger R, Saito K, Yamashita S, Sawanoi Y, Beime B, et al. Comparison of the Omron RS6 wrist blood pressure monitor with the positioning sensor on or off with a standard mercury sphygmomanometer. Blood Press Monit 2014 Oct;19(5):306-313. [doi: 10.1097/MBP.0000000000000063] [Medline: 25000543]
11. Omboni S, Riva I, Giglio A, Caldara G, Groppelli A, Parati G. Validation of the Omron M5-I, R5-I and HEM-907 automated blood pressure monitors in elderly individuals according to the International Protocol of the European Society of Hypertension. Blood Press Monit 2007 Aug;12(4):233-242. [doi: 10.1097/MBP.0b013e32813fa386] [Medline: 17625396]
12. Takahashi H, Yoshika M, Yokoi T. Validation of Omron RS8, RS6, and RS3 home blood pressure monitoring devices, in accordance with the European Society of Hypertension International Protocol revision 2010. Vasc Health Risk Manag 2013;9:265-272 [FREE Full text] [doi: 10.2147/VHRM.S44569] [Medline: 23745050]
13. Alihanoglu YI, Kayrak M, Ulgen MS, Yazici M, Yazici M, Yilmaz R, et al. The impact of central blood pressure levels on the relationship between oscillometric and central blood pressure measurements: a multicenter invasive study. J Clin Hypertens (Greenwich) 2013 Sep;15(9):681-686 [FREE Full text] [doi: 10.1111/jch.12166] [Medline: 24034662]
14. Mireles SA, Jaffe RA, Drover DR, Brock-Utne JG. A poor correlation exists between oscillometric and radial arterial blood pressure as measured by the Philips MP90 monitor. J Clin Monit Comput 2009 Jun;23(3):169-174. [doi: 10.1007/s10877-009-9178-8] [Medline: 19396553]
15. Umana E, Ahmed W, Fraley MA, Alpert MA. Comparison of oscillometric and intraarterial systolic and diastolic blood pressures in lean, overweight, and obese patients. Angiology 2006;57(1):41-45. [doi: 10.1177/0003319706605700106] [Medline: 16444455]
16. Minister of Justice of Canada. 2016 Mar 16. Consolidation: medical device regulations, SOR/98-282 URL: http://laws-lois.justice.gc.ca/ [WebCite Cache ID 6x5MlzlXi]
17. U.S. Department of Health and Human Services Food and Drug Administration. 2014. The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)] URL: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM284443.pdf [WebCite Cache ID 6xQuKtL6X]
18. Wan Y, Heneghan C, Stevens R, McManus RJ, Ward A, Perera R, et al. Determining which automatic digital blood pressure device performs adequately: a systematic review. J Hum Hypertens 2010 Jul;24(7):431-438 [FREE Full text] [doi: 10.1038/jhh.2010.37] [Medline: 20376077]
19. Association for the Advancement of Medical Instrumentation. 2013. ANSI/AAMI/ISO 81060-2 URL: http://www.aami.org/ [WebCite Cache ID 6xQukpofO]
20. Sharman JE, Avolio AP, Baulmann J, Benetos A, Blacher J, Blizzard CL, et al. Validation of non-invasive central blood pressure devices: ARTERY Society task force consensus statement on protocol standardization. Eur Heart J 2017 Oct 1;38(37):2805-2812. [doi: 10.1093/eurheartj/ehw632] [Medline: 28158489]
21. Miller RG. The jackknife--a review. Biometrika 1974;61(1):1-15.

22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
23. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 1995 Oct 21;346(8982):1085-1087. [Medline: 7564793]
24. Mancia G, Ferrari A, Gregorini L, Parati G, Pomidossi G, Bertinieri G, et al. Blood pressure and heart rate variabilities in normotensive and hypertensive human beings. Circ Res 1983 Jul;53(1):96-104 [FREE Full text] [Medline: 6861300]
25. Avolio AP, Xu K, Butlin M. Application of cardiovascular models in comparative physiology and blood pressure variability. Conf Proc IEEE Eng Med Biol Soc 2013;2013:217-220. [doi: 10.1109/EMBC.2013.6609476] [Medline: 24109663]
26. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]
27. Dewitte K, Fierens C, Stöckl D, Thienpont LM. Application of the Bland-Altman plot for interpretation of method-comparison studies: a critical investigation of its practice. Clin Chem 2002 May;48(5):799-801 [FREE Full text] [Medline: 11978620]
28. Munir S, Guilcher A, Kamalesh T, Clapp B, Redwood S, Marber M, et al. Peripheral augmentation index defines the relationship between central and peripheral pulse pressure. Hypertension 2008 Jan;51(1):112-118 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.107.096016] [Medline: 17998476]
29. SPRINT Research Group, Wright Jr JT, Williamson JD, Whelton PK, Snyder JK, Sink KM, et al. A randomized trial of intensive versus standard blood pressure control. N Engl J Med 2015 Nov 26;373(22):2103-2116 [FREE Full text] [doi: 10.1056/NEJMoa1511939] [Medline: 26551272]
30. Boubouchairopoulou N, Kollias A, Chiu B, Chen B, Lagou S, Anestis P, et al. A novel cuffless device for self-measurement of blood pressure: concept, performance and clinical validation. J Hum Hypertens 2017 Jul;31(7):479-482. [doi: 10.1038/jhh.2016.101] [Medline: 28124684]
31. Yoder JC, Yuen TC, Churpek MM, Arora VM, Edelson DP. A prospective study of nighttime vital sign monitoring frequency and risk of clinical deterioration. JAMA Intern Med 2013 Sep 9;173(16):1554-1555 [FREE Full text] [doi: 10.1001/jamainternmed.2013.7791] [Medline: 23817602]
32. Sleight P, La Rovere MT, Mortara A, Pinna G, Maestri R, Leuzzi S, et al. Physiology and pathophysiology of heart rate and blood pressure variability in humans: is power spectral analysis largely an index of baroreflex gain? Clin Sci (Lond) 1995 Jan;88(1):103-109. [Medline: 7677832]
33. Kollias A, Lagou S, Zeniodi ME, Boubouchairopoulou N, Stergiou GS. Association of central versus brachial blood pressure with target-organ damage: systematic review and meta-analysis. Hypertension 2016 Jan;67(1):183-190 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.115.06066] [Medline: 26597821]
34. Stergiou GS, Parati G, Vlachopoulos C, Achimastos A, Andreadis E, Asmar R, et al. Methodology and technology for peripheral and central blood pressure and blood pressure variability measurement: current status and future directions - position statement of the European Society of Hypertension Working Group on blood pressure monitoring and cardiovascular variability. J Hypertens 2016 Sep;34(9):1665-1677. [doi: 10.1097/HJH.0000000000000969] [Medline: 27214089]
35. Williams B, Lacy PS, Thom SM, Cruickshank K, Stanton A, Collier D, CAFE Investigators, Anglo-Scandinavian Cardiac Outcomes Trial Investigators, CAFE Steering Committee and Writing Committee. Differential impact of blood pressure-lowering drugs on central aortic pressure and clinical outcomes: principal results of the Conduit Artery Function Evaluation (CAFE) study. Circulation 2006 Mar 7;113(9):1213-1225 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.105.595496] [Medline: 16476843]
36. Miyashita H. Clinical assessment of central blood pressure. Curr Hypertens Rev 2012 May;8(2):80-90 [FREE Full text] [doi: 10.2174/157340212800840708] [Medline: 22866025]
37. McEniery CM, Cockcroft JR, Roman MJ, Franklin SS, Wilkinson IB. Central blood pressure: current evidence and clinical importance. Eur Heart J 2014 Jul;35(26):1719-1725 [FREE Full text] [doi: 10.1093/eurheartj/eht565] [Medline: 24459197]
38. Appel ML, Berger RD, Saul JP, Smith JM, Cohen RJ. Beat to beat variability in cardiovascular variables: noise or music? J Am Coll Cardiol 1989 Nov 1;14(5):1139-1148 [FREE Full text] [Medline: 2681319]
39. Stergiou GS, Lourida P, Tzamouranis D, Baibas NM. Unreliable oscillometric blood pressure measurement: prevalence, repeatability and characteristics of the phenomenon. J Hum Hypertens 2009 Dec;23(12):794-800. [doi: 10.1038/jhh.2009.20] [Medline: 19322203]
40. Koohi I, Batkin I, Groza VZ, Shirmohammadi S, Dajani HR, Ahmad S. Metrological characterization of a method for blood pressure estimation based on arterial lumen area model. IEEE Trans Instrum Meas 2017 Apr;66(4):734-745. [doi: 10.1109/TIM.2017.2657978]
41. Kumar N, Khunger M, Gupta A, Garg N. A content analysis of smartphone-based applications for hypertension management. J Am Soc Hypertens 2015 Feb;9(2):130-136. [doi: 10.1016/j.jash.2014.12.001] [Medline: 25660364]
42. O'Brien E, Atkins N, Stergiou G, Karpettas N, Parati G, Asmar R, Working Group on Blood Pressure Monitoring of the European Society of Hypertension. European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults. Blood Press Monit 2010 Feb;15(1):23-38. [doi: 10.1097/MBP.0b013e3283360e98] [Medline: 20110786]
43. O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, Altman DG, et al. The British Hypertension Society protocol for the evaluation of blood pressure measuring devices. J Hypertens 1993;11(suppl 2):S43-S62. [Medline: 2168451]

44. Padwal RS, McLean D, Ringrose J. Should patients with higher blood pressure variability be excluded from validation studies? An assessment of the '12/8' rule. Blood Press Monit 2017 Apr;22(2):112-114. [doi: 10.1097/MBP.0000000000000235] [Medline: 28072598]
45. Myers MG. Limitations of home blood pressure monitoring in clinical practice. Can J Cardiol 2015 May;31(5):583-584. [doi: 10.1016/j.cjca.2015.03.008] [Medline: 25843166]
46. Jones DW, Weatherly L, Hall JE. SPRINT: what remains unanswered and where do we go from here? Hypertension 2016 Feb;67(2):261-262 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.115.06723] [Medline: 26553235]

Abbreviations

AAMI: Association for the Advancement of Medical Instrumentation
CHEP: Canadian Hypertension Education Program
ESH: European Society of Hypertension
FDA: Food and Drug Administration
SPRINT: Systolic Blood Pressure Intervention Trial

Edited by G Eysenbach; submitted 10.05.17; peer-reviewed by D Dimitrov, S Omboni; comments to author 15.08.17; revised version received 10.11.17; accepted 16.11.17; published 25.04.18.

Please cite as:
Melville S, Teskey R, Philip S, Simpson JA, Lutchmedial S, Brunt KR
A Comparison and Calibration of a Wrist-Worn Blood Pressure Monitor for Patient Management: Assessing the Reliability of Innovative Blood Pressure Devices
J Med Internet Res 2018;20(4):e111
URL: https://www.jmir.org/2018/4/e111/
doi:10.2196/jmir.8009
PMID:29695375

Review

Web-Based Patient Education in Orthopedics: Systematic Review

Tessa Dekkers[1], MSc; Marijke Melles[1], PhD; Bob Sander Groeneveld[1], MSc; Huib de Ridder[1], PhD

Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

Corresponding Author:
Tessa Dekkers, MSc
Faculty of Industrial Design Engineering
Delft University of Technology
Landbergstraat 15
Delft, 2628 CE
Netherlands
Phone: 31 15 27 84640
Email: t.dekkers@tudelft.nl

Abstract

Background: Patients with orthopedic conditions frequently use the internet to find health information. Patient education that is distributed online may form an easily accessible, time- and cost-effective alternative to education delivered through traditional channels such as one-on-one consultations or booklets. However, no systematic evidence for the comparative effectiveness of Web-based educational interventions exists.

Objective: The objective of this systematic review was to examine the effects of Web-based patient education interventions for adult orthopedic patients and to compare its effectiveness with generic health information websites and traditional forms of patient education.

Methods: CINAHL, the Cochrane Library, EMBASE, MEDLINE, PsycINFO, PUBMED, ScienceDirect, Scopus, and Web of Science were searched covering the period from 1995 to 2016. Peer-reviewed English and Dutch studies were included if they delivered patient education via the internet to the adult orthopedic population and assessed its effects in a controlled or observational trial.

Results: A total of 10 trials reported in 14 studies involving 4172 patients were identified. Nine trials provided evidence for increased patients' knowledge after Web-based patient education. Seven trials reported increased satisfaction and good evaluations of Web-based patient education. No compelling evidence exists for an effect of Web-based patient education on anxiety, health attitudes and behavior, or clinical outcomes.

Conclusions: Web-based patient education may be offered as a time- and cost-effective alternative to current educational interventions when the objective is to improve patients' knowledge and satisfaction. However, these findings may not be representative for the whole orthopedic patient population as most trials included considerably younger, higher-educated, and internet-savvy participants only.

(J Med Internet Res 2018;20(4):e143) doi:10.2196/jmir.9013

KEYWORDS

patient education as topic; health education; orthopedics; internet; humans; preoperative care; patient satisfaction

Introduction

Background

Patient education is a valuable part of care that enables patients to be informed, active participants in their own treatment [1-3]. Traditionally, it is provided through face-to-face teaching methods by health care professionals (HCPs) [3-5]. These methods are often supplemented with written booklets or pamphlets [4,6], or multimedia channels such as audiotapes, digital versatile disc, and video [7,8]. However, as both internet access and the availability of health information on public websites increases, it is now common for patients to also use the internet to learn about health and illness [9]. People with orthopedic conditions such as osteoarthritis, rheumatic arthrosis, or trauma form no exception to this trend. Internet use among this group increases rapidly: 79% of patients had internet access in 2012, and among them, 23% in 2010 to 65% in 2012 had used the internet to research their orthopedic condition or upcoming treatment [10,11].

Patients themselves are positive about using the internet to find health information. They perceive online health information to produce health benefits and social benefits (eg, improved self-care behavior and better social support) in a manner that is easily accessible, cost-effective, and time-effective [12]. Reactions of HCPs, however, have been mixed. It is recognized that health information that is distributed online can incorporate unique features such as tailored information, multimedia, and interactivity to keep patients engaged with the educational material [12,13]. For example, McKay and colleagues incorporated interactive elements in their internet-based diabetes self-management support intervention by allowing patients to live chat with each other and HCPs [14]. That such elements can ultimately enhance the education's effectiveness is demonstrated, for example, in the fields of breast cancer and general surgery: Web-based patient education increases patients' knowledge and satisfaction [15,16], improves the physician-patient relationship [17], and creates awareness about health issues in the general population [18]. Despite these initial successes, concerns with Web-based education have been voiced in orthopedic practice as well. Most of these stress the poor quality of online health information, which is deemed overly commercialized and poorly readable even when produced by qualified HCPs [19-21]. Furthermore, despite increasing internet access in the population as a whole, clinicians fear the generalizability of previous findings to elderly patients who may be inexperienced with internet usage [13,17,22]. To acknowledge these potential downsides while meeting patients' demands for online patient education, it is important to systematically examine and evaluate the effects of Web-based educational interventions that are currently in place.

This review follows the definition of Roter and colleagues [23] in defining educational interventions as "pedagogic interventions, verbal or written, with a knowledge-based emphasis designed to convey information." This distinguishes educational interventions from behavioral and affective interventions that focus on shaping behavioral patterns and appealing to feelings and emotions, respectively. The core aim of educational interventions is knowledge acquisition by patients [8,18,19,22]. With knowledge, the patient can participate in decision making and build skills for self-care [20]. In this way, increased knowledge can result in better clinical outcomes and ultimately improve the patient' quality of life [24].

Web-Based Patient Education in Comparison With Traditional Patient Education

When evaluating Web-based patient education, it is inevitable to compare its effectiveness with that of traditional patient education. Therefore, the first aim of this review was to compare the effectiveness of Web-based patient education with the more traditional methods for patient education, such as face-to-face teachings or the use of print materials. To make an accurate comparison between the two, we will provide a brief overview of the effectiveness of traditional patient education as identified in previous systematic reviews below.

In orthopedic practice, positive effects following traditional patient education include increased knowledge regarding surgical procedures and the informed consent process, improved self-management skills, and reduced length of stay [25-28]. Yet, educational interventions are no more effective than other interventions such as attention control or physiotherapy [29]. Furthermore, clinical outcomes such as pain and functioning do not improve following patient education [26,28,30], just as patient education also does not decrease anxiety in a clinically meaningful way [26,28]. Finally, there is insufficient evidence currently available to determine the effect of education on patients' empowerment and self-efficacy [26], and no systematic reviews have examined the effect of patient education on patient satisfaction. From these findings, we hypothesize the following:

> Hypothesis 1: Web-based patient education interventions have a positive effect on patients' knowledge, but not on anxiety or clinical outcomes.

Web-Based Patient Education in Comparison With Generic Health Information Websites

As outlined earlier in this introduction, educational interventions are no longer the sole source of knowledge for patients, as an abundance of health information is also freely available on the internet. When patients make use of generic health information while included in the experimental arm of a Web-based patient education intervention trial, online health information forms a potential strong co-intervention [13]. Thus, to accurately evaluate Web-based patient education, it is important to not only compare its effect with that of traditional interventions but also with that of public health information websites. Therefore, the second aim of this study was to compare Web-based patient education interventions with health information websites.

Health information websites are often broader in scope than educational interventions, as they typically target the general population as well as patients, whereas patient education targets patients or other members of the health care system only [31]. This means these websites are also unlikely to involve HCPs, or make use of clinical measurements or other information about patients that is derived from the health care system. Furthermore, health information websites are generally not theory-based. In contrast, patient education interventions are often developed and implemented using various theoretical frameworks [32]. Although we recognize that use of theory in intervention development is varied and may be absent from some patient education interventions as well [31-33], embedment of theory in general does set apart educational interventions from generic health information websites. Therefore, we expect and hypothesize the following:

> Hypothesis 2: theory-based, or professionally facilitated, Web-based patient education interventions perform better than generic health information websites.

Review Objective

Concluding, promising results of Web-based patient education interventions have been reported, but a systematic review of Web-based patient education specifically for orthopedic practice has not yet been carried out. The effects of Web-based patient education can be evaluated in itself but should also be compared with other interventions currently in place: (1) to traditional patient education interventions that are theory-based or

professionally facilitated but are provided through different channels (such as verbally, written, or by using multimedia) and (2) to publicly accessible, generic health information websites that share the same channel of information provision (the internet) but are generally not theory-based or professionally facilitated. The overall aim of this systematic review was to tackle these comparisons by examining the effects of Web-based patient education interventions on patients with orthopedic conditions as reported in controlled and observational trials in comparison with traditional patient education and health information websites. The questions that guided us in examining the comparative effectiveness were as follows: (1) "what are the effects of Web-based patient education on adult patients with orthopedic conditions?" and (2) "what are the effects of Web-based patient education in comparison with the effects of traditional patient education and generic health information websites?"

Methods

Protocol and Registration

This systematic review has been written according to the requirements of the Preferred Reporting Items for Systematic Review and Meta-Analyses statement [34,35]. The review's protocol has not been published.

Eligibility Criteria

We included peer-reviewed, controlled, and observational trials reported in English or Dutch that self-defined as studying the effects of patient education interventions delivered via an online environment, including mobile devices, websites, and online systems, to adult people with any orthopedic illness or condition and currently receiving treatment for such conditions. Following our definition of educational interventions, we excluded behavioral or affective interventions. These may include educational components but differ from educational interventions as they specifically target behavioral patterns or appeal to feelings or social relationships to change patients' outcomes [23]. As our focus lay with studying interventions, we did not include studies that only discussed generic, not theory-based, not professionally facilitated health information websites and did not compare their effectiveness with Web-based patient interventions. No mandatory principal outcomes were defined for studies to be eligible for inclusion in the review. No restrictions on publication date were imposed in the search for eligible studies. However, in the final selection of studies, we excluded studies that were published before 1995 to ensure the review represented current evidence.

Information Sources

Studies were initially identified by searching the electronic databases CINAHL, Cochrane Central Register of Controlled Trials, EMBASE, MEDLINE, PsycINFO, PubMed, ScienceDirect, Scopus, and Web of Science from September 1, 2015 to November 30, 2015. As an example, the search strategy for the PubMed database can be found in Textbox 1. Search strategies for the other databases are available in Multimedia Appendix 1. The search was repeated in September 2017 to ensure the latest evidence was included. This search strategy was complemented by reviewing the bibliographies of included studies to identify additional studies of interest. We contacted one author for a full-text copy of an eligible study that was subsequently provided to the review team. For all other articles, full-text copies were available, and no further contact with the original authors was made.

Study Selection

The first author assessed the identified studies for eligibility by title and abstract. The predefined selection criteria were applied to full-text reports of potentially eligible studies primarily by the first author in discussion with two review authors (MM and HdR) until consensus was reached. A third review author (BSG) was available for arbitration, but this was not required.

Data Collection Process

A structured data extraction sheet was employed to extract data from included studies. The data extracted included (1) Study characteristics (ie, author, year of publication, design, population, and timing of outcome measures); (2) Intervention characteristics (ie, content and duration of intervention and control intervention, total sample size, and sample sizes in separate conditions); (3) Patient characteristics (ie, sociodemographic variables, health status, and experience with internet); and (4) outcomes (ie, type of outcome measure, instrument, and effect). For each study, the effect of the intervention was coded as (1) significant result (positive + or negative −), (2) nonsignificant result (=); or (3) not reported (×).

To provide a structured overview of the components in each intervention, we employed Barak and colleagues' [36] framework for internet-supported interventions. This framework provides guiding definitions for four components that make up a Web-based education intervention, including (1) Program content (educational or behavior change content), (2) Multimedia use (type of media used to convey program content), (3) Interactive online activities (activities offered to increase patient interest, understanding, and engagement), and (4) Guidance and supportive feedback (if and how patients can obtain automated or human support and feedback).

Textbox 1. PubMed search strategy for the identification of studies assessing the effects of Web-based patient education interventions for the adult orthopedic population.

(internet OR "world wide web" OR online OR web-based OR "computer assisted" OR e-health OR network OR "web services") AND ("patient education" OR "patient education as topic" [MeSH Terms] OR "consumer health informati*" OR "medical education" OR "health education" OR "health knowledge, attitudes, practice"[MeSH Terms]) AND (orthopedic* OR orthopaedic* OR "joint replacement" or "arthroplasty" OR "hip" OR "knee") AND (Adult OR Aged) AND (Effect OR efficacy OR performance OR result OR outcome)

Risk of Bias in Individual Studies

To appraise the risk of bias in included studies, data regarding reporting, external validity, internal validity, and statistical power were extracted independently by two review authors (TD and BSG) using a modified version of Downs and Black tool for assessment of methodological quality [37]. This tool was selected for its high internal consistency and reliability and its applicability to both randomized and observational studies [37,38]. In line with previous studies, the ambiguous item regarding statistical power was modified to indicate the presence of a statistical power analysis or sample group calculation by allocating 1 (present) or 0 (absent) points [39-41]. The range of the modified tool is 0 to 28, with higher scores indicating higher methodological quality. Studies were not excluded on the basis of their methodological quality; however, findings from medium- and poor-quality studies were given less weight in the qualitative synthesis than studies of high methodological quality.

Synthesis of Results

We examined the effectiveness of Web-based patient education interventions by describing and comparing the characteristics and results of the included studies, as summarized in the structured data extraction sheet (*see Data Collection Process*) through qualitative synthesis [42]. No meta-analysis was attempted because of the small number of included studies and considerable variability in the outcome measures employed.

Results

Study Selection

The search identified 1032 eligible studies of which 10 trials, reported in 14 papers, met the inclusion criteria and were included in the review (Figure 1). Five of the included studies [43-47] concern separate reports of the same trial. To account for potential inconsistencies in reporting, all five reports of the trial were included in the review [35].

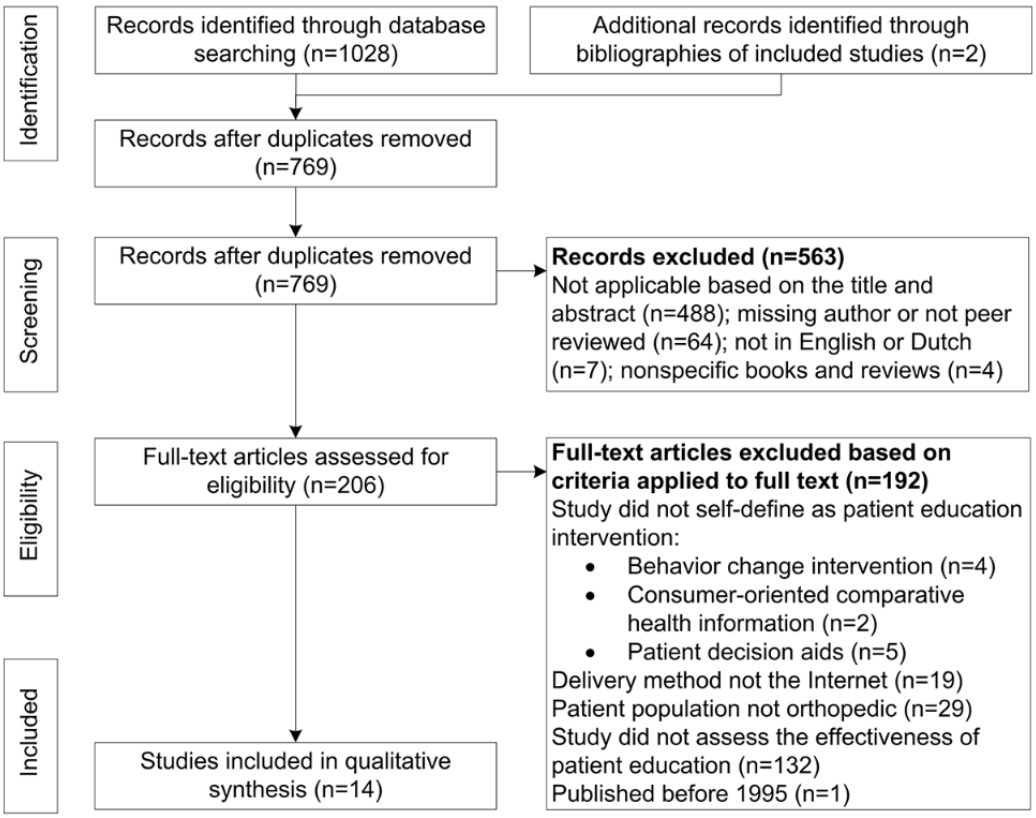
Study Characteristics

Seven of the 10 trials employed a randomized controlled design, two an observational design, and one a quasi-experimental design. Four trials assessed the effect of Web-based patient education in comparison with traditional patient education channels, including face-to-face education with a nurse or physician [43-49] and patient information sheets [50]. Three trials compared Web-patient education with health information websites [51-53], and three assessed the interventions' effects but did not compare these with either traditional patient education or health information websites [54-56].

Patient Characteristics

Most of the studies provided Web-based patient education to patients undergoing surgical treatment, including total knee arthroplasty [48,50], total hip arthroplasty [48,50,55], knee arthroscopy [43-49], shoulder arthroscopy [43-48], anterior cruciate ligament reconstruction [48], and unspecified ambulatory orthopedic surgery [52].

Figure 1. Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) flow diagram presenting identification and selection of articles for the systematic review of effectiveness of Web-based patient education in orthopedics.



Two studies provided Web-based patient education to patients with chronic conditions, including: rheumatoid arthritis [54] and osteoarthritis [56]. Two studies provided Web-based patient education to populations at risk for orthopedic conditions such as osteoporosis [51] and hip fracture [53]. The mean age of participants across studies was 56.3 years, and the sample was predominantly female (average 71.3% females in studies reporting gender). Most studies (70%) reported access to the internet as an explicit inclusion criterion, and some also required participants to also have an unspecified level of comfortableness [49] or skill [43-47,52,53] in using the internet.

Intervention Characteristics

The intervention characteristics of all included studies are described in Multimedia Appendix 2. Most interventions consisted of a single website that was developed specifically for study purposes, whereas 1 study provided patient education by sharing multiple websites that are publically available [50]. We did not identify any studies that used mobile devices for patient education.

Program content was specific to each intervention. Most interventions offered practical information about the orthopedic condition or treatment, such as the procedures planned for the day of surgery or instructions for postoperative monitoring [43-50,52,55,56]. Others focused on providing information regarding behavioral determinants [51,52] and local health care services [54]. Only 2 studies explicitly reported using content that was not primarily educational: patient testimonials or narratives [53,56].

Half of the interventions conveyed content in a moderate to highly dynamic manner, meaning that they used three or more multimedia formats such as text, pictures, videos, animations, or audio [43-47,51,53,55,56]. The other interventions primarily used text and pictures to convey the content. We did not find consistent evidence for the obvious assumption dynamic multimedia use increases the intervention's success. For example, both the static Orthoanswer website (primarily text) used by Fraval and colleagues [48] and the highly dynamic social cognitive theory website (text, graphics, audio, animation, and video) of Nahm and colleagues [56] increased patients' knowledge. On the other hand, the similarly dynamic website of Drieling and colleagues [51] did not do so.

While half of the interventions could be considered dynamic in terms of multimedia use, only one provided highly dynamic activities (meaning, more than three interactive online activities were offered) [51]. Activities offered to the patient on the dynamic Bone Health Improvement Project website included problem-solving exercises, goal-setting exercises, and self-assessment. Among the more static websites, self-assessment was the most common interactive activity [53,56]. Due to the limited use of interactive online activities, we were not able to assess the influence interactivity might have on patient outcomes.

Most websites offered some human support or feedback as part of the intervention [43-47,51,53,54,56]. Examples of extensive support include a moderated message board [53] and highly tailored automated feedback [51,56]. Other interventions offered fairly limited support by only sharing contact details of a nurse or other health professional [43-47,54]. Again, there was no clear evidence that the level of support or feedback provided had an influence on the interventions' success.

In terms of duration and frequency of website usage, we observed considerable variation. This ranged from single 20-min visits [49] to 18 repeated 60- to 90-min visits over the course of 6 months [51]. As duration and frequency were not consistently reported, we were not able to assess a dose-response relationship between usage of a Web-based intervention and outcomes.

Methodological Quality of Included Studies

The methodological quality of the studies was moderate, based on a mean Downs and Black score of 17.67±5.42 out of 28 (Table 1) [37,38]. Most studies adequately reported intervention and sample characteristics, but the external validity was often problematic, as was the lack of power analyses.

Outcome Measures of Included Studies

Most studies assessed knowledge acquisition [44,45,48-51,53-56] (90% of trials) and patient satisfaction, sometimes through qualitative feedback [46,48,49,51-53,55] (70% of trials). Other reoccurring outcome measures included anxiety [48-50], functional outcomes [45,52], and self-efficacy [46,51,53].

Many studies employed custom instruments that were designed by the researchers to assess the outcomes of their specific intervention. This resulted in a broad assortment of instruments that are difficult to interpret and compare (Table 2). To illustrate this diversity, consider instruments used to assess knowledge acquisition. Only one validated instrument (the Osteoporosis Health Belief Survey) was used in more than one study [51,53]. Four other studies also employed validated instruments, but not the same ones, as the topics of study (informed consent, anesthesia, and empowerment) differed considerably [44,48,50,56]. Four other studies employed instruments that had been developed specifically for each intervention, though the authors had pilot-tested or used these before [44,45,53,54]. Finally, 2 studies did not report anything regarding the validity or testing of their custom instruments [49,55].

The Effects of Web-Based Patient Education Interventions in Orthopedics

A summary of the effects of Web-based patient education interventions is provided in Multimedia Appendix 3.

Table 1. Methodological quality of included studies (ordered by quality).

Study	Downs and Black [37] subscales[a]					
	Reporting	External validity	Bias	Confounding	Power	Overall study quality[b]
Heikkinen et al [44]	10	1	5	6	1	High
Fraval et al [48]	8	2	5	6	1	High
Drieling et al [51]	10	1	5	5	0	High
Nahm et al [53]	9	1	6	4	1	High
Heikkinen et al [43]	10	1	4	6	0	High
Heikkinen et al [47]	10	1	4	6	0	High
Heikkinen et al [45]	9	1	4	6	0	High
Umapathy et al [56]	10	1	5	3	1	High
Yin et al [49]	9	1	5	5	0	High
Groves et al [50]	6	1	6	5	1	High
Meesters et al [54]	9	1	5	1	0	Medium
Goldsmith and Safran [52]	7	0	5	2	0	Medium
Heikkinen et al [46]	5	1	3	3	0	Medium
Sobel and Popp [55]	3	0	1	0	0	Low
Median study quality	9/11	1/3	5/7	5/6	0/1	High

[a]Lowest to highest possible score for reporting (0-11), external validity (0-3), bias (0-7), confounding (0-6), power (0-1), and overall quality (0-28).

[b]Percentage scores were calculated by dividing the final score by the maximum score and multiplication by 100. The percentage scores were used for ordinal categorization of the studies as low quality (≤33%), medium quality (33.4%-66.7%), and high quality (≥66.8%) [38].

Knowledge Acquisition

Web-based patient education significantly increased patients' knowledge about orthopedic conditions and orthopedic treatment [44,45,48-51,53-55]. Web-based interventions were more effective than interventions provided through traditional channels [44,45,48-50], and these effects persisted over 2 weeks [45]. Increased knowledge levels also resulted in patients feeling more knowledgeable [44,45,49,54,55]. However, feelings of knowledgeability did not significantly increase more after Web-based education [44,45], except when provided in addition to face-to-face sessions [49].

Patients who received educational interventions did not acquire more knowledge than those who independently reviewed health information websites. One trial reported that a theory-based intervention produced higher knowledge levels regarding osteoporosis than a health information website in healthy older females [51], but another found no significant difference between both interventions in the same target group [53].

Patient Satisfaction and Patient Feedback

Patient satisfaction was a main outcome in 2 studies [48,49]. Both found that Web-based patient education had a positive effect on patients' satisfaction. Yin and colleagues report a persistent increase in satisfaction with information and teaching on the day of surgery (M_i=8.7 vs M_c=7.7, P=.03) and at the first postoperative visit (M_i=9.2 vs M_c=8.1, P=.01) after exposing knee arthroscopy patients to a custom online teaching module with explanations of anatomy, pathology, and perioperative instructions [49]. Fraval and colleagues report that satisfaction increased more in orthopedic outpatients who consulted both the online module and received verbal counseling with their surgeon compared with those who had only received the latter [48].

Seventy percent of trials investigated patient satisfaction or collected qualitative patient feedback but had not defined it as a principal outcome. Feedback on the online interventions was generally positive: patients described them as "very effective" [52], "easy to use" [43,50,55], and "worth the time" [49]. Compared with face-to-face education and health information websites, Web-based education was mostly evaluated better [48,53]. Only Heikkinen and colleagues report worse evaluations in terms of clarity of the content for the Web-based intervention (mean=79.75) compared with the face-to-face session with a nurse (mean=86.41), P=.001 [46]. However, both methods were considered clear enough to warrant further use.

Anxiety

In the 3 studies that assessed patients' anxiety following Web-based patient education, no significant effects on anxiety were found. Knee arthroscopy patients reported few distressing emotions in general, and anxiety was not influenced by Web-based patient education or verbal education [47]. After visiting a website providing an overview of the preoperative, intraoperative, and postoperative care processes, orthopedic outpatients were not less anxious about the planned surgery than patients who had discussed the same content with their surgeon [48]. For knee arthroscopy patients, using a Web-based educational tool did also not decrease anxiety about the surgery but did decrease anxiety about recovery [49].

Table 2. Patient outcomes and instruments used to assess the effect of Web-based patient education (alphabetical order).

Outcome measure	Instrument	Used in
Knowledge acquisition	Deaconess Informed Comprehension Test	[48]
	Hip Fractures Knowledge Test	[53]
	Knowledge Test	[44,45]
	Modified Standard Anaesthesia Learning Test	[50]
	Osteoporosis Health Belief Survey	[51,53]
	Orthopaedic Patients Knowledge Questionnaire	[44]
	Osteoarthritis Quality Indicator	[56]
	Sufficiency of Knowledge	[44,45]
	Custom instrument (no name provided)	[54,55,49]
Patient satisfaction and patient feedback[a]	Client Satisfaction Questionnaire	[48]
	Patients' Evaluations of Education	[46]
	Perceived Health Website Usability Questionnaire	[53]
	Custom instrument (no name provided)	[49]
Anxiety	Emotions Questionnaire	[47]
	State-Trait Anxiety Index	[48]
	Patients' Evaluations of Education	[46]
	Custom instrument (no name provided)	[49]
Empowerment, self-efficacy, and health attitudes	Calcium subscale of Osteoporosis Self-efficacy Scale	[53]
	Osteoporosis Health Belief Scale	[51,53]
	Outcome Expectations for Exercise Scale	[53]
	Patients' Evaluations of Education	[46]
	Self-efficacy for Exercise	[53]
	Web-based Learning Self-efficacy Measure	[53]
Self-management and behavior change	Behavioral Risk Factor Surveillance System	[51]
	Block-National Cancer Institute Health Habits and History Questionnaire	[53]
	Brief Physical Activity Survey	[51]
	Health Education Impact Questionnaire	[56]
	Yale Physical Activity Survey	[53]
Clinical outcomes	The Symptoms	[43]
	Verbal Rating Scale of McGill Pain Questionnaire	[52]

[a]Qualitative feedback methods [51,52,55] are not included in the table.

Empowerment, Self-Efficacy, and Health Attitudes

Two studies included self-efficacy as a primary outcome measure and reported contradicting evidence [51,53]. One study showed that both patients who used a structured social cognitive theory–based educational intervention and those who browsed health information websites had increased self-efficacy for calcium intake, the health behavior of interest [53]. In contrast, these effects were not replicated in a similar study that reported that self-efficacy was not influenced by patient education at all [51]. A lower quality report of the larger randomized controlled trial of Heikkinen and colleagues reported results that indicate that Web-based patient education may even adversely influence self-efficacy. When participants were asked how well they could act based on the knowledge received in the education, the intervention group perceived their abilities significantly lower (mean=82.77) than the control group (mean=88.86), *P*=.001 [46]. Thus, the extent to which Web-based educational interventions impact self-efficacy remains unclear.

Self-Management and Health Behavior Change

Only one study assessed the effect of Web-based patient education on self-management [56]. In Umapathy and colleagues' 2015 study, patients with self-assessed osteoarthritis used a tailored information tool to enhance self-management for 12 months. Users of the tool reported increased health-directed activity, engagement with life, self-monitoring, skill acquisition, and social integration but not significantly

more so than nonusers. Users did acquire more knowledge about self-management and lifestyle as measured with the Osteoarthritis Quality Indicator and showed a significant reduction in weight (change score: −6.3%) compared with nonusers (change score: 2.5%), *P*= . 03. Although these results are promising, confounds in the study's design contaminate its findings: participants in this study were not randomized to the conditions, and this opportunity for patients to self-select may have resulted in motivated users and demotivated nonusers.

Clinical Outcomes

The evidence for an effect of Web-based patient education on clinical outcomes is limited and contradictory: although access to a pain management section of an ambulatory surgery website resulted in a significant decrease in "discomforting" pain scores after ambulatory surgery [52], Web-based tutorials about knee arthroscopy had no effect on pain after surgery [43]. In fact, the second study's findings suggest that pain may be less effectively decreased after Web-based patient education in comparison with face-to-face education. Four weeks after the surgery, patients who had received Web-based education reported more pain in other areas (15.7% moderate-high pain) in comparison with the control group (7% moderate-high pain). However, three-way interactions between pain, group, and time failed to reach significance. The same study also reports that other postoperative symptoms (including tiredness, problems with digestion, and swelling of the operation area) decreased regardless of the patient education method used.

Discussion

Overall Findings

This review set out to examine the effects of Web-based patient education in the care for adult orthopedic patients. This is an important subject, as orthopedic patients are commonly using the internet to find health information [10,11] and perceive this to have an impact on both their health and social environment [12], although these effects have not yet been systematically examined. The comparative evaluation of Web-based educational interventions is especially relevant: to generic health information websites that potentially form a strong co-intervention [13] and to traditional patient education interventions that may be more effective but have higher costs [12].

This review identified 14 studies that reported the effects of ten different Web-based patient education interventions targeted toward the orthopedic patient population. Although the amount of studies is limited, the overall methodological quality of the included studies is high. Still, the different studies could not be compared on a meta-analytic level given the wide variety in scope, primary outcomes, and means of outcome assessment. Furthermore, the reported findings may be limited to patients who were already able to use the internet, as 70% of the studies included in this review established criteria that excluded inexperienced, less skilled patients with limited access to the internet to the trials. Hence, it is difficult to draw definitive conclusions about the effectiveness of Web-based patient education interventions.

While keeping these limitations in mind, the currently available evidence does suggest that patients who are offered Web-based patient education find the service both usable and satisfactory [43,48,49,51,52,53,55]. It increases their knowledge levels [44,45,48-50,53-55], which also results in patients who feel knowledgeable [44,45,49,54,55] and are able to participate in the informed consent process [48,49,55]. Web-based education appeared to be more effective in these aspects than traditional education methods [44,45,48-50]. Despite their knowledge gain, the provision of online information to patients does not subsequently reduce patients' anxiety [47-49]. These findings support our first hypothesis that Web-based patient education interventions would have a positive effect on patients' knowledge but not on anxiety. Contrary to second hypothesis, however, Web-based education was not found more effective than generic health information websites [49,51]. A possible explanation for this finding is that both Web-based patient education materials and generic health information websites suffer from issues such as poor readability [21,57-59].

There is still insufficient evidence to determine the effect of Web-based patient education on self-efficacy, self-management, or clinical outcomes. Only 2 studies investigated self-efficacy [51,53], 1 observational study investigated self-management [56], 2 studies investigated pain [43,52], and no studies have assessed patients' functioning using standardized patient-reported outcome measures for orthopedic practice, such as the Western Ontario and McMasters Universities Osteoarthritis Index or Hip Disability and Osteoarthritis Outcome Score. Therefore, we were unable to test our hypothesis that Web-based patient education would not have an effect on clinical outcomes.

This review illustrates the typical Web-based patient educational intervention that is currently offered to people with orthopedic conditions. These are mostly websites focused on practical, informational content that is presented using multiple media formats including text, pictures, and video. Most offer some form of (human) support to patients using the programs but are still static in terms of interactivity. Still, it seems that online self-assessment is being recognized as an appropriate strategy to make educational content more engaging. At this point, there was not enough evidence to conclude that either of these intervention characteristics—content, media use, support, interactivity, or duration—has a consistent effect on the interventions' success. However, regarding support provision, it should be noted that almost all studies that did not specify the level of support offered on the website did include some form of provider contact as part of the usual care given to both the experimental and control groups [48-50,52]. Patients may have received feedback and support during these meetings, which makes it difficult to estimate the effects that added online support or feedback may have. Therefore, future work should report whether (information and communication technology) support or feedback was provided as part of usual care.

Most of our findings are in line with previous reviews of Web-based patient education. We found further support for the idea that changing the channel of communication in patient education can increase patient satisfaction, as was tentatively hypothesized in Nguyen and colleagues' 2004 review [13].

Web-based patient education is also equally effective in orthopedics as in oncology practice [15]. Similarly to orthopedic patients, breast cancer patients' knowledge and satisfaction increased following Web-based education, whereas their anxiety was not affected. Furthermore, in both fields, a wide variety of study outcomes and corresponding instruments was identified. Thus, this review can only further endorse the need for standardized instruments in the evaluation of Web-based interventions as previously addressed by Ryhänen and colleagues in 2010 [15].

Despite the aforementioned replications, we could not determine whether self-care behavior of orthopedic patients increased because of Web-based patient education, an effect that has been identified in cardiovascular patients who were offered online educational interventions [60]. Because the internet can be used without constant professional supervision, online interventions may play a continuous role in the education and support of chronically ill orthopedic patients [61-63]. Despite this potential, we found only 1 study that specifically evaluated education within the context of an online self-management intervention [56]. This may have been because we have excluded behavioral or affective interventions from review. This narrow scope allowed us to precisely examine the effectiveness of education alone, but a next step for Web-based interventions would be report separately on educational, behavioral, and affective content. This will allow those who are tasked with developing interventions to study the interplay between these components to determine the "ideal" dose for a specific population or condition. Taxonomies to facilitate such in-depth examination of intervention components have already been developed for behavior change techniques [64] and computer tailoring [65]. Slowly, similar efforts are done for Web-based interventions as well, such as Barak and colleagues' internet-supported interventions model [36] used in this review to describe intervention components and Win and colleagues' online patient education features model [12]. Still, a consensus on an appropriate taxonomy has not yet been reached, and until this is in place, it will be difficult to estimate the specific role education can play in enhancing complex outcomes such as self-management capabilities.

Limitations

This review has several limitations that relate to the representativeness of the samples included in the studies, the limited number of included studies, and the lack of a meta-analysis.

First, the quality of the reported studies was higher than what previous reviews of Web-based interventions have documented [13,66]. Most studies provided an elaborate description of the control groups and interventions, including the specific interactive elements designed into the programs. Still, the external validity of the included studies is low; no studies provided evidence that the included sample was representative of the entire population. This is concerning considering that most studies had criteria in place that excluded participants with less internet use and experience. Compared with these selected samples, the entire population was likely older [10,67-69], lower educated [10,68,69], and more likely to receive public care [10,70]. On the other hand, younger patients are also the ones who expect more information [71,72], value online services [73,74], and are most likely to benefit from educational interventions [75]. Thus, although we cannot conclude that it serves the whole orthopedic population, Web-based patient education may be an excellent way to cater to this younger patients' specific needs.

Second, we were able to evaluate only a limited number of studies. Although the initial search identified over a thousand potential studies, only ten trials specifically evaluated Web-based patient education interventions in a sufficiently controlled setting. As a result, we were not able to draw any reliable conclusions about the effect of Web-based patient education on patient reported outcomes, including postoperative pain and functioning, whereas reviews of traditional patient education show that these outcomes may be affected [28-30].

Third, the studies employed a wide variety of outcome measures that did not allow for a meta-analysis of the findings. Though the qualitative synthesis does indicate that Web-based patient education increases patients' knowledge levels and satisfaction, we were not able to determine the extent of these effects. Therefore, their clinical relevance has yet to be determined.

Conclusions

In summary, offering patient education interventions via the internet to adult people with orthopedic conditions increases their knowledge about their condition and its treatment [44,45,47-51,53-55]. Online educational interventions are typically instructional websites that make use of multimedia but offer limited interactivity. They are considered usable and can increase patient satisfaction [43,48,49,50,52,53,55]. However, the provision of online information to patients does not subsequently reduce patients' anxiety [47-49].

Given these findings, we tentatively conclude that Web-based patient education may be offered as a time- and cost-effective alternative to current educational interventions when the primary aim of the intervention is to increase patients' knowledge and satisfaction. However, there is too little evidence to advocate for Web-based patient education to replace existing interventions that aim to improve other outcomes, including self-management skills, pain, and function. Furthermore, it should be kept in mind that Web-based interventions currently cater to younger patients who may not be comparable to the general patient population. A solution for hospital administrators or health care policy makers currently planning an educational intervention for orthopedics patients is to provide Web-based education in addition to verbal or written components, which allows patients to select the platform they are most comfortable with while ensuring satisfactory results.

Acknowledgments

This work is part of the research program Tailored healthcare through customer profiling (Project 314-99-118), which is financed by the Netherlands Organisation for Scientific Research (NWO) and Zimmer Biomet Inc.

Authors' Contributions

TD, MM, and HdR conceived the study and its design. TD, MM, and HdR selected articles for inclusion and interpreted the data. TD and BSG assessed the studies' quality. TD drafted the article. All authors provided critical input to drafts and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies for the identification of studies assessing the effects of Web-based patient education interventions for the adult orthopedic population.

[PDF File (Adobe PDF File), 100KB - jmir_v20i4e143_app1.pdf]

Multimedia Appendix 2

Intervention characteristics of studies evaluating the effects of Web-based patient education in orthopedics (alphabetical order).

[PDF File (Adobe PDF File), 86KB - jmir_v20i4e143_app2.pdf]

Multimedia Appendix 3

Summary of the effects of Web-based patient education (ordered by comparison then by quality).

[PDF File (Adobe PDF File), 148KB - jmir_v20i4e143_app3.pdf]

References

1. Gruman J, Rovner MH, French ME, Jeffress D, Sofaer S, Shaller D, et al. From patient education to patient engagement: implications for the field of patient education. Patient Educ Couns 2010 Mar;78(3):350-356. [doi: 10.1016/j.pec.2010.02.002] [Medline: 20202780]
2. Fernsler JI, Cannon CA. The whys of patient education. Semin Oncol Nurs 1991 May;7(2):79-86. [Medline: 1882153]
3. Hoving C, Visser A, Mullen PD, van den Borne B. A history of patient education by health professionals in Europe and North America: from authority to shared decision making education. Patient Educ Couns 2010 Mar;78(3):275-281. [doi: 10.1016/j.pec.2010.01.015] [Medline: 20189746]
4. Pellino T, Tluczek A, Collins M, Trimborn S, Norwick H, Engelke ZK, et al. Increasing self-efficacy through empowerment: preoperative education for orthopaedic patients. Orthop Nurs 1998;17(4):48-51, 54-9. [Medline: 9814337]
5. Lübbeke A, Suvà D, Perneger T, Hoffmeyer P. Influence of preoperative patient education on the risk of dislocation after primary total hip arthroplasty. Arthritis Rheum 2009 Apr 15;61(4):552-558 [FREE Full text] [doi: 10.1002/art.24340] [Medline: 19333972]
6. Cheung A, Finegan BA, Torok-Both C, Donnelly-Warner N, Lujic J. A patient information booklet about anesthesiology improves preoperative patient education. Can J Anesth 2007 May;54(5):355-360. [doi: 10.1007/bf03022657]
7. Lin PC, Lin LC, Lin JJ. Comparing the effectiveness of different educational programs for patients with total knee arthroplasty. Orthop Nurs 1997;16(5):43-49. [Medline: 9369734]
8. Daltroy LH, Morlino CI, Eaton HM, Poss R, Liang MH. Preoperative education for total hip and knee replacement patients. Arthritis Care Res 1998 Dec;11(6):469-478. [doi: 10.1002/art.1790110607]
9. Colledge A, Car J, Donnelly A, Majeed A. Health information for patients: time to look beyond patient information leaflets. J R Soc Med 2008 Sep;101(9):447-453 [FREE Full text] [doi: 10.1258/jrsm.2008.080149] [Medline: 18779246]
10. Fraval A, Ming CY, 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]
11. Baker JF, Devitt BM, Kiely PD, Green J, Mulhall KJ, Synnott KA, et al. Prevalence of internet use amongst an elective spinal surgery outpatient population. Eur Spine J 2010 Oct;19(10):1776-1779 [FREE Full text] [doi: 10.1007/s00586-010-1377-y] [Medline: 20361342]
12. Win KT, Hassan NM, Oinas-Kukkonen H, Probst Y. Online patient education for chronic disease management: consumer perspectives. J Med Syst 2016 Apr;40(4):88. [doi: 10.1007/s10916-016-0438-0] [Medline: 26846749]

13. Nguyen HQ, Carrieri-Kohlman V, Rankin SH, Slaughter R, Stulbarg MS. Internet-based patient education and support interventions: a review of evaluation studies and directions for future research. Comput Biol Med 2004 Mar;34(2):95-112 [FREE Full text] [doi: 10.1016/S0010-4825(03)00046-5] [Medline: 14972630]
14. McKay HG, Glasgow RE, Feil EG, Boles SM, Barrera Jr MJ. Internet-based diabetes self-management and support: initial outcomes from the Diabetes Network project. Rehabil Psychol 2002;47(1):31-48. [doi: 10.1037/0090-5550.47.1.31]
15. Ryhänen AM, Siekkinen M, Rankinen S, Korvenranta H, Leino-Kilpi H. The effects of Internet or interactive computer-based patient education in the field of breast cancer: a systematic literature review. Patient Educ Couns 2010 Apr;79(1):5-13. [doi: 10.1016/j.pec.2009.08.005] [Medline: 19744817]
16. Hering K, Harvan J, D'Angelo M, Jasinski D. The use of a computer website prior to scheduled surgery (a pilot study): impact on patient information, acquisition, anxiety level, and overall satisfaction with anesthesia care. AANA J 2005 Feb;73(1):29-33. [Medline: 15727280]
17. Wald HS, Dube CE, Anthony DC. Untangling the Web--the impact of Internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
18. Idriss NZ, Alikhan A, Baba K, Armstrong AW. Online, video-based patient education improves melanoma awareness: a randomized controlled trial. Telemed J E Health 2009 Dec;15(10):992-997. [doi: 10.1089/tmj.2009.0055] [Medline: 20028190]
19. Hungerford DS. Internet access produces misinformed patients: managing the confusion. Orthopedics 2009 Oct;32(9):pii. [doi: 10.3928/01477447-20090728-04] [Medline: 19751023]
20. Brooks BA. Using the internet for patient education. Orthop Nurs 2001;20(5):69-77. [Medline: 12025306]
21. Cassidy JT, Baker JF. Orthopaedic patient information on the World Wide Web: an essential review. J Bone Joint Surg Am 2016 Feb 17;98(4):325-338. [doi: 10.2106/JBJS.N.01189] [Medline: 26888683]
22. Jariwala AC, Kandasamy MS, Abboud RJ, Wigderowitz CA. Patients and the internet: a demographic study of a cohort of orthopaedic out-patients. Surgeon 2004 Apr;2(2):103-106. [Medline: 15568435]
23. Roter DL, Hall JA, Merisca R, Nordstrom B, Cretin D, Svarstad B. Effectiveness of interventions to improve patient compliance: a meta-analysis. Med Care 1998;36(8):1138-1161. [doi: 10.1097/00005650-199808000-00004] [Medline: 9708588]
24. Feudtner C. What are the goals of patient education? West J Med 2001 Mar;174(3):173-174 [FREE Full text] [Medline: 11238347]
25. Schenker Y, Fernandez A, Sudore R, Schillinger D. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: a systematic review. Med Decis Making 2011;31(1):151-173. [doi: 10.1177/0272989X10364247] [Medline: 20357225]
26. Johansson K, Nuutila L, Virtanen H, Katajisto J, Salanterä S. Preoperative education for orthopaedic patients: systematic review. J Adv Nurs 2005 Apr;50(2):212-223. [doi: 10.1111/j.1365-2648.2005.03381.x] [Medline: 15788086]
27. Ronco M, Iona L, Fabbro C, Bulfone G, Palese A. Patient education outcomes in surgery: a systematic review from 2004 to 2010. Int J Evid Based Healthc 2012 Dec;10(4):309-323. [doi: 10.1111/j.1744-1609.2012.00286.x] [Medline: 23173656]
28. McDonald S, Page MJ, Beringer K, Wasiak J, Sprowson A. Preoperative education for hip or knee replacement. Cochrane Database Syst Rev 2014 May 13(5):CD003526. [doi: 10.1002/14651858.CD003526.pub3] [Medline: 24820247]
29. Kroon FP, van der Burg LR, Buchbinder R, Osborne RH, Johnston RV, Pitt V. Self-management education programmes for osteoarthritis. Cochrane Database Syst Rev 2014;1:CD008963. [doi: 10.1002/14651858.CD008963.pub2] [Medline: 24425500]
30. Louw A, Diener I, Butler DS, Puentedura EJ. Preoperative education addressing postoperative pain in total joint arthroplasty: review of content and educational delivery methods. Physiother Theory Pract 2013 Apr;29(3):175-194. [doi: 10.3109/09593985.2012.727527] [Medline: 23035767]
31. Padilla GV, Bulcavage LM. Theories used in patient/health education. Semin Oncol Nurs 1991 May;7(2):87-96. [Medline: 1882154]
32. Glanz K, Rimer BK, Viswanath K, editors. Health Behaviour and Health Education: Theory, Research, and Practice. San Francisco: Jossey-Bass; 2008.
33. Syx RL. The practice of patient education: the theoretical perspective. Orthop Nurs 2008;27(1):50-54. [doi: 10.1097/01.NOR.0000310614.31168.6b] [Medline: 18300691]
34. 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]
35. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
36. 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]

37. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384 [FREE Full text] [Medline: 9764259]
38. Hootman JM, Driban JB, Sitler MR, Harris KP, Cattano NM. Reliability and validity of three quality rating instruments for systematic reviews of observational studies. Res Synth Methods 2011 Jun;2(2):110-118. [doi: 10.1002/jrsm.41] [Medline: 26061679]
39. Cindy Ng LW, Mackney J, Jenkins S, Hill K. Does exercise training change physical activity in people with COPD? A systematic review and meta-analysis. Chron Respir Dis 2012 Feb;9(1):17-26. [doi: 10.1177/1479972311430335] [Medline: 22194629]
40. Robbins SM, Houghton PE, Woodbury MG, Brown JL. The therapeutic effect of functional and transcutaneous electric stimulation on improving gait speed in stroke patients: a meta-analysis. Arch Phys Med Rehabil 2006 Jun;87(6):853-859. [doi: 10.1016/j.apmr.2006.02.026] [Medline: 16731222]
41. Eng JJ, Teasell R, Miller WC, Wolfe DL, Townson AF, Aubut J, et al. Spinal Cord Injury Rehabilitation Evidence: methods of the SCIRE systematic review. Top Spinal Cord Inj Rehabil 2007;13(1):1-10 [FREE Full text] [doi: 10.1310/sci1301-1] [Medline: 22767989]
42. Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Management 2003 Sep;14(3):207-222. [doi: 10.1111/1467-8551.00375]
43. Heikkinen K, Leino-Kilpi H, Vahlberg T, Salanterä S. Ambulatory orthopaedic surgery patients' symptoms with two different patient education methods. Int J Orthop Trauma Nurs 2012 Feb;16(1):13-20. [doi: 10.1016/j.ijotn.2011.04.004]
44. Heikkinen K, Helena LK, Taina N, Anne K, Sanna S. A comparison of two educational interventions for the cognitive empowerment of ambulatory orthopaedic surgery patients. Patient Educ Couns 2008 Nov;73(2):272-279. [doi: 10.1016/j.pec.2008.06.015] [Medline: 18678461]
45. Heikkinen K, Leino-Kilpi H, Salanterä S. Ambulatory orthopaedic surgery patients' knowledge with internet-based education. Methods Inf Med 2012;51(4):295-300. [doi: 10.3414/ME11-02-0004] [Medline: 22476362]
46. Heikkinen K, Salanterä S, Leino-Kilpi H. How do patients evaluate their education? - a comparison of two education methods. Stud Health Technol Inform 2009;146:850-851. [Medline: 19593017]
47. Heikkinen K, Salanterä S, Leppänen T, Vahlberg T, Leino-Kilpi H. Ambulatory orthopaedic surgery patients' emotions when using different patient education methods. J Perioper Pract 2012 Jul;22(7):226-231. [Medline: 22919767]
48. Fraval A, Chandrananth J, Chong YM, Tran P, Coventry LS. Internet based patient education improves informed consent for elective orthopaedic surgery: a randomized controlled trial. BMC Musculoskelet Disord 2015 Feb 07;16:14 [FREE Full text] [doi: 10.1186/s12891-015-0466-9] [Medline: 25885962]
49. Yin B, Goldsmith L, Gambardella R. Web-based education prior to knee arthroscopy enhances informed consent and patient knowledge recall: a prospective, randomized controlled study. J Bone Joint Surg Am 2015 Jun 17;97(12):964-971. [doi: 10.2106/JBJS.N.01174] [Medline: 26085529]
50. Groves ND, Humphreys HW, Williams AJ, Jones A. Effect of informational internet web pages on patients' decision-making: randomised controlled trial regarding choice of spinal or general anaesthesia for orthopaedic surgery. Anaesthesia 2010 Mar;65(3):277-282. [Medline: 20336817]
51. Drieling RL, Ma J, Thiyagarajan S, Stafford RS. An internet-based osteoporotic fracture risk program: effect on knowledge, attitudes, and behaviors. J Womens Health (Larchmt) 2011 Dec;20(12):1895-1907. [doi: 10.1089/jwh.2010.2515] [Medline: 21970565]
52. Goldsmith DM, Safran C. Using the web to reduce postoperative pain following ambulatory surgery. Proc AMIA Symp 1999:780-784 [FREE Full text] [Medline: 10566466]
53. Nahm ES, Barker B, Resnick B, Covington B, Magaziner J, Brennan PF. Effects of a social cognitive theory-based hip fracture prevention web site for older adults. Comput Inform Nurs 2010;28(6):371-379. [doi: 10.1097/NCN.0b013e3181f69d73] [Medline: 20978408]
54. Meesters JJL, de Boer IG, van den Berg MH, Fiocco M, Vliet Vlieland TP. Evaluation of a website providing information on regional health care services for patients with rheumatoid arthritis: an observational study. Clin Rheumatol 2012 Apr;31(4):637-645 [FREE Full text] [doi: 10.1007/s10067-011-1897-9] [Medline: 22160606]
55. Sobel D, Popp PL. Informed consent and expectation management: a case study. J Healthc Risk Manag 2006;26(4):21-26. [doi: 10.1002/jhrm.5600260406] [Medline: 19606763]
56. Umapathy H, Bennell K, Dickson C, Dobson F, Fransen M, Jones G, et al. The web-based osteoarthritis management resource My Joint Pain improves quality of care: a quasi-experimental study. J Med Internet Res 2015 Jul 07;17(7):e167 [FREE Full text] [doi: 10.2196/jmir.4376] [Medline: 26154022]
57. Eltorai AE, Sharma P, Wang J, Daniels AH. Most American Academy of Orthopaedic Surgeons' online patient education material exceeds average patient reading level. Clin Orthop Relat Res 2015 Apr;473(4):1181-1186 [FREE Full text] [doi: 10.1007/s11999-014-4071-2] [Medline: 25475715]
58. Nassiri M, Bruce-Brand RA, O'Neill F, Chenouri S, Curtin PT. Surfing for hip replacements: has the "internet tidal wave" led to better quality information. J Arthroplasty 2014 Jul;29(7):1339-44.e1. [doi: 10.1016/j.arth.2014.01.009] [Medline: 24559520]

59. Keller S. How readable is the patient education material found on top-rated hospital web sites? J Hosp Librariansh 2014 Nov 04;14(4):366-373. [doi: 10.1080/15323269.2014.950160]
60. Fredericks S, Martorella G, Catallo C. A systematic review of web-based educational interventions. Clin Nurs Res 2015 Feb;24(1):91-113. [doi: 10.1177/1054773814522829] [Medline: 24571963]
61. Wilkinson A, Whitehead L. Evolution of the concept of self-care and implications for nurses: a literature review. Int J Nurs Stud 2009 Aug;46(8):1143-1147. [doi: 10.1016/j.ijnurstu.2008.12.011] [Medline: 19200992]
62. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
63. Trudeau KJ, Pujol LA, DasMahapatra P, Wall R, Black RA, Zacharoff K. A randomized controlled trial of an online self-management program for adults with arthritis pain. J Behav Med 2015 Jun;38(3):483-496 [FREE Full text] [doi: 10.1007/s10865-015-9622-9] [Medline: 25686538]
64. 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]
65. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
66. Bessell TL, McDonald S, Silagy CA, Anderson JN, Hiller JE, Sansom LN. Do internet interventions for consumers cause more harm than good? A systematic review. Health Expect 2002 Mar;5(1):28-37. [Medline: 11906539]
67. Thorne E, Mackenzie PD, Wilson M. Expanding role of the internet in the orthopaedic outpatient setting. Bull Am Coll Surg 2017 Jan;99(1):32-35. [doi: 10.1308/rcsbull.2017.32]
68. Baker JF, Green J, Synnott KA, Stephens MM, Poynton AR, Mulhall KJ. Internet use in an orthopaedic outpatient population. Curr Orthop Pract 2013;24(3):308-311. [doi: 10.1097/BCO.0b013e31828e542b]
69. Walsh KP, Rehman S, Goldhirsh J. Disparities in internet use among orthopedic outpatients. Orthopedics 2014 Feb;37(2):e133-e140. [doi: 10.3928/01477447-20140124-14] [Medline: 24679198]
70. Gutierrez N, Kindratt TB, Pagels P, Foster B, Gimpel NE. Health literacy, health information seeking behaviors and internet use among patients attending a private and public clinic in the same geographic area. J Community Health 2014 Feb;39(1):83-89. [doi: 10.1007/s10900-013-9742-5] [Medline: 23900880]
71. Rankinen S, Salanterä S, Heikkinen K, Johansson K, Kaljonen A, Virtanen H, et al. Expectations and received knowledge by surgical patients. Int J Qual Health Care 2007 Apr;19(2):113-119. [doi: 10.1093/intqhc/mzl075] [Medline: 17277008]
72. Klemetti S, Leino-Kilpi H, Charalambous A, Copanitsanou P, Ingadottir B, Istomina N, et al. Information and control preferences and their relationship with the knowledge received among European joint arthroplasty patients. Orthop Nurs 2016;35(3):174-182. [doi: 10.1097/NOR.0000000000000246] [Medline: 27187223]
73. Ackerman IN, Bucknill A, Page RS, Broughton NS, Roberts C, Cavka B, et al. Preferences for disease-related education and support among younger people with hip or knee osteoarthritis. Arthritis Care Res (Hoboken) 2017;69(4):499-508. [doi: 10.1002/acr.22950] [Medline: 27273912]
74. Berendsen AJ, de Jong GM, Schuling J, Bosveld HE, de Waal MW, Mitchell GK, et al. Patient's need for choice and information across the interface between primary and secondary care: a survey. Patient Educ Couns 2010 Apr;79(1):100-105. [doi: 10.1016/j.pec.2009.07.032] [Medline: 19713065]
75. Fredericks S, Guruge S, Sidani S, Wan T. Postoperative patient education: a systematic review. Clin Nurs Res 2010 May;19(2):144-164. [doi: 10.1177/1054773810365994] [Medline: 20435785]

Abbreviations

HCP: health care professional

Edited by G Eysenbach; submitted 21.09.17; peer-reviewed by B Ingadottir, P Copanitsanou, KT Win, R Pankomera, R Robinson; comments to author 02.11.17; revised version received 20.12.17; accepted 23.12.17; published 23.04.18.

Please cite as:
Dekkers T, Melles M, Groeneveld BS, de Ridder H
Web-Based Patient Education in Orthopedics: Systematic Review
J Med Internet Res 2018;20(4):e143
URL: http://www.jmir.org/2018/4/e143/
doi:10.2196/jmir.9013
PMID:29685869

Original Paper

Using Plain Language and Adding Communication Technology to an Existing Health-Related Questionnaire to Help Generate Accurate Information: Qualitative Study

Marlies Welbie[1], MSc; Harriet Wittink[1], PhD; Marjan J Westerman[2], PhD; Ilse Topper[1], MSc; Josca Snoei[1], MSc; Walter LJM Devillé[3,4,5], MD, PhD

[1]Research Group Lifestyle and Health, Faculty of Health Care, University of Applied Sciences, Utrecht, Netherlands
[2]Institute of Health Sciences, Amsterdam Public Health Research Institute, VU University, Amsterdam, Netherlands
[3]Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, Netherlands
[4]Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, Netherlands
[5]National Knowledge and Advisory Centre on Migrants, Refugees and Health (Pharos), Utrecht, Netherlands

Corresponding Author:
Marlies Welbie, MSc
Research Group Lifestyle and Health
Faculty of Health Care
University of Applied Sciences
Heidelberglaan 7
Utrecht, 3584 CS
Netherlands
Phone: 31 638763239
Email: marlies.welbie@hu.nl

Abstract

Background: Low-educated patients are disadvantaged in using questionnaires within the health care setting because most health-related questionnaires do not take the educational background of patients into account. The Dutch Talking Touch Screen Questionnaire (DTTSQ) was developed in an attempt to meet the needs of low-educated patients by using plain language and adding communication technology to an existing paper-based questionnaire. For physical therapists to use the DTTSQ as part of their intake procedure, it needs to generate accurate information from *all* of their patients, independent of educational level.

Objective: The aim of this study was to get a first impression of the information that is generated by the DTTSQ. To achieve this goal, response processes of physical therapy patients with diverse levels of education were analyzed.

Methods: The qualitative Three-Step Test-Interview method was used to collect observational data on actual response behavior of 24 physical therapy patients with diverse levels of education. The interviews included both think-aloud and retrospective probing techniques.

Results: Of the 24 respondents, 20 encountered one or more problems during their response process. The use of plain language and information and communication technology (ICT) appeared to have a positive effect on the comprehensibility of the DTTSQ. However, it also had some negative effects on the interpretation, retrieval, judgment, and response selection within the response processes of the participants in this study. No educational group in this research population stood out from the rest in the kind or number of problems that arose. All respondents recognized themselves in the outcomes of the questionnaire.

Conclusions: The use of plain language and ICT within the DTTSQ had both positive and negative effects on the response processes of its target population. The results of this study emphasize the importance of earlier recommendations to accompany any adaption of any questionnaire to a new mode of delivery by demonstrating the difference and equivalence between the two different modes and to scientifically evaluate the applicability of the newly developed mode of the questionnaire in its intended setting. This is especially important in a digital era in which the use of plain language within health care is increasingly being advocated.

(J Med Internet Res 2018;20(4):e140) doi:10.2196/jmir.7940

KEYWORDS

educational status; surveys and questionnaires; physical therapy specialty; qualitative research

Introduction

Background

It is widely known and accepted that patient-centered care has the potential to increase the effectiveness of health care in general [1]. Unfortunately low-educated patients are not always able to benefit from a patient-centered care approach. A possible explanation for this can be found in the fact that patient-centered care demands of patients to take an active mutual partnership in the patient-provider interaction [2,3]. Patient-centered care puts a relatively strong emphasis on communication and information and takes the patient's perspective as a starting point [4,5]. Low-educated people have trouble providing information about their health problems to health care professionals. It is often hard for them to determine which information their health care provider (HCP) needs. The majority of them lack the health care vocabulary to report symptoms accurately, and they tend to provide information in a way that is illogical and difficult to comprehend by their HCP [6]. Having trouble providing information causes problems in patient-provider interaction, which impacts health outcomes negatively [7]. Evidence shows that the use of standardized health-related questionnaires contributes to the quality and patient centeredness of patient-provider interaction [8-13]. However, as most health-related questionnaires are not designed in ways that meet the needs of low-educated patients, these patients are disadvantaged in using them effectively within the health care setting [11-13]. In 2016, 9.5% of the Dutch population in the age range of 15 to 75 years had an educational level of primary school at most [14]. These people specifically are at risk when it comes to understanding and using health information [15]. If low-educated patients would be able to complete standardized health-related questionnaires independently and accurately, this might help them to provide relevant information on their health problem in a way that is logical and understandable to their HCP.

The content of the most frequently used questionnaire in Dutch physical therapy practice [16], the Patient Specific-Complaint (PSC) questionnaire [17], fits the goal of helping patients to provide relevant information regarding their health problem to their physical therapist. It is aimed at making the patient select his main limitations in functioning and formulate his own specific treatment goals. This paper-based questionnaire is responsive and sensitive to change to complaints that are highly relevant to the individual patient [17,18]. However, all members of the Dutch study sample of a recent study on the PSC questionnaire had problems completing it independently. All these 25 respondents, whose education levels varied from primary education to doctoral degrees, had trouble comprehending and interpreting one or more parts of the questionnaire. Six of them had difficulties finding a well-fitting answer to one or more questions. Due to these problems, the questionnaire generated invalid information in thirteen cases. Within the group respondents who provided invalid information, patients with no or primary education only were overrepresented [19].

The Dutch PSC questionnaire [17] was used as a starting point for the development of a user-friendly health-related questionnaire that meets the needs of low-educated physical therapy patients. This aim was met by using plain language and taking advantage of the possibilities of information and communication technology (ICT) by offering alternatives to text (eg, audio, pictures, and movies), self-explanatory scales, and easily accessible background information on the questionnaire's rationale. This resulted in the prototype of a new interactive questionnaire called the Dutch Talking Touch Screen Questionnaire (DTTSQ). The codesign process that led to the development of this prototype was described in detail by Cremers et al in 2015 [20].

Objective

The aim of this study was to get a first impression of the validity of the prototype of the DTTSQ by analyzing the response processes of patients with diverse levels of education. The research question that underlay this study was, "What problems occur during the response process of physical therapy patients with diverse levels of education while they complete the Dutch Talking Touchscreen Questionnaire?"

Methods

Design

A qualitative study was conducted. The Three-Step Test-Interview (TSTI) method [21] was used to collect observational data on actual response behavior of the respondents. The interviews included both think-aloud and retrospective probing techniques. Qualitative pretesting of questionnaires using cognitive methods such as the TSTI [21] is a well-known step within the development process of health-related questionnaires [22-25]. It enables researchers to give answers to questions such as the following: do all respondents understand the questions in the same way, do the questions ask for information that the respondents have and can retrieve, and does the wording of the questions provide respondents with all necessary information they require to be able to answer them in the way that was intended by its developers [22]?

Device

The DTTSQ was developed during a user-centered design process [20], which meant that low-educated persons were closely involved in designing the questionnaire. As a result, questions about pain location and pain intensity were added to the original questions that addressed the nature and severity of limitations in activities of daily living and the priority in which these limitations should be focused on during physical therapy. Needs regarding ease of use were met by the use of visual (pictures and videos) and auditory (speech) support, which was added to the questions. Respondents could insert their answers by tapping on the touch screen. The DTTSQ started with an

introductory video clip in which a host explains the purpose of the questionnaire and gives instructions on how to use the questionnaire (see Multimedia Appendix 1, screenshot 1 "Welcome"). All the questions were shown on separate screens. The application did not have a back function, so respondents could only move forward. After the first three questions, a new clip was shown to introduce and operationalize the term "activities" and to give instructions on an additional navigation function within the activity screens (see Multimedia Appendix 1, screenshot 6 "Activities"). The questionnaire finished with a video clip in which the host thanked the respondent for completing the questionnaire, explained what the physical therapist would do next, and announced that the questionnaire would end and close down automatically (Multimedia Appendix 1, screenshot 16 "Thank you"). To help patients keeping track of their answering process, overviews of their answers were shown regularly during the response process (see Multimedia Appendix 1; screenshot 5 "Overview location of the health problems," screenshot 8 "Overview activities," screenshot 10 "Overview most important activities," screenshot 14 "Overview most important activities and effort," and screenshot 15 "Overview all outcomes of the questionnaire"). For respondents who needed help or wanted more information on questions and/or answering options, a help function was provided. When the help function was activated, the question and answering options, as well as instructions on operation and background information on the questions, were given in spoken word [20].

Recruitment Strategy and Participants

Recruitment took place in eleven primary care practices in deprived areas of Utrecht, The Netherlands. Potential participants were invited by their physical therapists to participate in this study. The physical therapists shortly explained the goal of the study and provided the patient with an information letter that was written in plain Dutch language. If patients were interested, the physical therapist asked permission to give the patients' telephone number to researcher IT. Then researcher IT (1) Contacted the patient by telephone, (2) Again shortly explained the aim of the study, (3) Made sure the patient understood what was asked of him/her, (4) Answered any question the potential participant may have had, and (5) Checked the inclusion criteria. Inclusion criteria for participants were as follows: aged 18 years or older, Dutch as their first language, and both parents born in The Netherlands. The sampling procedure was aimed at getting a broad variation in levels of education and age, plus balance in our sample regarding gender. Throughout the recruitment process, the recruiting physical therapists were constantly kept informed about the profiles of participants the researchers were looking for. In total, 24 physical therapy patients were included in this study. Characteristics of study population can be found in Tables 1 and 2.

Data Collection and Procedures

Data collection took place at the respondents' homes or at the physical therapy practice of the respondent's physical therapist. The choice of location depended on the preference of the respondent. Two researchers were present (researchers IT and JS). Researcher IT conducted the interviews. When researcher JS missed information, she asked complementary questions.

The TSTI method was conducted as follows [21]:

Step 1

Researchers IT and JS observed each respondent as they completed the DTTSQ while thinking out loud. This step was aimed at collecting observational data regarding the respondent's response behavior. The data collected consists of two types: (1) observations of respondent's behavior and (2) think-aloud data. The data were recorded in the form of videotapes as well as audiotapes for later analysis and real-time notes by the researchers for use during the interview itself and later analysis. The researchers wrote their notes down on hardcopies of print screens of the DTTSQ.

Step 2

After the respondent finished completing the DTTSQ, researcher IT conducted an in-depth interview to clarify and complete the observational data. During this step, researcher IT only focused on those actions or thoughts she felt not fully informed about or were not fully clear to her. This step was aimed at filling gaps in the observational data and check information.

Step 3

During the final step, researcher IT conducted a semistructured interview aimed at eliciting experiences and opinions of the respondent. In this part of the interview, the respondent was stimulated to add secondary data such as accounts and reports of feelings, explanations, preferences, recommendations, etc. Researcher IT asked the respondent to paraphrase questions and to explain in his own words how he interpreted the question and why he chose the answering options he chose. When a respondent encountered problems in responding to a question, he was asked what he thought the exact nature of the problem was and why he behaved as he did in response to the question. He also was asked for suggestions for improvement of the question in terms of wording, layout, instructions, etc. Additionally, the respondent was asked to describe his health problem(s) and treatment goal(s) in his own words. Comparing these descriptions to the respondent's responses to the questionnaire during step 1 of the TSTI provided useful information as indicators of the validity of the data collected by the DTTSQ. Finally, the respondent was asked if he recognized himself in the outcomes of the questionnaire that were shown at the end of the questionnaire (see Multimedia Appendix 1, screenshot 15 "Overview all outcomes of the questionnaire"). After the TSTI was finished, researcher IT collected the demographic data through a brief structured interview.

Table 1. Characteristics of the respondents subdivided according to their level of education.

Characteristic	Low-educated[a] respondents (n=6)	Moderately educated[b] respondents (n=13)	Highly educated[c] respondents (n=5)
Mean age (range), years	65.8 (47-79)	50.5 (18-73)	56 (32-76)
Gender, n			
Male	2	5	2
Female	4	8	3

[a]Low means no education or primary education.

[b]Moderately means lower secondary education, (upper) secondary education, or postsecondary nontertiary education (including vocational education).

[c]Highly means tertiary education (bachelor's degree or higher).

Table 2. Characteristics per respondent.

Pseudonym	Age (year)	Educational level[a]	Last occupation
Jerome	47	Low	Truck driver
Michelle	56	Low	Cleaning lady
Ida	66	Low	Cleaning lady
Ronald	70	Low	Home painter
Dora	77	Low	Cleaning lady
Ilene	79	Low	Cleaning lady
Peter	18	Moderate	Student
Jude	18	Moderate	Student
Joline	19	Moderate	Photographer
Sandra	39	Moderate	Graphic designer
Christine	39	Moderate	Nurse for mentally disabled people
Lydia	56	Moderate	Domiciliary care
Rose	60	Moderate	Saleswoman
Francine	61	Moderate	Administrative officer
Henry	64	Moderate	Project coordinator
Bob	68	Moderate	Cashier
Roger	70	Moderate	Home painter
Bill	72	Moderate	Order picker
Mia	73	Moderate	Administrative officer
Ellen	32	High	Management assistant
Helga	54	High	Artist
Jill	55	High	Management assistant
Harald	63	High	Financial controller
Bernie	76	High	Lecturer chemistry

[a]Low refers to no education or primary education; moderate refers to lower secondary education, (upper) secondary education, or postsecondary nontertiary education (including vocational education); and high refers to tertiary education (bachelor's degree or higher).

Data Analysis

Data were analyzed using a thematic content analysis approach [26]. Four types of data were analyzed: (1) video recordings of the first two steps of the interview, (2) Dutch transcriptions of the third step of the interview, (3) observed respondent behavior in field notes, and (4) background information regarding the educational level, age, gender, and occupation of each respondent. Researcher MW started with open coding, coding all fragments of the twenty-four transcripts of step three of each interview using MAXQDA 10 of VERBI Software GmbH, Berlin.

The codes and fragments of seven randomly selected transcripts were validated by two peer researchers by independently coding each transcript with the coding scheme developed by researcher

MW. Differences in fragmentation or coding were discussed during consensus meetings.

To get more familiar with the data and to create an overview, researcher MW made a descriptive summary of each case on the basis of all four types of generated data after she finished open coding. Each summary contained all emerging themes regarding problems that occurred during the four phases of the response process as described by Tourangeau: (1) *comprehension*: (a) comprehension of text and wording and (b) interpretation of the meaning of the text, (2) *retrieval*: gathering relevant information, (3) *judgment*: assessing the retrieved information to judge its adequacy in relation to the meaning of the question, and (4) *response selection*: selecting the best fitting answering option [27]. The emerged themes in the summaries were supplemented with related field notes and background information regarding the educational level, age, gender, and occupation of the respondent. Then researcher MW listed all emerging "themes" from the descriptive summaries regarding problems that arose during the four steps of the response process. She established which themes recurred or were common and which were less common or stood alone. Then she structured the earlier created coding scheme by arranging all open codes by labeling them as, "problem with comprehension," "problem with interpretation," "problem with retrieval," "problem with judgment," or "problem with response selection."

The following step in analyzing the data was comparing the description of the limitations in functioning and treatment goals described by respondents during the semistructured interview (interview step 3) to the answering options the respondent selected in the DTTSQ during the think-aloud phase of the data collection (interview step 1). If the chosen answer during step 1 did not fit the description in step 3, researcher MW closely watched the video again to see which actions or thoughts during the four steps of Tourangeau [27] during the response process of the question led the respondent to select the chosen answering option.

As a last step, researcher MW compared the analyzed interviews of low, moderately, and highly educated respondents to see whether or not the problems that occurred during the response processes differed between these groups of respondents. Transcripts were made in Dutch language. Only quotes used in this paper were translated from Dutch to English by researcher MW and checked by researcher HW, who is a bilingual speaker. During the whole course of the study, procedures and results were checked and discussed with researchers HW, MJW, and WD.

Ethics

No external funding was received by the Utrecht University of Applied Sciences to conduct this study. The study was registered with the Medical Ethics Commity of the Acadamic Medical Centre of Amsterdam, which declared that it does not fall under the scope of the "Medical Research Involving Human Subjects Act." The study was conducted according to the principles of the Declaration of Helsinki. All respondents provided written informed consent. The respondents names used in this paper are all fictitious to protect their privacy.

Results

Encountered Problems

Of the 24 respondents, 20 encountered one or more problems during their response process. Low-educated Michelle and moderately educated Christine, Lydia, and Sandra did not encounter any problem. All members of the total study population stated that they recognized themselves in the overall outcomes of the questionnaire. Bernie stated:

> *If I would have developed this questionnaire so it would have fitted my health problem I would have done it differently. Instead of selecting specific points on the body chart, for instance, I would have enabled people to select regions. In my case that would have enabled me to select the whole lower part of my body instead of a few specific points in it. But even though I would have done it differently, I recognize myself in the summary of my limitations in functioning. That is mainly due to the pictures of the activities in which I am impaired. When I look at all the outcomes as a whole, it is right. I recognize my own health situation.*

Most problems concerned interpretation of questions and answering options. Questions 1 and 4 generated the most problems. Question 3 generated no problems at all (see Table 3).

Problems With Comprehension of Text and Wording, and Interpretation

There were no problems with comprehension of text and wording. A total of 13 respondents of all educational groups encountered problems with interpretation. Ronald and Bob encountered this problem with three questions and Helga and Jerome with two different questions. The other 9 respondents encountered this problem with one question.

Table 3. Number of respondents having problems per question for each step of the response process. Hyphen indicates nonapplicabilty.

Question or assignment	Comprehension problems	Interpretation problems	Retrieval problems	Judgment problems	Response selection problems
1. Do you have pain? (Multimedia Appendix 1: screenshot 2 "Pain")	-	6	-	-	-
2. Tap on the location of your health problem. You can tap on multiple locations. (Multimedia Appendix 1: screenshot 3 "Location of the health problem")	-	-	1	1	2
3. This is the location of your pain. Rate the severity of your pain on the scale below. (Multimedia Appendix 1: screenshot 4 "Pain severity")	-	-	-	-	-
4. Select the activities in which you are impaired. (Multimedia Appendix 1: screenshot 7 "Activity "Lying")	-	9	6	8	3
5. Select the three activities which are most important to you. (Multimedia Appendix 1: screenshot 9 "most important activities")	-	-	-	-	2
6. Select the activity which is most important to you. (Multimedia Appendix 1: screenshot 11 "Most important activity 1")	-	1	-	-	-
7. Which of these two activities is most important to you now? (Multimedia Appendix 1: screenshot 12 "Most important activity 2")	-	1	-	-	-
8. Rate the effort it takes to carry out this activity. (Multimedia Appendix 1: screenshot 13 "Effort activity 1")	-	2	-	-	2

Interpreting Pictures

A total of 7 respondents interpreted pictures that were used as answering options in question 4 differently than was intended by the developers of the questionnaire. Ilene, for instance, selected "dressing and undressing" (Figure 1) and going to the toilet (Figure 2) because the way in which the person in the picture carried out the activity and the context in which he did it were different from theirs. This is illustrated in the following conversation:

Ilene:

> *I selected "dressing and undressing" because the person on the photo is standing up while he is dressing himself. I cannot do that. I have to sit down.*

Interviewer:

> *Would you have selected this activity if the person on the photo was sitting down while he dressed and undressed himself?*

Ilene:

> *No, because that is no problem for me. That is the way I do it. It is the same with going to the toilet. I selected that photo because the person on the photo does not use the support arms while he is using the toilet.*

Interviewer:

> *Would you have selected the photo if he would have used the support arms?*

Ilene:

> *No of course not! I do not have any problem going to the toilet because I have these support arms. I have everything I need in my house.*

Interpreting Categories

The answering options in the form of pictures of question 4 were put into eight different activity categories. These categories were shown on eight separate screens. The use of categories influenced the response process of two respondents negatively. Rose, for instance, recognized her impairment in the activity "reaching for something above the head," but did not select it:

> *I really was in great doubt with "reaching!" Because I thought: yes indeed that is problematic for my shoulder so I should select that activity. But the activity was placed in the category "standing" which I associated with using the legs and back, not with arm movements. In hindsight I probably should have selected it, but when I was completing the questionnaire I chose not to.*

Interpreting Plain Language

Six respondents misinterpreted question 2 "Do you have pain?" Four of them mentioned the short and simple way in which the question was formulated as the reason for this misinterpretation. All six respondents selected the answer "no," whereas in fact they were seeking help with their physical therapist because of pain complaints. Henry stated:

> *Well I am not in pain at this moment. But when I go photographing I take long walks carrying heavy lenses. And then my hip hurts sometimes. This is something my physical therapist needs to know because it should be the aim of the treatment. But I interpreted the question as "are you in pain at this moment." And that is why I answered "no." The sentence, the question, is very short. It is not specific enough. It should have said: "Are you in pain during certain activities" or something.*

Figure 1. Activity: "dressing and undressing".

Figure 2. Activity: "going to the toilet".

Differences in Professional and Layman Interpretations

Although "getting up and sitting down" and "getting in and out of a car" are different activities from a physical therapist's perspective, these are very similar movements from the perspective of moderately educated Bob, who stated:

> *Well "getting up and sitting down" and "getting in and out of a car" are kind of the same activities to me. So it is hard for me to say which one is more important in answer to question 6 and 7. I know I selected "getting in and out of the car" as the most important activity when I was completing the questionnaire. But when I would have to choose again I would go with "getting up and sitting down," because that is more generic and therefore it occurs more frequently in daily life.*

Interpreting the Numeric Rating Scale

Low-educated Dora and Ronald scored the numeric rating scale of question 8 backwards. They interpreted 10 as "no effort" and 0 as "the most effort possible."

Problems With Retrieval

A total of 7 respondents of all educational groups had problems retrieving information during their response processes.

Lack of Retrieval Because of the Form of Answering Options

A total of 4 respondents did not retrieve information because of the lack of answering options. They looked at the body chart of question 2 and the pictures of question 4 and searched their memory for any health problems related to the answering options. As a result, existing health problems that were *not* associated with the given answering options were *not* retrieved from memory. After Harald finished completing the questionnaire, he told the researcher that he was impaired in pulling objects, which is not a given answering option in question 4. Harald stated:

> *I did not miss it while I was completing the questionnaire. I probably thought that that picture would come later or in another category or something. I don't know. I did not really notice that it wasn't there.*

Lack of Retrieval Because of Memory Issues

In three cases, the root of the problem seemed to be a memory issue, which was not related to the content or form of the questionnaire. Ellen, for instance, described to the interviewer why she selected the activities "lifting" and "carrying" in answer to question 4. During this description, her recollection of the health problem became clearer. This made her realize in hindsight that "picking something up from the floor" would have been a better answer.

Problems With Judgment

A total of 9 respondents of all educational levels encountered problems with judgment.

Retrieved Information Judged as "Adequate to Answer the Question" Was Not Related to Physical Therapy (Anymore)

All 8 respondents indicated health problems that were not part of their treatment goal for physical therapy (anymore). Bob, for instance, indicated on the body chart that he had pain in his neck and shoulders, *and* he had low back pain. During the interview, he told the researcher that his neck and shoulder pain were chronic and existed for many years now. He did not believe it would be of any use for the physical therapist to put effort into trying to ease this pain. Therefore, it was not a part of his treatment goals. He was seeking help from his physical therapist for his acute low back pain.

Problems With Response Selection

A total of 8 respondents of all educational levels had problems with response selection. Bernie encountered this problem with two different questions of the questionnaire.

Not Able to Select the Right Answering Option Because These Options Do Not Match the Respondent's Response to the Question

All 8 respondents had problems with response selection because the response items did not match their answer(s). Bernie for instance had a complaint that was not "touchable" or located at a particular part of the body. But he was forced to place a dot on the body chart to be able to go on to the next question. Bernie stated:

> *This is not right at all! It says: "tap on the location of your health problem." But then one has to be able to locate his complaints. I can't. The way I walk does not feel normal to me, it does not feel the way it used to feel. I cannot say that I feel it "in my legs." It really is the movement itself that feels "off." I go to the physical therapist to find out what causes this. So at this moment I don't know where the root of the problem is located. Because I am forced to point out a location and the legs are clearly involved in walking, I have put a dot on the legs. But it is just not right. I mean, when I would have had pain in my hand I could have answered this question. If I would have felt it in my foot I would have tapped on the foot. But in my case it is about the movement...*

Discussion

Principal Findings

Of the 24 respondents, 20 encountered one or more problems during their response process. No problems were experienced with comprehension of text or wording. Most problems arose with (1) Interpretation of pictures and plain language, (2) Respondents not retrieving health problems that were *not* associated with the given answering options, and (3) Respondents judging retrieved health problems as relevant, although these were not related to their physical therapy treatment goals. No educational group in this research population stood out from the rest in the kind or number of problems that arose.

Despite the fact that 20 respondents did not respond to each question in the way that was intended by its developers, all respondents recognized themselves in the outcomes of the questionnaire shown in a screen summary.

Comparison With Prior Work

The clarity of text and wording seems to be better in the DTTSQ than in the PSC questionnaire [17], which was used as a starting point for development. In the study on the response process of the PSC questionnaire, "comprehension" and "interpretation" were put together into one category called "problems with reading and comprehending the questionnaire" [19]. Due to the way in which the data was collected in the PSC questionnaire study (lacking a think-aloud component), even in hindsight it is not always possible to determine if the source of each "problem with reading and comprehending the questionnaire" was comprehension or interpretation. This makes the PSC questionnaire and DTTSQ studies not fully comparable in this respect. Still, little over half of the respondents in the DTTSQ study versus all respondents in the PSC questionnaire study had comprehension and/or interpretation problems.

Invalid answers were reported in 52% (13/25) of the Dutch subjects in the PSC questionnaire study [19]. In this study, the percentage of respondents that gave one or more invalid answers was much higher: 83% (20/24) cases. Again the data of these two studies are not fully comparable. The PSC questionnaire study did not contain a think-aloud component. Having a think-aloud component in a study tends to add data on validity of answers, while at the same time there is no loss of data in comparison to studies without a think-aloud component [21]. This may be an explanation for the considerable difference between the amount of invalid answers found between the two studies.

Except for the problems caused by the use of plain language, using pictures as answering options and showing questions on separate screens without a back function, the problems found in this study were not new or exclusive for the DTTSQ. Problems such as "differences in layman and professional perspective" and "memory issues" are commonly seen in comparable studies and well documented in Tourangeau's book "The Psychology of Survey Response" [23,24,27,28].

Problems Caused by the Use of Plain Language

Four out of 6 respondents that misinterpreted question 1 of the DTTSQ "Do you have pain" mentioned the short and simple formulation of this question as the root of the problem. The formulation of question 1 and the layout of the screen on which it was shown was in line with the "European Easy-to-Read Guidelines" [29]. With the formulation of this question, however, the developers of the questionnaire may not have done enough justice to the complex concept of pain. It may be necessary to provide more detailed background information on the purpose and focus of the question [30]. Considering that understanding *spoken* language is easier to people than understanding *written* language [31], it might be recommended to add information by using a voice-over. In this way, information on the purpose and focus of the question and/or answering option(s) can be given without making the reading task more difficult [32].

Problems Caused by the Design of the User Interface

Use of Pictures

In addition to plain language, pictures were used to contribute to the comprehensibility of the questionnaire. Respondents' interpretation of the pictures did not always match the intended meaning by its developer. Optimizing this match by testing the interpretation of newly developed pictures in the target population before they are used in the questionnaire is recommended during the further development of the DTTSQ.

Showing All Questions on Separate Screens

The questions of the DTTSQ were shown in separate screens, and respondents were not able to go back to earlier screens. This makes the response process different from that of paper-based questionnaires in which respondents are able to oversee the whole questionnaire, choose the order in which they answer questions, and go back and forth between questions. The answering options of question 4 were subdivided into eight categories shown on eight separate screens. Lacking the complete overview of all answering options may have complicated the decision on whether or not to select an activity because the respondent was not able to see whether or not pictures in coming screens would be a better fit. Giving a complete overview of all answering options, for instance by presenting them as thumbnail images [33] and providing a back option, may help to reduce the amount of problems with response selection.

Limitations

This study was not designed to reach data saturation. The goal was to get a first impression of the response processes of respondents with diverse educational levels completing the prototype of the questionnaire to be able to make informed choices in further development of the questionnaire. Because twenty-four cases were included in this study, it can be assumed that the most common problems have been exposed [34].

Conclusions

The use of plain language and ICT within the DTTSQ has had positive and negative influences on the response processes of the research population.

Results of recent reviews and articles on the comparability of paper-based and electronic versions of questionnaires may give the impression that digitalizing questionnaires can be done without influencing psychometric properties [35-39] and response rates [40-44]. This is true when the digital version is a near copy of the paper-based questionnaire in terms of content and layout. But in an era in which the use of plain language and "inclusive design" or "electronic health for all" [45,46] is being advocated increasingly [47,48], copying the content and layout of the original into the digital version may not be enough.

The results of this study emphasize the importance of two basic recommendations, which are as follows: (1) accompany any adaption of any questionnaire to a new mode of delivery by evidence, demonstrating the difference and equivalence between

the two different modes [49] and (2) scientifically evaluate the applicability of the newly developed mode of the questionnaire in its intended setting, to assess if it meets the standard criteria of validity, reproducibility, and feasibility [50]. Such studies should be designed and executed in a way that suits the (in)abilities of the target population of the questionnaire that is being evaluated. Like the qualitative method chosen in this study suited the (in)abilities of low-educated and/or low-literate participants by not demanding any reading or writing skills from study participants.

Acknowledgments

The authors would like to thank Kees van der Veer, PhD, and Jan Pool, PhD, for helping them think about the application of the TSTI method to low-educated populations, and Anita Cremers, PhD, for her reflections on the data from a designer's perspective. They are grateful to Vera Cornelissen and Diane Pijl for their contributions as peer reviewers.

Conflicts of Interest

None declared.

Multimedia Appendix 1

An overview of the screenshots of the Dutch Talking Touch Screen Questionnaire.

[PDF File (Adobe PDF File), 1MB - jmir_v20i4e140_app1.pdf]

References

1. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769 [FREE Full text] [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
2. de Boer D, Delnoij D, Rademakers J. The importance of patient-centered care for various patient groups. Patient Educ Couns 2013 Mar;90(3):405-410. [doi: 10.1016/j.pec.2011.10.002] [Medline: 22079211]
3. Saha S, Beach MC. The impact of patient-centered communication on patients' decision making and evaluations of physicians: a randomized study using video vignettes. Patient Educ Couns 2011 Sep;84(3):386-392 [FREE Full text] [doi: 10.1016/j.pec.2011.04.023] [Medline: 21600723]
4. Bensing J. Bridging the gap. The separate worlds of evidence-based medicine and patient-centered medicine. Patient Educ Couns 2000 Jan;39(1):17-25. [Medline: 11013544]
5. Mead N, Bower P. Patient-centredness: a conceptual framework and review of the empirical literature. Soc Sci Med 2000 Oct;51(7):1087-1110. [Medline: 11005395]
6. Williams MV, Davis T, Parker RM, Weiss BD. The role of health literacy in patient-physician communication. Fam Med 2002 May;34(5):383-389. [Medline: 12038721]
7. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31 Suppl 1:S19-S26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]
8. Nelson EC, Hvitfeldt H, Reid R, Grossman D, Lindblad S, Mastanduno MP, et al. Using Patient-Reported Information to Improve Health Outcomes and Health Care Value: Case studies from Dartmouth, Karolinska and Group Health. New Hampshire: The Darthmouth Institute for Health Policy and Clinical Practice; 2012.
9. Calvert M, Brundage M, Jacobsen PB, Schünemann HJ, Efficace F. The CONSORT patient-reported outcome (PRO) extension: implications for clinical trials and practice. Health Qual Life Outcomes 2013 Oct 29;11:184 [FREE Full text] [doi: 10.1186/1477-7525-11-184] [Medline: 24168680]
10. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di DD, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014 May 10;32(14):1480-1501. [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]
11. Hahn EA, Cella D. Health outcomes assessment in vulnerable populations: measurement challenges and recommendations. Arch Phys Med Rehabil 2003 Apr;84(4 Suppl 2):S35-S42. [doi: 10.1053/apmr.2003.50245] [Medline: 12692770]
12. Paiva CE, Siquelli FA, Zaia GR, de Andrade DA, Borges MA, Jácome AA, et al. Development of a new multimedia instrument to measure cancer-specific quality of life in Portuguese-speaking patients with varying literacy skills. Springer Plus 2016;5(1):972 [FREE Full text] [doi: 10.1186/s40064-016-2675-6] [Medline: 27429882]
13. El-Daly I, Ibraheim H, Rajakulendran K, Culpan P, Bates P. Are patient-reported outcome measures in orthopaedics easily read by patients? Clin Orthop Relat Res 2016 Jan;474(1):246-255 [FREE Full text] [doi: 10.1007/s11999-015-4595-0] [Medline: 26472587]
14. Onderwijsincijfers. Highest level of education achieved URL: https://www.onderwijsincijfers.nl/kengetallen/sectoroverstijgend/nederlands-onderwijsstelsel/hoogst-behaalde-onderwijsniveau [WebCite Cache ID 6vLM3S0Hc]

15. 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]
16. Swinkels RAHM, van Peppen RP, Wittink H, Custers JWH, Beurskens AJHM. Current use and barriers and facilitators for implementation of standardised measures in physical therapy in the Netherlands. BMC Musculoskelet Disord 2011 May 22;12:106. [doi: 10.1186/1471-2474-12-106] [Medline: 21600045]
17. Beurskens AJ, de Vet HC, Köke AJ, Lindeman E, van der Heijden GJ, Regtop W, et al. A patient-specific approach for measuring functional status in low back pain. J Manipulative Physiol Ther 1999;22(3):144-148. [Medline: 10220712]
18. Beurskens AJ, de Vet HC, Köke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 1996 Apr;65(1):71-76. [Medline: 8826492]
19. Welbie M, Wittink H, Westerman M, Devillé W. Completing the patient specific-complaint questionnaire in physical therapy practice is problematic for high and low literate patients: a qualitative study. Int J Pers Cent Med 2016;6(1):-. [doi: 10.5750/ijpcm.v6i1.566]
20. Cremers AH, Welbie M, Kranenborg K, Wittink H. Deriving guidelines for designing interactive questionnaires for low-literate persons: development of a health assessment questionnaire. Univ Access Inf Soc 2015 Sep 30;16(1):161-172. [doi: 10.1007/s10209-015-0431-2]
21. Hak T, van der Veer K, Jansen H. The three-step test-interview (TSTI): an observation-based method for pretesting self-completion questionnaires. Surv Res Methods 2008;2(3):50. [doi: 10.18148/srm/2008.v2i3.1669]
22. Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res 2003 May;12(3):229-238. [Medline: 12769135]
23. ten Velden M, Couldrick L, Kinébanian A, Sadlo G. Dutch children's perspectives on the constructs of the child occupational self-assessment (COSA). OTJR 2012 Aug 24;33(1):50-58. [doi: 10.3928/15394492-20120817-01]
24. Liu RD, Buffart LM, Kersten MJ, Spiering M, Brug J, van Mechelen W, et al. Psychometric properties of two physical activity questionnaires, the AQuAA and the PASE, in cancer patients. BMC Med Res Methodol 2011 Mar 16;11:30 [FREE Full text] [doi: 10.1186/1471-2288-11-30] [Medline: 21410979]
25. Johnson C, Aaronson N, Blazeby JM, Bottomley A, Fayers P, Koller M, et al. EORTC. Brussels: EORTC Quality of Life Group; 2011. Guidelines for developing questionnaire modules URL: http://groups.eortc.be/qol/sites/default/files/archives/guidelines_for_developing_questionnaire-_final.pdf [accessed 2018-03-29] [WebCite Cache ID 6yHAB7sHU]
26. Braun V, Clarke V. Using thematic analysis in psychology. Quol Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Tourangeau R, Rasinski K. The psychology of survey response. Cambridge: Cambridge University Press; 2000.
28. Pool JJM, Hiralal SR, Ostelo RWJG, van der Veer K, de Vet HC. Added value of qualitative studies in the development of health related patient reported outcomes such as the Pain Coping and Cognition List in patients with sub-acute neck pain. Man Ther 2010 Feb;15(1):43-47. [doi: 10.1016/j.math.2009.05.010] [Medline: 19632876]
29. Freyhoff G, Hess G, Kerr L, Menzel E, Tronbacke B, Van der Veken K. ILR. 1998. Make it simple, European guidelines for the production of easy-to-read information for people with learning disability; for authors, editors, information providers, translators and other interested persons URL: https://digitalcommons.ilr.cornell.edu/gladnetcollect/270/ [accessed 2018-03-13] [WebCite Cache ID 6xtDxINY9]
30. MSKTC. 2014. Writing and testing plain language URL: http://www.msktc.org/lib/docs/KT_Toolkit/MSKTC_Plain_Lang_Tool_508.pdf [accessed 2018-03-13] [WebCite Cache ID 6xtELICLY]
31. Sousa DA. How the brain learns to read. California: Corwin - A SAGE compagny; 2014.
32. Johnson J. Designing with the mind in mind. Waltham, MA: Elsivier Inc; 2014.
33. Johnson J, Roberts T, Verplank W, Smith D, Irby C, Beard M, et al. The Xerox star: a retrospective. Computer 1989 Sep;22(9):11-26. [doi: 10.1109/2.35211]
34. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. New York: ACM Press; 1993 Presented at: CHI'93 Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems; April 24-29, 1993; Amsterdam, The Netherlands p. 206-213 URL: http://portal.acm.org/citation.cfm?doid=169059.169166 [doi: 10.1145/169059.169166]
35. Norquist J, Chirovsky D, Munshi T, Tolley C, Panter C, Gater A. Assessing the comparability of paper and electronic versions of the EORTC QOL module for head and neck cancer: a qualitative study. JMIR Cancer 2017 May 12;3(1):e7 [FREE Full text] [doi: 10.2196/cancer.7202] [Medline: 28500019]
36. Lundy JJ, Coons SJ, Aaronson NK. Testing the measurement equivalence of paper and interactive voice response system versions of the EORTC QLQ-C30. Qual Life Res 2014 Feb;23(1):229-237. [doi: 10.1007/s11136-013-0454-1] [Medline: 23765449]
37. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. Qual Life Res 2015 Aug;24(8):1949-1961. [doi: 10.1007/s11136-015-0937-3] [Medline: 25702266]
38. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015;13:167. [doi: 10.1186/s12955-015-0362-xMedline:]

39. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008 Mar;11(2):322-333 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
40. Horevoorts NJ, Vissers PA, Mols F, Thong MS, van de Poll-Franse LV. Response rates for patient-reported outcomes using web-based versus paper questionnaires: comparison of two invitational methods in older colorectal cancer patients. J Med Internet Res 2015;17(5):e111 [FREE Full text] [doi: 10.2196/jmir.3741] [Medline: 25953059]
41. van den Berg MH, Overbeek A, van der Pal HJ, Versluys AB, Bresters D, van Leeuwen FE, et al. Using web-based and paper-based questionnaires for collecting data on fertility issues among female childhood cancer survivors: differences in response characteristics. J Med Internet Res 2011;13(3):e76 [FREE Full text] [doi: 10.2196/jmir.1707] [Medline: 21955527]
42. Kilsdonk E, van Dulmen-den Broeder E, van der Pal HJ, Hollema N, Kremer LC, van den Heuvel-Eibrink MM, et al. Effect of web-based versus paper-based questionnaires and follow-up strategies on participation rates of Dutch childhood cancer survivors: a randomized controlled trial. JMIR Cancer 2015 Nov 24;1(2):e11 [FREE Full text] [doi: 10.2196/cancer.3905] [Medline: 28410161]
43. Duracinsky M, Lalanne C, Goujard C, Herrmann S, Cheung-Lung C, Brosseau J, et al. Electronic versus paper-based assessment of health-related quality of life specific to HIV disease: reliability study of the PROQOL-HIV questionnaire. J Med Internet Res 2014;16(4):e115 [FREE Full text] [doi: 10.2196/jmir.3330] [Medline: 24769643]
44. Hohwü L, Lyshol H, Gissler M, Jonsson SH, Petzold M, Obel C. Web-based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. J Med Internet Res 2013;15(8):e173 [FREE Full text] [doi: 10.2196/jmir.2595] [Medline: 23978658]
45. World Health Organization. 2006. Building foundations of ehealth: progress of member states: report of the WHO global observatory for ehealth URL: http://www.who.int/goe/publications/bf_FINAL.pdf [accessed 2018-03-13] [WebCite Cache ID 6xtFMk4gy]
46. McHattie L, Cumming G, French T. Transforming patient experience: health web science meets medicine 2.0. Med 2 0 2014;3(1):e2 [FREE Full text] [doi: 10.2196/med20.3128] [Medline: 25075246]
47. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. 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. [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
48. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998 Feb;25(2):198-199. [Medline: 9489805]
49. Stableford S, Mettger W. Plain language: a strategic response to the health literacy challenge. J Public Health Policy 2007;28(1):71-93. [doi: 10.1057/palgrave.jphp.3200102] [Medline: 17363939]
50. Trudeau C. Plain language in healthcare: what lawyers need to know about health literacy. Mich B J 2016;10:36-39 [FREE Full text]

Abbreviations

DTTSQ: Dutch Talking Touch Screen Questionnaire
HCP: health care provider
ICT: information and communication technology
PSC: Patient Specific-Complaint
TSTI: Three-Step Test-Interview

Edited by G Eysenbach; submitted 28.04.17; peer-reviewed by S Rubinelli, JR Bautista; comments to author 29.07.17; revised version received 22.12.17; accepted 31.01.18; published 23.04.18.

Please cite as:
Welbie M, Wittink H, Westerman MJ, Topper I, Snoei J, Devillé WLJM
Using Plain Language and Adding Communication Technology to an Existing Health-Related Questionnaire to Help Generate Accurate Information: Qualitative Study
J Med Internet Res 2018;20(4):e140
URL: http://www.jmir.org/2018/4/e140/
doi:10.2196/jmir.7940
PMID:29685873

Original Paper

ComprehENotes, an Instrument to Assess Patient Reading Comprehension of Electronic Health Record Notes: Development and Validation

John P Lalor[1], MS; Hao Wu[2], PhD; Li Chen[2], MS; Kathleen M Mazor[3], EdD; Hong Yu[1,4,5,6], PhD, FACMI

[1]College of Information and Computer Sciences, University of Massachusetts, Amherst, MA, United States
[2]Psychology Department, Boston College, Chestnut Hill, MA, United States
[3]Meyers Primary Care Institute, University of Massachusetts Medical School / Reliant Medical Group / Fallon Health, Worcester, MA, United States
[4]Department of Computer Science, University of Massachusetts, Lowell, MA, United States
[5]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States
[6]Bedford Veterans Affairs Medical Center, Center for Healthcare Organization and Implementation Research, Bedford, MA, United States

Corresponding Author:
Hong Yu, PhD, FACMI
Department of Medicine
University of Massachusetts Medical School
55 Lake Avenue North, AC7-059
Worcester, MA, 01605
United States
Phone: 1 508 856 3474
Email: hong.yu@umassmed.edu

Abstract

Background: Patient portals are widely adopted in the United States and allow millions of patients access to their electronic health records (EHRs), including their EHR clinical notes. A patient's ability to understand the information in the EHR is dependent on their overall health literacy. Although many tests of health literacy exist, none specifically focuses on EHR note comprehension.

Objective: The aim of this paper was to develop an instrument to assess patients' EHR note comprehension.

Methods: We identified 6 common diseases or conditions (heart failure, diabetes, cancer, hypertension, chronic obstructive pulmonary disease, and liver failure) and selected 5 representative EHR notes for each disease or condition. One note that did not contain natural language text was removed. Questions were generated from these notes using Sentence Verification Technique and were analyzed using item response theory (IRT) to identify a set of questions that represent a good test of ability for EHR note comprehension.

Results: Using Sentence Verification Technique, 154 questions were generated from the 29 EHR notes initially obtained. Of these, 83 were manually selected for inclusion in the Amazon Mechanical Turk crowdsourcing tasks and 55 were ultimately retained following IRT analysis. A follow-up validation with a second Amazon Mechanical Turk task and IRT analysis confirmed that the 55 questions test a latent ability dimension for EHR note comprehension. A short test of 14 items was created along with the 55-item test.

Conclusions: We developed ComprehENotes, an instrument for assessing EHR note comprehension from existing EHR notes, gathered responses using crowdsourcing, and used IRT to analyze those responses, thus resulting in a set of questions to measure EHR note comprehension. Crowdsourced responses from Amazon Mechanical Turk can be used to estimate item parameters and select a subset of items for inclusion in the test set using IRT. The final set of questions is the first test of EHR note comprehension.

(J Med Internet Res 2018;20(4):e139) doi:10.2196/jmir.9380

KEYWORDS

electronic health records; health literacy; psychometrics; crowdsourcing

Introduction

Background and Significance

Providing patients access to their medical records through personal health records (PHRs) is becoming more common as physicians move to electronic health record (EHR) systems. PHRs are defined as "electronic, lifelong resource of health information needed by individuals to make health decisions" [1]. Providing patients direct access to their EHR clinical notes can enhance patients' understanding of their clinical conditions and improve their health care outcomes [2-4]. For example, the Veterans Health Administration offers the My Health*e*Vet PHR through a Web-based patient portal, which allows millions of veterans to view their EHRs [5]. These records include both structured (eg, patient vitals) and unstructured data (eg, discharge summaries and clinical notes). However, patients with limited health literacy may struggle to understand the content of their medical notes, which can include visit summaries with medical terms, lab reports, and terms and phrases that are not common outside of medicine. A patient's health literacy can have an impact on their desire to engage with their own PHR [6,7].

Low health literacy can impact a patient's ability to communicate with their health care providers and to navigate and understand complex EHR information. Health literacy is defined by the Institute of Medicine as "the degree to which individuals have the capacity to obtain, process, and understand basic information and services needed to make appropriate decisions regarding their health" [8]. According to the National Assessment of Adult Literacy, only 12% of adults are proficient in health literacy [9]. The average American reads at or below an eighth grade level, and over 90 million Americans have limited health literacy [9]. Moreover, 50% of patients do not understand at least one term in their medical problem list [8,10,11]. In addition, EHR notes do not align well with existing readability prediction formulas, making it difficult to estimate EHR note readability [12]. Consider the following example, taken from a de-identified EHR clinical note: "The *monitor* has not shown any *dysrhythmias* or *arrhythmia* either *prior to or during any of his spells."* A patient might struggle to understand the medical terms *dysrhythmias* and *arrhythmia* and might not understand what the *monitor* is or what *prior to or during any of his spells* is referring to.

Low health literacy can lead to serious problems. For example, low health literacy was shown to be independently associated with an increase in mortality among the elderly [13]. A recent assessment of health literacy involving over 400 Veterans found that 87% of Veterans have low health literacy [14]. Most health care consumers do not understand phrases often used in cancer consultations [15]. Patients understand less than 30% of medical terms commonly used in the emergency department [16]. Patients with low health literacy are more likely to lack awareness of their atrial fibrillation diagnosis [17] and are at higher risk for increased fear of cancer progression [18].

Objective

Given the prevalence of low health literacy in the population, tools that effectively assess a patient's health literacy are needed for both research and practice. Of the existing instruments, 3 that are widely used are the Rapid Estimate of Adult Literacy in Medicine (REALM), the Test of Functional Health Literacy in Adults (TOFHLA), and the Newest Vital Sign (NVS) [19-21]. Each of these has value, but also limitations. For example, REALM can be administered in 2 to 3 min, but it assesses word recognition, not comprehension [19]. TOFHLA assesses reading comprehension and numeracy using passages from health care–related documents, hospital forms, and prescription labels [20]; a short version of TOFHLA reduced the administration time from 22 min to 12 min [22]. NVS contains 6 items tied to a single stimulus (a food label) and can be administered in 3 min. It was intended as a screening tool and is less appropriate for generating scores that discriminate between different levels of health literacy in patients [21,23]. Taken together, these tests can provide information on a patient's general health literacy, but none assesses a patient's ability to comprehend EHR notes.

The purpose of this study was to create an instrument to measure EHR note comprehension in patients. We first identified a set of representative EHR notes for 6 diseases and conditions from a large hospital EHR system. From these notes, a group of physicians and medical researchers generated questions using the Sentence Verification Technique (SVT) [24-26]. We obtained responses for these questions from the crowdsourcing platform Amazon Mechanical Turk (AMT) and analyzed the results using item response theory (IRT) [27-30] to select a subset of questions for a test of EHR note comprehension. To the best of our knowledge, the ComprehENotes question set is the first instrument to assess EHR note comprehension.

Methods

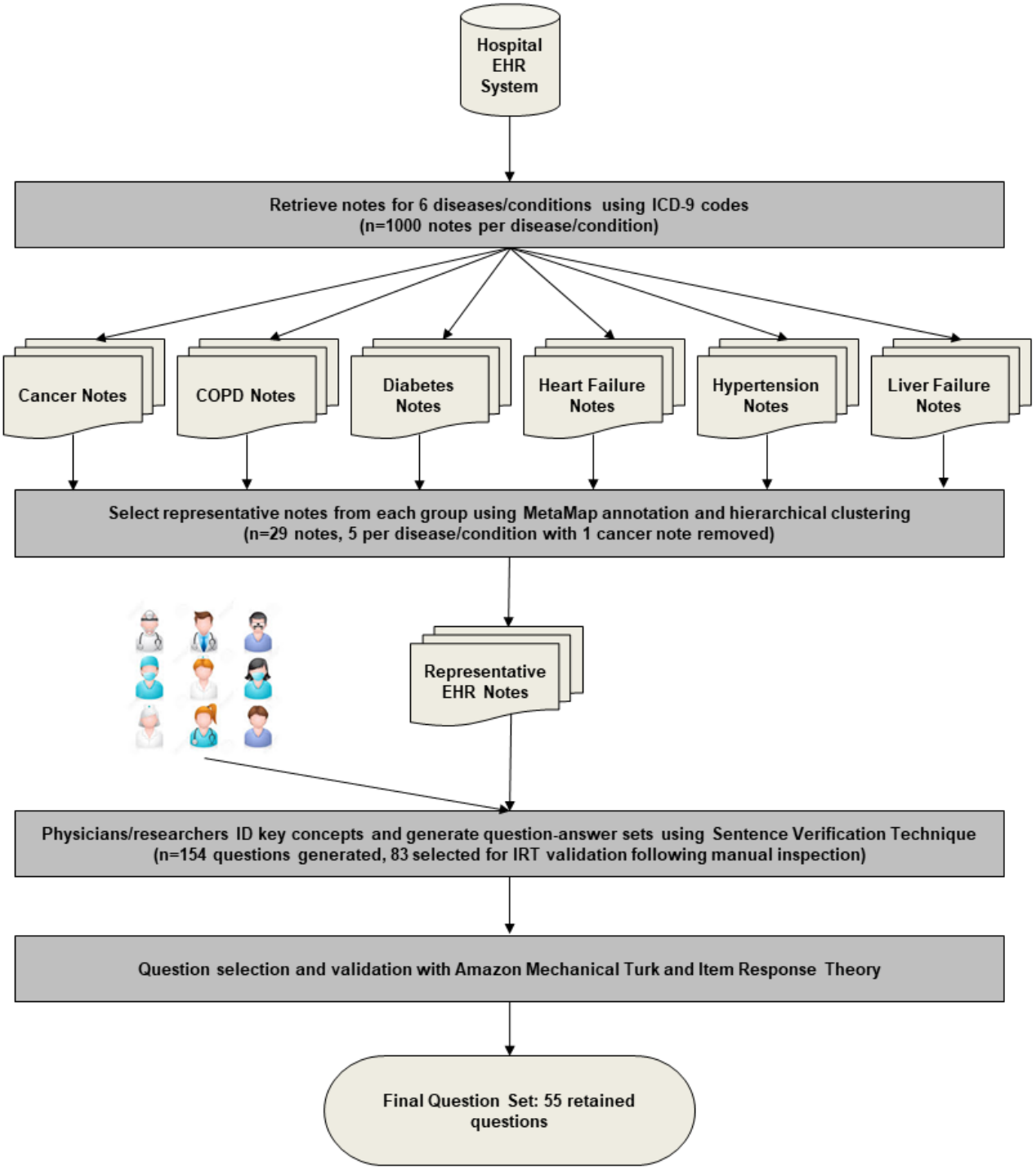
Overview

The goal of this work was to develop a set of questions that could be used to test patient EHR note comprehension. To that end, we developed a process for note selection, question generation, and question selection and validation (Figure 1). We discuss each step in detail in the following sections.

Electronic Health Record Note Selection

We selected notes according to the International Classification of Disease codes associated with 6 important and common diseases: heart failure (428), hypertension (401), diabetes (249, 250), chronic obstructive pulmonary disease (COPD; 493.2, 491, 492, 494, 496, 506), liver failure (571), and cancer (140-239). By selecting notes from multiple diseases, our goal was to obtain a variety of notes associated with common diseases to generate questions across multiple topics.

Figure 1. Visualization of the question generation and validation process. COPD: chronic obstructive pulmonary disease; EHR: electronic health record; ICD-9: International Classification of Disease-9; IRT: item response theory.



We retrieved EHR discharge summary and progress notes from the University of Massachusetts Memorial Hospital EHR system. Progress notes provide information regarding a patient's conditions and treatments, whereas discharge notes may include a summary of the patient's visit, necessary patient follow-up, and other information. These types of notes include information that is relevant to patients and are good candidates for question generation. For each disease, we randomly selected 1000 notes. As the EHR notes vary significantly in length (anywhere from 50 words to over 1500 words), we limited the note selection to notes between 300 and 1000 words long. Notes that are longer than 1000 words often contain duplicate information or large tables of lab results, with few free-text section from which we can generate questions. We annotated each note with the MetaMap [31], a toolkit developed by the National Library of Medicine, to map the note to Unified Medical Language System (UMLS) concepts [32]. For each category, we ran topic modeling on the 1000 notes using the UMLS concepts that were identified by MetaMap and hierarchically clustered the notes into 5 clusters based on topic similarities. Finally, we selected 1 representative note (the note with the most UMLS concepts) from each cluster. By selecting the note with the most concepts,

our goal was to identify those notes with the most information that could be used as part of the question generation process. This procedure resulted in a total of 30 notes, with 5 notes per disease. We discarded 1 cancer note because the physicians identified it as a pure lab test report that did not include any natural language text.

Generating Questions With Sentence Verification Technique

We asked experts to create question-answer sets by following these 2 steps: (1) identifying important content in the notes and (2) creating comprehension test questions. Specifically, the selected 29 de-identified notes were provided to 5 groups. Each group included 1 physician and 2 to 3 nonclinician researchers (a total of 4 physicians and 13 researchers, where 1 physician participated in 2 groups). The groups were given an introduction to the SVT methodology before taking part in the exercise. Each member read every assigned EHR note and then identified important content (usually a sentence). Each member then followed the SVT protocol to create question-answer sets for the identified content.

SVT is a procedure for generating reading comprehension items to evaluate whether an individual has understood a passage of text [24,33,34]. SVT has been applied in many different reading comprehension environments, such as basic language research [35], evaluating the effect of prior beliefs on comprehension [36], and assessing language skills of non-native English speakers [37]. In addition, SVT has been used to develop tests to assess comprehension of cancer screening and prevention messages [25,26]. SVT tests are sensitive to both differences in reading skill and text difficulty. Tests using SVT questions have been shown to be effective for measuring reading comprehension and for assessing comprehension of written and spoken health messages [25,26].

An SVT test is designed by taking a sentence or phrase from a passage of text (the *original*) and generating 3 additional sentences or phrases: (1) a *paraphrase*, where as much of the sentence or phrase is changed as possible while preserving the original meaning, (2) a *meaning change*, where the original sentence or phrase is changed slightly but enough that the original meaning is changed, and (3) a *distractor*, which is unrelated to the original but still consistent with the passage theme [24].

Once generated, the question-answer sets were then discussed in the group and a final question-answer set was agreed upon. From the 29 EHR notes, 154 question-answer sets were generated. Table 1 shows examples of question-answer sets generated by the groups, and Textbox 1 shows how these questions would be presented to patients in a test scenario. We selected 83 of the 154 questions for further analysis. Questions were selected based on their content. We manually selected questions that were generally relevant to the main topic (eg, diabetes) over questions that were very specific to a patient's note to keep the question set general enough to be given to future patients. We retained 11 to 13 question-answer sets for 4 of the 6 topics and 18 question-answer sets for COPD and diabetes.

Data Collection

To gather enough human responses to fit the IRT model, we recruited participants from AMT. AMT is a Web-based microtask crowdsourcing platform where individuals (called Turkers) perform Human Intelligence Tasks (HITs) in exchange for payment. HITs are usually pieces of larger, more complex tasks that have been broken up into multiple, smaller subtasks. AMT and other crowdsourcing platforms are used to build large corpora of human-labeled data at low cost compared with using expert annotators [38,39]. Researchers' projects have used AMT to complete a variety of tasks [40,41]. Recent research has shown that AMT and other crowdsourcing platforms can be used to generate corpora for clinical natural language processing and disease mention annotation [41,42]. AMT was used to detect errors in a medical ontology, and it was found that the crowd was as effective as the domain experts [43]. In addition, AMT workers were engaged in identifying disease mentions in PubMed abstracts [42] and rank adverse drug reactions in order of severity [44] with good results.

We created 6 comprehension tasks on AMT, 1 per disease topic, to analyze each topic separately. Each task was completed by 250 Turkers, who were presented with the test questions, 1 question at a time. This sample size is large enough to satisfy the accepted standards for IRT models based on the noncentral chi-square distribution [45]. We collected demographic information from the Turkers before administering the test questions, and we implemented several quality control mechanisms to ensure the quality of the Turker results. Only Turkers with approval rates above 95% and located in the United States were able to participate. The 95% approval rate identifies Turkers who have been approved most of the time according to their completion of other tasks on AMT and is indicative of the high quality of their previous tasks. Restricting the task to users located in the United States is used as a proxy for English proficiency. In addition, in each test, 1 question was randomly selected as a quality-check question and was presented to the Turker twice during the course of the evaluation. If the Turker gave 2 different answers to the repeated question, their responses were not included in later analyses. Two simple questions were also added to the test as quality control. If the Turker answered 1 or both of the quality control questions incorrectly, their responses were rejected from consideration and not included in later analyses.

For the COPD and diabetes tests, the 18 questions were split into 3 groups of 6 questions. Each Turker was given a random selection of 2 of the 3 groups. In this way, the test lengths were similar to the other disease topic tests, and the conditions in which Turkers provided responses were consistent across the groups. For the COPD and diabetes tasks, we recruited 400 Turkers so that the number of responses per question was consistent with the other topics.

Table 1. Examples of questions generated from the researcher and physician groups.

Original statement from EHR[a] notes	Paraphrase	Meaning change	Distractor
The monitor has not shown any dysrhythmias or arrhythmia either before or during any of his spells	His heart rhythm is normal before and during his fainting spells	He has had abnormal rhythm before or during his spells of chest pain	The monitor has shown abnormal heart rhythms before and during his spells
Patient recently presented to the hospital with shortness of breath	She went to the hospital for trouble breathing	She visited the clinic due to shortness of iron	Shortness of breath has many causes

[a]EHR: electronic health record.

Textbox 1. Examples of how the generated questions would be displayed as a questionnaire, using the examples from Table 1.

Please read the following question and then examine the answer choices and choose the answer that best represents the question text.

What does the following sentence mean? "The monitor has not shown any dysrhythmias or arrhythmia either prior to or during any of his spells."

1. He has had abnormal rhythm before or during his spells of chest pain.
2. The monitor has shown abnormal heart rhythms before and during his spells.
3. His heart rhythm is normal before and during his fainting spells.

What does the following sentence mean? "Patient recently presented to the hospital with shortness of breath."

1. Shortness of breath has many causes.
2. She went to the hospital for trouble breathing.
3. She visited the clinic due to shortness of iron.

Item Analysis and Selection Using Item Response Theory

After data collection, the Turker responses were analyzed using a 3-parameter logistic (3PL) IRT model. IRT [27,46] is one of the most widely used approaches for item evaluation and test construction [29,30,47]. For example, the Patient Reported Outcomes Measurement Information System funded by the National Institutes of Health has used IRT to characterize item banks and to support computerized adaptive testing [28].

In IRT, a statistical model jointly models an individual's responses to individual test items with a person's ability level and the item's features [27]. IRT models make several assumptions: (1) people differ from each other on an unobserved latent dimension of interest (usually called *ability*); (2) the probability of correctly answering a particular item is a function of the latent ability dimension (the item characteristic curve, ICC); (3) responses to individual items are independent of each other for a given ability level of a person (the *local independence assumption*); and (4) responses from different individuals are independent of each other. There are a variety of IRT models; one of the models widely used is the 3PL model. In the 3PL model, ICC is assumed to follow a logistic function with a nonzero lower asymptote:

In the above equation, p_{ij} is the probability that person j answers item i correctly, and θ_j is the ability level of individual j. In this work, θ represents the ability of an individual on the task of EHR note comprehension. As individual persons are assumed to be sampled from a population, their ability levels are assumed to be a random effect with a normal distribution. There are also 3 item parameters: the guessing parameter c_i is the lower asymptote of the ICC curve and represents the probability of guessing, the difficulty parameter b_i is the level of ability that produces a chance of correct response equal to the average of the upper and lower asymptotes, and the slope or discrimination parameter a_i is related to the steepness of the curve.

The 3PL model was fit to data for each set of questions using the open source software R packages *mirt* and *ltm* [48,49]. Marginal residuals of each pair of items and each triplet of items were checked, and items that gave large residuals were removed for violation of local independence. Items with a negative slope were also removed. Guessing parameters not significantly different from 0 were set to 0. A key parameter used to identify a good question for future evaluations is the slope of ICC. If the slope is flat, then the item cannot distinguish between individuals of high ability levels and individuals of low ability levels. After refitting the remaining items, items with a slope parameter not significantly greater than 0 or less than 0.71 were removed. The value 0.71 corresponds to a communality of 0.15 in an exploratory factor analysis, which means that 15% of the variance of the item would be explained by the latent ability factor if the item were continuous. We retained 55 items in this analysis for further validation. From the 55 items, we also identified 14 of the 55 items with the largest slopes (discrimination parameters) and highest average information for inclusion in the short form of the test. The short test should be as informative as possible while reducing the length of the test, making it more practical to administer.

Confirmatory Evaluation of Item Quality Using Item Response Theory

The questions retained from the initial IRT analysis were combined into a single test and deployed in a new AMT task to validate the item parameters. For this task, we split the 55 retained questions into 3 groups (each of 18-19 questions) and created 3 AMT tasks in which Turkers were shown 2 of the 3 groups and asked for responses as above. Quality checks were included as in the first set of AMT tasks. For these tasks, Turkers who participated in the initial data collection were excluded. Responses were generated and a second round of IRT analysis was performed to confirm that the questions retained from the first round could be considered a cohesive test of EHR note comprehension as a whole.

Results

Amazon Mechanical Turk Responses and Turker Demographics

We first report descriptive statistics and demographic information about the Turkers who completed the per-topic and validation AMT tasks (Figure 2; Table 2). Responses for both the per-topic and validation tasks covered a wide range of correctly answered questions. Mean scores for the cancer, COPD, diabetes, heart failure, hypertension, liver failure, and validation tasks were 69% (7.6/11), 78% (9.4/12), 88% (10.6/12), 70% (8.4/12), 78% (8.6/12), 79% (10.3/13), and 85% (31.4/37), respectively. Across all tasks, no more than 10.8% (27/250 for the heart failure task) of responses were removed because of quality control checks.

We also looked at raw scores and estimated ability in the validation task to see whether there were patterns in the responses that matched expected behavior (Table 3). As expected, mean scores for individuals with more education are higher than for individuals with less education. In addition, Turkers over 45 years score higher on average than Turkers under 45 years. There is a slight drop in mean scores for Turkers aged over 65 years, which makes sense given that adults aged 65 years and older have lower health literacy on average [9].

Item Response Theory Analysis

Item Selection Using Item Response Theory

Of the 83 questions provided to Turkers in the per-topic AMT tasks, 55 (66%) were retained after the initial IRT analysis (Figure 3). Items were identified for removal according to the procedure identified in the Methods section.

Figure 2. Box plots of Turker scores on the AMT per-topic and validation tasks. The center rectangles span the range from the first quartile to the third quartile of responses, and the bolded line inside each box represents the median score. Open circles indicate outlier scores. In the cancer plot, the upper and lower horizontal lines indicate the maximum and minimum scores, respectively. For all others, the lower horizontal line is 1.5 times the interquartile range below the first quartile. Average raw score is above 69% in all cases. Counts indicate the number of AMT responses retained after quality-control. AMT: Amazon Mechanical Turk; COPD: chronic obstructive pulmonary disease.

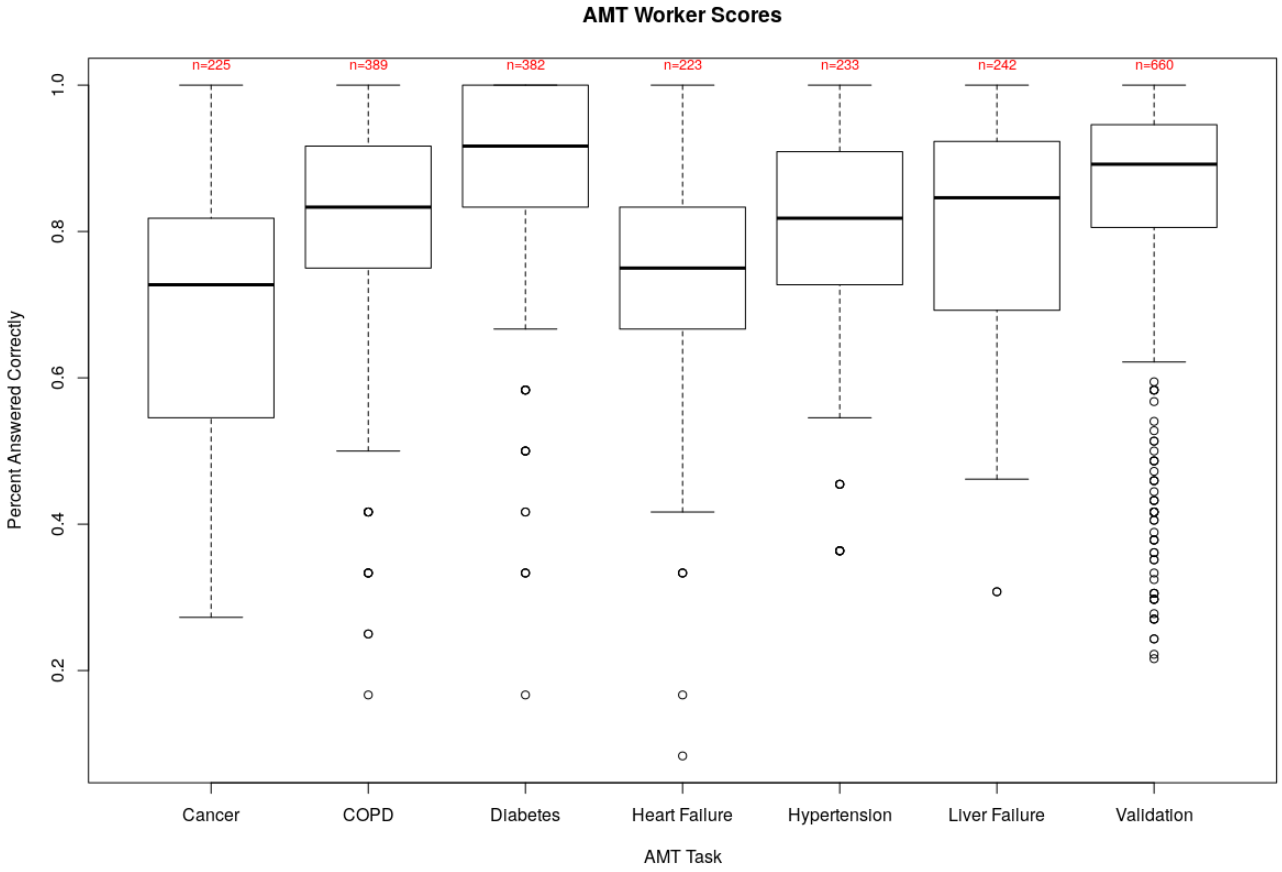


Table 2. Demographic information of Turkers from the per-topic and validation Amazon Mechanical Turk tasks.

Demographic characteristic	Per-topic tasks count (N=1694), n (%)	Validation task count (N=660), n (%)
Gender		
Male	880 (51.95)	250 (37.9)
Female	814 (48.05)	411 (62.1)
Race		
African American	107 (6.32)	58 (8.8)
Asian	163 (9.62)	51 (7.7)
Hispanic	89 (5.25)	32 (4.8)
American Indian	7 (0.41)	12 (1.8)
Pacific Islander	9 (0.53)	0 (0)
White	1319 (77.86)	507 (76.8)
Highest level of education		
Less than high school	17 (1.00)	4 (0.6)
High school degree	504 (29.75)	189 (28.6)
Associate's degree	283 (16.71)	108 (16.4)
Bachelor's degree	697 (41.15)	256 (38.8)
Master's degree or higher	193 (11.39)	103 (15.6)
Age in years[a]		
18-21	N/A[b]	12 (1.8)
22-34	N/A	330 (50.0)
35-44	N/A	158 (23.9)
45-54	N/A	106 (16.1)
55-64	N/A	39 (5.9)
65 and older	N/A	15 (2.3)

[a]Age demographic information was not collected as part of the per-topic Amazon Mechanical Turk tasks.

[b]N/A: not applicable.

Table 4 shows examples of retained and removed items. In the case of the removed item, the question simply defining the term *Osteoporosis* was too easy for the Turker population. That is, most of the Turkers answered the question correctly, and thus, the probability of answering the question correctly is very high even at low levels of ability. A question like this does not give us any information about an individual's ability and therefore is not needed in the test set.

The test information curve is presented in Figure 4. Test information is defined as the reciprocal of the squared SE of the ability estimate: $I=1/\sigma^2$, where σ is the SE [27]. Test information measures how accurate the ability estimates are at varying levels of ability. Given that most items have negative difficulty, the information curve has high values in the negative ability levels. That is, estimates of ability for negative ability levels are more accurate. Test information is greater than 4 for the range of ability levels between −2.8 and 0.7, which means for this range of ability levels (from 2.8 SDs below to 0.7 SD above the average of the population of AMT users), SE of an ability estimate is smaller than 0.5. The full test is most informative in ability around −2 with maximum information of 44.2 (Figure 4, red dotted line). This maximum is mostly because of a single item (44) with the largest slope of 11.3. Due to the very large slope parameter, this item is very informative around ability of −2 but is not informative at other areas of ability. As one goal of the test is to identify individuals with low ability, this item may be useful and is therefore included in our test set. However, we also wanted to confirm that the other test questions are still informative in their own right. To do this, we plotted the test information curve without item 44. Without this item, the item information curve is most informative around −1.5, with a maximum of 30.6 (Figure 4, black solid line). The maximum information of each item, its location in the ability spectrum, and the average information in the range between −4 and 4 are also summarized in Multimedia Appendix 1. The test information curve of the short test is also presented in Figure 4. The short test includes item 44, and thus, we also plot information for a 13-item test without item 44. For the short test, test information is greater than 4 (ie, SE of ability estimate is smaller than 0.5) in the range between −2.4 and −0.5, or 2.4 SDs to 0.5 SD below the average AMT user, again appropriate for a population of low literacy.

Table 3. Average estimated ability of Turkers according to demographic information for the validation task.

Demographic characteristic	Mean correct, %	Average estimated ability
Education		
Less than high school	64.7	−0.899
High school degree	84.9	−0.038
Associate's degree	83.8	−0.013
Bachelor's degree	83.8	−0.034
Master's degree or higher	88.1	0.199
Age in years		
18-21	77.4	−0.493
22-34	83.7	−0.042
35-44	83.6	−0.066
45-54	0.222	
88.3		
55-64	89.4	0.212
65 and older	85.9	−0.122
Gender		
Male	80.6	−0.236
Female	87.2	0.143

Figure 3. Results of analysis to identify useful items from the question sets. Items were removed according to the reasons outlined in the Methodology section. COPD: chronic obstructive pulmonary disease.

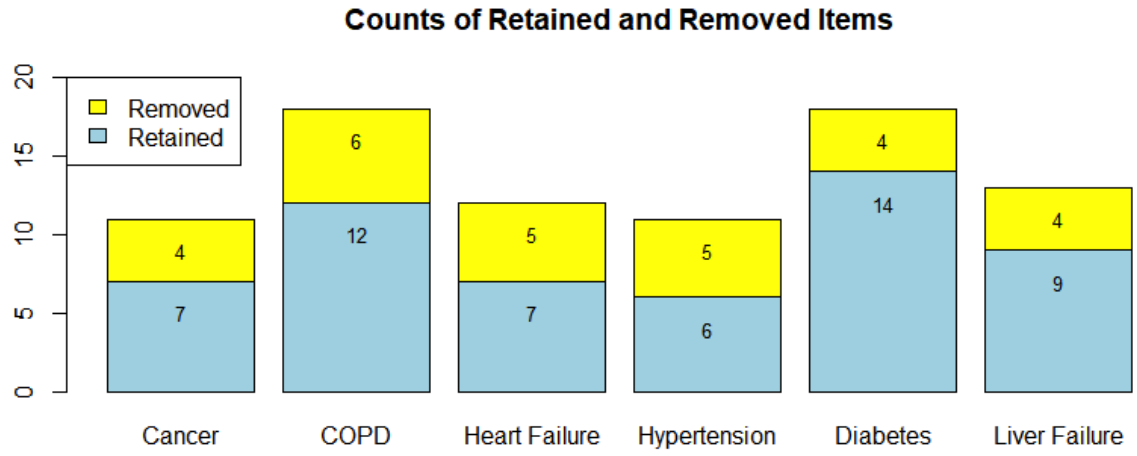
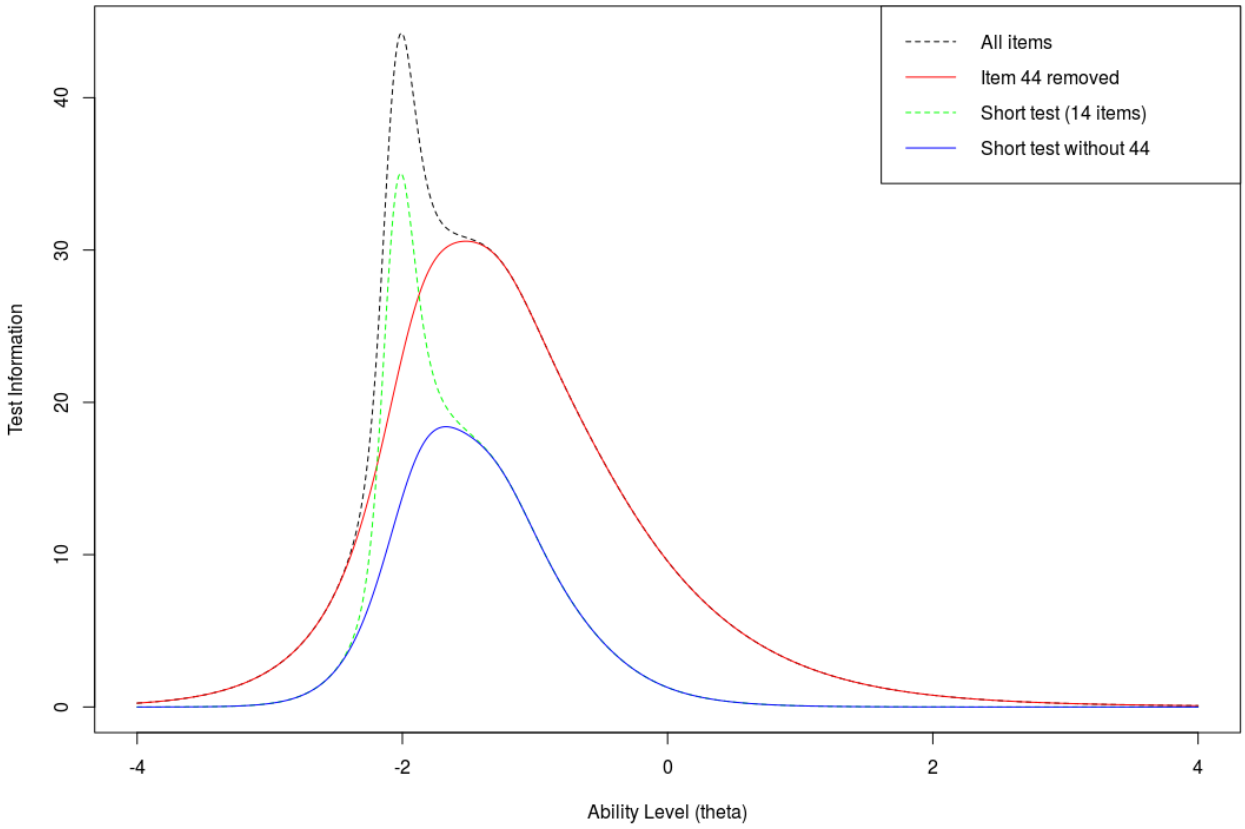


Table 4. Examples of retained and removed questions following item response theory analysis.

Item Retention Decision	Question	Paraphrase	Meaning change	Distractor
Retained	Pegfilgrastim 6 mg subcutaneous one dose	Do an under skin injection of one dose of 6 mg pegfilgrastim	Pegfilgrastim 6 mg epidermal one dose	Pegfilgrastim may prevent neutropenia
Removed	Osteoporosis	Weakness in bones	Hardening of bones as we get older	Some bones get hard and some weak

Figure 4. Test information curve for the full ComprehENotes instrument (55 items) and various subsets.



Discussion

Principal Findings

The goal of this project was to develop an instrument to assess patients' ability to comprehend content in EHR notes. To that end, we developed a process for identifying relevant EHR notes, creating a large question set and reducing the question set to a reasonable size using IRT. We generated questions from EHR notes using SVT and administered them to a population of crowd workers using AMT. We then used IRT to estimate the item parameters and select a subset of items for our instrument. The final test measures a patient's ability to read and comprehend EHR notes. These questions are general enough to be applicable to a wide variety of individuals while still being grounded in specific medical concepts as a result of the hierarchical clustering process.

In contrast with existing tests of health literacy, ComprehENotes was developed by generating questions directly from real patient de-identified EHR notes. Key concepts from the notes were identified by physicians and medical researchers as part of the question generation process. These concepts were deemed important for patients to understand, and the test questions were designed to assess comprehension of these concepts. The ComprehENotes test is the first to directly assess a key element of health literacy, that is, the ability to read and comprehend EHR notes. (To obtain the test, please contact the authors.)

The test is most informative at low levels of ability (Figure 4), which is consistent with our long-term goal of identifying patients with low EHR note comprehension ability. Although the test was easy for the AMT workers, the demographics show that those individuals are not representative of demographics at higher risk of low health literacy (eg, low education and the elderly). Those AMT workers who did fit in the demographics that are more likely to have low health literacy did perform worse in terms of average ability (Table 3). The number of Turkers in those groups was low compared with other demographic groups (Table 2), and thus, more evaluation with individuals with higher likelihood of low health literacy is required. Most of the questions have low difficulty estimates, which makes the test appropriate for screening for low health literacy. It is important to note that the ability estimates are based on the responses of the AMT workers. If we were to fit a new IRT model using response patterns from a patient population, ability estimates of future test takers would be with respect to the patient population. This does not affect the test itself but only how the ability estimates are interpreted. Using the test as developed here, new response patterns are scored and compared with the average AMT user.

We also identified items from our instrument that can be used in a short test to reduce administration time while still being informative. The short test reduces the number of items from 55 to 14 while still being very informative at low levels of ability. This short test can be administered more quickly than the full test while still being informative at low levels of ability.

Limitations

There are limitations with this work. Fitting IRT models requires a large number of human responses to a relatively small number of questions. The length of the question set must be short to

avoid a drop in response quality due to boredom or fatigue. Although the cost of gathering a large number of responses is reduced by using AMT or other crowdsourcing platforms, scaling the number of questions that can be analyzed with IRT remains a challenge.

The groups of physicians and medical researchers who generated our question sets are not experts in question generation using SVT. However, before the task, they all received training on what SVT is and how to construct questions using the methodology. In addition, we manually selected a subset of the questions that were generated for IRT analysis and validation. In this way, we were able to identify a set of questions that could be generalized to a test set. The IRT validation confirmed that a set of questions was appropriate as a test of EHR note comprehension.

The demographics of Turkers that took part in our tasks are not representative of the entire US population, and in particular, do not cover groups with low average health literacy (eg, minorities, people with less than a high school degree, older adults) [9]. However, all but 1 of the questions included in the final question set have difficulty parameters less than or equal to 0. These questions therefore will be appropriate to test ability for individuals with low EHR note comprehension ability. Future work should validate that the questions are in fact appropriate for individuals with low health literacy.

The full ComprehENotes test is long at 55 questions. The length makes it impractical to administer in clinical settings because of the time needed to complete the test. However, we have also identified a short test of 14 items that can be administered in a short period of time. The 14-item test includes items with the largest slope parameters and average information. The short test is still informative at levels of ability below 0, which is appropriate given that the goal of developing this test was to identify individuals with poor EHR note comprehension ability.

Conclusions and Future Work

The ComprehENotes question set is an instrument for measuring EHR note comprehension. Validation of the metric as compared with existing tests of health literacy is still required. During a pilot version of our AMT task, we asked participants to complete the Short Test of Functional Health Literacy in Adults (STOFHLA) as well as our test and found that all the respondents scored a perfect score (36) or answered 1 question wrong on the STOFHLA and were therefore considered to have Adequate Health Literacy according to the STOFHLA scoring. Comparing this metric to existing tests such as REALM or TOFHLA in a population with low health literacy is an important future work to validate the metric as a valid measure of health literacy. In addition, further analysis of how different groups perform on this question set can inform how EHR notes are provided to patients and what types of educational materials should be provided to patients.

The ComprehENotes test can be administered to patients as is to assess EHR note comprehension ability. As the questions are associated with certain diseases and conditions, subsets of the test can also be administered independently to test EHR note comprehension in specific patient populations. For example, the questions associated with liver failure can be extracted and administered as a standalone test to assess EHR note comprehension in liver failure patients. In this way, questions specific to certain diseases can be used to test comprehension among patient populations where the terms are more likely to appear.

Finally, this work is a first step toward being able to evaluate patients' understanding of their health based on information directly contained in their own EHR. We have shown that it is possible to develop a test of health literacy from questions obtained from EHR notes. Automating steps of the question generation and validation processes with clinical natural language processing tools are interesting directions for future work. For example, one such step would be to build an NLP model to generate questions for a specific patient given his or her own EHR note text. The model can be trained on the ComprehENotes questions to identify information that would be relevant for generating good questions. These personalized questions can be administered to patients to evaluate their ability to read and comprehend their own notes.

Acknowledgments

The authors would like to thank Yang Liu from the University of Maryland for his suggestions regarding IRT analysis. The authors thank Jinying Chen for selecting EHR notes and the following researchers and doctors who participated in the question generation: Elaine Freund, Weisong Liu, Jinying Chen, Victoria Wang, Shreya Makkapati, Don Liang, Tianyang Wang, Jonathan Gatley, Abhyuday Jagannatha, Dong Chen, Jiaping Zheng, Francisco Garcia, Sonali Harchandani, Barinder Hansra, Andrew Hsu, German Chiriboga, Edgard Granillo, and Benjamin Helfand. The authors also thank the anonymous Turkers who participated in the AMT tasks. This work was supported in part by the Investigator Initiated Research 1I01HX001457 from the Health Services Research & Development Program of the US Department of Veterans Affairs and in part by a startup fund from the University of Massachusetts Medical School, both to Hong Yu. The content is solely the responsibility of the authors and do not represent the views of the US Department of Veterans Affairs, the US Government, or the University of Massachusetts Medical School.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Table of item parameter estimates and item information in the validation sample.

[PDF File (Adobe PDF File), 44KB - jmir_v20i4e139_app1.pdf]

References

1. Burrington-Brown J, Fishel J, Fox L, Friedman B, Giannangelo K, Jacobs E, AHIMA e-HIM Personal Health Record Work Group. Defining the personal health record. AHIMA releases definition, attributes of consumer health record. J AHIMA 2005 Jun;76(6):24-25. [Medline: 15986557]
2. Ross SE, Lin CT. The effects of promoting patient access to medical records: a review. J Am Med Inform Assoc 2003 Apr;10(2):129-138. [Medline: 12595402]
3. Honeyman A, Cox B, Fisher B. Potential impacts of patient access to their electronic care records. Inform Prim Care 2005;13(1):55-60 [FREE Full text] [Medline: 15949176]
4. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470. [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
5. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to electronic health records: results from a survey of veterans. Med Care 2013 Mar;51(3 Suppl 1):S52-S56. [doi: 10.1097/MLR.0b013e31827808db] [Medline: 23407012]
6. Noblin AM, Wan TT, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13. [Medline: 23209454]
7. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
8. Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health literacy: a prescription to end confusion. Washington, DC: National Academies Press; 2004.
9. Kutner M, Greenburg E, Jin Y, Paulson C. Nces.ed. 2006. The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy URL: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483 [accessed 2017-05-15] [WebCite Cache ID 6yCzjJCua]
10. Jones RB, McGhee SM, McGhee D. Patient on-line access to medical records in general practice. Health Bull (Edinb) 1992 Mar;50(2):143-150. [Medline: 1517087]
11. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, Kim EH, et al. Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc 2006:514-518. [Medline: 17238394]
12. Zheng J, Yu H. Readability formulas and user perceptions of electronic health records difficulty: a corpus study. J Med Internet Res 2017 Mar 2;19(3):e59 [FREE Full text] [doi: 10.2196/jmir.6962] [Medline: 28254738]
13. Sudore RL, Yaffe K, Satterfield S, Harris TB, Mehta KM, Simonsick EM, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. J Gen Intern Med 2006 Aug;21(8):806-812. [doi: 10.1111/j.1525-1497.2006.00539.x] [Medline: 16881938]
14. Schapira MM, Fletcher KE, Hayes A, Eastwood D, Patterson L, Ertl K, et al. The development and validation of the hypertension evaluation of lifestyle and management knowledge scale. J Clin Hypertens (Greenwich) 2012 Jul;14(7):461-466. [doi: 10.1111/j.1751-7176.2012.00619.x] [Medline: 22747619]
15. Chapman K, Abraham C, Jenkins V, Fallowfield L. Lay understanding of terms used in cancer consultations. Psychooncology 2003 Sep;12(6):557-566. [doi: 10.1002/pon.673] [Medline: 12923796]
16. Lerner EB, Jehle DV, Janicke DM, Moscati RM. Medical communication: do our patients understand? Am J Emerg Med 2000 Nov;18(7):764-766. [doi: 10.1053/ajem.2000.18040] [Medline: 11103725]
17. Reading SR, Go AS, Fang MC, Singer DE, Liu IA, Black MH, Anticoagulation and Risk Factors in Atrial Fibrillation–Cardiovascular Research Network (ATRIA‐CVRN) Investigators. Health literacy and awareness of atrial fibrillation. J Am Heart Assoc 2017 Apr 11;6(4):- [FREE Full text] [doi: 10.1161/JAHA.116.005128] [Medline: 28400367]
18. Halbach SM, Enders A, Kowalski C, Pförtner TK, Pfaff H, Wesselmann S, et al. Health literacy and fear of cancer progression in elderly women newly diagnosed with breast cancer--a longitudinal analysis. Patient Educ Couns 2016 May;99(5):855-862. [doi: 10.1016/j.pec.2015.12.012] [Medline: 26742608]
19. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]
20. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995 Oct;10(10):537-541. [Medline: 8576769]
21. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
22. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1):33-42. [Medline: 14528569]

23. Osborn CY, Weiss BD, Davis TC, Skripkauskas S, Rodrigue C, Bass PF, et al. Measuring adult literacy in health care: performance of the newest vital sign. Am J Health Behav 2007;31 Suppl 1:S36-S46. [doi: 10.5555/ajhb.2007.31.supp.S36] [Medline: 17931135]
24. Royer JM, Hastings CN, Hook C. A sentence verification technique for measuring reading comprehension. J Lit Res 1979;11(4):355-363. [doi: 10.1080/10862967909547341]
25. Mazor KM, Roblin DW, Williams AE, Greene SM, Gaglio B, Field TS, et al. Health literacy and cancer prevention: two new instruments to assess comprehension. Patient Educ Couns 2012 Jul;88(1):54-60 [FREE Full text] [doi: 10.1016/j.pec.2011.12.009] [Medline: 22244323]
26. Mazor KM, Rogers HJ, Williams AE, Roblin DW, Gaglio B, Field TS, et al. The cancer message literacy tests: psychometric analyses and validity studies. Patient Educ Couns 2012 Oct;89(1):69-75 [FREE Full text] [doi: 10.1016/j.pec.2012.06.018] [Medline: 22789147]
27. Baker FB, Kim SH, editors. Item Response Theory: Parameter Estimation Techniques, Second Edition. Boca Raton, FL: CRC Press; 2004.
28. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. Clin Exp Rheumatol 2005 Oct;23(5 Suppl 39):S53-S57. [Medline: 16273785]
29. 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;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]
30. 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]
31. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
32. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
33. Royer JM, Greene BA, Sinatra GM. The sentence verification technique: a practical procedure for testing comprehension. J Reading 1987;30(5):414-422.
34. Royer JM. Uses for the sentence verification technique for measuring language comprehension URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.505.1224&rep=rep1&type=pdf [WebCite Cache ID 6uqvG932K]
35. Kardash CA, Royer JM, Greene BA. Effects of schemata on both encoding and retrieval of information from prose. J Educ Psychol 1988;80(3):324-329. [doi: 10.1037/0022-0663.80.3.324]
36. Kardash CM, Scholes RJ. Effects of preexisting beliefs and repeated readings on belief change, comprehension, and recall of persuasive text. Contemp Educ Psychol 1995 Apr;20(2):201-221. [doi: 10.1006/ceps.1995.1013]
37. Royer JM, Carlo MS. Assessing the language acquisition progress of limited English proficient students: problems and new alternative. Appl Meas Educ 1991 Apr;4(2):85-113. [doi: 10.1207/s15324818ame0402_1]
38. Snow R, O'Connor B, Jurafsky D, Ng AY. Cheap and Fast, But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. In: Association for Computational Linguistics. Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing; October 25-27, 2008; Honolulu, Hawaii p. 254-263 URL: http://aclweb.org/anthology/D08-1027
39. Sabou M, Bontcheva K, Scharl A. Crowdsourcing Research Opportunities: Lessons from Natural Language Processing. 2012 Presented at: 12th International Conference on Knowledge Management and Knowledge Technologies (iKnow-2012); September 2012; New York, NY. [doi: 10.1145/2362456.2362479]
40. Demartini G, Difallah DE, Cudré-Mauroux P. ZenCrowd: Leveraging Probabilistic ReasoningCrowdsourcing Techniques for Large-scale Entity Linking. 2012 Presented at: WWW '12 Proceedings of the 21st international conference on World Wide Web; April 16-20, 2012; Lyon, France p. 496-478. [doi: 10.1145/2187836.2187900]
41. Zhai H, Lingren T, Deleger L, Li Q, Kaiser M, Stoutenborough L, et al. Web 2.0-based crowdsourcing for high-quality gold standard development in clinical natural language processing. J Med Internet Res 2013;15(4):e73 [FREE Full text] [doi: 10.2196/jmir.2426] [Medline: 23548263]
42. Good BM, Nanis M, Wu C, Su AI. Microtask crowdsourcing for disease mention annotation in PubMed abstracts. Pac Symp Biocomput 2015:282-293 [FREE Full text] [Medline: 25592589]
43. Mortensen JM, Minty EP, Januszyk M, Sweeney TE, Rector AL, Noy NF, et al. Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT. J Am Med Inform Assoc 2015 May;22(3):640-648 [FREE Full text] [doi: 10.1136/amiajnl-2014-002901] [Medline: 25342179]
44. Gottlieb A, Hoehndorf R, Dumontier M, Altman RB. Ranking adverse drug reactions with crowdsourcing. J Med Internet Res 2015;17(3):e80 [FREE Full text] [doi: 10.2196/jmir.3962] [Medline: 25800813]
45. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods 1996;1(2):130-149. [doi: 10.1037//1082-989X.1.2.130]
46. Novick MR. The axioms and principal results of classical test theory. J Math Psychol 1966;3(1):1-18. [doi: 10.1016/0022-2496(66)90002-2]

47. Brady CJ, Mudie LI, Wang X, Guallar E, Friedman DS. Improving consensus scoring of crowdsourced data using the Rasch model: development and refinement of a diagnostic instrument. J Med Internet Res 2017 Jun 20;19(6):e222 [FREE Full text] [doi: 10.2196/jmir.7984] [Medline: 28634154]
48. Rizopoulos D. ltm: an R package for latent variable modeling and item response theory analyses. J Stat Soft 2006;17(5):1-25. [doi: 10.18637/jss.v017.i05]
49. Chalmers RP. mirt: a multidimensional item response theory package for the R environment. J Stat Soft 2012;48(6):1-29. [doi: 10.18637/jss.v048.i06]

Abbreviations

AMT: Amazon Mechanical Turk
COPD: chronic obstructive pulmonary disease
EHR: electronic health record
HIT: Human Intelligence Task
ICC: item characteristic curve
IRT: item response theory
NVS: Newest Vital Sign
PHR: personal health record
REALM: Rapid Estimate of Adult Literacy in Medicine
SVT: Sentence Verification Technique
TOFHLA: Test of Functional Health Literacy in Adults
UMLS: Unified Medical Language System
3PL: 3-parameter logistic

Edited by G Eysenbach; submitted 10.11.17; peer-reviewed by S Kim, G Baumblatt, L Petersson, S Garcia; comments to author 07.12.17; revised version received 06.02.18; accepted 20.02.18; published 25.04.18.

Please cite as:
Lalor JP, Wu H, Chen L, Mazor KM, Yu H
ComprehENotes, an Instrument to Assess Patient Reading Comprehension of Electronic Health Record Notes: Development and Validation
J Med Internet Res 2018;20(4):e139
URL: http://www.jmir.org/2018/4/e139/
doi:10.2196/jmir.9380
PMID:29695372

Review

Evaluation of Technology-Enhanced Learning Programs for Health Care Professionals: Systematic Review

Pam Nicoll[1], BSc (Hon), Dip APS; Sandra MacRury[2], MBChB, MD, FRCP; Hugo C van Woerden[2,3], MBChB, PhD, FFPH; Keith Smyth[2], PhD

[1]National Health Service Education for Scotland, Inverness, United Kingdom

[2]University of the Highlands and Islands, Inverness, United Kingdom

[3]Cardiff University, Heathpark Campus, Cardiff, United Kingdom

Corresponding Author:
Pam Nicoll, BSc (Hon), Dip APS
National Health Service Education for Scotland
Old Perth Road
Inverness, IV2 3JH
United Kingdom
Phone: 44 01463255 ext 7910
Email: pam.nicoll@nes.scot.nhs.uk

Abstract

Background: Technology-enhanced learning (TEL) programs are increasingly seen as the way in which education for health care professionals can be transformed, giving access to effective ongoing learning and training even where time or geographical barriers exist. Given the increasing emphasis on this mode of educational support for health care practitioners, it is vital that we can effectively evaluate and measure impact to ensure that TEL programs are effective and fit for purpose. This paper examines the current evidence base for the first time, in relation to the evaluation of TEL programs for health care professionals.

Objective: We conducted a systematic review of the current literature relating to the evaluation of TEL programs for health care professionals and critically appraised the quality of the studies.

Methods: This review employed specific search criteria to identify research studies that included evaluation of TEL for health care professionals. The databases searched included Medline Ovid, Cumulative Index of Nursing and Allied Health Literature Plus Advanced, Applied Social Sciences Index and Abstracts, ZETOC, Institute of Electrical and Electronics Engineers Explore Digital Library, Allied and Complementary Medicine, and Education Resources Information Center between January 2006 and January 2017. An additional hand search for relevant articles from reference lists was undertaken. Each of the studies identified was critically appraised for quality using the Crowe Critical Appraisal Tool. This approach produced a percentage total score for each study across specified categories. A proportion of the studies were independently assessed by an additional two reviewers.

Results: The review identified 21 studies that met the inclusion criteria. The studies included scored totals across eight categories within a range of 37%-95% and an average score of 68%. Studies that measured TEL using learner satisfaction surveys, or combined pretest and posttest knowledge score testing with learner satisfaction surveys, were found to be the most common types of TEL evaluations evident in the literature. The studies reviewed had low scores across reporting on ethical matters, design, and data collection categories.

Conclusions: There continues to be a need to develop effective and standard TEL evaluation tools, and good quality studies that describe effective evaluation of TEL education for health care professionals. Studies often fail to provide sufficient detail to support transferability or direct future TEL health care education programs.

(J Med Internet Res 2018;20(4):e131) doi:10.2196/jmir.9085

KEYWORDS

technology-enhanced learning; evaluation; e-learning; blended learning; digital learning; program evaluation; effectiveness

Introduction

The term technology-enhanced learning (TEL) is often used to describe a broad field of digital technologies used to support and mediate educational activities [1]. In this review, the term TEL is used to describe activities that are totally digitally mediated and those that are blended with more traditional educational approaches. The last two decades have seen considerable growth in the use of technology within higher education at undergraduate and postgraduate levels across the world [2]. Effective and ongoing continuous professional development (CPD) and education are essential to the delivery of high quality health and care services. TEL is increasingly presented as a means by which learners can be provided with enhanced or transformed educational experiences.

A range of published reports have highlighted TEL as an effective method to support health care education [3,4]. In their e-learning strategy, The Higher Education Funding Council England [5] summarized three levels of potential benefits of TEL: (1) e *fficiency*, whereby existing processes can be carried out in a more cost-effective, time-effective, sustainable, or scalable manner; (2) *enhancement*, which improves existing processes and outcomes; and (3) t *ransformation*, representing radical change in existing processes or the introduction of new processes. The recognition of the need for continuing education and effective work-based training to support health care professionals to deliver good quality, safe, and effective care is widely accepted [6,7]. The increasing demands for effective and affordable health care education in light of resource and time constraints, together with improved access to hardware, software, and the popularity of blended learning formats, means that TEL is increasingly considered an ideal approach within health care education.

The general availability of mobile and flexible technologies enables learners to minimize time away from health care settings to undertake training and to engage with learning resources when and where they are most suitable to their needs [8]. TEL offers a range of specific advantages for health care education, given the flexibility to update learning resources in a fast-changing field, and the scope offered for learners to share knowledge and learn critical clinical skills and decision making safely in nonclinical environments [9].

The ability to demonstrate the added value and impact of TEL for health care education remains challenging. Previous authors have captured the nature of the challenges in the review and evaluation of TEL within medical education [10-12]. Pickering and Joynes [13] highlighted the lack of robust evidence and meaningful evaluation to support widespread implementation of TEL resources. The main challenges concern a lack of clarity around the purpose of evaluation, comprehensiveness, depth, and methodology choice to support development of the required evidence base, upon which to build effective future TEL health care programs. We need reliable, practical mechanisms to evaluate: the value for money; equity of access; and learner, service, and organizational benefits that TEL may bring [14]. There is a need to critically examine the literature on TEL implementation and evaluation within health care to guide production and implementation of effective TEL health care education programs now and in the future.

A range of studies exist in the TEL literature which document implementation of TEL within medical and health care educational approaches. However, studies demonstrating a comprehensive TEL evaluation or use of standardized TEL evaluation tools in practice are fewer in number. Previous authors such as Ellaway [10,11], Cook and Ellaway [12], and Pickering and Joynes [13] (amongst others) have highlighted the need for robust evaluation, and have set out to develop both TEL educational standards and frameworks for evaluating TEL in medical education. Cook and Ellaway [12] have proposed a general model for evaluation. Pickering and Joynes [13] have proposed what they consider to be a more holistic TEL evaluation model for medical education.

This systematic review of the literature aimed to identify studies that have implemented TEL evaluation for CPD and postgraduate or work-based TEL health care education programs, and to assess these using a published critical appraisal tool. The studies identified provide an evidence base for the evaluation and development of future TEL programs for health care professionals.

Methods

Design

The review was carried out using a systematic integrative review method. This method allows for the inclusion of empirical and theoretical literature and quantitative and qualitative studies. This method enabled an increased number of studies to be included in the review and is appropriate for the review of evidence to highlight gaps in the literature [15].

Eligibility and Inclusion

The inclusion and exclusion criteria for this review were developed using the Population, Intervention, Comparison, and Outcomes (PICOS) [16] framework for systematic reviews, which is illustrated in Table 1. In devising the search strategy for this study, the PICOS framework has been used as a search tool and as an organizing framework to list terms by the main concepts in the search question. This framework is commonly used to identify components of clinical evidence for systematic reviews in evidence-based medicine and is endorsed by the Cochrane Collaboration [16].

Search Strategy

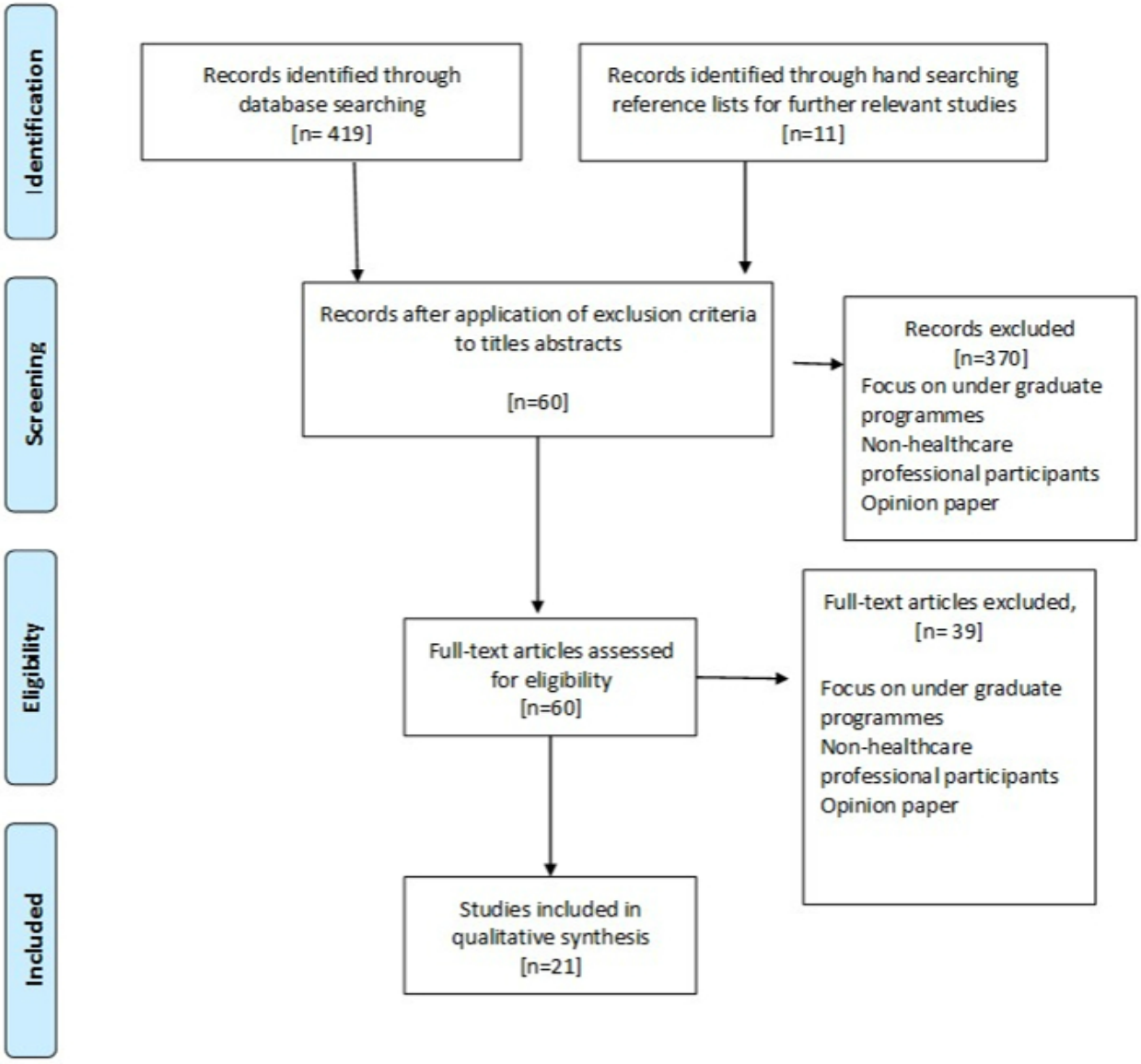
The following electronic databases were searched: Medline Ovid, Cumulative Index of Nursing and Allied Health Literature (CINAHL) Plus Advanced, Applied Social Sciences Index and Abstracts (ASSIA), ZETOC, Institute of Electrical and Electronics Engineers (IEEE) Explore Digital Library, Allied and Complementary Medicine (AMED), and Education Resources Information Center (ERIC) between January 2006 and January 2017 (see Multimedia Appendix 1 for details). The search was conducted using three concepts (and appropriate synonyms for each) across the selected databases: technology-enhanced learning, health care, and educational measurement. A total of 13 synonyms were used in the literature

search. These phrases included technology-enhanced learning, technology-enhanced education, e-learning, e-education, blended learning, blended education, digital learning, digital education, evaluation studies, program evaluation, effectiveness, validation studies, and intervention. Reference lists were hand searched for relevant studies. The flow diagram in Figure 1 illustrates the search strategy.

Table 1. Application of the Population, Intervention, Comparison, and Outcomes (PICOS) Framework to the research question. TEL: technology-enhanced learning.

Parameter	Details
Participants	Health care professionals in full-time or part-time employment undertaking continuing professional development that is delivered using TEL (full time higher and further education students, school learners excluded).
Intervention	Studies using a TEL evaluation tool or framework to evaluate technology-enhanced health care education programs. The evaluation tool or framework must be used to evaluate a program for health care professionals.
Comparison	Some studies will have no comparison or comparator; others will examine one type of TEL approach against another.
Outcomes	Study must include: (1) evaluation of effective use of TEL, (2) the techniques being evaluated must be sufficiently specified, (3) assessment of learning outcomes, and (4) assessment of educational content.
Study design	Both empirical and theoretical research published in English between 2006-2017 from peer reviewed journals, conference papers. Systematic reviews and meta-analyses will not be included. Opinion papers will be excluded.

Figure 1. Literature search exclusion chart.



Data Extraction

A specific data extraction tool was developed based on the inclusion and exclusion criteria. Each article was reviewed and information was extracted in relation to participant type, study design, sample size, types of TEL used, TEL evaluation tool used, and key study findings.

Quality Assessment

Critical appraisal is a standardized way of assessing research so that decisions can be made based on the best evidence available [17]. The Crowe Critical Appraisal Tool (CCAT) was developed to provide a tool that can reliably assess a range of research designs, provide a comprehensive appraisal approach, and provide a suitable scoring method [17]. Quality assessment was conducted for all articles included within the study. The CCAT was used to assess quality across reporting items in eight categories: Preamble, Introduction, Design, Sampling, Data collection, Ethical matters, Results, and Discussion. The items were rated on a nominal scale (Present/Absent/Not applicable). The CCAT was selected as an appropriate critical appraisal tool for this study, as it can be reliably applied across a range of study designs and has been applied to a range of both quantitative and qualitative studies [17]. The CCAT emphasizes the importance of measuring and recording scores for each of the categories rather than simply the final score for each study. This approach prevents papers that score high overall, but poorly in one or more categories, from becoming less visible than papers that score highly throughout all categories. A subset of the studies [14,18-20] was randomly selected for scoring by additional reviewers (SM and HVW).

Results

Search Outcome

A total of 430 articles were identified in the initial titles search of the online databases and hand search of the literature. Following the application of the inclusion criteria to the titles and abstracts, 60 articles remained. The full text of the 60 articles were evaluated using the inclusion criteria and a total of 21 articles [14,18-37] were identified as suitable for the review.

Multimedia Appendix 2 shows a summary of the information extracted from the studies. Ten of the studies measured TEL using learner satisfaction surveys, 8 combined pretest and posttest knowledge score testing with learner satisfaction surveys (1 with pretest and posttest knowledge score testing only), and 2 used qualitative frameworks. The selected studies described the use of a variety of different TEL modes or combinations of each, including: 12 used Web-based e-learning, 4 used learning management systems, 2 used video simulation, and 6 used blended learning formats.

Quality Assessment

The results of the quality assessment are summarized within Table 2. The CCAT scores for the studies selected using the inclusion criteria indicate that the quality of studies varied greatly across the range of research parameters examined. Of the 21 papers, 19 presented sufficient information to be included in the CCAT evaluation. Two papers [14,18] were identified as TEL articles rather than research papers with CCAT scores in early sections too low to be viable to continue scoring, in accordance with CCAT user guidance. Ten studies scored above the average score of 68% overall and 9 scored less than this average (these are italicized within Table 2). The lowest average-scored sections for the set of 19 studies were in Ethical Matters, with a score of 2/5. Design, Data Collection, and Results each produced average scores of 3/5. Preliminaries, Discussion, Introduction, and Sampling sections each produced the higher average scores of 4/5 for the studies examined.

These elements give a useful starting point in describing a requisite information set for inclusion within all good quality research studies. The value of studies that fail to include and compliment these basic elements with additional standard or sufficient research information data is significantly reduced. The studies examined had low scores for the Ethics section overall. This section looked for information related to consideration of standard research ethics, such as participant ethics and researcher ethics, even where formal ethical approval had not been required. Information that conveys ethical considerations is a prerequisite of all research studies.

The relatively low scores achieved in relation to Design, Data Collection, and Results categories are also concerning, as this renders many of the studies difficult (if not impossible) to replicate. This section looked for inclusion of information on interventions, outcomes, or treatment measures, in addition to sufficient descriptions of the research design and rationale. A key requisite of effective TEL evaluation research is ensuring that the intervention is sufficiently described to support others, who may wish to make comparisons, or to confidently apply the research to their own practice or education program development [11]. Only in this way will we begin to establish a reliable evidence base around TEL evaluation. The higher average quality assessment scores were in relation to the Discussion section and Preliminary and Introductory elements such as title, abstract, background, and objectives information.

A subset of the studies [14,18-20] was randomly selected for additional scoring by PN. The studies were provided to two of the coauthors of this paper (SM and HVW) for them to provide an additional score using the CCAT scoring template and guide. The average difference between the original and additional scoring across 4 of the papers was found to be 19%.

Table 2. Crowe Critical Appraisal Tool (CCAT) scores summary. Italics indicate studies that scored less than the average score of 68%. N/A: not applicable.

Authors	CCAT category								Raw score n (%)
	Preliminaries	Introduction	Design	Sampling	Data collection	Ethical matters	Results	Discussion	
Akroyd et al [14]	2	1	N/A	N/A	N/A	N/A	N/A	N/A	0
Lotrecchiano et al [18]	1	1	N/A	N/A	N/A	N/A	N/A	N/A	0
Westbrook [37]	2	4	1	1	3	0	1	3	15 (37.5)
Konstantinidis et al [31]	4	3	1	2	1	0	1	4	16 (40)
Walsh et al [35]	4	3	2	2	2	0	2	4	19 (47.5)
Heartfield et al [29]	2	4	2	2	2	1	3	4	20 (50)
Chuo et al [24]	3	4	3	3	4	0	2	3	21 (52.5)
Wang [36]	2	2	3	2	2	5	4	2	22 (55)
Gill [27]	3	4	4	2	3	2	3	4	23 (57.5)
Ingelbeen et al [30]	4	5	3	4	1	2	2	4	25 (62.5)
Safwat and Pourabdollah [33]	3	4	3	4	4	0	4	4	26 (65)
Goldberg Goetz et al [28]	5	5	3	5	4	0	3	5	30 (75)
Byrne et al [22]	3	5	4	4	4	4	4	4	32 (80)
Bekkers et al [21]	5	5	4	5	2	3	4	4	32 (80)
Popescu et al [32]	5	5	4	4	4	2	4	4	32 (80)
Sranacharoenpong et al [34]	5	5	4	5	5	1	4	3	32 (80)
Chang et al [23]	4	5	3	4	4	4	5	4	33 (82.5)
Moreira et al [19]	5	5	4	5	5	0	5	5	34 (85)
Fontaine et al [26]	5	5	3	5	5	2	5	5	35 (87.5)
Schneiderman et al [20]	5	5	4	4	4	4	4	5	35 (87.5)
Cortese-Peske [25]	5	5	4	5	4	5	5	5	38 (95)

Discussion

Principal Findings

Despite the growth in popularity and types of TEL education programs produced over the last two decades, this review was only able to identify a small pool of studies that met the inclusion criteria for TEL evaluation of a health care professional education program. Many of the included studies described the TEL methods evaluated as virtual learning environments, online or e-learning modules, platforms, or blended formats. There was little evidence provided within the selected studies regarding evaluation of bidirectional TEL approaches or newer types of TEL approaches such as Personal Learning Environments (PLEs). PLEs are activity spaces in which students interact and communicate with one another, and with experts, by using Web 2.0 tools. The ultimate result of using Web 2.0 tools is the development of collective learning approaches such as “just-in-time” and “at-your-fingertips” learning opportunities that can support a wide range of teaching and learning activities [38]. The evaluation studies identified in this review relied heavily on measuring TEL using learner satisfaction measures. Only one study [37] cited use of a structured approach to evaluating the more interactive two-way learning process between learner and tutor that is offered by models such as the Salmon-5 stage model [39]. The most widely cited types of learner outcome measurements used within educational evaluations are Kirkpatrick’s [40] models. While these methods of measurement may often provide a useful starting point for effective assessment of TEL evaluations, the model itself may not be ideally suited to the evaluation of TEL health care education. That is, where the Kirkpatrick model emphasizes increased confidence in newly acquired knowledge as being important, effective health care education would want evidence that this new knowledge is both learned and implemented in practice beyond a practitioner’s own perception of knowledge gain or confidence [13]. For health care education to truly support health care practice we need to be able to accurately measure the added knowledge, skills, or awareness that TEL programs may or may not provide.

This review of the literature concerning evaluations of TEL highlights the pattern that previous authors have noted for studies to employ a narrow focus in evaluation on either the technology equipment itself, measurement of learner satisfaction, or preknowledge and postknowledge scores [41,42]. Some of the studies included aspects of intralearner activity within their evaluation [18,22,27,28,30,33,37]. However, the

dominance in the literature of evaluation studies measuring largely (or only) individual pretest and posttest knowledge presents a number of concerns within TEL for health care professionals. That is, where program assessments are relied upon to determine added value, learner gain, or improvement it is necessary to consider the extent to which they were matched to the actual TEL aims or enhancement being sought.

Testing methods can heavily influence the learner's focus and how they approach learning. If we consider that one of the key functions of employing technology in learning is to help people connect more effectively with each other and the learning materials, and to inspire learner interaction in accordance with a social constructivist learning approach, then assessments that focus solely on the work of each individual (ie, cognitivist style, pretest and posttest scoring) may have a considerable impact on each learner's behavior and the efficacy of the program overall [43]. There is an ongoing need for more TEL evaluation studies to detail the purpose of TEL interventions and the assessment and overall approaches adopted, and to demonstrate how the technology is enhancing the learning experience [2].

There is a need to be able to identify high quality TEL for health care education research studies and to be able to compare the outcomes from these sources to produce practical TEL evaluation tools. This need has also been highlighted throughout the last two decades in the literature on TEL in other contexts [44]. A more specific and standard approach to TEL evaluation in health care education with common measurements or tools would enable health care professionals, employers, and program designers to measure and then document effective TEL for health care education. In this way, a reliable TEL evidence base can grow and be progressively used to its full effect to influence or even transform educational programs for health care education now and in the future.

The CCAT tool used in this study enabled various forms of evidence presented within the literature to be explored in terms of TEL evaluation and the quality of the evidence presented. While some of the studies did present a full range of detailed research information, a number of those examined lacked information on the fundamental elements of good quality research. Ellaway [11] has highlighted similar issues related to published research for online learning where, as with TEL study interventions, they are often inadequately described. A more complex analysis of interassessor consistency in scoring using the CCAT tool could have been undertaken within this study. However, although the interassessor scoring was limited in this case by available resources, it was already building on the robust methods outlined by Crowe et al [17].

Limitations

This review has not measured standard educational quality parameters or set out to identify the requisite elements of a robust TEL evaluation guide (or tool) for health care professionals' education. Instead, it focused on identifying what evidence of TEL evaluation for health care professional education already exists within the literature and examined whether it was of sufficient quantity and quality as an evidence base for organizations to use to develop increasingly effective and transformative TEL education programs. Although the subject of TEL dates to at least the 1990s in the context of further and higher education, its application within health care education is much more recent. On that basis, the decision was made to restrict the literature search to 2006-2017. Other databases such as Web of Science were checked but yielded few references that fully met the inclusion criteria, and such databases were therefore not included within the literature review methods for this study.

Conclusions

This review found limited published evidence of standard tools being implemented to measure TEL in health care education programs. Developing and implementing TEL health care education can require organizations to make considerable financial, human, and infrastructure investment. There is a mismatch between the scale of uptake of TEL in health care education and availability of a sufficiently robust evidence base of meaningful TEL evaluations in health care education. The outcomes of the systematic review and critical appraisal of this study support the views of Kirkwood and Price [2] who stated that there is a scarcity of published studies of practical TEL education programs that generate evidence that is appropriate to the interventions described, and that can be drawn upon.

A review of the TEL evaluation literature to help identify an evidence-based list of essential parameters to include within TEL health care education evaluation reports and studies would be a useful focus for further research. There continues to be a need to develop effective and standard TEL evaluation tools and for the publication of good quality studies that describe effective evaluation of TEL education for health care professionals . Studies often fail to provide sufficient detail to support transferability or direct future TEL health care education programs' design and implementation. The use of a standard, practical, and quality approach to TEL evaluation, recording, and reporting with the same tools (or even parameters) across a variety of health care education programs would address this gap over time. This type of approach would also reduce duplication of efforts for organizations in creating or recreating tools, and importantly support cross-program and cross-organizational comparisons.

There is a range of guides, frameworks, and standards emerging in the literature and across practice to guide the design of TEL within health care and higher education institutions, programs, and resources. The models that have been proposed require widespread implementation, rigorous in-practice testing, and effective reporting to ensure that TEL education programs for health care professionals are evaluated in a more robust manner than is currently evident in the literature [13]. In this way, such programs can then usefully shape the emerging field of TEL evaluation for health care education.

Acknowledgments

The authors wish to acknowledge funding from National Health Service Highland and the University of the Highlands and Islands that supported this research study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Electronic database searches.

[PDF File (Adobe PDF File), 540KB - jmir_v20i4e131_app1.pdf]

Multimedia Appendix 2

Summary of information from studies.

[PPTX File, 58KB - jmir_v20i4e131_app2.pptx]

References

1. Goodyear P, Retalis S. Technology-Enhanced Learning: Design Patterns and Pattern Language. Rotterdam, Netherlands: Sense Publishers; 2010:1-27.
2. Kirkwood A, Price L. Technology-enhanced learning and teaching in higher education: what is 'enhanced' and how do we know? A critical literature review. Learn Media Technol 2013 Feb 20;39(1):6-36. [doi: 10.1080/17439884.2013.770404]
3. Carroll C, Booth A, Papaioannou D, Sutton A, Wong R. UK health-care professionals' experience of on-line learning techniques: a systematic review of qualitative data. J Contin Educ Health Prof 2009;29(4):235-241. [doi: 10.1002/chp.20041] [Medline: 19998474]
4. Childs S, Blenkinsopp E, Hall A, Walton G. Effective e-learning for health professionals and students--barriers and their solutions. A systematic review of the literature--findings from the HeXL project. Health Info Libr J 2005 Dec;22 Suppl 2:20-32 [FREE Full text] [doi: 10.1111/j.1470-3327.2005.00614.x] [Medline: 16279973]
5. The Higher Education Funding Council for England. Digital Education Resource Archive (DERA). England: Higher Education Funding Council for England (HEFCE); 2009. Enhancing learning and teaching through the use of technology: a revised approach to HEFCE's strategy for e-learning URL: http://dera.ioe.ac.uk/140/ [accessed 2018-03-12] [WebCite Cache ID 6xrT9sXor]
6. Schostak J, Davis M, Hanson J, Schostak J, Brown T, Driscoll PS, et al. 'Effectiveness of Continuing Professional Development' project: a summary of findings. Med Teach 2010;32(7):586-592. [doi: 10.3109/0142159X.2010.489129] [Medline: 20653382]
7. Health Education England. 2016 Mar. Improving Safety Through Education and Training Report by the Commission on Education and Training for Patient Safety URL: http://http/www.hee.nhs.uk/the-commission-on-education-and-training-for-patient-safety [accessed 2018-03-12] [WebCite Cache ID 6xrULLzzD]
8. Lupton D. Digital sociology: an introduction. SSRN Journal 2012 Aug 15:1-16 [FREE Full text] [doi: 10.2139/ssrn.2273418]
9. Santy J, Smith L. Being an e-learner in health and social care: a student's guide. Abingdon, England: Routledge; 2007:1-33.
10. Ellaway R. Weaving the 'e's together. Med Teach 2006 Nov;28(7):587-590. [doi: 10.1080/01421590600909070] [Medline: 17594547]
11. Ellaway R. Apples and architraves: a descriptive framework for e-learning research. Med Teach 2010 Jan;32(1):95-97. [doi: 10.3109/01421590903573966] [Medline: 20095783]
12. Cook D, Ellaway RH. Evaluating technology-enhanced learning: a comprehensive framework. Med Teach 2015;37(10):961-970. [doi: 10.3109/0142159X.2015.1009024] [Medline: 25782599]
13. Pickering JD, Joynes VCT. A holistic model for evaluating the impact of individual technology-enhanced learning resources. Med Teach 2016 Dec;38(12):1242-1247. [doi: 10.1080/0142159X.2016.1210112] [Medline: 27584048]
14. Akroyd M, Jordan G, Rowlands P. Interprofessional, simulation-based technology-enhanced learning to improve physical health care in psychiatry: the recognition and assessment of medical problems in psychiatric settings course. Health Informatics J 2016 Jun;22(2):312-320. [doi: 10.1177/1460458214557098] [Medline: 25425630]
15. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
16. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014 Nov 21;14:579 [FREE Full text] [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]

17. Crowe M, Sheppard L, Campbell A. Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs. J Clin Epidemiol 2012 Apr;65(4):375-383. [doi: 10.1016/j.jclinepi.2011.08.006] [Medline: 22078576]
18. Lotrecchiano G, McDonald P, Lyons L, Long T, Zajicek-Farber M. Blended learning: strengths, challenges, and lessons learned in an interprofessional training program. Matern Child Health J 2013 Nov;17(9):1725-1734. [doi: 10.1007/s10995-012-1175-8] [Medline: 23291875]
19. Moreira I, 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 05;17(1):e3 [FREE Full text] [doi: 10.2196/jmir.3344] [Medline: 25560547]
20. Schneiderman J, Corbridge S, Zerwic JJ. Demonstrating the effectiveness of an online, computer-based learning module for arterial blood gas analysis. Clin Nurse Spec 2009;23(3):151-155. [doi: 10.1097/NUR.0b013e3181a075bc] [Medline: 19395892]
21. Bekkers M, Simpson SA, Dunstan F, Hood K, Hare M, Evans J, STAR Study Team. Enhancing the quality of antibiotic prescribing in primary care: qualitative evaluation of a blended learning intervention. BMC Fam Pract 2010 May 07;11:34 [FREE Full text] [doi: 10.1186/1471-2296-11-34] [Medline: 20459655]
22. Byrne E, Donaldson L, Manda-Taylor L, Brugha R, Matthews A, MacDonald S, et al. The use of technology enhanced learning in health research capacity development: lessons from a cross country research partnership. Global Health 2016 May 10;12(1):19 [FREE Full text] [doi: 10.1186/s12992-016-0154-z] [Medline: 27160242]
23. Chang W, Hsiao SS, Chang PC, Lee PH. Developing an E-learning education programme for staff nurses: processes and outcomes. Nurse Educ Today 2008 Oct;28(7):822-828. [doi: 10.1016/j.nedt.2008.02.003] [Medline: 18384917]
24. Chuo Y, Liu C, Tsai C. Effectiveness of e-learning in hospitals. Technol Health Care 2015;23 Suppl 1:S157-S160. [doi: 10.3233/thc-150949] [Medline: 26410320]
25. Cortese-Peske M. Icahn School of Medicine at Mount Sinai.: ERIC; 2013. Improving cultural competency and disease awareness among oncology nurses caring for adult T-cell leukemia and lymphoma patients URL: http://www.proquest.com/en-US/products/dissertations/individuals.shtml [accessed 2017-03-09] [WebCite Cache ID 6xriqNXiK]
26. Fontaine G, Cossette S, Heppell S, Boyer L, Mailhot T, Simard MJ, et al. Evaluation of a Web-based E-learning platform for brief motivational interviewing by nurses in cardiovascular care: a pilot study. J Med Internet Res 2016 Aug 18;18(8):e224 [FREE Full text] [doi: 10.2196/jmir.6298] [Medline: 27539960]
27. Gill A. E-learning and professional development--never too old to learn. Br J Nurs 2007;16(17):1084-1088. [doi: 10.12968/bjon.2007.16.17.27255] [Medline: 18026055]
28. Goldberg GD, Clement DG, Cotter JJ. Development and alumni assessment of an interdisciplinary PhD program offered through a blended learning environment. J Allied Health 2011;40(3):137-142. [Medline: 21927779]
29. Heartfield M, Morello A, Harris M, Lawn S, Pols V, Stapleton C, et al. E-learning competency for practice nurses: an evaluation report. Aust J Prim Health 2013;19(4):287-291. [doi: 10.1071/PY13033] [Medline: 24134876]
30. Ingelbeen B, Aryeetey R, Khanna J, Mahamoudane N, Vincent D, Kiyan C, et al. Exploring three approaches to offer distance learning courses through a social network of health researchers in three African countries. In: IEEE Xplore.: IEE Explore; 2014 May 9 Presented at: International School of Tanganyika-Africa Conference Proceedings, 2014; 2014; Le Meridien Ile Maurice, Mauritius URL: http/ieeexplore.ieee.org [doi: 10.1109/ISTAFRICA.2014.6880657]
31. Konstantinidis S, Kitsou A, Bamidis P, Mikos T, Zafrakas M, Nikolaidou M, et al. A SCORM compliant e-learning platform for cervical cancer pPrevention. IEEE Xplore; 2008 Jun 19 Presented at: Computer-Based Medical Systems, 2008; June 19, 2008; Jyvaskyla, Finland. [doi: 10.1109/CBMS.2008.88]
32. Popescu D, Avramescu ET, Pascual LJ, Pratikakis I, Perantonis S, Lupescu O. A Web-based e-training platform for biomedical engineering education. 2014 Jun 19 Presented at: 2014 International Conference and Exposition on Electrical and Power Engineering; 2014; Iasi, Romania URL: http://ieeexplore.ieee.org/document/6969889/ [doi: 10.1109/ICEPE.2014.6969889]
33. Safwat A, Pourabdollah A. E-learning for healthcare professionals towards HIS in Egypt. 2009 Dec 16 Presented at: 7th International Conference on Communications and Information Technology (ICICT); December 16, 2009; Cairo, Egypt URL: http://ieeexplore.ieee.org
34. Sranacharoenpong K, Hanning RM, Sirichakwal PP, Chittchang U. Process and outcome evaluation of a diabetes prevention education program for community healthcare workers in Thailand. Educ Health (Abingdon) 2009 Dec;22(3):335. [Medline: 20029769]
35. Walsh K, Rafiq I, Hall R. Online educational tools developed by Heart improve the knowledge and skills of hospital doctors in cardiology. Postgrad Med J 2007 Jul;83(981):502-503 [FREE Full text] [doi: 10.1136/pgmj.2007.059501] [Medline: 17621623]
36. Wang F. Valuation of Online Continuing Medical Education and Telemedicine in Taiwan. J Educ Techno Soc 2008 Jan 01;11(4):190 [FREE Full text]
37. Westbrook C. Online collaborative learning in health care education. EURODL 2012 Jan 26;1:1-6 [FREE Full text]
38. Rahimi E, van den Berg J, Veen W. A learning model for enhancing the student's control in educational process using Web 2.0 personal learning environments. Br J Educ Technol 2014 May 13;46(4):780-792. [doi: 10.1111/bjet.12170]
39. Salmon G. E-moderating: the kiey to teaching and learning online. London: Routledge; Jan 22, 2004.

40. Kirkpatrick Partners. 2018. The new world Kirkpatrick model URL: https://www.kirkpatrickpartners.com/ [accessed 2016-05-12] [WebCite Cache ID 6xrqgx10w]
41. Koehler N. Medical students' use of and attitudes towards medical applications. Journal MTM 2012 Dec 22;1(4):16-21. [doi: 10.7309/jmtm.73]
42. Wallace S, Clark M, White J. 'It's on my iPhone': attitudes to the use of mobile computing devices in medical education, a mixed-methods study. BMJ Open 2012;2(4):2012-00109 [FREE Full text] [doi: 10.1136/bmjopen-2012-001099] [Medline: 22923627]
43. Cubric M. Wiki-based process framework for blended learning. In: Proceeding of International Symposium on Wikis, 2007. 2007 Oct 25 Presented at: International Symposium on Wikis; October 21-25, 2007; Montreal, Canada p. 11-24. [doi: 10.1145/1296951.1296953]
44. Roblyer M. Educational technology research that makes a differenceries introduction. CITE 2005(Online serial, 5(2)):746-748 [FREE Full text]

Abbreviations

CCAT: Crowe Critical Appraisal Tool
CPD: continuous professional development
PICOS: Population, Intervention, Comparison, and Outcomes
PLE: Personal Learning Environment
TEL: technology-enhanced learning

Edited by G Eysenbach; submitted 29.09.17; peer-reviewed by S Manca, KC Wong, R Pankomera; comments to author 26.10.17; revised version received 18.12.17; accepted 21.12.17; published 11.04.18.

Please cite as:
Nicoll P, MacRury S, van Woerden HC, Smyth K
Evaluation of Technology-Enhanced Learning Programs for Health Care Professionals: Systematic Review
J Med Internet Res 2018;20(4):e131
URL: http://www.jmir.org/2018/4/e131/
doi:10.2196/jmir.9085
PMID:29643049

Original Paper

Patient Continued Use of Online Health Care Communities: Web Mining of Patient-Doctor Communication

Bing Wu[1], PhD

School of Economics and Management, Tongji University, Shanghai, China

Corresponding Author:
Bing Wu, PhD
School of Economics and Management
Tongji University
1239 Siping Road
Shanghai,
China
Phone: 86 65982200
Email: ww_bing@163.com

Abstract

Background: In practice, online health communities have passed the adoption stage and reached the diffusion phase of development. In this phase, patients equipped with knowledge regarding the issues involved in health care are capable of switching between different communities to maximize their online health community activities. Online health communities employ doctors to answer patient questions, and high quality online health communities are more likely to be acknowledged by patients. Therefore, the factors that motivate patients to maintain ongoing relationships with online health communities must be addressed. However, this has received limited scholarly attention.

Objective: The purpose of this study was to identify the factors that drive patients to continue their use of online health communities where doctor-patient communication occurs. This was achieved by integrating the information system success model with online health community features.

Methods: A Web spider was used to download and extract data from one of the most authoritative Chinese online health communities in which communication occurs between doctors and patients. The time span analyzed in this study was from January 2017 to March 2017. A sample of 469 valid anonymous patients with 9667 posts was obtained (the equivalent of 469 respondents in survey research). A combination of Web mining and structural equation modeling was then conducted to test the research hypotheses.

Results: The results show that the research framework for integrating the information system success model and online health community features contributes to our understanding of the factors that drive patients' relationships with online health communities. The primary findings are as follows: (1) perceived usefulness is found to be significantly determined by three exogenous variables (ie, social support, information quality, and service quality; R^2=0.88). These variables explain 87.6% of the variance in perceived usefulness of online health communities; (2) similarly, patient satisfaction was found to be significantly determined by the three variables listed above (R^2=0.69). These variables explain 69.3% of the variance seen in patient satisfaction; (3) continuance use (dependent variable) is significantly influenced by perceived usefulness and patient satisfaction (R^2=0.93). That is, the combined effects of perceived usefulness and patient satisfaction explain 93.4% of the variance seen in continuance use; and (4) unexpectedly, individual literacy had no influence on perceived usefulness and satisfaction of patients using online health communities.

Conclusions: First, this study contributes to the existing literature on the continuance use of online health communities using an empirical approach. Second, an appropriate metric was developed to assess constructs related to the proposed research model. Additionally, a Web spider enabled us to acquire objective data relatively easily and frequently, thereby overcoming a major limitation of survey techniques.

(J Med Internet Res 2018;20(4):e126) doi:10.2196/jmir.9127

KEYWORDS

health information management; health communication; information literacy; social networking

Introduction

Background

Research indicates that online health communities (OHCs) have become major sources of health information for the general public and have dramatically changed health information seeking and exchange [1]. Individuals seek online health information for various reasons such as quick and easy access, anonymity or privacy, the variety of available information, cost-effectiveness, and improved communication. Thus, health information acquired online can increase people's knowledge and further develop personal skills and abilities [2].

In practice, OHCs have passed the adoption stage to reach the diffusion phase of development. In this phase, patients equipped with knowledge regarding health care issues are capable of switching between different communities to maximize their OHC activities. OHCs employ doctors to answer patient questions, and high quality OHCs are more likely to be acknowledged by patients. Therefore, one important question that must be addressed is: what factors drive patients to maintain ongoing relationships with OHCs in which communication occurs between doctors and patients? However, this important question has received limited scholarly attention [3].

Objective

The goal of this study was to examine the factors that drive patients to perpetuate ongoing relationships with OHCs in which communication occurs between doctors and patients, by integrating the information system success model (ISSM) and OHC features. Methodologically, the study uses a combination of Web mining and structural equation modeling (SEM) [4] to analyze data captured by a Web spider from one of the most authoritative OHCs in China. This approach can circumvent many of the problems associated with survey data, such as expense and the need for an acceptable response rate.

Methods

Research Model

An OHC is a group of individuals who interact in health and wellness-related virtual communities to seek information, assistance, emotional support, and communication opportunities [5]. Many benefits of OHC participation have been identified, including increased support, perceived empathy, optimism, reduced levels of stress, depression, and psychological trauma [6]. Compared with other online communities, OHCs aim to provide various types of social support [7]. Thus, individual literacy and social support are two important OHC features [8]. In recent years, OHC research has increasingly begun to emphasize relationship-building [9] and psychological empowerment [10]. Although these studies have revealed various interesting findings, the factors that govern the patient continuance use of OHCs remain unknown [3].

As a critical indicator of information system (IS) success, continuance use is essential for realizing the value of IS within organizations [11]. Therefore, the measurement of IS success is important for assessing the effectiveness of IS governance and IS investments. Previous research has identified IS success measures 12. It has been proposed that a system can be evaluated in terms of quality measures such as information quality, system quality, and service quality. These characteristics affect subsequent use and user satisfaction, and contribute to the belief that certain benefits will be achieved by using the system [12].

The validity of the ISSM in various contexts, which has been demonstrated through a variety of empirical studies, indicates that ISSM are well accepted by IS scholars and useful for practitioners [13]. So, it is assumed that the updated ISSM can be adapted for system success measurement in the OHC context [14]. Considering that system quality has been an essential prerequisite for information systems, especially for OHCs [15], we used measures related to quality (ie, information quality and service quality).

In summary, the aim of this study was to develop a research model integrating the ISSM and OHC features to study which factors drive patients to maintain ongoing relationships with OHCs in which communication occurs between doctors and patients. The proposed model identifies several attributes as predictors of a patient's OHC continuance use. The relationships between the constructs are depicted in the conceptual model demonstrated in Figure 1. The continuance use expresses the likelihood that patients will continue to visit and participate in an OHC. Patient satisfaction and perceived usefulness are primarily subjective concepts influenced by individual literacy, social support, information quality, and service quality. Because no reliable measures developed by Web mining are found in the literature [16], we adjusted and developed a metric to measure constructs in the proposed model.

Individual Literacy–Related Hypotheses

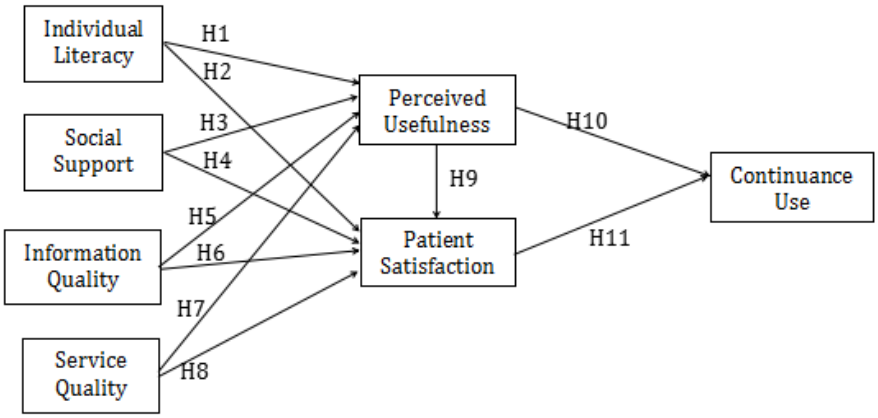
Although the current use of OHCs is extensive and their possible benefit for the public is substantial, website limitations prevent patients from exploiting the full potential of OHCs. Problems associated with retrieving information from OHCs tend to involve language, literacy, education, technology, and cultural barriers [17].

The relationship between individual literacy and perceived usefulness is of interest to many scholars [18]. Three types of literacy (ie, information literacy, media literacy, and health literacy) are related to understanding and acting on individual literacy in this study. They can be measured by Web mining patients' posts [19]. We define patient post as a new thread posted by a patient that can be followed by replies from doctors and additional questions from the patient.

Information Literacy

Information literacy is defined as a set of skills that are necessary to identify, access, evaluate, and use information effectively [20]. It is important for individuals who wish to access health information and relates to their ability to share content [21]. The text length of a patient's post [22] is an important indicator of the patient's information literacy.

Figure 1. Proposed research model.



Thus, the information literacy of a patient *i* (ie, $IL1_i$) could be evaluated by the average text length of each post (equation 1, Figure 2), where $length(p_{ij})$ is the text length of the j^{th} post p_{ij} by patient *i* and *n* is the total number of posts by patient *i* during a certain time interval.

Media Literacy

Media literacy is defined as the ability to evaluate and judge the accuracy of media information [23]. The sentiment of a patient's post [24, which refers to feelings and emotions expressed by a patient through their post, can be used as a critical indicator of the patient's media literacy. Sentiment analysis can be used to systematically quantify subjective information from a patient's post. We use the Chinese sentiment polarity dictionary, National Taiwan University Semantic Dictionary (NTUSD), developed by Taiwan University, to identify the sentiment of each post p_{ij} of patient *i* as a negative, neutral, or positive emotion (equation 2, Figure 2) [25], where the sentiment score $senti(p_{ij})$ is the j^{th} post p_{ij} by patient *i*, $length(p_{ij})$ is the text length of the j^{th} post p_{ij} by patient *i*,*s* is number of sentences in the j^{th} post p_{ij} by patient *i*, and *t* is sentiment expression terms in sentences, $pos(t,s)$ is part of *t* in sentence *s*, $pol(t,s)$ polarity of *t* in sentence *s*, and $ntusd(t, pos(t,s), pol(t,s))$ is the NTUSD intensity value for term *t* based on its sentence part $pos(t,s)$ and polarity $pol(t,s)$. The media literacy of patient *i* (ie, $IL2_i$) could be calculated as the average sentiment of each post (equation 3, Figure 2), assuming the total number of posts by patient *i* is *n* during a certain time interval.

Health Literacy

Health literacy has been defined as the ability to read, comprehend, and act upon health information [26]. The number of medical terms in a patient's post can be an indicator of the patient's proficiency in health literacy [27]. We began by extracting the medical terminology from the tag in the OHC. Next, we counted the examples of these terminologies contained in each post of patient *i*,. Health literacy of patient *i* (ie, $IL3_i$) could then be calculated as the average number of medical terms in each post (equation 4; Figure 2), where $medt(p_{ij})$ is the number of medical terms in the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Individual Literacy

Generally, individual literacy is directly related to the ability to communicate with others [28] and to express individual needs. This, in turn, affects the patient's perceived usefulness of and satisfaction with OHCs. Accordingly, we propose the following research hypotheses:

> Hypothesis 1: individual literacy has a positive effect on perceived usefulness of OHCs.
>
> Hypothesis 2: individual literacy has a positive effect on patient satisfaction when using OHCs.

Social Support–Related Hypotheses

The social support aspect is used to assess whether patients can benefit from social relationships in the OHC context [29]. Social support in OHCs refers to the verbal and nonverbal communication exchanged between recipients and providers that reduces uncertainty regarding the situation or the relationship, and functions to enhance a perception of personal control in one's experience [30].

Three types of social support (ie, companionship support, informational support, and emotional support) [31] are examined in this study. They can be measured by Web mining of doctors' replies to patients' posts.

Companionship Support

Companionship support gives someone a sense of social belonging, through the presence of companions to engage in shared social activities [32]. Companionship support in OHCs refers to communications of opinions or facts relevant to current health-related difficulties (eg, advice, personal feedback, and information) that may provide a solution to a problem. Such support also increases information sharing, which in turn leads to a better patient-doctor experience. Thus, companionship support to patient *i* (ie, $SS1_i$) could be calculated as the average number of doctors who replied to each post by the patient (equation 5, Figure 2), where $num(drp_{ij})$ is the total number of doctors who replied to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Figure 2. Equations.

$$IL1_i = \frac{1}{n}\sum_{j=1}^{n} length(p_{ij}) \quad (1)$$

$$sent(p_{ij}) = \frac{1}{length(p_{ij})}\sum_{s=1}^{length(p_{ij})}\sum_{t \in s} ntusd(t, pos(t,s), pol(t,s)) \quad (2)$$

$$IL2_i = \frac{1}{n}\sum_{j=1}^{n} sent(p_{ij}) \quad (3)$$

$$IL3_i = \frac{1}{n}\sum_{j=1}^{n} med(p_{ij}) \quad (4)$$

$$SS1_i = \frac{1}{n}\sum_{j=1}^{n} num(drp_{ij}) \quad (5)$$

$$SS2_i = \frac{1}{n}\sum_{j=1}^{n} num(brp_{ij}) \quad (6)$$

$$SS3_i = \frac{1}{n}\sum_{j=1}^{n} sent(rp_{ij}) \quad (7)$$

$$IQ1_i = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{num(drp_{ij})}\sum_{k=1}^{num(drp_{ij})} title_k(drp_{ij}) \quad (8)$$

$$IQ2_i = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{num(drp_{ij})}\sum_{k=1}^{num(drp_{ij})}\frac{tag_k(drp_{ij})}{tag(p_{ij})} \quad (9)$$

$$IQ3_i = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{num(drp_{ij})}\sum_{k=1}^{num(drp_{ij})} med_k(rp_{ij}) \quad (10)$$

$$SQ1_i = \frac{1}{n}\sum_{j=1}^{n}[t(rp_{ij}) - t(p_{ij})] \quad (11)$$

$$SQ2_i = \frac{1}{n}\sum_{j=1}^{n} length(rp_{ij}) \quad (12)$$

$$SQ3_i = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{num(drp_{ij})}\sum_{k=1}^{num(drp_{ij})} count_k(drp_{ij}) \quad (13)$$

$$PU1_i = \frac{1}{n}\sum_{j=1}^{n} sent(fp_{ij}) \quad (14)$$

$$PU2_i = \frac{1}{n}\sum_{j=1}^{n}\frac{length(fp_{ij})}{num(fp_{ij})} \quad (15)$$

$$PU3_i = \frac{1}{n}\sum_{j=1}^{n}[t(fp_{ij}) - t(p_{ij})] \quad (16)$$

$$PS1_i = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{num(drp_{ij})}\sum_{k=1}^{num(drp_{ij})} pic_k(drp_{ij}) \quad (17)$$

$$PS2_i = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{num(drp_{ij})}\sum_{k=1}^{num(drp_{ij})} ove_k(drp_{ij}) \quad (18)$$

$$PS3_i = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{num(drp_{ij})}\sum_{k=1}^{num(drp_{ij})} vge_k(drp_{ij}) \quad (19)$$

$$CU1_i = t(p_{i1}) - t_0 \quad (20)$$

$$CU2_i = \frac{n}{t(p_{ij}) - t(p_{i1})} \quad (21)$$

Informational Support

Informational support refers to advice, guidance, suggestions, or useful information provided to someone. Informational support in OHCs involves a doctor-patient interaction where the former assists latter with managing particular questions and uncertainties regarding a given situation in a highly supportive manner [33]. The number of best answers provided by doctors is an important indicator of valuable support to patients. Valuable support to patient *i* (ie, $SS2_i$) could be calculated as the average number of best answers provided by doctors to each post by the patient (equation 6, Figure 2), where $num(brp_{ij})$ is the total number of best answers provided by doctors to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Emotional Support

Patients in an OHC are influenced by the emotional support they receive and the quality of information and companionship they are provided [34]. Interaction through OHCs can increase emotional support and lead to a sense of self-esteem in patients [35]. Studies have reprted on the positive impact of such support on overall health outcomes, health education, and patient empowerment [36]. Emotional support to patient *i* (ie, $SS3_i$) could be calculated as the average reply sentiment of the doctors to each post (equation 7, Figure 2), where sentiment is calculated according to equation 2 in Figure 2, $senti(rp_{ij})$ is the reply sentiment of the doctors to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Social Support

The social support of OHCs is credited with many positive health outcomes such as increased immune system strength and reduced stress and depression [37]. Here, we suggest that positive social support will translate into positive perceptions of OHCs. Thus, the following hypotheses are proposed:

> Hypothesis 3: social support has a positive effect on perceived usefulness of OHCs.
>
> Hypothesis 4: social support has a positive effect on patient satisfaction when using OHCs.

Information Quality–Related Hypotheses

Information quality in OHCs refers to content issues and includes the completeness, accuracy, format, and currency aspects of the information [38]. Three types of information quality (ie, information reliability, information relevancy, and information consistency) [39] are identified in this study that can be measured by Web mining doctors' replies to patients' posts.

Information Reliability

Information reliability is, literally, the extent to which an individual can rely on the source of the information [40]. The official title of doctors certified by a national agency with uniform standards is a guarantee of reliable information sources. Thus, the reliability of information sources for patient *i* (ie, $IQ1_i$) could be calculated as the average title of the doctors who replied to each post (equation 8, Figure 2), where $title(drp_{ij})$ is the academic title of the k^{th} doctors who replied to the j^{th} post p_{ij} by patient *i*, $num(drp_{ij})$ is the total number of doctors who replied to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval. A scale of 1 to 5 was used to normalize the job titles (1=assistant doctor, 2=resident doctor, 3=chief doctor, 4=associate archiater, and 5=archiater).

Information Relevancy

Information relevancy in OHCs means that information is applicable and helpful for health-related goals of patients [41]. Thus, we calculated the degree of relevancy between the tags of each post field of patient *i* replied to by doctors and the expertise tag of each replying doctor. Then, the information relevancy to patient *i* (ie, $IQ2_i$) could be calculated as the average number of the relevancy degree (equation 9, Figure 2), where $tag_k(drp_{ij})$ divided by $tag(p_{ij})$ is the number of the relevancy degree of the k^{th} doctors who replied to the j^{th} post p_{ij} by patient *i*, $num(drp_{ij})$ is the total number of doctors who replied to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Information Consistency

Information consistency in OHCs denotes that information is always presented in the same format and is in line with previous information [42]. The number of examples of medical terminology in a doctor's reply could be a key indicator of information consistency in OHCs. On the basis of the medical terminology extracted from the tag in the OHC, we counted the number of medical terms in each reply to a post. Therefore, the information consistency to patient *i* (ie, $IQ3_i$)could be calculated as the average number of medical terms in the replies of doctors to each post (equation 10, Figure 2), where $med(rp_{ij})$ is the number of medical terms in the replies of k^{th} doctors who answered the j^{th} post p_{ij} by patient *i*, $num(drp_{ij})$ is the total number of doctors who replied to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Information Quality

As a necessary measure of success, OHC information quality is a highly vital factor that affects the patient's online behavior. This factor indicates the importance of quality online information and references [43]; and information quality significantly affects perceived usefulness [12]. Thus, we propose the following research hypotheses:

> Hypothesis 5: information quality has a positive effect on perceived usefulness of OHCs.
>
> Hypothesis 6: information quality has a positive effect on patient satisfaction when using OHCs.

Service Quality–Related Hypotheses

OHC service quality is defined by the feelings of patients regarding the level of service while browsing and interacting within the community [44]. Three types of service quality (ie, service responsiveness, service effort, and service empathy) are identified in this study that can be measured by Web mining doctors' replies to patients' posts.

Service Responsiveness

Service responsiveness in OHCs refers to the willingness of doctors to help patients with prompt service. Thus, service responsiveness [45] of doctors to patient *i* (ie, $SQ1_i$) could be measured as the average time taken for the first reply to be posted to the j^{th} post $[t(rp_{ij})-t(p_{ij})]$ as shown in Figure 2 (equation 11), assuming the total number of posts by patient *i* is *n* during a certain time interval.

Service Effort

Service effort is defined in OHCs as the amount of time and energy the doctor invests in replying to patients [46]. Thus, service effort exerted in responsiveness by doctors to patient *i* (ie, $SQ2_i$) could be evaluated by the average text length of replies to each post [22], as shown in Figure 2 (equation 12), where $length(rp_{ij})$ is the text length of replies to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Service Empathy

Service empathy in the OHC means that the doctors provide individualized attention to the patients [47]. The average number of patients to whom doctors have replied could be used as a critical indicator of service empathy. Thus, the service empathy [48] of doctors to patient *i* (ie, $SQ3_i$) could be evaluated by the average number of patients to whom doctors have replied (equation 13, Figure 2), where $count_k(drp_{ij})$ is the total number of patients to whom the k^{th} doctor replied, in which the k^{th} doctor is the one who replied to j^{th} post p_{ij} by patient *i*, $num(drp_{ij})$ is the total number of doctors who replied to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval *i*.

Service Quality

Service quality at OHCs differs from the traditionally studied service quality in IS research, which focuses on the services of organizations [49]. OHCs not only support the creation, exchange, and perception of information but also create a virtual community to support the use of collaborative and interactive services. In recent years, there have been a number of studies on how service quality affects perceived usefulness and patient satisfaction in OHCs [50]. Several studies have observed that service quality has a significant effect on patient satisfaction [50].

This study proposes that service quality influences the perceived usefulness and patient satisfaction of OHCs, which, in turn,

influences continuance use. Thus, the following hypotheses are proposed:

> Hypothesis 7: service quality has a positive effect on perceived usefulness of OHCs.
>
> Hypothesis 8: service quality has a positive effect on patient satisfaction when using OHCs.

Perceived Usefulness–Related Hypotheses

On the basis of its definition as capable of being used advantageously, perceived usefulness is explained as the degree to which a person believes that using a particular system would enhance his or her job performance [51]. The perceived usefulness of an OHC can be described as the extent to which a patient believes that an OHC can be useful in achieving health-related goals.

A patient post can be followed by replies from doctors, and then be followed by additional questions from the patient. The feedback provided through this back-and-forth dialogue adds to the perceived usefulness of patients [17,52]. Thus, three types of perceived usefulness (ie, feedback emotion, feedback effort, and feedback time) are identified in this study that can be measured by Web mining patients' additional questions.

Feedback Emotion

As indicators of whether or not motives are satisfied in OHCs, emotions in additional questions show how patients feel about replies from doctors. Thus, emotions in patient feedback *i* (ie, $PU1_i$) could be calculated as the average sentiment of additional questions to each post (equation 14, Figure 2), where $senti(fp_{ij})$ is the sentiment of each additional question of the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Feedback Effort

The patient feedback effort in OHCs is the amount of time and energy the patient invests in feedback to doctors. The patients can signal their perceived usefulness by feedback efforts. Thus, the feedback effort of patient *i* (ie, $PU2_i$) could be calculated as the average ratio of length to number of additional questions to each post (equation 15, Figure 2), where $num(fp_{ij})$ is the total number of additional questions of the j^{th} post p_{ij} by patient *i*, $length(fp_{ij})$ the text length of additional questions of the j^{th} post p_{ij} by patient *i* and, *n* is the total number of posts by patient *i* during a certain time interval.

Feedback Time

Feedback time in OHCs is the total amount of time it takes for patients to respond to replies from doctors. Thus, the feedback time of patient *i* (ie, $PU3_i$) could be evaluated by the average time taken before posting the first additional question following a post (equation 16, Figure 2), where $t(fp_{ij})-t(p_{ij})$ is the time interval until the first additional question to the j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Perceived Usefulness

A system high in perceived usefulness is one that a user believes offers a positive use-performance relationship [53]. Several prior studies suggest that perceived usefulness is the antecedent of overall customer satisfaction. Additionally, perceived usefulness is a construct that has been repeatedly shown to influence attitude and is a direct determinant of continued IS use [54]. Thus, the following hypotheses are proposed:

> Hypothesis 9: perceived usefulness has a positive effect on patient satisfaction when using OHCs.
>
> Hypothesis 10: perceived usefulness has a positive effect on the continuance use of OHCs.

Patient Satisfaction–Related Hypotheses

In the OHC context, patient satisfaction describes the feelings, attitudes, and expectations of patients who perceive that they have received good services from an OHC and are likely to revisit. Since the 1980s, user satisfaction has received considerable attention in the IS literature and remains a focus of considerable research interest [55]. The quality of a user's experience and satisfaction with a website has been used in recent research as determinants of success [56].

Patients can express opinions to doctors conveniently and easily in OHCs; thus, their satisfactions can be broadly recognized by the total number of public thank-you letters, online votes, and virtual gifts to doctors. Therefore, three types of patient satisfaction are identified in this study (ie, satisfaction via thank-you letters, satisfaction via online votes, and satisfaction via virtual gifts).

Satisfaction via Thank-You Letters

Thank-you letters are written by patients to doctors to express appreciation in OHCs. Thus, satisfaction via thank-you letters of patient *i* (ie, $PS1_i$) could be evaluated by the average number of thank-you letters received by doctors who replied to a post (equation 17, Figure 2), where $tyl_k(drp_{ij})$ is the total number of thank-you letters for the k^{th} doctor who replied to the j^{th} post p_{ij} by patient *i*, $num(drp_{ij})$ is the total number of doctors who replied to j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Satisfaction via Online Votes

Online votes show praise given by patients to doctors in OHCs. Thus, satisfaction via online votes of patient *i* (ie, $PS2_i$) could be evaluated by the average number of online votes to doctors who replied to posts (equation 18, Figure 2), where $ov_k(drp_{ij})$ is the total number of online votes for the k^{th} doctor who replied to the j^{th} post p_{ij} by patient *i*, $num(drp_{ij})$ is the total number of doctors who replied to *the* j^{th} post p_{ij} by patient *i*, and *n* is the total number of posts by patient *i* during a certain time interval.

Satisfaction via Virtual Gifts

Virtual gifts are sent to doctors by patients in OHCs. Thus, satisfaction via virtual gifts of patient *i* (ie, $PS3_i$) could be evaluated by the average number of virtual gifts given to doctors who replied to posts (equation 19, Figure 2), where $vg_k(drp_{ij})$

is the total number of virtual gifts for the k^{th} doctor who replied to the j^{th} post p_{ij} by patient i, $num(drp_{ij})$ is the total number of doctors who replied to the j^{th} post p_{ij} by patient i, and n is the total number of posts by patient i during a certain time interval.

Patient Satisfaction

As DeLone and McLean [12] suggest, a positive use experience will result in increased user satisfaction. Similarly, increased patient satisfaction will result in increased continuance use [57]. Because user satisfaction is a significant influencer of user retention, it is often regarded as the most useful and easiest way of evaluating an IS [14]. Thus, the following hypothesis is proposed:

> Hypothesis 11: patient satisfaction has a positive effect on the continuance use of OHCs.

Continuance Use

Typically, continuance use illustrates patients' behavior to continue using OHCs [58,59]. Thus, two types of patient continuance use are identified in this study (ie, membership length and frequency of posts from patients).

The membership length of patient i (ie, $CU1_i$) could be calculated as the time span from registration to the present (equation 20, Figure 2).

The posting frequency of patient i (ie, $CU2_i$) could be calculated as n posts during $t(p_{ij})-t(p_{i1})$ as shown in Figure 2 (equation 21), assuming the total number of posts by patient i is n during a certain time interval.

Data Collection

The objective of this study was to understand the factors that drive patients to maintain ongoing relationship with OHCs in which communication occurs between doctors and patients. Although survey instruments are often used in such circumstances, data are difficult to obtain, and data availability and sample size problems may occur. Therefore, this study collected objective data directly from OHCs by using a Web spider.

To test our hypotheses, we captured data from a Chinese OHC (the club.xywy website). This OHC is one of the most authoritative forums in China that connects individuals with health problems with doctors who can help them solve these problems.

In this study, Locoy Spider (Hefei loy Information Technology Co, Ltd China) was employed to download and extract data. This software is freely available and can be downloaded from the Locoy website. The time span analyzed in our study was from January 2017 to March 2017. We preprocessed the data according to the following rules: (1) delete records with no doctor's reply; (2) delete records not in accordance with the required format; and (3) delete records with advertisements. Finally, 469 valid anonymous patients with 9667 posts were obtained, which is equal to 469 respondents in survey research. Table 1 summarizes the demographics of the patients in the dataset.

Measure Items in the Dataset

Table 2 summarizes the measure items in the dataset according to research hypotheses. We used z-score to normalize each measure item before data analysis.

Data Analysis

First, we examined the fitness and construct validity of the proposed measurement model by assessing reliability, convergent validity, and discriminant validity. Then, we examined the structural model to investigate the strength and direction of the relationships among the theoretical constructs.

Construct Validity

Evaluation of Reliability and Convergent Validity

The reliability of all multi-item constructs should exhibit a Cronbach alpha larger than .70. Convergent validity should meet the following criterion: the indicator's estimated coefficient should be significant with respect to its posited underlying construct factor.

The measurement scales were evaluated using three criteria: all item-factor loadings (k) should be significant and exceed 0.7, the composite reliability (CR) for each construct should exceed 0.7, and the average variance extracted (AVE) for each construct should be larger than 0.5 [60]. Table 3 shows that the item loading, AVE, CR, and Cronbach alpha values for all constructs in the measurement model exceeded the recommended threshold values. In sum, the adequacy of the measurement model indicated that all items were reliable indicators of the hypothesized constructs.

Table 1. Demographics characteristics of the patients.

Items	Frequency (n=469), n (%)
Gender	
Male	250 (53.3)
Female	219 (46.7)
Age (years)	
Under 30	114 (24.3)
30 or above	355 (75.7)

Table 2. Measure items in the dataset.

Construct code and measures		Mean (SD)
Individual literacy (IL) [20,23,26]		
IL1	Patient post text length (to measure information literacy from low to high)	18.8 (15.3)
IL2	Patient post sentiment (to measure media literacy from low to high)	8.2 (2.4)
IL3	Number of medical terms in a patient post (to measure health literacy from low to high)	7.7 (3.0)
Social support (SS) [16,32,34]		
SS1	Number of doctors who reply to a patient post (to measure social contact from low to high)	852.7 (309.4)
SS2	Number of best answers to a patient post (to measure valuable assistance from low to high)	15.1 (9.9)
SS3	Sentiment of the doctor's reply to a patient post (to measure emotional support from low to high)	10.7 (3.8)
Information quality (IQ) [40,41,42,59]		
IQ1	Job titles of doctors who reply to a patient post (to measure information reliability from low to high)	3.5 (1.8)
IQ2	Relevance degree between the expertise of the doctors who reply to a patient post and the field addressed by a patient post (to measure information relevancy from low to high)	5.5 (3.6)
IQ3	Number of medical terms in the doctor's reply (to measure information consistency from low to high)	10.2 (8.1)
Service quality (SQ) [45,46,47]		
SQ1	Responsiveness of the doctors to a patient post (to measure service responsiveness from low to high)	4.8 (2.3)
SQ2	Text length of doctors' replies (to measure service effort from low to high)	20.9 (17.3)
SQ3	Number of patients replied to by doctors (to measure service empathy from low to high)	189.6 (148.3)
Perceived usefulness (PU) [17,52,54]		
PU1	Sentiment of the patient's additional questions (to measure feedback emotion from low to high)	8.7 (4.2)
PU2	Ratio of length to number of the patient's additional questions (to measure feedback effort from low to high)	13.8 (8.1)
PU3	Time interval until the patient's first additional question (to measure feedback time from low to high)	15.1 (13.3)
Patient satisfaction (PS) [55,56]		
PS1	Number of thank-you letters to doctors (to measure satisfaction via thank-you letters from low to high)	3.5 (1.2)
PS2	Number of online votes awarded to doctors (to measure satisfaction via online votes from low to high)	12.1 (22.3)
PS3	Number of virtual gifts sent to doctors (to measure satisfaction via virtual gifts from low to high)	16.2 (33.5)
Continuance use (CU) [58]		
CU1	The time span from registration to the present (to measure membership length from low to high)	20.6 (17.1)
CU2	Frequency of the patient's posts (to measure posting frequency from low to high)	10.1 (8.1)

Discriminant Validity

To test discriminant validity, the average variance shared between a construct and its measures should be larger than the variance shared by the construct and all other constructs in the model [61]. The correlation analysis is shown in Table 4.

Goodness of Fit

The structural model was tested to assess how well the model represented the data. We evaluated the following indices [62]: the chi-square test statistic, the goodness-of-fit index (GFI), the normed fit index, the comparative fit index, the Tucker-Lewis index, and the root mean square residual. Table 5 presents the results and recommended values. We conclude that the GFIs displayed the recommended levels, which suggests that the model provided a good fit with the data.

Structural Model for Hypothesis Testing

The purpose of the analysis was to determine which factors and how these factors affect the continuance use. SEM using partial least squares (PLS) was applied to test the hypothesized model. Use of SEM is commonly justified in the social sciences because of its ability to impute relationships between unobserved constructs (latent variables) from observable variables, and PLS is a latent SEM technique that uses a component-based approach to estimation [63]. These techniques facilitate the analysis of a structural model that assesses relationships among theoretical constructs and a measurement model that assesses the reliability and validity of measures. The test of the structural model includes estimates of the path coefficients, which indicate the strengths of the relationships between the dependent and independent variables, and values, which represent the amount of variance explained by the independent variables. Together, the path coefficients (ie, the loadings and significance and R^2) indicate how well the data support the hypothesized model.

Table 3. Construct reliability and convergent validity.

Construct code	Item loadings	Average variance extracted	Composite reliability	Cronbach alpha
Individual literacy (IL)		0.77	0.94	.89
IL1	0.95			
IL2	0.89			
IL3	0.89			
Social support (SS)		0.74	0.96	.91
SS1	0.95			
SS2	0.91			
SS3	0.86			
Information quality (IQ)		0.76	0.92	.89
IQ1	0.95			
IQ2	0.89			
IQ3	0.87			
Service quality (SQ)		0.74	0.96	.85
SQ1	0.93			
SQ2	0.85			
SQ3	0.85			
Perceived usefulness (PU)		0.73	0.92	.90
PU1	0.95			
PU2	0.89			
PU3	0.89			
Patient satisfaction (PS)		0.82	0.87	.88
PS1	0.93			
PS2	0.87			
PS3	0.86			
Continuance use (CU)		0.77	0.93	.86
CU1	0.91			
CU2	0.92			

Table 4. Interconstruct correlations and discriminant validity.

Constructs	Individual literacy (IL)	Social support (SS)	Information quality (IQ)	Service quality (SQ)	Perceived usefulness (PU)	Patient satisfaction (PS)	Continuance use (CU)
IL	0.85[a]						
SS	0.46	0.83[a]					
IQ	0.26	0.34	0.68[a]				
SQ	0.05	0.22	0.18	0.77[a]			
PU	0.18	0.38	0.24	0.37	0.61[a]		
PS	0.36	0.12	0.10	0.27	0.53	0.74[a]	
CU	0.27	0.19	0.16	0.18	0.15	0.19	0.63[a]

[a]The average variance extracted for the reflective variables is consistently larger than the off-diagonal squared correlations, which suggests satisfactory discriminant validity among variables.

Table 5. Overall model t indices for the research model.

Model t indices	Results value	Recommended value
Chi-square or degrees of freedom	1.88	≤3
Goodness-of-fit index	0.93	≥0.9
Normed fit index	0.96	≥0.9
Comparative fit index	0.92	≥0.9
Tucker-Lewis index	0.92	≥0.9
Root mean square residual	0.07	≤0.08

Results

Path Analysis

Figure 3 illustrates the resulting path coefficients of the proposed research model: (1) perceived usefulness is found to be significantly determined by social support, information quality, and service quality (R^2=0.88). These variables explain 87.6% of the variance seen in perceived usefulness; (2) similarly, patient satisfaction is found to be significantly determined by social support, information quality, and service quality (R^2=0.69). These variables explain 69.3% of the variance seen in patient satisfaction; and (3) the dependent variable (continuance use) is significantly determined by perceived usefulness and patient satisfaction (R^2=0.93). That is, the combined effects of perceived usefulness and patient satisfaction explain 93.4% of the variance in continuance use.

A summary of the hypothesis test results of the standardized path coefficients and path significances is provided in Table 6. Most of the paths are significant in the expected direction. Overall, 9 out of 11 hypotheses are supported by the data. The insignificant variables are individual literacy and perceived.

Influence of Exogenous Variables on Perceived Usefulness and Patient Satisfaction

Unexpectedly, two hypotheses (hypothesis 1 and hypothesis 2), that were supported in previous empirical research, were not supported by our data. The results indicate that individual literacy has no influence on perceived usefulness and patient satisfaction. There are two plausible reasons for this result. First, this may be because nearly 75.7% (355/469) of the sample are older than 30 years, and they seek health information on behalf of others, such as their parents or children. Second, although individual literacy provides the ability to use an OHC, perceived usefulness and patient satisfaction depend on the functionality and characteristics of OHCs. From this perspective, the results are logical.

Social support, information quality, and service quality exert significant direct effects on perceived usefulness and patient satisfaction. Thus, hypothesis 3 to 8 were supported. Generally, social support is an important component of OHCs. In terms of the main effects of social support on perceived usefulness and patient satisfaction, our findings agree with previous research results [64]. Because patients who visit OHCs may regard such communities as repositories of information, information quality and service quality significantly influence perceived usefulness and patient satisfaction. That is, if patients perceive a high quality of information and service, what matters the most when determining their continuance use is whether they have perceived usefulness and satisfaction.

Influence of Perceived Usefulness and Patient Satisfaction on Continuance Use

Complementing the findings of prior studies [65], perceived usefulness and patient satisfaction had significant effects on continuance use. In addition, the effects of exogenous variables on patient satisfaction are mediated by perceived usefulness.

Overall, these results suggest that increasing social support, informational quality, and service quality may improve the perceived usefulness and patient satisfaction of OHCs, which in turn fosters continuance use within OHCs.

Figure 3. Path analysis. The asterisks *** next to path coefficient values signify *P*<.001.

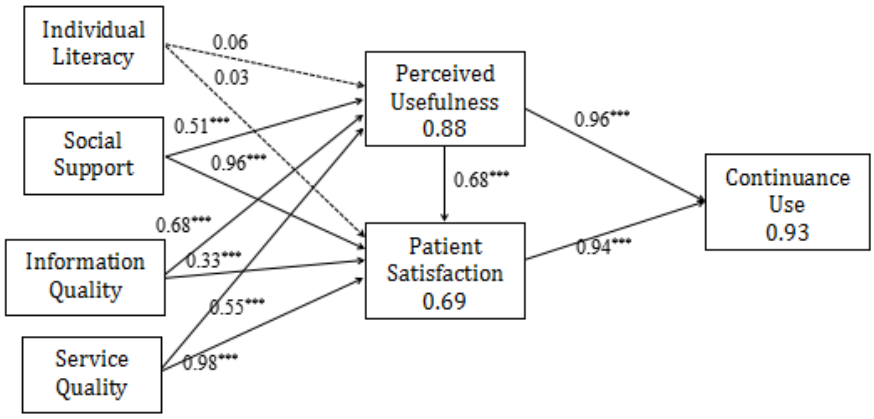


Table 6. Model path analysis. OHC: online health community.

Hypotheses (H)	Path coefficient	Support
H1: individual literacy has a positive effect on perceived usefulness	0.06	No
H2: individual literacy has a positive effect on patient satisfaction when using OHCs	0.03	No
H3: social support has a positive effect on perceived usefulness	0.5[a]	Yes
H4: social support has a positive effect on patient satisfaction when using OHCs	0.96[a]	Yes
H5: information quality has a positive effect on perceived usefulness	0.68[b]	Yes
H6: information quality has a positive effect on patient satisfaction when using OHCs	0.33[a]	Yes
H7: service quality has a positive effect on perceived usefulness	0.55[a]	Yes
H8: service quality has a positive effect on patient satisfaction when using OHCs	0.98[a]	Yes
H9: perceived usefulness has a positive effect on patient satisfaction when using OHCs	0.68[a]	Yes
H10: perceived usefulness has a positive effect on the continuance use of OHC	0.96[a]	Yes
H11: patient satisfaction has a positive effect on the continuance use of OHC	0.93[a]	Yes

[a]$P<.001$.

[b]$P<.01$.

Discussion

Principal Findings

The results show that the research framework for integrating the ISSM and OHC features contributes to our understanding of the roles certain constructs play in motivating patients to revisit OHCs. Social support, information quality, and service quality exert significant direct effects on perceived usefulness and patient satisfaction. Regarding the influences of perceived usefulness and patient satisfaction, both have significant effects on continuance use of OHC patients. In addition, the effects of social support, information quality, and service quality on patient satisfaction are mediated by perceived usefulness. Overall, these results suggest that increasing social support, informational quality, and service quality may improve the perceived usefulness and patient satisfaction of OHCs, which could in turn positively influence the continuance use of OHC patients. Unexpectedly, individual literacy has no influence on perceived usefulness and satisfaction of OHC patients.

Limitations and Future Research

Although the results of this study are promising, several limitations must be considered. First, the data analyzed in the study were collected from an OHC in China, in which communications are between patients and doctors. It is unclear whether the results of this study can be generalized to all OHCs, due to the limited number of Chinese OHCs. Therefore, future research should include additional OHCs. Second, this study used cross-sectional data to examine the critical factors that influence patient continuance use of OHCs, which may not fully capture OHC evolution. Future research may conduct longitudinal designs to better understand OHC success.

Despite these limitations, this study has several important implications for future OHC research and interventions. First, Web-tracking techniques can provide more detailed and accurate information on patient behavior patterns, and in turn contribute to improving our understanding of OHC communication mechanisms [66]. The model presented here provides a foundation for future researchers to build on and refine. Second, although we designed a novel metric for the constructs in our model using Web mining, future studies should try to develop more tailored metrics for OHCs. For example, the structural properties of OHCc (eg, size and membership structure) can also be important factors in hypothesis testing.

Comparison With Prior Work

First, our insights support a reconsideration of the impact of individual literacy on perceived usefulness and patient satisfaction. Because individual literacy does not contribute to perceived usefulness and patient satisfaction, practitioners should invest substantial effort in promoting continuous activity, as patients may seek health information on behalf of others. Because social support [37] contributes to the development of perceived usefulness and patient satisfaction, OHC practitioners should encourage doctors via suitable policies. Similarly, since OHC information quality and service quality strongly influence the development of perceived usefulness and patient satisfaction, it is important to accelerate the accumulation of information resources in terms of defining and archiving existing knowledge [67].

Second, our findings have implications for practitioners concerned with constructing sustainable OHCs that are active and supportive. By showing that perceived usefulness mediates the influence of social support, information quality, and service quality on patient satisfaction and, in turn, on continuance use of OHCs, we emphasize the importance of useful OHC information resources. Therefore, entities responsible for information-resource allocation should use the powerful role that online collectives can play [68].

Conclusions

The goal of this study was to examine the factors that drive patients to maintain ongoing relationships with OHCs in which communication occurs between doctors and patients. First, this study contributes to the literature by identifying which factors and how these factors influence patient continuance use. Moreover, the measures for the constructs in the proposed model have intuitive explanations and display good performance. Second, we developed an appropriate metric for constructs in the proposed model. In addition, a Web spider to collect data enabled us to acquire objective data frequently and with relative ease, thereby overcoming a major limitation of survey-based techniques. For example, if a questionnaire alone had been used, we could not control for all variables or claim causal effects regarding sentiment change [69]. Thus, the data collection technique and the SEM method provide new opportunities for measuring people's health behavior in the OHC context [70].

Conflicts of Interest

None declared.

References

1. Kanthawala S, Vermeesch A, Given B, Huh J. Answers to health questions: internet search results versus online health community responses. J Med Internet Res 2016 Apr 28;18(4):e95 [FREE Full text] [doi: 10.2196/jmir.5369] [Medline: 27125622]
2. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Comput Human Behav 2017 Apr;69:256-267. [doi: 10.1016/j.chb.2016.12.032]
3. Zhang Y. Understanding the sustained use of online health communities from a self-determination perspective. J Assn Inf Sci Tec 2015 Jul 07;67(12):2842-2857. [doi: 10.1002/asi.23560]
4. Jacobucci R, Grimm KJ, McArdle JJ. Regularized structural equation modeling. Struct Equ Modeling 2016 Apr;23(4):555-566 [FREE Full text] [doi: 10.1080/10705511.2016.1154793] [Medline: 27398019]
5. Van Oerle S, Mahr D, Lievens A. Coordinating online health communities for cognitive and affective value creation. J Serv Manag 2016 Aug 15;27(4):481-506. [doi: 10.1108/JOSM-09-2015-0264]
6. Hardyman W, Daunt KL, Kitchener M. Value co-creation through patient engagement in health care: a micro-level approach and research agenda. Public Manage Rev 2014 Feb 07;17(1):90-107. [doi: 10.1080/14719037.2014.881539]
7. Hibbard JH. Patient activation and the use of information to support informed health decisions. Patient Educ Couns 2017 Jan;100(1):5-7. [doi: 10.1016/j.pec.2016.07.006] [Medline: 27432014]
8. Petrič G, Atanasova S, Kamin T. Ill literates or illiterates? Investigating the eHealth literacy of users of online health communities. J Med Internet Res 2017 Oct 04;19(10):e331 [FREE Full text] [doi: 10.2196/jmir.7372] [Medline: 28978496]
9. Goh JM, Gao GD, Agarwal R. The creation of social value: can an online health community reduce rural-urban health disparities? MIS Q 2016 Jan 1;40(1):247-263. [doi: 10.25300/MISQ/2016/40.1.11]
10. Petrič G, Atanasova S, Kamin T. Impact of social processes in online health communities on patient empowerment in relationship with the physician: emergence of functional and dysfunctional empowerment. J Med Internet Res 2017 Mar 13;19(3):e74 [FREE Full text] [doi: 10.2196/jmir.7002] [Medline: 28288953]
11. Wang YS. Assessing e-commerce systems success: a respecification and validation of the DeLone and McLean model of IS success. Inf Syst J 2007;18(5):529-557 [FREE Full text] [doi: 10.1111/j.1365-2575.2007.00268.x]
12. DeLone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
13. Gorla N, Somers TM. The impact of IT outsourcing on information systems success. Inf Manag 2014 Apr;51(3):320-335. [doi: 10.1016/j.im.2013.12.002]
14. Hadji B, Degoulet P. Information system end-user satisfaction and continuance intention: a unified modeling approach. J Biomed Inform 2016 Jun;61:185-193 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.021] [Medline: 27033175]
15. Bossen C, Jensen LG, Udsen FW. Evaluation of a comprehensive EHR based on the DeLone and McLean model for IS success: approach, results, and success factors. Int J Med Inform 2013 Oct;82(10):940-953. [doi: 10.1016/j.ijmedinf.2013.05.010] [Medline: 23827768]
16. Bronstein J. An examination of social and informational support behavior codes on the Internet: the case of online health communities. Libr Inf Sci Res 2017 Jan;39(1):63-68. [doi: 10.1016/j.lisr.2017.01.006]
17. Myrick JG. The role of emotions and social cognitive variables in online health information seeking processes and effects. Comput Human Behav 2017 Mar;68:422-433. [doi: 10.1016/j.chb.2016.11.071]
18. Diviani N, van den Putte B, Meppelink CS, van Weert JC. Exploring the role of health literacy in the evaluation of online health information: insights from a mixed-methods study. Patient Educ Couns 2016 Jun;99(6):1017-1025. [doi: 10.1016/j.pec.2016.01.007]
19. Myneni S, Cobb N, Cohen T. In pursuit of theoretical ground in behavior change support systems: analysis of peer-to-peer communication in a health-related online community. J Med Internet Res 2016 Feb 02;18(2):e28. [doi: 10.2196/jmir.4671]

20. Peter J, Leichner N, Mayer AK, Krampen G. Making information literacy instruction more efficient by providing individual feedback. Stud High Educ 2015 Sep 17;42(6):1110-1125. [doi: 10.1080/03075079.2015.1079607]
21. Albers P, Pace CL, Odo DM. From affinity and beyond: a study of online literacy conversations and communities. J Lit Res 2016 Jul 28;48(2):221-250. [doi: 10.1177/1086296X16659069]
22. Wu B, Jiang S, Chen HC. Effects of individuals' motivations on communications in online health forums. Soc Behav Personal 2016 Mar 23;44(2):299-312. [doi: 10.2224/sbp.2016.44.2.299]
23. Nupairoj N. El ecosistema de la alfabetización mediática: un enfoque integral y sistemático para divulgar la educomunicación [The ecosystem of media literacy: a holistic approach to media education]. Comunicar 2016 Oct 01;24(49):29-37. [doi: 10.3916/C49-2016-03]
24. Abirami AM, Askarunisa A. Sentiment analysis model to emphasize the impact of online reviews in healthcare industry. Online Inf Rev 2017 Aug 14;41(4):471-486. [doi: 10.1108/OIR-08-2015-0289]
25. Zimbra D, Chen H, Lusch RF. Stakeholder analyses of firm-related web forums. ACM Trans Manage Inf Syst 2015 Apr 03;6(1):1-38 [FREE Full text] [doi: 10.1145/2675693]
26. 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]
27. Suri VR, Majid S, Chang YK, Foo S. Assessing the influence of health literacy on health information behaviors: a multi-domain skills-based approach. Patient Educ Couns 2016 Jun;99(6):1038-1045. [doi: 10.1016/j.pec.2015.12.017]
28. Bright P, Hambly K, Tamakloe S. What is the profile of individuals joining the KNEEguru online health community? A cross-sectional mixed-methods study. J Med Internet Res 2016 Apr 18;18(4):e84. [doi: 10.2196/jmir.5374]
29. Loane SS, Webster CM, D'Alessandro S. Identifying consumer value co-created through social support within online health communities. J Macromarketing 2014 May 29;35(3):353-367. [doi: 10.1177/0276146714538055]
30. Choi MJ, Kim SH, Lee SL, Kwon BC, Yi JS, Choo J, et al. Toward predicting social support needs in online health social networks. J Med Internet Res 2017 Aug 02;19(8):e272. [doi: 10.2196/jmir.7660]
31. Rains SA, Peterson EB, Wright KB. Communicating social support in computer-mediated contexts: a meta-analytic review of content analyses examining support messages shared online among individuals coping with illness. Commun Monogr 2015 Mar 17;82(4):403-430. [doi: 10.1080/03637751.2015.1019530]
32. Cavallo DN, Brown JD, Tate DF, DeVellis RF, Zimmer C, Ammerman AS. The role of companionship, esteem, and informational support in explaining physical activity among young women in an online social network intervention. J Behav Med 2013 Oct 1;37(5):955-966. [doi: 10.1007/s10865-013-9534-5]
33. McKinley CJ, Wright PJ. Informational social support and online health information seeking: examining the association between factors contributing to healthy eating behavior. Comput in Human Behav 2014 Aug;37:107-116. [doi: 10.1016/j.chb.2014.04.023]
34. Pan WJ, Shen CH, Feng B. You get what you give: understanding reply reciprocity and social capital in online health support forums. J Health Commun 2016 Dec 27;22(1):45-52. [doi: 10.1080/10810730.2016.1250845]
35. Yagil D, Shnapper-Cohen M. When authenticity matters most: physicians' regulation of emotional display and patient satisfaction. Patient Educ Couns 2016 Oct;99(10):1694-1698 [FREE Full text] [doi: 10.1016/j.pec.2016.04.003] [Medline: 27085519]
36. Petrovčič A, Petrič G. Differences in intrapersonal and interactional empowerment between lurkers and posters in health-related online support communities. Comput Human Behav 2014 May;34:39-48. [doi: 10.1016/j.chb.2014.01.008]
37. Ruppel EK, McKinley CJ. Social support and social anxiety in use and perceptions of online mental health resources: exploring social compensation and enhancement. Cyberpsychol Behav Soc Netw 2015 Aug;18(8):462-467. [doi: 10.1089/cyber.2014.0652] [Medline: 26252931]
38. Lambert K, Mullan J, Mansfield K, Koukomous A, Mesiti L. Evaluation of the quality and health literacy demand of online renal diet information. J Hum Nutr Diet 2017 Oct;30(5):634-645. [doi: 10.1111/jhn.12466] [Medline: 28211108]
39. Fehrenbacher DD. Perceptions of information quality dimensions from the perspective of commodity theory. Behav Inf Technol 2016 Feb 15;35(4):254-267. [doi: 10.1080/0144929x.2015.1128974]
40. Adams SA. Revisiting the online health information reliability debate in the wake of "web 2.0": an inter-disciplinary literature and website review. Int J Med Inform 2010 Jun;79(6):391-400. [doi: 10.1016/j.ijmedinf.2010.01.006] [Medline: 20188623]
41. Storino A, Castillo-Angeles M, Watkins AA, Vargas C, Mancias JD, Bullock A, et al. Assessing the accuracy and readability of online health information for patients with pancreatic cancer. JAMA Surg 2016 Sep 01;151(9):831-837. [doi: 10.1001/jamasurg.2016.0730] [Medline: 27144966]
42. Ghasemaghaei M, Hassanein K. Online information quality and consumer satisfaction: the moderating roles of contextual factors - a meta-analysis. Inf Manage 2015 Dec;52(8):965-981. [doi: 10.1016/j.im.2015.07.001]
43. Greene J, Hibbard JH, Sacks RM. Summarized costs, placement of quality stars, and other online displays can help consumers select high-value health plans. Health Aff (Millwood) 2016 Apr;35(4):671-679. [doi: 10.1377/hlthaff.2015.1367] [Medline: 27044968]

44. Lu N, Wu H. Exploring the impact of word-of-mouth about physicians' service quality on patient choice based on online health communities. BMC Med Inform Decis Mak 2016 Nov 26;16(1):151 [FREE Full text] [doi: 10.1186/s12911-016-0386-0] [Medline: 27888834]
45. Nambisan P, Gustafson DH, Hawkins R, Pingree S. Social support and responsiveness in online patient communities: impact on service quality perceptions. Health Expect 2016 Feb;19(1):87-97 [FREE Full text] [doi: 10.1111/hex.12332] [Medline: 25556856]
46. Nguyen BV, Burstein F, Fisher J. Improving service of online health information provision: a case of usage-driven design for health information portals. Inf Syst Front 2014 May 27;17(3):493-511. [doi: 10.1007/s10796-014-9507-4]
47. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 01;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
48. Lee K, Jung SY, Hwang H, Yoo S, Baek HY, Baek RM, et al. A novel concept for integrating and delivering health information using a comprehensive digital dashboard: an analysis of healthcare professionals' intention to adopt a new system and the trend of its real usage. Int J Med Inform 2017;97:98-108 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.10.0011]
49. Masserini L, Liberati C, Mariani P. Quality service in banking: a longitudinal approach. Qual Quant 2016 Oct 4;51(2):509-523. [doi: 10.1007/s11135-016-0420-4]
50. Sharma G, Lijuan W. The effects of online service quality of e-commerce websites on user satisfaction. Electron Libr 2015 Jun;33(3):468-485. [doi: 10.1108/el-10-2013-0193]
51. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989;35(8):982-1003 [FREE Full text] [doi: 10.1287/mnsc.35.8.982]
52. James TL, Villacis Calderon ED, Cook DF. Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. Expert Syst Appl 2017 Apr;71(C):479-492. [doi: 10.1016/j.eswa.2016.11.004]
53. Felbermayr A, Nanopoulos A. The role of emotions for the perceived usefulness in online customer reviews. J Interact Mark 2016 Nov;36:60-76. [doi: 10.1016/j.intmar.2016.05.004]
54. Sun Y, Jeyaraj A. Information technology adoption and continuance: a longitudinal study of individual's behavioral intentions. Inf Manage 2013 Nov;50(7):457-465. [doi: 10.1016/j.im.2013.07.005]
55. Ghobakhloo M, Tang SH. Information system success among manufacturing SMEs: case of developing countries. Inf Technol Dev 2015 Jan 19;21(4):573-600. [doi: 10.1080/02681102.2014.996201]
56. Dalip DH, Gonçalves MA, Cristo M, Calado P. A general multiview framework for assessing the quality of collaboratively created content on web 2.0. J Assoc Inf Sci Technol 2017;68(2):286-308.
57. Mou J, Cohen JF. Trust and online consumer health service success: a longitudinal study. Inf Dev 2016;33(2):169-189. [doi: 10.1177/0266666916642507]
58. Snead KC, Magal SR, Christensen LF, Ndede-Amadi AA. Attribution theory: a theoretical framework for understanding information systems success. Syst Pract Action Res 2014 Jul 24;28(3):273-288. [doi: 10.1007/s11213-014-9328-x]
59. Tilly R, Posegga O, Fischbach K, Schoder D. Towards a conceptualization of data and information quality in social information systems. Bus Inf Syst Eng 2017;59(1):3-21. [doi: 10.1007/s12599-016-0459-8]
60. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39-50. [doi: 10.2307/3151312]
61. Nagase M, Kano Y. Identifiability of nonrecursive structural equation models. Stat Probab Lett 2017 Mar;122:109-117. [doi: 10.1016/j.spl.2016.11.010]
62. Cangur S, Ercan I. Comparison of model fit indices used in structural equation modeling under multivariate normality. J Mod App Stat Meth 2015 May 01;14(1):152-167. [doi: 10.22237/jmasm/1430453580]
63. Richter NF, Cepeda G, Roldán JL, Ringle CM. European management research using partial least squares structural equation modeling (PLS-SEM). Eur Manag J 2016 Dec;34(6):589-597. [doi: 10.1016/j.emj.2016.08.001]
64. Kipenis L, Askounis D. Assessing e-participation via user's satisfaction measurement: the case of OurSpace platform. Ann Oper Res 2015 Sep 14;247(2):599-615. [doi: 10.1007/s10479-015-1911-8]
65. Hess TJ, McNab AL, Basoglu KA. Reliability generalization of perceived ease of use, perceived usefulness, and behavioral intentions. MIS Q 2014 Jan 1;38(1):1-28. [doi: 10.25300/MISQ/2014/38.1.01]
66. Michail T. Designing a successful collaborative Wiki: the choice between outcome quality and online community needs. Int J Technol Hum Interact 2017;13(2):22-39 [FREE Full text] [doi: 10.4018/IJTHI.2017040102]
67. Jacobson D, Pehlivan E, Vilvovsky S, Wong W. Combining web mining techniques and structural equations modeling for measuring e-commerce perceptions. Case Studies Bus Ind Gov Stat 2009;2(2):99-108 [FREE Full text]
68. Atanasova S, Kamin T, Petrič G. Exploring the benefits and challenges of health professionals' participation in online health communities: emergence of (dis)empowerment processes and outcomes. Int J Med Inform 2017 Feb;98:13-21. [doi: 10.1016/j.ijmedinf.2016.11.005] [Medline: 28034408]
69. Paige SR, Krieger JL, Stellefson ML. The influence of eHealth literacy on perceived trust in online health communication channels and sources. J Health Commun 2017 Jan;22(1):53-65. [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]

70. Zhao K, Yen J, Greer G, Qiu B, Mitra P, Portier K. Finding influential users of online health communities: a new metric based on sentiment influence. J Am Med Inform Assoc 2014 Oct;21(e2):e212-e218 [FREE Full text] [doi: 10.1136/amiajnl-2013-002282] [Medline: 24449805]

Abbreviations

AVE: average variance extracted
CR: composite reliability
GFI: goodness-of-fit index
IS: information system
ISSM: information system success model
NTUSD: National Taiwan University Semantic Dictionary
OHC: online health community
PLS: partial least squares
SEM: structural equation modeling

Edited by G Eysenbach; submitted 07.10.17; peer-reviewed by L Zhao, Z Liu, L Huan, K Zhao, N Bragazzi, H Yu, S Aggarwal; comments to author 14.11.17; revised version received 15.12.17; accepted 01.02.18; published 16.04.18.

Please cite as:
Wu B
Patient Continued Use of Online Health Care Communities: Web Mining of Patient-Doctor Communication
J Med Internet Res 2018;20(4):e126
URL: http://www.jmir.org/2018/4/e126/
doi:10.2196/jmir.9127
PMID:29661747

Original Paper

Multidisciplinary Collaboration in the Treatment of Patients With Type 2 Diabetes in Primary Care: Analysis Using Process Mining

Tania Conca[1], MSc; Cecilia Saint-Pierre[1], MSc; Valeria Herskovic[1], PhD; Marcos Sepúlveda[1], PhD; Daniel Capurro[2], MD, PhD; Florencia Prieto[3], MD; Carlos Fernandez-Llatas[4], PhD

[1]Computer Science Department, School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile
[2]Department of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile
[3]Department of Family Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile
[4]Institute of Information and Communication Technologies, Universitat Politècnica de València, Valencia, Spain

Corresponding Author:
Marcos Sepúlveda, PhD
Computer Science Department
School of Engineering
Pontificia Universidad Católica de Chile
Av Vicuña Mackenna 4860
Macul
Santiago,
Chile
Phone: 56 2 2534 4447
Email: marcos@ing.puc.cl

Abstract

Background: Public health in several countries is characterized by a shortage of professionals and a lack of economic resources. Monitoring and redesigning processes can foster the success of health care institutions, enabling them to provide a quality service while simultaneously reducing costs. Process mining, a discipline that extracts knowledge from information system data to analyze operational processes, affords an opportunity to understand health care processes.

Objective: Health care processes are highly flexible and multidisciplinary, and health care professionals are able to coordinate in a variety of different ways to treat a diagnosis. The aim of this work was to understand whether the ways in which professionals coordinate their work affect the clinical outcome of patients.

Methods: This paper proposes a method based on the use of process mining to identify patterns of collaboration between physician, nurse, and dietitian in the treatment of patients with type 2 diabetes mellitus and to compare these patterns with the clinical evolution of the patients within the context of primary care. Clustering is used as part of the preprocessing of data to manage the variability, and then process mining is used to identify patterns that may arise.

Results: The method is applied in three primary health care centers in Santiago, Chile. A total of seven collaboration patterns were identified, which differed primarily in terms of the number of disciplines present, the participation intensity of each discipline, and the referrals between disciplines. The pattern in which the three disciplines participated in the most equitable and comprehensive manner had a lower proportion of highly decompensated patients compared with those patterns in which the three disciplines participated in an unbalanced manner.

Conclusions: By discovering which collaboration patterns lead to improved outcomes, health care centers can promote the most successful patterns among their professionals so as to improve the treatment of patients. Process mining techniques are useful for discovering those collaborations patterns in flexible and unstructured health care processes.

(J Med Internet Res 2018;20(4):e127) doi:10.2196/jmir.8884

KEYWORDS

process assessment (health care); interprofessional relations; primary health care; type 2 diabetes mellitus; data mining

Introduction

Background

Type 2 diabetes mellitus (T2DM) is a chronic disease that can cause complications and serious health-related consequences [1]. This disease affects around 9.1% of the global adult population and is expected to reach 10% by 2040 [2]. In Latin America, the prevalence of T2DM is 9.4%, whereas in Chile it is 11.4% [2-5]. Consequently, T2DM constitutes a significant public health problem at the global level, affecting medium- and low-income countries to a greater extent [6]. As a consequence, considerable resources are allocated to its treatment, particularly in Chile [7].

When T2DM is not controlled, a series of complications can arise that impact the quality of life of the patient and which can increase their risk of mortality, for example, cardiovascular complications, diabetic retinopathy, peripheral neuropathy, and kidney failure [8,9]. It is possible to prevent these complications with good metabolic control [10] through the use of medication and an appropriate diet [1,11]. In the Chilean primary health care system, only 36% of the population with T2DM is classified as well-controlled or stable [4].

Multidisciplinary Collaboration in Type 2 Diabetes Mellitus Care

Patients diagnosed with T2DM have multiple biomedical, psychological, and social needs that must be met in a coordinated manner by professionals from several disciplines. In this sense, multidisciplinary collaboration, which refers to joint work conducted by professionals from different disciplines who interact around a particular patient, has become critical for enhancing clinical outcomes [4,6,12-14]. Although clinical protocols establish treatment guidelines, health care processes for each patient vary and often deviate from standard indications [15]. The organization and composition of a treatment team can be an influential factor in patient evolution, as well as in the coordination between the different disciplines. Evidence suggests that the most successful interventions in relation to chronic diseases include certain key functions being carried out by nonphysicians, and interventions by well-integrated teams have been linked to greater patient satisfaction [16]. In the particular case of Chile, there is a significant shortage of physicians in primary health care [17]. Understanding how to organize multidisciplinary collaboration can facilitate the design of more efficient and effective treatment protocols.

Information systems record data related to services provided to patients. Process mining is a relatively new research discipline that has been used in health care to extract knowledge from information systems to analyze process design [18,19]. The algorithms developed in this discipline create graphical representations of models from the real execution of processes, which can be easily understood by individuals from a wide range of disciplines [20]. These models frequently demonstrate that reality differs to the perceptions, opinions, and beliefs held by parties directly involved in health care processes [21]. Process mining also facilitates the analysis of processes from an organizational perspective, which may help improve the understanding of how collaboration occurs within treatment teams [22].

Research Goals

This study seeks to verify whether it is possible to determine certain patterns of collaboration using data from electronic clinical records (ECR) and to study if these patterns are related to the clinical outcomes of patients. Accordingly, we propose a methodology based on the application of process mining tools to analyze collaboration between health care professionals (HCPs) to (1) Identify collaboration patterns in the treatment of patients with T2DM in primary care, that is, the distinct interaction networks within the treatment teams and (2) evaluate the performance of the discovered patterns, confirming whether they relate to the clinical evolution of patients (represented by glycated hemoglobin, HbA_{1c} measurements). This approach uses information that is already being recorded in the relevant health care institutions and therefore, does not require the collection of new data.

Related Work

Process Mining Applied to Health Care Processes

The inherent variability of health care processes has been addressed in a number of different ways. Some traditional control-flow discovery algorithms help to understand the different pathways that can be executed on a model and to distinguish the most common behaviors by managing the thresholds that indicate the frequency of activity sequences. The heuristic miner and fuzzy miner algorithms have been used to identify and study the main flow of the model based on data from the information systems of a hospital in Seoul, South Korea [23]. However, this approach generates a single model and, to discover different behaviors, it is necessary to test distinct thresholds, which can result in the analysis of unstructured processes becoming particularly complex. Another approach for creating simpler models for unstructured processes involves grouping several low-level activities with the same name at a higher level [24]. The proposed procedure is useful when the event log consists of large amounts of different activities and the traces differ not only in the sequence of activities but also in terms of the presence or absence thereof.

A different perspective is to create groups of patients according to certain preselected characteristics and to subsequently generate models for each group to capture the variability of the associated health care processes [25]. Once these groups of patients have been established, it is possible to generate models to represent the clinical flow followed by patients, including their progression across departments, specialists, and types of medical appointments, or over the natural course of an illness. This method has been applied to patients with T2DM using variables associated with related complications, including those concerning HbA_{1c}, blood pressure, and cholesterol [26]. The results were used to analyze the circulation of the different groups and the probabilities of passing from one state to another.

Similar to the previous approach, other studies have successfully generated several simple models to represent highly flexible processes by applying different clustering techniques before the

execution of discovery algorithms. The purpose of this additional step is to ensure that the logs with highly variable records become more manageable by grouping cases according to behavior similarity. Sequential clustering has been used during log preprocessing to identify regular behaviors, process variants, and exceptional cases [27]. While sequential clustering groups traces according to sequences of similar activities [28], trace clustering provides a set of grouping techniques based on distance that seeks to differentiate traces according to certain characteristics such as the frequency of activity occurrence, the number of events, or the number of events executed by each resource in a trace [29].

Collaboration in Medical Teams

Several studies have addressed the issue of collaboration in the treatment of patients with T2DM. One qualitative study from the patient perspective found that patients with T2DM and asthma consider that, to obtain the best outcome, it is necessary to receive treatment from a multidisciplinary team, despite them stating that they did not require such a team for their own treatment [30]. Another study found patients were satisfied with collaborative treatment [31]. No relationship between collaboration and clinical outcome was found in either of these studies.

Conversely, quantitative studies have given rise to evaluations of collaboration by considering the evolution of HbA_{1c}, hospitalization costs, and readmission rates as clinical outcomes. One study found that the hospitalization costs and readmission rates decreased as the health care team became more integrated in terms of collaboration between physicians [32]. T2DM patients receiving treatment from multidisciplinary teams achieved better outcomes than those that did not, demonstrating improvements in HbA_{1c}, *low-density lipoprotein* cholesterol, and an increased use of statins, as well as progress in statin and antiplatelet therapy [33]. Furthermore, differences have been identified in outcomes related to HbA_{1c}, blood pressure, and cholesterol according to a report compiled by treating professionals [34]. A controlled study into the 2-year treatment of geriatric patients with T2DM by a multidisciplinary team, in comparison with a control group that received no collaborative treatment, found differences in the outcomes during the second year of collaborative treatment [35].

The organization and composition of the treating team can be an influential factor in patient evolution, as well as in terms of the coordination between the different clinical disciplines present within the team. For example, in one study, referrals to T2DM educators and dietitians were minimal, even among overweight and obese patients [12].

Process Mining, Social Network Analysis, and Collaboration

In the health context, some studies have used social network analysis to deepen understanding of the organization of professionals in patient care. It should be noted that although the methodology presented in this paper differs from the perspectives outlined in the previous section, the latter should not be discounted. The techniques can complement one another.

One study used visual graphics to demonstrate the structure of the referral networks and appointments that link physicians, defining four physician subgroups with similar referral, appointment, discussion, and attention coverage patterns [36]. The role of each physician was classified as the emitter, transmitter, or receiver according to the proportion of interactions that he or she initiated in relation to those initiated by others. A similar analysis was undertaken to gauge the structure of a team of nurses [37]. A further approach was based on the analysis of egocentric networks [38] by studying the network formed around a central actor and the actors with whom he or she interacts [39]. In this instance, the relationships between the secondary actors were not specified. Networks were used to analyze referrals and counter referrals among nurses and other disciplines to understand how they collaborate with other professionals and their contribution to multidisciplinary care in a primary health care setting. Accordingly, one important factor considered in this paper is referral flow, that is, whether this flow is unidirectional or bidirectional.

Study Setting and Context

To better understand the context of this study and frame the implications of our results, next, we briefly describe the primary health care system in Chile.

The Chilean health care system is composed by both public and private insurers. The law mandates that each employed person must pay at least 7% of their income to a health insurer. Approximately 74.4% of the population is insured by the public system, and people who earn less than minimum wage, as well as children, students, and unemployed individuals, have free health care in the public provider network [40].

Primary health care centers, called Centro de Salud Familiar (CESFAM), are the first point of contact of users with the public health care network. These centers use a family medicine model in which patients are grouped into zones that treat at most 10,000 patients with a multidisciplinary team of health care workers. CESFAM treat acute morbidities that may be solved or referred to a more complex center and chronic morbidities that require periodic assessment, for example, diabetes, hypertension, and chronic pulmonary disease. The Chilean Ministry of Health establishes a treatment protocol for each of these conditions, published as a clinical guideline (eg, [4] for T2DM). The main problem faced by the CESFAM is a lack of resources, namely, not having enough HCPs. In Chile, the average number of physicians per 1000 inhabitants is 1.9, whereas for member states of the Organization for Economic Co-operation and Development it is 3.3 [17].

The Chilean Ministry of Health established in 2005 a program that prioritizes a set of 80 health conditions, guaranteeing timely and free access to treatment [40]. T2DM is one of such conditions. This, along with the Ministry guidelines for T2DM treatment, means that for beneficiaries of the public health care system, treatment is homogeneous: all patients have access to the same protocols and medications.

This paper is based on historical information from patients diagnosed with T2DM, collected in three CESFAM. These three

centers are located in low income, high social vulnerability districts of Santiago, Chile. Overall, they have treated an average of 8000 people per year over the last 5 years, and approximately 30% of patients treated on an annual basis have T2DM. All three centers use the same flowchart, based on the clinical guidelines [4], to determine treatment for T2DM patients. The data collected correspond to the period from 2012 to 2016.

Methods

Data Source

After obtaining institutional review board approval, the dataset was extracted from the information system used in the three health care centers. Its database stores information related to patients, including their appointments, diagnoses, and test results. For each visit to the health care center, the system records the date, type of appointment, and the professional in charge of the episode. Importantly, this work only considers patients with T2DM and activities associated with periodic cardiovascular appointments (Cardiovascular Periodic Appointment, CVPA) that are performed by specialists from the professional triad team consisting of physician, nurse, and dietitian. Every time one of these professionals completes a CVPA, they must specify the discipline and approximate date of the patient's next appointment.

The percentage of HbA_{1c} was selected as a metric to represent patient evolution. The HbA_{1c} test measures the glycemic history of the patient over the preceding 120 days [41] and is one of the tests used to monitor diabetic patients. The frequency of the test depends on the state of compensation of the patient, the treatment used, and medical judgment. Although the specific treatment objectives should be individualized for each patient, the American Diabetes Association recommends that the goal of therapy should be to reduce HbA_{1c} below 7%. For values higher than this, the clinical guidelines of the Chilean Ministry of Health clinical guidelines and the internal guidelines of the health care centers included in this study establish two categories of decompensation for patients: moderately decompensated, for values between 7% and 9% (included) and highly decompensated, for values higher than 9% [4]. The date on which a patient undergoes a test and its result are both logged in the records.

Patient Selection

A total of 3369 patients with T2DM were identified across the three health care centers. Subsequently, to measure their respective evolution, we included individuals who had at least two recorded HbA_{1c} test results. In total, 2843 patients met these conditions.

To isolate external factors that might influence a patient's evolution beyond the clinical team's collaboration patterns, we included diabetic patients with no comorbidities or diabetes-related complications and good adherence to prescribed appointments and tests. We used the diabetes complication severity index and the chronic illness with complexity index count as measures of comorbidities and complications [1,42,43] and an interval under 4 months between the prescribed appointment and the actual appointment as a reasonable proxy for adherence.

Adherence to follow-up appointments is important for the evolution of patients, as through them professionals can intervene in the habits and self-care of the patient [44]. Greater rates of missed appointments are associated with significantly higher HbA_{1c} measurements [45]. Moreover, if patients do not show adherence to their treatment, the effectiveness of treatment is compromised, and they might develop complications [46]. In general, the diabetic population presents a low adherence both to medications and timely attendance to scheduled appointments [47]. To isolate the influence of the adherence to appointments, those patients who do not adhere to the HbA_{1c} tests were excluded. Nevertheless, a margin of time must be considered to determine that the patient attended the appointment on time [48]. The protocol of the health care centers stipulates that patients with higher states of decompensation should undergo more regular HbA_{1c} tests. In the context of the period under analysis, the following was considered acceptable by the health care centers studied: that patients in a state of compensation took the HbA_{1c} test up to 1 year after their last measurement, that patients who were moderately decompensated did the same after up to 6 months, and that highly decompensated patients did so after up to 3 months.

Given the context of the health care centers studied, in particular, their scarce resource availability, their restrictions for taking appointments (in general, patients cannot schedule appointments more than 1 month in advance), and the availability of hours for taking exams, it is normal that there is a delay that goes beyond the responsibility of the patient. To address these restrictions that depend on the health care center, a tolerance of up to 4 months for taking the HbA_{1c} test was considered. This time frame was discussed with and suggested by the HCPs.

Finally, patients who were tested for HbA_{1c} at intervals greater than those established for the clinical protocols according to their degree of compensation, considering a 4-month tolerance, were not considered in the analysis. This restriction ensures more complete information and greater consistency in terms of data evolution because the longer the time elapsed between tests, the more difficult it becomes to determine the variability in terms of patient compensation during that period. Of the 579 patients with neither severe conditions nor comorbidities, 319 had acceptable levels of adherence for inclusion in this study.

To normalize the period of study for all included cases, this paper considered a horizon of one and a half years to analyze the impact of multidisciplinary collaboration on the treatment of patients. The first measurement of HbA_{1c} that is available for a patient marks point zero of the period of study. To determine the end of the period, a subsequent HbA_{1c} measurement was sought as close as possible to 18 months after point zero. A tolerance period of 8 months was considered before and following the year-and-a-half mark, that is, the final measurement included had to fall within a range running from month 10 to month 26 (18±8), factoring in the possibility that other previous measurements may have been taken during this period. Of the 319 patients, 231 had a minimum acceptable

study period of 10 months. As the study analyzed the response of the patient to the intervention and organization of the triad of professionals within a defined time frame, detailed information related to the complete evolution during the lifetime of the patient was not required. Figure 1 shows the patient selection process.

Regarding the sample, 133 out of 231 patients were men (57.6%), and 98 out of 231 were women. Overall, 50% (115/231) were aged 60 years or above, and 81% (187/231) were aged 50 years or above (average age 59.7; range: 20-89). The mean amount of CVPAs was 4.8 per patient (range: 1-15). Table 1 outlines this information.

Figure 2 outlines the duration of the periods of study considered. The X-axis shows the number of months included in the analysis, whereas the Y-axis shows the number of patients related to the corresponding horizon.

Collaboration Through Network Analysis

Clinical records were used to build a log, with the following information for each CVPA: the ID of the patient, a time stamp, and the relevant attending discipline (physician, nurse, or dietitian). The time stamps were used to identify the sequence or order in which the disciplines intervened.

Definition 1 (discipline log). Let V represent the set of the three disciplines: physician (P), nurse (N), and dietitian (D), and H={ h_1,..., h_n} the set of patients. Let c_{hi} be the sequence of disciplines in V who attend to patient h_i. We define L, the discipline log, as L={ c_{hi}}∀ h_i∈ H.

With information from the discipline log, a collaborative network can be created to show the relationship between the clinical disciplines in the treatment of patients. Specifically, a collaborative network related to a group of patients is defined as a directed graph in which the nodes refer to different clinical disciplines that intervene in the treatment of the disease, whereas the arcs represent the existing derivations among the disciplines in the case of each patient. It should be noted that following a CVPA of a patient, the professional may assign the subsequent appointment to either a professional from a distinct discipline or from the same discipline. Therefore, the graphs generated could include self-loops.

Definition 2 (collaborative network). Let V represent the set of the three disciplines: physician (P), nurse (N), and dietitian (D), and H={ h_1,…, h_n} the set of patients. A collaborative network will be the graph G^W=(V, E), where V={P, D, N} is the set of distinct disciplines that constitute part of the treatment of the patients in H and E ⊆{V × V} are the arcs that represent all the different derivations that occur when considering all the patients of set H.

Figure 1. Criteria applied during patient selection process. T2DM: type 2 diabetes mellitus; HbA_{1c}: glycated hemoglobin.

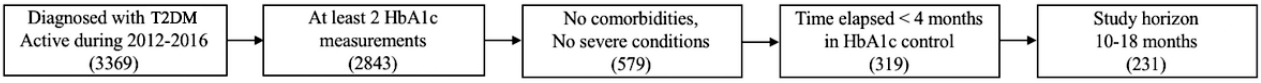


Table 1. Description of the studied population.

Variable	Average (SD)
Age	59.7 (12.6)
Years with type 2 diabetes mellitus	4.6 (3.8)
Number of glycated hemoglobin (HbA_{1c}) measurements	3.7 (0.95)
Number of cardiovascular periodic appointments	4.8 (2.3)

Figure 2. Periods of study used in the analysis.

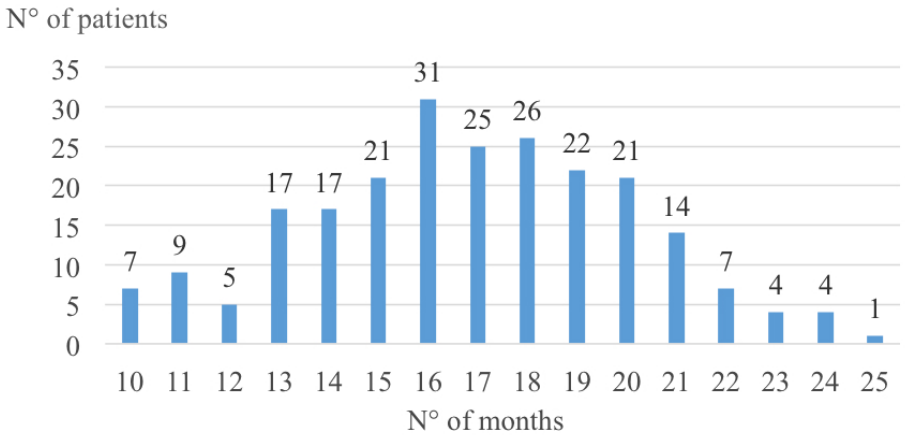


Figure 3 shows an example of a collaborative network that represents the multidisciplinary structure of treatment received by a group of patients. In this case, the three disciplines make referrals among themselves, and the dietitian (D) makes self-referrals for certain CVPAs.

Three metrics were defined for both the nodes and the arcs of the collaborative network:

Participation index: the proportion of CVPAs performed by a specific discipline in relation to the total number of CVPAs. It is calculated for each discipline (node). It ranges from 0%, which represents no participation of the discipline in the treatment, to 100%, whereby all appointments were undertaken by the particular discipline. This value can be interpreted as the prominence of a clinical discipline with regard to the patient intervention.

Self-referral index: the proportion of CVPAs that are referrals to the same discipline in relation to the total number of CVPAs that a particular discipline refers in total. It is calculated for each discipline (self-loops). It ranges from 0%, which represents no self-referrals made by the discipline, to 100%, whereby all referrals made by the professionals of one discipline are to the same discipline.

Referral index: the proportion of CVPAs that one discipline refers to a different discipline in relation to the total number of CVPA referrals made by that particular discipline. It is calculated for each pair of disciplines (arcs between different nodes). It ranges from 0%, which represents no referrals to the other discipline, to 100%, whereby all referrals made by the professionals of one discipline are to the other discipline. This value can be interpreted as the level of support among different disciplines.

Pattern Identification

The process mining algorithm selected for discovery was PALIA, implemented by the Institute of Information and Communication Technologies (ITACA) of the Universitat Politècnica de València, Valencia, Spain [49], which was applied using the PALIA Web [20] application. PALIA Web is a process discovery application created for the analysis of flexible and unstructured workflows. This tool was chosen because it receives an event log as input and outputs visualizations that are easy to understand for people who are not experts in process mining. In addition, it has filters that can be applied to the data before performing discovery, including, for example, trace clustering for the creation of different models based on groups of patients showing similar behavior.

The first step to identify the distinct forms of treating patients was to apply the PALIA process mining algorithm to the discipline log, complemented by trace clustering. This included the use of the flow disintegration functionality, which groups similar traces (sequences of disciplines that attend to each patient), and the application of the PALIA algorithm to create a visualization of the different groups or trace clusters.

The PALIA algorithm was executed with the following parameters for the flow disintegrations: similarity of 15% and outliers of 3%. The similarity percentage indicates that by conducting trace clustering, individuals from the same group are unable to differentiate by more than 15% according to the measurement of dissimilarity used by the algorithm, which is based on a heuristic topological editing distance [50]. Therefore, individuals from distinct groups differed by more than 15%. Conversely, the percentage of outliers indicates the minimum proportion of individuals that can be grouped under a single cluster. If the algorithm identifies a smaller group than the one established under that parameter, those patients are grouped together with the outliers. In this case, the 3% identified is equivalent to 7 patients. In addition, a heat map was applied to the diagrams (arcs and nodes) with a scale of red to green, where red indicates high frequency and green low frequency.

Clinical Outcome

The test used to monitor and control the state of the patient was HbA_{1c}. The clinical guidelines of the Chilean Ministry of Health, used by the health care centers, propose three categories for HbA_{1c} values: below 7% (called *compensated*, the clinical goal for patients), between 7% and 9%, and above 9%, each corresponding to a different course of action and time frame for follow-up [4]. As in our data we had several measurements for each patient, we propose a segmentation of patients considering their temporal trend [51]. We tried several different segmentations with the help of HCPs until we achieved a four-segment categorization that is described below. The HCPs stated that the most interesting category of patients for them were patients with high HbA_{1c} who, within the time frame, managed to reach the clinical goal (below 7%).

Figure 3. Collaborative network model. P: physician; N: nurse; D: dietitian.

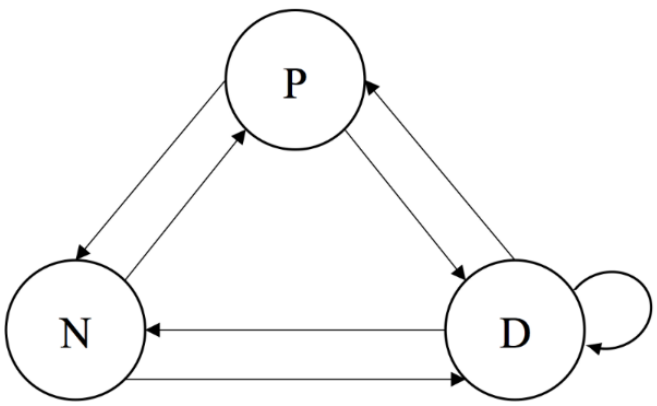
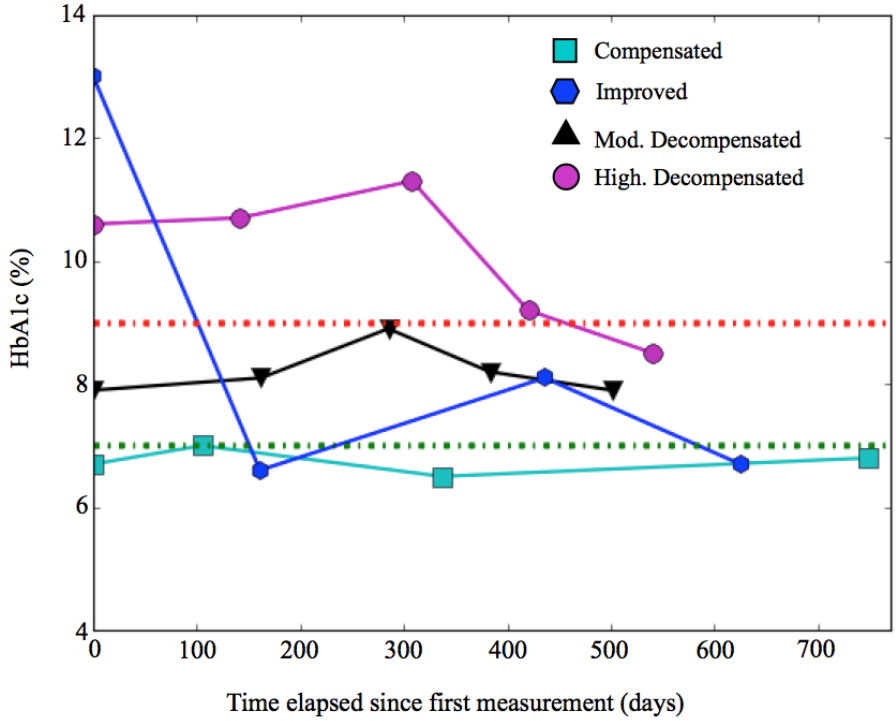


Figure 4. Example of the graphic representation of the clinical evolution of a patient of each segment. HbA_{1c}: glycated hemoglobin.



Patients were separated into four segments according to the evolution of their HbA_{1c} results, as follows:

1. *Compensated*: patients with all measurements under 7%, or at the most one measurement between 7% and 9%, inclusive, but with an average in terms of all measurements under 7%, that is, it was accepted that these patients had exceeded the compensation limit once, but their average remained at a compensated level.
2. *Improved:* patients with a negative HbA_{1c} slope, and whereby their final measurement of the period was less than 7%, that is, regardless of their initial value; such patients showed a tendency to reduce their HbA_{1c} and end the study period in a compensated state.
3. *Moderately decompensated:* patients who did not reach or exceed 9% in any of their measurements, but who do not belong to the compensated or improved segments.
4. *Highly decompensated:* patients who recorded some measurements over 9%, and who do not belong to the improved segment.

For example, Figure 4 shows a graphical representation of one patient from each defined segment. The HbA_{1c} values of 7% and 9% are marked with dotted green and red lines, respectively. It can be seen that even though the compensated patient has one measurement equal to 7%, the average of his or her measurements is below 7%. In the case of the improved patient, despite that his or her HbA_{1c} increased at one point, the overall trend for HbA_{1c} was to decrease, and the patient completed the period of study in a compensated state. The moderately decompensated patient never exceeded 9% but failed to qualify as either improved or compensated. Finally, the highly decompensated patient spent the majority of the time with values in excess of 9% and failed to achieve compensated status by the end of the period.

Statistical Analysis

The CVPA data of the 231 patients were collected, with a total of 1116 CVPAs. The analysis to study whether there is a statistically significant relationship between the identified patterns and patient evolution was undertaken using a proportion test. Fisher test was used when the evaluation sample proved to be too small. For each pattern, the proportion of patients who evolved in a specific manner was compared with those who evolved in the same manner in the total population studied. For all tests, the statistical significance was set to 0.05, and analysis was undertaken using R.

Results

Collaboration Patterns

PALIA created 12 different models and 7.8% (18 patients out of 231) of outliers (see Figures 5-7). As is the norm in health care processes, there is a high variability in the results obtained. The most frequently occurring behavior is present in 23.4% of cases (54 out of 231), followed by 11.3% (26 out of 231) with respect to the second group, decreasing to 3.0% for the final group (7 out of 231). Of the 12 models, six (models A, B, C, D, E, and F) have three nodes, four (models G, H, I, and J) have two nodes, and two (models K and L) have only one node.

By reviewing the models in greater detail, certain similarities between the identified behaviors can be observed. To identify the collaboration patterns, differences and similarities regarding the participation and self-referral indexes were analyzed for the 12 clusters created by the algorithm. This analysis was conducted separately according to the number of nodes present (disciplines that participated in the intervention) in each model.

Figure 5. Models with three nodes created by the PALIA Web application with parameters: similarity=15% and outliers=3%. P: physician; N: nurse; D: dietitian.

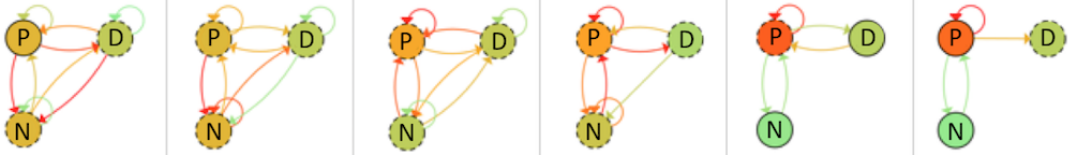
Models with Three Nodes						
Name	Model A	Model B	Model C	Model D	Model E	Model F
Diagram						
N° of Patients (% - out of 231)						
	23 (10.0%)	12 (5.2%)	21 (9.1%)	14 (6.1%)	9 (3.9%)	8 (3.5%)
Participation Index						
Physician	38%	37%	51%	54%	76%	76%
Nurse	37%	40%	23%	30%	4%	3%
Dietitian	25%	23%	26%	16%	20%	21%
Self-referral Index						
Physician	9%	24%	46%	38%	76%	68%
Nurse	17%	41%	6%	41%	0%	0%
Dietitian	4%	10%	9%	0%	0%	-
Referral Index						
P → N	49%	52%	32%	26%	7%	4%
N → P	38%	24%	63%	59%	100%	100%
P → D	42%	24%	22%	36%	17%	28%
D → P	28%	80%	54%	64%	100%	-
N → D	46%	35%	31%	0%	0%	0%
D → N	68%	10%	37%	36%	0%	-

Figure 6. Models with two nodes created by the PALIA Web application with parameters: similarity=15% and outliers=3%. P: physician; N: nurse.

Models with Two Nodes				
Name	Model G	Model H	Model I	Model J
Diagram				
N° of Patients (% - out of 231)				
	54 (23.4%)	7 (3.0%)	13 (5.6%)	11 (4.8%)
Participation Index				
Physician	68%	34%	41%	41%
Nurse	32%	66%	59%	59%
Dietitian	-	-	-	-
Self-referral Index				
Physician	52%	25%	22%	0%
Nurse	13%	44%	38%	22%
Dietitian	-	-	-	-
Referral Index				
P → N	48%	75%	78%	100%
N → P	87%	56%	62%	78%

Figure 7. Models with one node created by the PALIA Web application with parameters: similarity=15% and outliers=3%. P: physician; N: nurse.

Models with One Node		
Name	**Model K**	**Model L**
Diagram		
N° of Patients (% - out of 231)		
	26 (11.3%)	15 (6.5%)
Participation Index		
Physician	100%	-
Nurse	-	100%
Dietitian	-	-
Self-referral Index		
Physician	100%	-
Nurse	-	100%
Dietitian	-	-

There are two groups of patients that are treated by only one discipline during the entire period of study. One of these is treated solely by a physician (model K) and another solely by a nurse (model L). Both behaviors were classified under one pattern we called *self-contained*, as, in this instance, only a single discipline attends to the patient.

Conversely, in the models with two nodes, it is possible to observe that clusters G and H have one node that takes the lead in treatment, with a participation percentage that exceeds 65% in both cases (68% for the physician and 66% for the nurse, respectively). In addition, the leader makes self-referrals in approximately half of all their CVPAs (52% and 44%, respectively) and refers the other half to a distinct discipline. In turn, the second node plays an important ancillary role in relation to the first by referring the majority of their CVPAs to the discipline leader (over 70% of cases) while making very few self-referrals. These behaviors are classified under the *tacit leader* pattern, as one discipline has a greater participation because the other discipline refers the majority of their cases to the former. In particular, it can be observed that the physician is the leader in model G, whereas the role of leader is performed by the nurse in model H.

In the other two clusters with two nodes (clusters I and J), evident similarities also enable to group them into a single pattern. The participation index is the same in each node for both clusters, and the participation index for both disciplines are in the range 50%±10% (59% and 41%), that is, the disciplines participate in an equitable way. By reviewing the referrals of the CVPAs, it can be seen that the level of self-referral is lower than in the aforementioned cases. Upon receiving a CVPA, each discipline prefers to refer the patient to the other discipline (over 60% of cases in each model). This pattern was called *shared*, as the participation of both disciplines is more equitable, with no clear leader, and whereby referrals among different disciplines are more prevalent than self-referrals.

Subsequently, the diagrams with three nodes underwent comparison. In clusters A and B, even though the dietitian participates to a lesser extent, the three disciplines have a more equitable participation according to their participation indexes. The node with the highest participation has a participation index of 40%. Therefore, there is no single discipline that acts as leader. There is interaction across all disciplines regardless of the direction of the interaction. In general, both clusters work in an integrated way and make referrals in a more equitable manner than the rest. Consequently, these clusters are grouped under a single pattern called *participatory*.

Clusters C and D are characterized by the physician occupying the central role in the collaboration, with a participation index of 51% and 54%, respectively, compared with the nurse and dietitian who have a lower participation index. However, it can be seen that there is more integration in cluster C than in cluster D, in which there is almost no interaction between nurse and dietitian (in either direction). Rather, the nurse makes self-referrals in the majority of the CVPAs and refers almost no cases to the dietitian, compared with cluster C. Therefore, they are deemed two distinct patterns. Cluster C is called *equitably centered*, as the physician is at the center of collaboration and the other two disciplines participate equitably. The physician occupies the role of the sole leader, as he or she self-refers a significant proportion of cases (46%, compared with 6% for the nurse and 9% for the dietitian). The cluster D is identified as a *hierarchically centered* pattern, as the nurse occupies the role of secondary leader after the physician by self-referring a significant portion of CVPAs. Furthermore, there is almost no interaction between nurse and dietitian.

Finally, clusters E and F are related in that the physician in both possesses almost complete control over all treatment. The physician presents a participation index that exceeds all other cases (76% in both clusters), which can be explained by the high self-referral index the discipline has (over 65%). In both clusters, the main interaction is between physician and dietitian, whereas the nurse's participation index is below 5% in both clusters. Clusters E and F have been grouped under a *self-referred leader* pattern. The seven identified patterns are summarized in Table 2.

Clinical Outcome

By applying the aforementioned segmentation by clinical outcome to the 231 studied patients, four segments of patients were obtained. These segments are outlined in Table 3.

Figure 8 shows the evolution of the segments, displaying the values of the HbA_{1c} test for each patient over time. Each patient is represented by a line that corresponds to the value of his or her respective tests. The X-axis shows the time elapsed since the first measurement of each patient according to the period of study (which is not the same calendar date for all patients).

Relationship Between Collaboration and Clinical Outcomes

The final step of the proposed methodology is to conduct a statistical analysis to evaluate whether there is a statistically significant relationship between the different patterns identified and the evolution of patients. A proportion test was performed, and Fisher test was used when the evaluation sample proved to be too small. For each pattern, the proportion of patients who evolved in a specific manner was compared with those who evolved in the same manner in the total population studied. For all tests, the statistical significance was set to 0.05, and analysis was undertaken using R.

In performing the proportion test, patients grouped under each collaboration pattern were considered as different subpopulations. For each subpopulation, the proportion of patients who evolved in a specific way in accordance with the four aforementioned compensation segments was calculated. As the number of patients is different in each subpopulation, the previous proportions were compared with the proportions of the total population. Table 4 outlines the overall frequencies and percentages obtained for each subpopulation and the total population, respectively.

The proportion test showed statistically significant differences for certain patterns vs the total population studied. The greatest difference was observed on patients treated under the self-contained pattern, in which 24% more remained compensated compared with the total population (73% vs 49%, $P<.01$). The second difference occurs in the participatory pattern, in which a lower proportion of highly decompensated patients is recorded compared with the total population (3% vs 16%, $P=.03$). Finally, treatment under the self-referred leader pattern shows the worst result, with 19% more patients highly decompensated than the total population (35% vs 16%, $P=.05$).

Evaluation of Professionals

The results obtained were initially presented to a primary care physician. The physician noted that the self-contained and participatory patterns coincided with her experience. The relationship of compensated patients with the self-contained pattern may be understood as the following: if a patient remains in a stable condition, it is more common for fewer disciplines to provide the medical attention. Conversely, the participatory pattern provides evidence that suggests the importance of having a multidisciplinary team that oversees patient care and also supports the significant role played by the dietitian in the treatment process, as they are the main promoters of change in terms of patient lifestyles.

After this preliminary evaluation, we evaluated the results with three different groups of primary care physicians (one for each of the health care centers involved in the study). During each session, one of the researchers presented the results and answered questions. Then, the physicians filled out an informed consent form and answered a questionnaire aimed at understanding whether our results matched their experience and gathering their perceptions, observations, and concerns about the results. Finally, we conducted a brief discussion in which participants were free to voice their opinion, and one researcher took notes.

Table 2. Identified collaboration patterns.

Pattern	Description
Self-contained	Only one discipline (either nurse or physician) intervenes in patient treatment.
Tacit leader	Two disciplines, nurse and physician, one of whom is the leader of the treatment.
Shared	Two disciplines, without a leader. Each discipline refers the majority of their cardiovascular periodic appointments to another discipline.
Participatory	Three disciplines participate equitably. There is no leader.
Equitably centered	Three disciplines, in which the physician is the leader. The nurse and the dietitian respond primarily to the physician, but they also interact among themselves (to a lesser extent).
Hierarchically centered	Three disciplines, in which the physician is the leader. The nurse and the dietitian respond primarily to the physician, and they do not interact among themselves.
Self-referred leader	Three disciplines, in which the physician has almost complete control over treatment, receiving only minimal support from the other disciplines, primarily the dietitian.

Table 3. Patient segments according to their glycated hemoglobin (HbA_{1c}) evolution.

Segments	Patients, n (%)
Compensated	114 (49.4)
Improved	37 (16.0)
Moderately decompensated	45 (18.6)
Highly decompensated	37 (16.0)
Total	231 (100)

Figure 8. Clinical evolution of the patients in the different segments. HbA_{1c}: glycated hemoglobin.

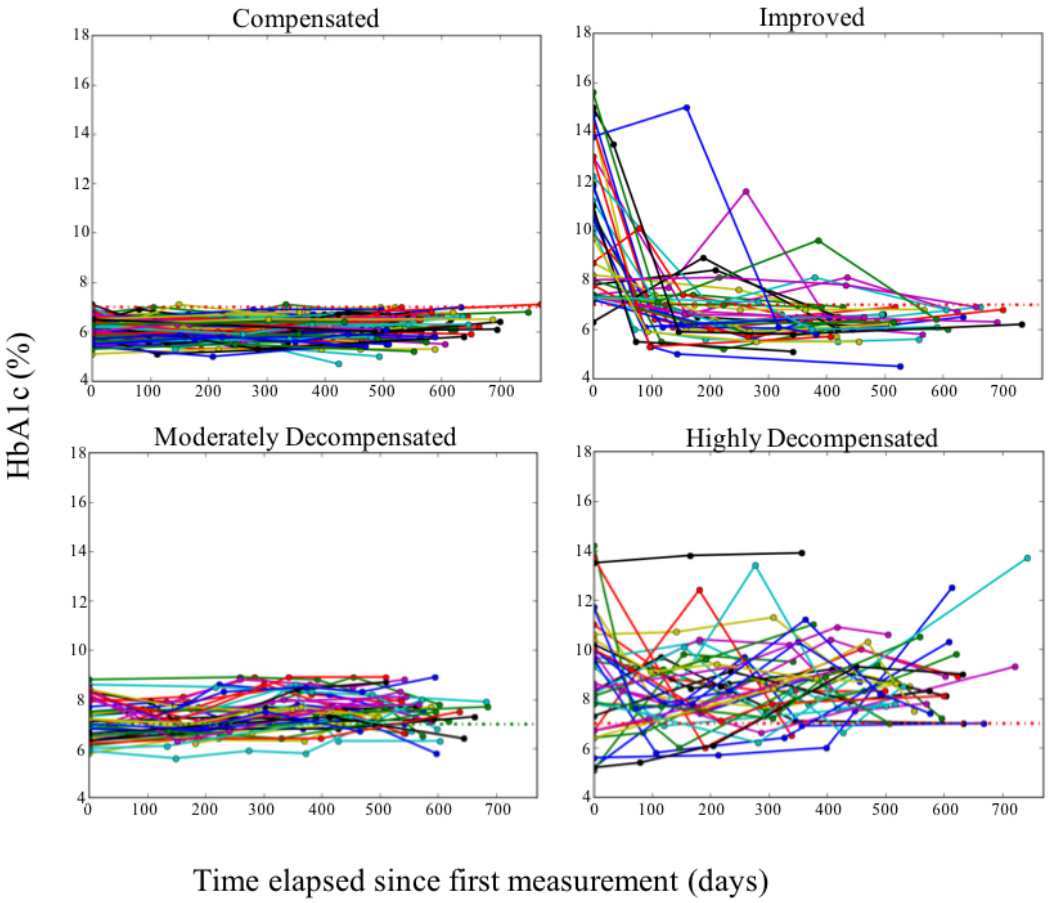


Table 4. Number of patients according to collaboration pattern and clinical evolution segment.

Segment	Compensated n (%)	Improved n (%)	Moderately decompensated n (%)	Highly decompensated n (%)	Total n (%)
Self-contained	30 (73)	3 (7)	4 (10)	4 (10)	41 (100)
Tacit leader	23 (38)	12 (20)	15 (25)	11 (18)	61 (100)
Shared	14 (58)	2 (8)	4 (17)	4 (17)	24 (100)
Participatory	17 (49)	10 (29)	7 (20)	1 (3)	35 (100)
Equitably centered	7 (33)	3 (14)	5 (24)	6 (29)	21 (100)
Hierarchically centered	5 (36)	1 (7)	3 (21)	5 (36)	14 (100)
Self-referred leader	5 (29)	2 (12)	4 (24)	6 (35)	17 (100)
Outliers	13 (72)	4 (22)	1 (6)	0 (0)	18 (100)
Total	114 (49.4)	37 (16.0)	43 (18.6)	37(16.0)	231(100)

Table 5. Evaluation results (N=23).

Pattern and segment	Observed by, n (%)
Pattern	
Self-contained	5 (22)
Tacit leader	6 (26)
Shared	11 (48)
Participatory	20 (87)
Equitably centered	14 (61)
Hierarchically centered	10 (43)
Self-referred leader	8 (35)
Segment	
Compensated	20 (87)
Moderately decompensated	21 (91)
Highly decompensated	18 (78)
Improved	21 (91)

Table 6. Statement results (N=23). T2DM: type 2 diabetes mellitus.

Statement	Agreement
The patterns describe the main ways of collaboration in T2DM[a] treatment in this Center	4.4
The patterns allow a correct classification of the ways of collaboration in T2DM treatment in this Center	4.2
Knowing these patterns may allow a better treatment of T2DM in this Center	4.3
The segments describe the main behaviors of T2DM patients in this Center	4.2
The segments allow a correct classification of the groups of patients treated for T2DM in this Center	4.3
It would be useful to treat differently patients classified in each segment	4.4

In total, 23 physicians participated in the evaluation. First, the participants stated whether they had observed each pattern and each segment. They were also able to name patterns and segments that had not been identified by the research. Then, participants answered a series of statements in a 5-point Likert Scale and several open questions (eg, "How useful are the relationships between patterns and segments?"). Some answers are provided below, translated from Spanish. A summary of the results is presented in Tables 5 and 6.

The participants stated that, in their experience, treatment of T2DM patients was participatory or equitably centered. One participant wrote, *I was surprised by the existence of a self-contained pattern.* They did not identify additional patterns, but did state the existence of other interventions in their centers, for example, educational workshops (to teach patients about how to handle T2DM), as well as other factors that may explain that a patient is always treated by a physician (One patient wrote as follows: *The disobedient pattern. Patients that go back to a physician although they were explicitly told to go to another professional [or that reject referrals]*).

Regarding the patient segments, these were more commonly identified by the participants—all of the segments had been observed by at least 18 out of 23 (78%) of participating physicians. When asked whether other segments were missing from our proposal (according to their experience), 4 mentioned the *worsened* segment, and 4 mentioned a *fluctuating* segment (which are both covered by the highly decompensated segment).

Overall, the evaluation by HCPs was positive. The physicians mentioned that the results could help them improve their protocols, patient treatment, and the management of human resources. Some mentioned that as there was no causal link established, they could not be certain that these changes would bring improvement.

Discussion

Principal Findings

The application of process mining techniques to the ECRs of the health care centers enables the analysis of the collaboration among HCPs. The advantage of the chosen algorithm is that it creates models that are easy to understand for HCPs. With these visualizations, the professionals in question may be able to view the work undertaken in the health care centers and comprehend how their protocols are actually taking place.

Leveraging the availability of data to control and improve processes can facilitate finding deviations from established protocols. One concrete example of this is the noncompliance with the norm that establishes appointments across three disciplines over the course of a year. It can be seen that 36.8% (85 out of 231) of patients had appointments with two disciplines

during the period of study, whereas 17.7% (41 out of 231) with only one. Understanding how professionals collaborate may be useful in allocating resources according to existing requirements. It should be noted that to monitor processes, it is necessary to establish metrics that are aligned to that which professionals are seeking to control and improve.

In particular, the use of clustering techniques by graph topology facilitated overall management of the variability inherent to health care processes to further understand the process. PALIA helped enable the observation of the variability with which patient care was undertaken and, within this variability, those characteristics that differentiate certain behaviors from others.

Comparison With Prior Work

By contrasting clinical evolution with the identified patterns, a number of differences can be observed. Certain comparisons showed statistically significant differences, which may signal a relationship between the collaboration pattern and overall patient evolution. Specifically, the self-contained pattern has a higher proportion of compensated patients compared with the total population. It may be possible to explain this because compensated patients are treated by a sole specialist, and this type of medical appointment is generally sufficient for them to remain stable, given the state of their condition. Other papers have published results that correlate collaboration with clinical outcomes, as randomized collaboration studies vs a control group [33,35,52], or via classifications of collaboration [34]. However, these studies have been unable to provide objective evidence of the collaboration patterns by means of quantitative analysis and their correlation with patient evolution.

Treatment under the participatory pattern is positively associated with patients who experience improvements in their evolution, and it is also associated with a lower proportion of patients who remain highly decompensated. If the level of significance is relaxed to 0.1, both characteristics are statistically significant. This outcome contrasts with the findings of Uddin [32], whereby he considers only integration between physicians in his analysis, in contrast to this paper, which incorporates evaluation by discipline.

The hierarchically centered and self-referred leader patterns are associated with more significant proportions of patients who are highly decompensated. This differs from the findings of Bosch [34], who has found no statistically significant differences between the types of hierarchical collaboration. The typology introduced by Bosch is based on qualitative analysis and is self-reported, which could produce a bias that explains the discrepancy. Teams working under these patterns may view themselves as providing treatments that are failing to reduce the number of decompensated patients to the desired extent. In the case of the self-referred leader, this discrepancy is statistically significant. Conversely, the shared and equitably centered patterns show no statistically significant differences compared with the total population. Therefore, the conclusion is that these particular types of treatment approaches cannot explain the evolution of patients with T2DM. Finally, the tacit leader pattern shows a lower proportion of compensated patients when compared with the total population studied (P=.06). It can be inferred that this type of treatment focuses on patients who show some type of decompensation.

In light of the foregoing, it can be observed that the main difference between the patterns with positive and negative performance is a distribution of participation between the distinct disciplines, which in the best-case scenario relates to more equitable participation. It also relates to the more integrated interactions between different disciplines. More participatory forms of work function better than those in which there is just one leader with significant control over treatment. The correlation between treatments that are primarily controlled by the physician (self-referred leader) and an increase in HbA_{1c}, compared with self-contained treatment and its respective correlation with compensated patients, is particularly interesting. It could be reasoned that one specialist is sufficient to monitor the state of patients who are well controlled, whereas a multidisciplinary intervention may be more beneficial for patients whose conditions are more serious.

This paper presents preliminary empirical results in relation to the organization of health care teams and the response of their patients to T2DM treatment. In contrast to the findings of Bosch [34], whereby the way in which teams organized themselves was self-reported, in this paper the approach used involved obtaining data from ECRs to ensure the type of collaboration was more objectively verifiable. However, despite their use of a completely different methodology, certain similarities arose with the Bosch typification, which divided groups into group culture, developmental culture, hierarchical culture, rational culture, cultural balance, and team climate. The relationships identified between the patterns and the evolution of patients does not necessarily constitute a causal link. A relationship in the opposite direction may, in fact, be possible. Indeed, certain external factors and individual characteristics of the patients could affect comparisons. For example, lifestyles, compliance with clinical indications, amount of exercise, or nutritional habits are all factors that may impact the evolution of patients with T2DM.

Limitations

The main limitation of the proposed methodology is that the analysis procedure and information used will be determined by the availability and quality of the data collected by information systems. The reception of incomplete and inconsistent data was one of the main problems faced. Although the professional expert in the field is the individual who is able to guide analysis objectives, the availability of data is the factor that determines the steps that must be taken to address these objectives. Some relevant data that may affect patient outcomes were not available, for example, some health care centers undertake education efforts in the form of workshops. Furthermore, there was no available information about medication adherence. Therefore, there may be certain steps described in this paper that require adaptation if they are to be extended to other cases with additional (or fewer) data. One available variable that was not used in the analysis was age (eg, the clinical goal for HbA_{1c} at the centers for patients over 80 is 8% instead of 7%).

A further limitation is that multidisciplinarity was only measured at the level of the cardiovascular team and did not include other types of professionals. Moreover, data were only analyzed at the discipline level regardless of the particular professional that provided care.

This study is based on data from 2012 onwards, so some results may not be representative to the present day. The clinical appointments included in this study relate to scheduled visits explicitly registered as CVPAs in the information systems. A comparison of collaboration between the different health care centers could also have been beneficial, although the proportion of patients in the related sample would have been extremely small to generate statistically significant results. In the future, it would be useful to execute this analysis using a larger sample. Finally, future studies are recommended to include the severity of the patient condition and their comorbidities as variables to measure patient evolution via changes in these indexes over time, rather than using them as filters.

Conclusions

The use of process mining represents an opportunity for health care centers to understand how their processes are being executed and which forms of collaboration lead to improved outcomes. The definition of simple clinical-based segmentations is critical for facilitating the interpretation of results. The process mining tool used in this research, PALIA Web, allows to analyze how different HCPs collaborate in flexible and unstructured processes such as the treatment of T2DM patients. By combining trace clustering (to manage the diversity of patients present in the log) and collaboration pattern discovery techniques, it is able to identify several collaboration patterns among HCPs in the treatment of patients. It also allows to easily visualize the obtained collaboration patterns so that HCPs can interpret the differences between them. The methodology used in this study made it possible to analyze the relationship between these collaboration patterns and the clinical evolution of patients, so as to identify the most successful patterns. Finally, health care centers can then promote the most successful patterns among their professionals so as to improve the treatment of T2DM patients.

Acknowledgments

This paper was partially funded by the National Commission for Scientific and Technological Research, the Formation of Advanced Human Capital Program and the National Fund for Scientific and Technological Development (CONICYT-PCHA/Doctorado Nacional/2016-21161705 and CONICYT-FONDECYT/1150365; Chile). The authors would like to thank Áncora UC primary health care centers for their help with this research. The founding sponsors had no role in the design of the study in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Authors' Contributions

TC was the main researcher, executing algorithms and analyzing their results. MS and VH together developed the concept of this work and provided guidance and substantial revisions. CS collaborated with TC in the execution of this research. DC aided in data acquisition and initial interpretation, and FP provided additional interpretations and collaboration in the evaluation sessions. Finally, CFL provided invaluable guidance and the PALIA tool, with special adaptations for this research in particular.

Conflicts of Interest

None declared.

References

1. Chen C, Tseng C, Cheng S. Continuity of care, medication adherence, and health care outcomes among patients with newly diagnosed type 2 diabetes: a longitudinal analysis. Med Care 2013 Mar;51(3):231-237. [doi: 10.1097/MLR.0b013e31827da5b9] [Medline: 23269110]
2. International Diabetes Federation. IDF. 2015. IDF Diabetes Atlas 7th Edition (2015) URL: https://www.idf.org/e-library/epidemiology-research/diabetes-atlas/13-diabetes-atlas-seventh-edition.html [accessed 2018-03-19] [WebCite Cache ID 6y2wN35rT]
3. Ministerio de Salud de Chile. minsal.cl. 2010. Encuesta Nacional de Salud ENS Chile 2009-2010 URL: http://www.minsal.cl/estudios_encuestas_salud/ [accessed 2018-03-23] [WebCite Cache ID 6y7khh3EZ]
4. Ministerio de Salud de Chile. minsal.cl. 2010. Guía Clinica Diabetes Mellitus Tipo 2 URL: http://www.minsal.cl/portal/url/item/72213ed52c3e23d1e04001011f011398.pdf [accessed 2018-03-20] [WebCite Cache ID 6y3vb8xtk]
5. Sapunar Z J. Epidemiología de la Diabetes mellitus en Chile. Revista Medica Clinica Las Condes 2016;27(2):146-151. [doi: 10.1016/j.rmclc.2016.04.003]
6. World Health Organization. who.int. Global Report on Diabetes URL: http://www.who.int/diabetes/global-report/en/ [accessed 2018-03-20] [WebCite Cache ID 6y3wciMhR]
7. Poblete F, Glasinovic A, Sapag J, Barticevic N, Arenas A, Padilla O. Social support and cardiovascular health: adaptation of a social support scale for hypertensive and diabetic patients in primary care, Chile. Aten Primaria 2015 Oct;47(8):523-531 [FREE Full text] [doi: 10.1016/j.aprim.2014.10.010] [Medline: 25795034]

8. Tuligenga RH, Dugravot A, Tabák AG, Elbaz A, Brunner EJ, Kivimäki M, et al. Midlife type 2 diabetes and poor glycaemic control as risk factors for cognitive decline in early old age: a post-hoc analysis of the Whitehall II cohort study. Lancet Diabetes Endocrinol 2014 Mar;2(3):228-235 [FREE Full text] [doi: 10.1016/S2213-8587(13)70192-X] [Medline: 24622753]
9. American Diabetes Association. Standards of Medical Care in Diabetes. Diabetes Care 2017;40(1).
10. Gamiochipi M, Cruz M, Kumate J, Wacher NH, DIMSS Study Group. Effect of an intensive metabolic control lifestyle intervention in type-2 diabetes patients. Patient Educ Couns 2016 Dec;99(7):1184-1189. [doi: 10.1016/j.pec.2016.01.017] [Medline: 26947102]
11. Wagner EH, Sandhu N, Newton KM, McCulloch DK, Ramsey SD, Grothaus LC. Effect of improved glycemic control on health care costs and utilization. J Am Med Assoc 2001 Jan 10;285(2):182-189. [Medline: 11176811]
12. McDonald J, Jayasuriya R, Harris MF. The influence of power dynamics and trust on multidisciplinary collaboration: a qualitative case study of type 2 diabetes mellitus. BMC Health Serv Res 2012 Mar 13;12:63 [FREE Full text] [doi: 10.1186/1472-6963-12-63] [Medline: 22413897]
13. Gucciardi E, Espin S, Morganti A, Dorado L. Exploring interprofessional collaboration during the integration of diabetes teams into primary care. BMC Fam Pract 2016 Feb 01;17:12 [FREE Full text] [doi: 10.1186/s12875-016-0407-1] [Medline: 26831500]
14. Appleyard J, Rawaf S. Person centered medicine, primary care, and public health. Int J Pers Cent Med 2015;5(3):97-100.
15. Caron F, Vanthienen J, Vanhaecht K, Limbergen EV, De Weerdt J, Baesens B. Monitoring care processes in the gynecologic oncology department. Comput Biol Med 2014 Jan;44:88-96. [doi: 10.1016/j.compbiomed.2013.10.015] [Medline: 24377692]
16. Rothman AA, Wagner EH. Chronic illness management: what is the role of primary care? Ann Intern Med 2003 Feb 04;138(3):256-261. [Medline: 12558376]
17. Organisation for Economic Co-operation and Development. OECD. 2016. OECD Health Policy Overview: Health Policy in Chile URL: http://www.oecd.org/els/health-systems/health-policy-in-your-country.htm [accessed 2018-03-20] [WebCite Cache ID 6y3zcHPAF]
18. Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: a literature review. J Biomed Inform 2016 Jun;61:224-236 [FREE Full text] [doi: 10.1016/j.jbi.2016.04.007] [Medline: 27109932]
19. Van Der Aalst W. In: Daniel F, Barkaoui K, Dustdar S, editors. Process mining manifesto. Berlin, Heidelberg: Springer; 2011:169-194.
20. Fernandez-Llatas C, Lizondo A, Monton E, Benedi J, Traver V. Process mining methodology for health process tracking using real-time indoor location systems. Sensors (Basel) 2015 Nov 30;15(12):29821-29840 [FREE Full text] [doi: 10.3390/s151229769] [Medline: 26633395]
21. Mans RS, van der Aalst W, Vanwersch RJ. Process Mining in Healthcare. Heidelberg: Springer International Publishing; 2015.
22. van der Aalst W. Process Mining. Berlin Heidelberg: Springer-Verlag; 2011.
23. Kim E, Kim S, Song M, Kim S, Yoo D, Hwang H, et al. Discovery of outpatient care process of a tertiary university hospital using process mining. Healthc Inform Res 2013 Mar;19(1):42-49 [FREE Full text] [doi: 10.4258/hir.2013.19.1.42] [Medline: 23626917]
24. Mans RS, Schonenberg MH, Song M, van der Aalst W, Bakker PJ. Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital. In: Fred A, Filipe J, Gamboa H, editors. Communications in Computer and Information Science. Berlin Heidelberg: Springer; 2008.
25. Harper P. Combining data mining tools with health care models for improved understanding of health processes and resource utilisation. Clin Invest Med 2005 Dec;28(6):338-341. [Medline: 16450628]
26. Harper PR, Sayyad MG, de Senna V, Shahani AK, Yajnik CS, Shelgikar KM. A systems modelling approach for the prevention and treatment of diabetic retinopathy. Eur J Oper Res 2003;150(1):81-91. [doi: 10.1016/S0377-2217(02)00787-7]
27. Rebuge Á, Ferreira DR. Business process analysis in healthcare environments: a methodology based on process mining. Inf Syst 2012;37(2):99-116. [doi: 10.1016/j.is.2011.01.003]
28. Ferreira D, Zacarias M, Malheiros M, Ferreira P. Approaching Process Mining with Sequence Clustering: Experiments and Findings. In: Business Process Management. Berlin Heidelberg: Springer; 2007:360-374.
29. Song M, Günther CW, van der Aalst W. Trace Clustering in Process Mining. In: Business Process Management. Berlin Heidelberg: Springer; 2009:109-120.
30. Cheong LH, Armour CL, Bosnic-Anticevich SZ. Multidisciplinary collaboration in primary care: through the eyes of patients. Aust J Prim Health 2013;19(3):190-197. [doi: 10.1071/PY12019] [Medline: 22951207]
31. Boyle E, Saunders R, Drury V. A qualitative study of patient experiences of type 2 diabetes care delivered comparatively by general practice nurses and medical practitioners. J Clin Nurs 2016 Jul;25(13-14):1977-1986. [doi: 10.1111/jocn.13219] [Medline: 27251784]
32. Uddin S, Hossain L. Effects of Physician Collaboration Network on Hospital Outcomes. 2012 Presented at: Fifth Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2012); 2012; Melbourne, Australia.
33. Borgermans L, Goderis G, Van Den Broeke C, Verbeke G, Carbonez A, Ivanova A, et al. Interdisciplinary diabetes care teams operating on the interface between primary and specialty care are associated with improved outcomes of care: findings

from the Leuven Diabetes Project. BMC Health Serv Res 2009;9:179 [FREE Full text] [doi: 10.1186/1472-6963-9-179] [Medline: 19811624]
34. Bosch M, Dijkstra R, Wensing M, van der Weijden T, Grol R. Organizational culture, team climate and diabetes care in small office-based practices. BMC Health Serv Res 2008 Aug 21;8:180 [FREE Full text] [doi: 10.1186/1472-6963-8-180] [Medline: 18717999]
35. Counsell SR, Callahan CM, Clark DO, Tu W, Buttar AB, Stump TE, et al. Geriatric care management for low-income seniors: a randomized controlled trial. J Am Med Assoc 2007 Dec 12;298(22):2623-2633. [doi: 10.1001/jama.298.22.2623] [Medline: 18073358]
36. Anderson JG. Evaluation in health informatics: social network analysis. Comput Biol Med 2002 May;32(3):179-193. [Medline: 11922934]
37. Gray JE, Davis DA, Pursley DM, Smallcomb JE, Geva A, Chawla NV. Network analysis of team structure in the neonatal intensive care unit. Pediatrics 2010 Jun;125(6):e1460-e1467. [doi: 10.1542/peds.2009-2621] [Medline: 20457681]
38. Mian O, Koren I, Rukholm E. Nurse practitioners in Ontario primary healthcare: referral patterns and collaboration with other healthcare professionals. J Interprof Care 2012 May;26(3):232-239. [doi: 10.3109/13561820.2011.650300] [Medline: 22256946]
39. Crossley N, Bellotti E, Edwards G, Everett MG, Koskinen J, Tranmer M. Social Network Analysis for Ego-Nets. Los Angeles: SAGE Publications Ltd; 2015.
40. Ministerio de Salud de Chile. Fondo Nacional de Salud URL: https://www.fonasa.cl/sites/fonasa/beneficiarios [accessed 2018-03-20] [WebCite Cache ID 6y4Ch2Y3c]
41. Goldstein DE, Little RR, Lorenz RA, Malone JI, Nathan D, Peterson CM, et al. Tests of glycemia in diabetes. Diabetes Care 2004 Jul;27(7):1761-1773. [Medline: 15220264]
42. Young BA, Lin E, Von Korff M, Simon G, Ciechanowski P, Ludman EJ, et al. Diabetes complications severity index and risk of mortality, hospitalization, and healthcare utilization. Am J Manag Care 2008 Jan;14(1):15-23 [FREE Full text] [Medline: 18197741]
43. Meduru P, Helmer D, Rajan M, Tseng C, Pogach L, Sambamoorthi U. Chronic illness with complexity: implications for performance measurement of optimal glycemic control. J Gen Intern Med 2007 Dec;22 Suppl 3:408-418 [FREE Full text] [doi: 10.1007/s11606-007-0310-5] [Medline: 18026810]
44. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther 2001 Oct;26(5):331-342. [Medline: 11679023]
45. Karter AJ, Parker MM, Moffet HH, Ahmed AT, Ferrara A, Liu JY, et al. Missed appointments and poor glycemic control: an opportunity to identify high-risk diabetic patients. Med Care 2004 Feb;42(2):110-115. [doi: 10.1097/01.mlr.0000109023.64650.73] [Medline: 14734947]
46. World Health Organization. 2003. Adherence to long-term therapies: evidence for action URL: http://www.who.int/chp/knowledge/publications/adherence_report/en/ [accessed 2018-03-20] [WebCite Cache ID 6y4CORM0i]
47. Toth EL, Majumdar SR, Guirguis LM, Lewanczuk RZ, Lee TK, Johnson JA. Compliance with clinical practice guidelines for type 2 diabetes in rural patients: treatment gaps and opportunities for improvement. Pharmacotherapy 2003 May;23(5):659-665. [Medline: 12741441]
48. Melnikow J, Kiefe C. Patient compliance and medical research: issues in methodology. J Gen Intern Med 1994 Feb;9(2):96-105. [Medline: 8164085]
49. Fernandez-Llatas C, Valdivieso B, Traver V, Benedi JM. Using process mining for automatic support of clinical pathways design. Methods Mol Biol 2015;1246:79-88. [doi: 10.1007/978-1-4939-1985-7_5] [Medline: 25417080]
50. Fernández-Llatas C, Benedi J, García-Gómez JM, Traver V. Process mining for individualized behavior modeling using wireless tracking in nursing homes. Sensors (Basel) 2013 Nov 11;13(11):15434-15451 [FREE Full text] [doi: 10.3390/s131115434] [Medline: 24225907]
51. Concaro S, Sacchi L, Cerra C, Bellazzi R. Mining administrative and clinical diabetes data with temporal association rules. Stud Health Technol Inform 2009;150:574-578. [Medline: 19745376]
52. Wishah RA, Al-Khawaldeh OA, Albsoul AM. Impact of pharmaceutical care interventions on glycemic control and other health-related clinical outcomes in patients with type 2 diabetes: randomized controlled trial. Diabetes Metab Syndr 2015;9(4):271-276. [doi: 10.1016/j.dsx.2014.09.001] [Medline: 25301007]

Abbreviations

CESFAM: Centro de Salud Familiar
CVPA: cardiovascular periodic appointment
T2DM: type 2 diabetes mellitus
ECR: electronic clinical record
HbA $_{1c}$: glycated hemoglobin
HCP: health care professional

Edited by G Eysenbach; submitted 01.09.17; peer-reviewed by R Mans, R Jayasuriya; comments to author 28.09.17; revised version received 31.01.18; accepted 18.02.18; published 10.04.18.

Please cite as:
Conca T, Saint-Pierre C, Herskovic V, Sepúlveda M, Capurro D, Prieto F, Fernandez-Llatas C
Multidisciplinary Collaboration in the Treatment of Patients With Type 2 Diabetes in Primary Care: Analysis Using Process Mining
J Med Internet Res 2018;20(4):e127
URL: http://www.jmir.org/2018/4/e127/
doi:10.2196/jmir.8884
PMID:29636315

Original Paper

Prevalence of Multiple Chronic Conditions Among Older Adults in Florida and the United States: Comparative Analysis of the OneFlorida Data Trust and National Inpatient Sample

Zhe He[1*], PhD; Jiang Bian[2*], PhD; Henry J Carretta[3], MPH, PhD; Jiwon Lee[4], PhD; William R Hogan[2], MD; Elizabeth Shenkman[2], PhD; Neil Charness[5], PhD

[1]School of Information, Florida State University, Tallahassee, FL, United States

[2]Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States

[3]Department of Behavioral Sciences and Social Medicine, Florida State University, Tallahassee, FL, United States

[4]Department of Statistics, Florida State University, Tallahassee, FL, United States

[5]Department of Psychology, Florida State University, Tallahassee, FL, United States

[*]these authors contributed equally

Corresponding Author:
Zhe He, PhD
School of Information
Florida State University
271 Shores Bldg
142 Collegiate Loop
Tallahassee, FL, 32306
United States
Phone: 1 8506445775
Email: zhe@fsu.edu

Abstract

Background: Older patients with multiple chronic conditions are often faced with increased health care needs and subsequent higher medical costs, posing significant financial burden to patients, their caregivers, and the health care system. The increasing adoption of electronic health record systems and the proliferation of clinical data offer new opportunities for prevalence studies and for population health assessment. The last few years have witnessed an increasing number of clinical research networks focused on building large collections of clinical data from electronic health records and claims to make it easier and less costly to conduct clinical research.

Objective: The aim of this study was to compare the prevalence of common chronic conditions and multiple chronic conditions in older adults between Florida and the United States using data from the OneFlorida Clinical Research Consortium and the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS).

Methods: We first analyzed the basic demographic characteristics of the older adults in 3 datasets—the 2013 OneFlorida data, the 2013 HCUP NIS data, and the combined 2012 to 2016 OneFlorida data. Then we analyzed the prevalence of each of the 25 chronic conditions in each of the 3 datasets. We stratified the analysis of older adults with hypertension, the most prevalent condition. Additionally, we examined trends (ie, overall trends and then by age, race, and gender) in the prevalence of discharge records representing multiple chronic conditions over time for the OneFlorida (2012-2016) and HCUP NIS cohorts (2003-2013).

Results: The rankings of the top 10 prevalent conditions are the same across the OneFlorida and HCUP NIS datasets. The most prevalent multiple chronic conditions of 2 conditions among the 3 datasets were—hyperlipidemia and hypertension; hypertension and ischemic heart disease; diabetes and hypertension; chronic kidney disease and hypertension; anemia and hypertension; and hyperlipidemia and ischemic heart disease. We observed increasing trends in multiple chronic conditions in both data sources.

Conclusions: The results showed that chronic conditions and multiple chronic conditions are prevalent in older adults across Florida and the United States. Even though slight differences were observed, the similar estimates of prevalence of chronic conditions and multiple chronic conditions across OneFlorida and HCUP NIS suggested that clinical research data networks such as OneFlorida, built from heterogeneous data sources, can provide rich data resources for conducting large-scale secondary data analyses.

(J Med Internet Res 2018;20(4):e137) doi:10.2196/jmir.8961

KEYWORDS

medical informatics; chronic disease; comorbidity; geriatrics

Introduction

Background

Chronic conditions (CCs) affect nearly half of the adult population in the United States. The prevalence of some CCs such as hypertension, asthma, cancer, and diabetes has increased over the last a few years [1-3]. Older patients with multiple chronic conditions (MCCs) are often faced with increased health care needs and subsequent higher medical costs, posing significant financial burden to patients, their caregivers, and the health care system.

Understanding the trends in the prevalence of MCC informs policy makers, health care providers, and payers about chronic disease management and prevention and helps to predict future health care needs [4]. The literature on MCC research mostly uses national claims data or national surveys to estimate the prevalence of MCCs [4-7]. Freid et al [4] presented the estimates of the population aged 45 and older with 2 or more self-reported CCs using the National Health Interview Survey (NHIS) data. They reported that the percentage of adults with MCCs increased in both 45 to 64 years and 65 and older age groups between 1999 and 2010. Ward and Schiller [5] analyzed the prevalence of MCCs among US adults also using the 2010 NHIS data and reported an increasing prevalence of MCCs from 2001 to 2010. Ashman and Beresovsky did an MCC analysis among US adults who visited physician offices, using the National Ambulatory Medical Care Survey data [6]. They found that hypertension was the most prevalent CC that appeared in the top 5 MCC dyads and triads. He et al [7] used the National Health and Nutrition Examination Survey data and a public clinical trial registry—ClinicalTrials.gov—to analyze the gap between the prevalence of MCCs and the clinical trials on the prevalent MCCs. They found that the current and past clinical trials rarely investigate the prevalent MCCs.

Recent years have witnessed a wide adoption of electronic health record (EHR) systems driven by the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 [8]. By 2015, over 90% of nonfederal acute care hospitals adopted a certified EHR [9]. By the end of 2017, about 90% of the office-based physicians have been using EHRs in the Unites States [10]. With public health reporting as part of the meaningful use criteria for hospitals to receive the incentive payments of the HITECH Act, EHRs have been recognized as an important data source for public health surveillance [11] (especially in chronic disease surveillance [12-14]), cohort identification for clinical studies [15], and disease-risk prediction [16]. The advantage of using EHRs over survey data is multifaceted. First, EHRs have fine-grained clinical data that are rarely collected and reported in the survey or claims data. Second, EHRs contain longitudinal patient data, whereas survey data mostly provide merely a snapshot of the health conditions for a person. However, as EHR data only contain patients who paid a visit to the health care facilities, they may not be as representative of the national population as the survey data. Therefore, it is necessary to investigate the extent to which EHR data can represent the broader population to inform researchers who are using EHRs for public health and chronic disease surveillance. Recently, Perlman et al created an EHR-based public health surveillance system in New York City [14]. They compared the CC estimates generated in this system with those from a population-based survey in New York and found that diabetes, hypertension, smoking, and obesity prevalence was close to the survey results, but depression and influenza vaccination estimates were substantially lower than the survey-based estimates [14].

The last few years have witnessed an increasing number of clinical research networks focused on building large collections of clinical datasets from EHRs and claims to offer a collaborative environment for researchers across disparate organizations. It is anticipated that the analysis of such data will lead to advances in medical knowledge, progress in health care delivery, and improvements in population health [17-21]. One notable example is the National Patient-Centered Clinical Research Network (PCORnet) [17,22], funded by the Patient-Centered Outcomes Research Institute (PCORI). PCORnet comprises a coordinating center and 33 partner networks, including 13 Clinical Data Research Networks (CDRNs) and 20 Patient-Powered Research Networks. PCORnet is “designed to make it faster, easier, and less costly to conduct clinical research than is now possible by harnessing the power of large amounts of health data and patient partnerships” [22]. It is a national “network of networks” that routinely collects data from a variety of health care organizations, including hospitals, community clinics, health plans, and national data registries (eg, cancer registries and vital statistics).

PCORnet empowers individuals and organizations to use this big dataset to answer practical questions that help patients, clinicians, and other stakeholders make informed health care decisions. For example, PCORnet provides an invaluable cohort discovery service that proves particularly useful for identifying cohorts of a variety of health conditions, especially for rare diseases. With such a large collection of electronic patient data, PCORnet can effectively support large-scale randomized clinical trials, comparative effectiveness research studies, and longitudinal observational studies. EHRs such as those warehoused in CDRNs have been widely used for comparative effectiveness analysis [23-26], cohort identification [27-29], and public health surveillance studies [25,30,31]. However, it is not yet known the extent to which the population in these CDRNs such as OneFloridais is representative of the national population. This is an important metric that needs to be examined to understand the comprehensiveness of the OneFlorida population now and to improve the interpretability and generalizability of the OneFlorida data and the reproducibility of the aforementioned studies.

Florida has the largest elderly population in the United States. OneFlorida is one of the 13 CDRNs contributing to the national PCORnet [32]. The OneFlorida Data Trust is a secure centralized data repository that integrates various data sources from contributing organizations in the OneFlorida research consortium, including 22 hospitals and 914 community-based clinical practices that provide care to 48% of Floridians. As of June 2017, the Data Trust contains 10.9 million patient records including data from partners' EHR systems, as well as claims data from Florida Medicaid. Ultimately, the Data Trust will include claims data for Florida Medicare beneficiaries, Florida Vital Statistics records, and Florida Cancer Data System records. The OneFlorida Data Trust employs the PCORNet Common Data Model (CDM) version 3.1 [33], which uses standard vocabularies to encode diagnoses (ie, International Classification of Diseases, ICD), procedures (ie, ICD procedure codes, Current Procedural Terminology, and Healthcare Common Procedure Coding System codes), laboratory observations (ie, Logical Observation Identifiers Names and Codes), and medications (ie, RxNorm and National Drug Code). The OneFlorida and PCORnet data only contains Health Insurance Portability and Accountability Act limited data, for which we obtained permission to use. Throughout this paper, OneFlorida refers to the inpatient data extracts used to conduct this analysis unless otherwise noted.

Objective

The purpose of this study is to estimate and compare the prevalence of common CCs and MCCs among older adults in Florida and United States from the OneFlorida Data Trust and a national data source—the National Inpatient Sample (NIS) from the Healthcare Cost and Utilization Project (HCUP) of Agency for Healthcare Research and Quality [34]. The NIS is a comprehensive source of inpatient hospital data in the United States. As NIS contains only the inpatient data, we also used the inpatient EHR records in the OneFlorida Data Trust to estimate Florida population. For this paper we define MCC as 2 or more CCs according to the Center for Medicare and Medicaid Services (CMS) algorithm [35].

We formulated 2 research questions (RQs) in this study:

> RQ1: What is the prevalence of common CCs in hospital discharge records for older adults in the OneFlorida Data Trust inpatient data and how does it compare with unweighted national estimates from the HCUP NIS?
>
> RQ2: Are the 10 most common CCs and the prevalence of MCC in hospital discharge records for older adults in the OneFlorida Data Trust consistent with the unweighted HCUP NIS national population?

Methods

Data Collection and Preparation

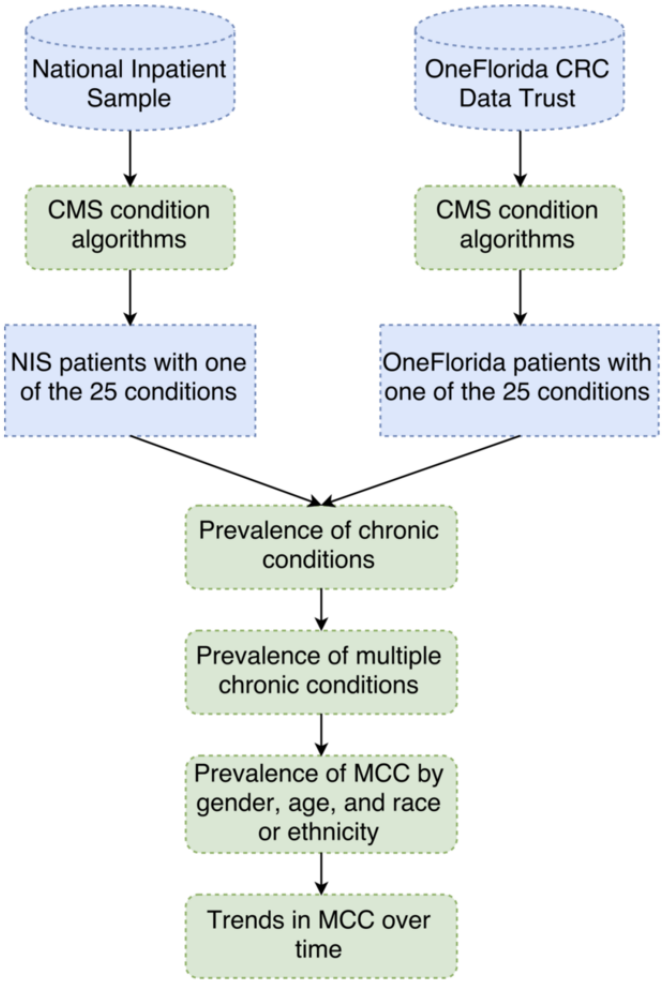
OneFlorida inpatient discharge records for 2012 to 2016 for 22 CCs were identified using the CMS Chronic Condition Warehouse (CCW) algorithm [35]. We included records with an admission source of home, another facility, or the emergency department. The 2013 discharge records were used for the cross-sectional analysis and the 2012 to 2016 records were used for a longitudinal comparison.

NIS is the largest publicly available all-payer inpatient health care database in the United States. Unweighted, it contains 7 million hospital discharge records each year and the weighted sample represents 25 million discharges. Beginning within the 2012 data year, the NIS approximates a 20% stratified sample of all discharges from US community hospitals, excluding rehabilitation and long-term acute-care hospitals. The 2013 NIS file was used for our cross-sectional analysis, and the 2003 to 2013 data were used for the longitudinal analysis. NIS includes information on all patients, including individuals covered by Medicare, Medicaid, or private insurance, uninsured. Researchers and policy makers use NIS to make national estimates of inpatient health care utilization [36], access to care [37], inpatient charges [36,38,39], quality of hospital care [37], and outcomes [39,40].

Figure 1 illustrates the process of data preparation and analysis. As the first step, we identified patients with CC using the CMS CCW algorithm [35]. The CMS CCW algorithm identifies cases for 27 condition categories using the criteria, such as (1) a validated list of ICD-9-CM and ICD-10 diagnosis codes, (2) the number of discharge record occurrences with diagnosis codes meeting the case definition within a year, (3) the number of consecutive years with confirming diagnoses in order to identify an individual case within a specific CC category in a given year to identify 27 conditions, and (4) the source type of service. We excluded 2 algorithm conditions that do not use inpatient records for case identification for cataracts or glaucoma, because those conditions are typically not associated with inpatient care. We modified the algorithm criteria for 7 other conditions, which were (1) rheumatoid arthritis and osteoarthritis, (2) chronic kidney disease, (3) heart failure, (4) diabetes, (5) Alzheimer disease, (6) Alzheimer disease and related conditions, and (7) ischemic heart disease. These 7 conditions require 2 or 3 consecutive years with the diagnosis to meet the case criteria or in the case of rheumatoid arthritis or osteoarthritis, 2 diagnoses within a year. Due to privacy concerns, the NIS does not assign unique patient identifiers that can be tracked across facilities or time. Therefore, we modified the criteria for those 7 conditions and identified cases based on a single inpatient discharge record. We limited the analysis to persons aged 65 years or older for the 25 remaining conditions defined by the CMS algorithm [35].

We identified older adults as those who were above 65 years at the time of inpatient discharge in both data sources. We stratified our analysis by age group, namely, 65 to 74, 75 to 84, and 85 and above. Besides age, we also extracted the gender and race or ethnicity variables of the patients. For OneFlorida analysis, we generated 2 datasets, one for a cross-sectional analysis (2013) and the other for a longitudinal analysis with data from all the years currently available in the OneFlorida Data Trust (2012-2016). For HCUP NIS, we used the 2013 data for the cross-sectional analysis and 2003 to 2013 data for the longitudinal analysis. The decision of using different year range for OneFlorida Data Trust and HCUP NIS was made based on the availability of the data and the richness of the analysis.

Figure 1. Workflow of data preparation and analysis. CRC: Clinical Research Consortium. CMS: Center for Medicare and Medicaid Services. NIS: National Inpatient Sample. MCC: multiple chronic conditions.



Nevertheless, both OneFlorida Data Trust and HCUP NIS have the 2012 to 2013 data.

OneFlorida Data Trust uses the PCORNet CDM version 3.1, which is a relational schema. The data are stored in a Microsoft SQL server hosted by the University of Florida Health Science Center. We included patients who had either direct inpatient admissions or emergency-to-inpatient admissions. The HCUP NIS data were released in the SAS format. We preprocessed the HCUP SAS datasets and loaded them into a Microsoft SQL server.

Data Analysis

The analysis included descriptive statistics for the 25 individual conditions and MCC in 3 analytic files, that is (1) the OneFlorida data for the year 2013 only (OneFlorida 2013), (2) the NIS data for the year 2013 only (NIS 2013), and (3) the OneFlorida data for 2012 to 2016 (OneFlorida 2012-2016). Chi-square tests were used to examine group differences in the proportions of interest across the 3 data files.

We first analyzed the basic demographic characteristics of the older adults in the two 2013 datasets and the OneFlorida 2012 to 2016 data. Then we analyzed the prevalence of each of the 25 CCs for each of the 3 datasets. We did a deep dive, stratified the analysis, of the older adults with hypertension, the most prevalent condition. The prevalence of hypertension in the 24 age-gender-race-ethnicity strata was compared across the 3 datasets. We also examined the number of conditions per hospital record in each dataset for 2013 and further stratified the prevalence of patients with MCCs in 2013 by gender and race or ethnicity.

Additionally, we examined trends (ie, overall trends and then by age, race, and gender) in the prevalence of discharge records representing MCCs across time for the OneFlorida and NIS cohorts. Pearson correlation coefficient was computed to compare the MCC trends stratified by age group, sex, and race or ethnicity.

Results

Basic Characteristics

Table 1 shows the basic characteristics of the older adult populations in the OneFlorida 2013 data, OneFlorida 2012 to 2016 data, and the HCUP NIS 2013 datasets. The average age of older adults in the OneFlorida 2013 and 2012 to 2016 data was similar (2-tailed *t* test, degrees of freedom=63435.3226, *P*>.05). The older adults in OneFlorida 2013 were slightly younger than those in NIS 2013. There were more elderly female patients than elderly male patients across all 3 datasets. OneFlorida 2012 to 2016 had a statistically significantly higher percentage of Hispanics, non-Hispanic (NH) blacks, and a lower

percentage of non-Hispanic whites and Asian or Pacific Islanders than the NIS 2013 (chi-square statistics 72587091891.83, $P<.001$).

Prevalence of Chronic Conditions

The rankings of the top 10 prevalent conditions were the same across the 3 datasets. These conditions were hypertension, hyperlipidemia, ischemic heart disease, diabetes, anemia, chronic kidney disease, atrial fibrillation, heart failure, chronic obstructive pulmonary disease, and RA. However, there were differences in the prevalence of each disease between the NIS and OneFlorida data. Comparing the NIS and OneFlorida data, one can observe that a higher percentage of older adults in OneFlorida had hypertension (80.97% vs 76.32%), hyperlipidemia (52.42% vs 45.94%), and diabetes (35.32% vs 33.93%) than in NIS; whereas a higher percentage of older adults in NIS had chronic kidney disease (33.22% vs 31.24%) and heart failure (25.36% vs 19.77%). The prevalence of arthritis was 43% in male and 54% in female respondents in a recent national survey of older adults (65 and older) with self-reported chronic medical conditions in 2013 to 2014 [3]. The numbers were nearly twice the prevalence of such a condition in the inpatient clinical data reported in Table 2. This likely reflects the fact that people with arthritis were mostly treated in outpatient settings and thus diagnosis of arthritis is irrelevant to most inpatient discharges.

Prevalence of Hypertension by Gender, Age Group, and Race or Ethnicity

Table 3 shows the prevalence of hypertension in older adults stratified by sex, age group, race and ethnicity in the NIS 2013, the OneFlorida 2013, and the pooled OneFlorida 2012 to 2016 data. Hypertension was chosen because it was the condition with the highest prevalence among the older persons we studied. The largest differences in the estimates between the 2013 files (OneFlorida and NIS) was about 3% for NH white females aged 85 years and older, and NH white males aged 65 to 74 years. We observed differences of more than 1% for females in the following 4 strata—NH black aged 65 to 74 years, NH white aged 65 to 74 years, NH white aged 75 to 84 years, and NH white aged 85 years and older. Among males, differences of greater than 1% were observed for the strata except for NH white aged 75 to 84 years. Estimates between OneFlorida 2013 and the pooled OneFlorida 2012 to 2016 data were largely similar with some increases in OneFlorida 2012 to 2016 data for hypertension prevalence, perhaps reflecting the increasing trends associated with obesity and sedentary life styles.

Prevalence of Multiple Chronic Conditions

Figure 2 illustrates the percentage of the population with one or more CCs, which is, MCCs in older adults in the HCUP NIS and OneFlorida for 2013. The 3 datasets exhibited similar characteristics. Out of the 25 CCs, more than 18% older adults had 4 conditions. More than 65% older adults had 4 or more conditions. Persons with MMCs were very prevalent among older Americans.

Prevalence of Multiple Chronic Conditions by Gender

Figure 3 illustrates the prevalence of MCC stratified by sex. With respect to the number of MCCs, male and female older adults did not exhibit notable difference in both the OneFlorida and NIS data. No statistical tests were performed to test the statistical difference among the groups. This contrasted with the population aged 18 to 64 years in which women had a higher prevalence of MCCs.

Prevalence of Multiple Chronic Conditions by Race or Ethnicity

Figure 4 illustrates the prevalence of MCCs by race or ethnicity. It appears that the distribution of records with one or more CCs were similar among race or ethnicity groups. Note that even though Hispanic was overrepresented and Asian was underrepresented in OneFlorida, their MCC distribution within each race or ethnicity was similar to the NIS.

Table 1. Descriptive statistics for the National Inpatient Sample and OneFlorida patient datasets of older adults. HCUP NIS: Healthcare Cost and Utilization Project National Inpatient Sample.

Characteristics	OneFlorida 2013 (N=40,087)	OneFlorida 2012-2016 (N=147,900)	HCUP NIS 2013 (N=2,447,640)
Age in years, mean (SD)	76.4 (8.04)	76.4 (8.03)	78.0 (7.80)
Gender, n (%)			
Male	19,094 (47.63)	71,642 (48.44)	1,084,593 (44.31)
Female	20,993 (52.37)	76,255 (51.56)	1,362,844 (55.68)
Unspecified	0 (0)	3 (0.0)	203 (0.01)
Ethnicity, n (%)			
Non-Hispanic white	27,881 (69.55)	101,871 (68.88)	1,817,861 (74.27)
Non-Hispanic black	5835 (14.56)	19,487 (13.18)	231,968 (9.48)
Asian and Pacific Islander	550 (1.37)	1819 (1.23)	50,768 (2.07)
Hispanic	3102 (7.74)	11,718 (7.92)	156,780 (6.41)
Other	2719 (6.78)	13,005 (8.79)	190,263 (7.77)
Average number of multiple chronic conditions, (SD)	4.7 (2.3)	4.9 (2.6)	4.5 (2.0)

Table 2. Prevalence of the 25 chronic conditions in the 2013 National Inpatient Sample, the 2013 OneFlorida, and the pooled 2012 to 2016 OneFlorida data. HCUP NIS: Healthcare Cost and Utilization Project National Inpatient Sample.

Condition	Number of patients, n (%)		
	OneFlorida 2013 (N=40,087)	OneFlorida 2012-2016 (N=147,900)	HCUP NIS 2013 (N=2,447,640)
Hypertension	32,460 (80.97)	123,640 (83.60)	1,868,149 (76.32)
Hyperlipidemia	21,013 (52.42)	82,046 (55.48)	1,124,402 (45.94)
Ischemic heart disease	15,191 (37.90)	57,235 (38.69)	911,199 (37.23)
Diabetes	14,158 (35.32)	53,362 (36.07)	830,551 (33.93)
Anemia	14,445 (36.03)	57,108 (38.61)	819,538 (33.48)
Chronic kidney disease	12,525 (31.24)	49,957 (33.78)	813,196 (33.22)
Atrial fibrillation	9973 (24.88)	38,347 (25.93)	625,467 (25.55)
Heart failure	7926 (19.77)	31,411 (21.23)	620,787 (25.36)
Chronic obstructive pulmonary disease and bronchiectasis	8063 (20.11)	31,658 (21.41)	559,336 (22.85)
Rheumatoid arthritis or osteoarthritis	9325 (23.26)	41,348 (27.96)	481,299 (19.66)
Acquired hypothyroidism	7622 (19.01)	29,731 (20.10)	45,6024 (18.63)
Alzheimer disease and related disorders or senile dementia	5351 (13.35)	21,349 (14.43)	370,502 (15.14)
Depression	5643 (14.08)	22,554 (15.25)	323,717 (13.23)
Osteoporosis	2677 (6.68)	10,526 (7.12)	161,620 (6.60)
Asthma	3071 (7.66)	17,136 (11.59)	152,557 (6.23)
Stroke or transient ischemic attack	3040 (7.58)	13,288 (8.98)	117,165 (4.79)
Acute myocardial infarction	1799 (4.48)	7946 (5.37)	107,079 (4.37)
Prostate cancer	2516 (6.27)	10,146 (6.86)	103,151 (4.21)
Breast cancer	1876 (4.68)	7229 (4.89)	99,430 (4.06)
Alzheimer disease[a]	1202 (3.00)	4574 (3.09)	89,683 (3.66)
Colorectal cancer	1496 (3.73)	5398 (3.65)	77,409 (3.16)
Lung cancer	1376 (3.43)	5408 (3.66)	75,982 (3.10)
Hip or pelvic fracture	1127 (2.81)	5153 (3.48)	69,693 (2.85)
Benign prostatic hyperplasia	919 (2.39)	6297 (4.26)	47,979 (1.96)
Endometrial cancer	361 (0.90)	1419 (0.96)	15,173 (0.62)

[a]The case counts for persons with Alzheimer disease are also included in the counts for the Alzheimer disease and related disorders or senile dementia category.

Table 3. Prevalence of hypertension stratified by gender-age-racial or ethnic groups in the 2013 Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS), the 2013 OneFlorida, and the pooled 2012 to 2016 OneFlorida data.

Sex	Age range in years	Race or ethnicity	Number of patients, n (%[a])		
			OneFlorida 2013 (N=40,087)	OneFlorida 2012-2016 (N=147,900)	HCUP NIS 2013 (N=2,447,640)
Female	65-74	Asian and Pacific Islander	115 (0.29)	381 (0.26)	7487 (0.31)
Female	65-74	Hispanic	704 (1.76)	2696 (1.82)	28,310 (1.16)
Female	65-74	Non-Hispanic black	1516 (3.78)	5280 (3.57)	53,625 (2.19)
Female	65-74	Non-Hispanic white	4931 (12.30)	18,653 (12.61)	258,444 (10.56)
Female	75-84	Asian and Pacific Islander	84 (0.21)	308 (0.21)	8,888 (0.36)
Female	75-84	Hispanic	518 (1.29)	1949 (1.32)	27,484 (1.12)
Female	75-84	Non-Hispanic black	995 (2.48)	3464 (2.34)	42,002 (1.72)
Female	75-84	Non-Hispanic white	3871 (9.66)	14,522 (9.82)	275,781 (11.27)
Female	≥85	Asian and Pacific Islander	36 (0.09)	112 (0.08)	6295 (0.26)
Female	≥85	Hispanic	216 (0.54)	843 (0.57)	16,218 (0.66)
Female	≥85	Non-Hispanic black	574 (1.43)	1797 (1.22)	24,602 (1.01)
Female	≥85	Non-Hispanic white	2465 (6.15)	8765 (5.93)	226,926 (9.27)
Male	65-74	Asian and Pacific Islander	136 (0.34)	417 (0.28)	7355 (0.30)
Male	65-74	Hispanic	597 (1.49)	2428 (1.64)	25,393 (1.04)
Male	65-74	Non-Hispanic black	1240 (3.09)	4591 (3.10)	42,773 (1.75)
Male	65-74	Non-Hispanic white	5697 (14.21)	21,889 (14.80)	258,218 (10.55)
Male	75-84	Asian and Pacific Islander	59 (0.15)	245 (0.17)	6966 (0.28)
Male	75-84	Hispanic	417 (1.04)	1595 (1.08)	20,110 (0.82)
Male	75-84	Non-Hispanic black	710 (1.77)	2441 (1.65)	26,263 (1.07)
Male	75-84	Non-Hispanic white	3677 (9.17)	13,854 (9.37)	221,772 (9.06)
Male	≥85	Asian and Pacific Islander	20 (0.05)	78 (0.05)	3908 (0.16)
Male	≥85	Hispanic	133 (0.33)	560 (0.04)	8817 (0.36)
Male	≥85	Non-Hispanic black	234 (0.58)	815 (0.55)	9913 (0.41)
Male	≥85	Non-Hispanic white	1375 (3.43)	5449 (3.68)	120,645 (4.93)

[a]The denominator is all the patients ≥65 years old.

Figure 2. Number of conditions in older adults in the Healthcare Cost and Utilization Project National Inpatient Sample (NIS) 2013 and OneFlorida2013 data.

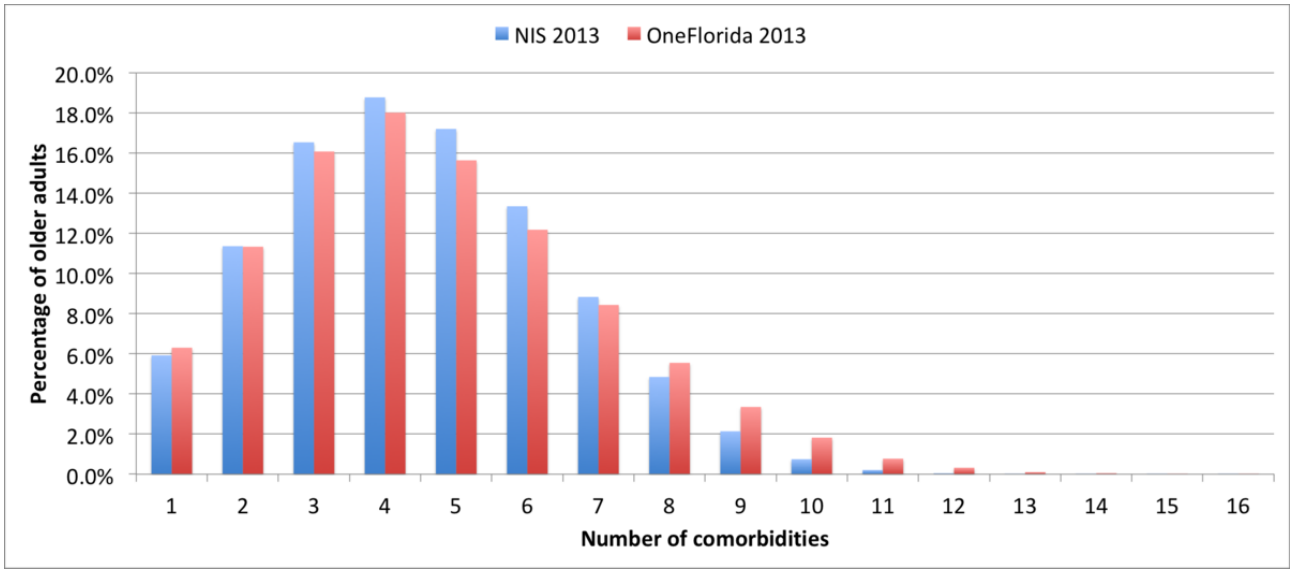


Figure 3. Prevalence of multiple chronic conditions in older adults by gender in Healthcare Cost and Utilization Project National Inpatient Sample (NIS) 2013 and OneFlorida 2013.

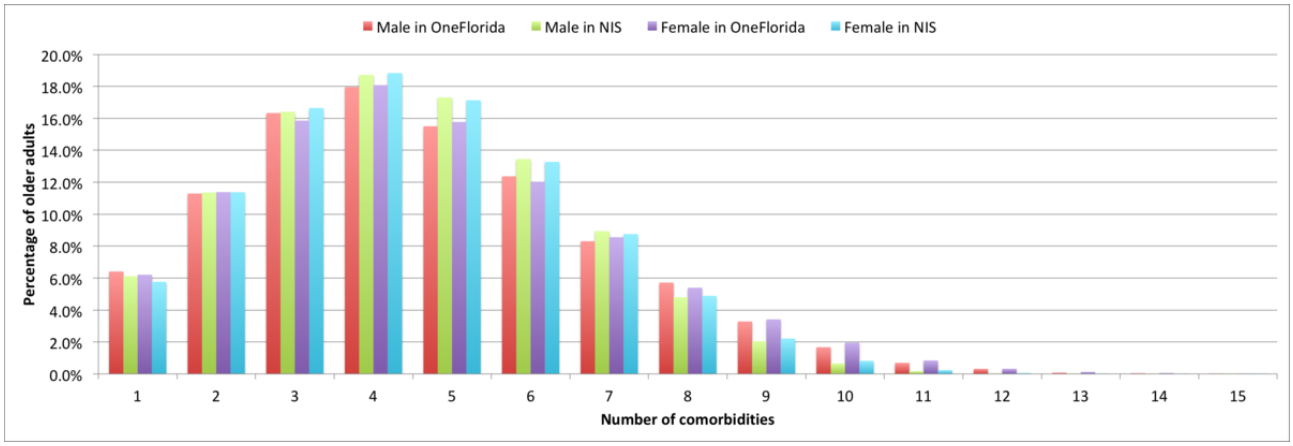
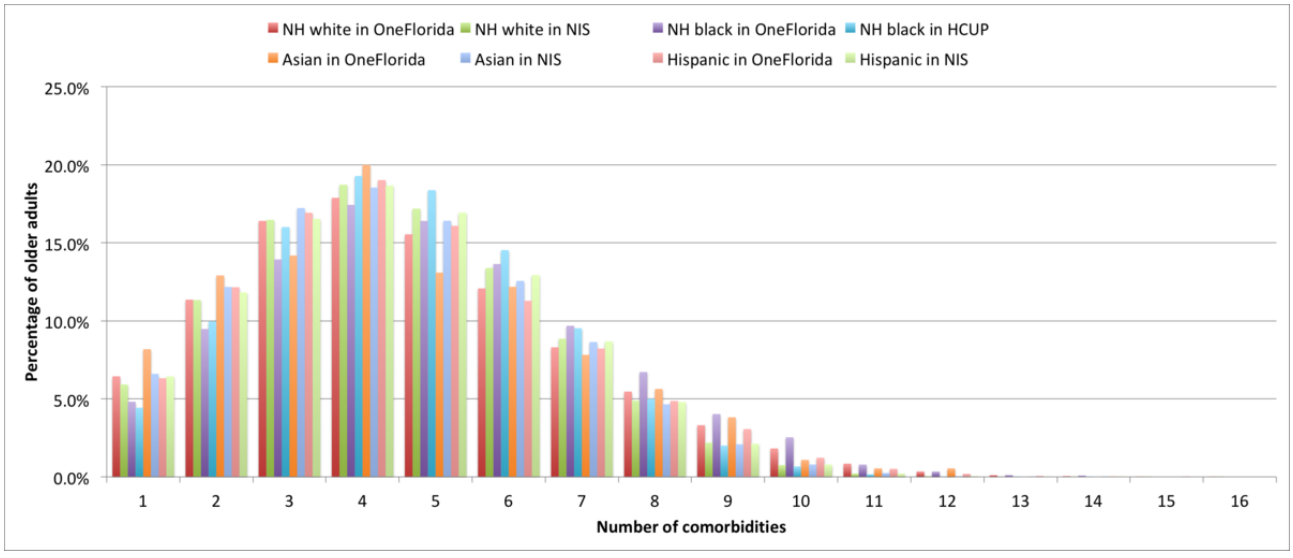


Figure 4. Prevalence of multiple chronic conditions in older adults by race and ethnicity groups in Healthcare Cost and Utilization Project National Inpatient Sample (NIS) 2013 and OneFlorida 2013 data. NH: non-Hispanic.



Prevalence of Multiple Chronic Conditions by Pairs of Conditions

Table 4 shows the prevalence of the 10 most common pairs of co-occurring chronic conditions. Even though the prevalent MCCs of 2 conditions were the same in both OneFlorida and NIS cohorts, their rankings were slightly different. OneFlorida cohort had a higher percentage of patients with anemia and hypertension than the NIS cohort (32.17% vs 25.79%). OneFlorida cohort had a slightly higher percentage of older adults with atrial fibrillation and hypertension than the NIS cohort (22.72% vs 19.88%). The most prevalent MCCs of 2 conditions among the 3 datasets were—hyperlipidemia and hypertension, hypertension and ischemic heart disease, diabetes and hypertension, chronic kidney disease and hypertension, anemia and hypertension, and hyperlipidemia and ischemic heart disease.

Trends in Multiple Chronic Conditions

The following 4 figures (Figures 5-8) present a longitudinal examination of the number of discharges reflecting 2 or more conditions for the period 2012 to 2016 for the OneFlorida data and 2003 to 2013 for the NIS data. In Figure 5, the overall prevalence of 2 or more CCs raised steadily from approximately 66% in 2003 to approximately 83% in 2013 in the NIS data. The OneFlorida data began in 2012 at approximately 81% prevalence of MCC and rose to approximately 84% by 2016. Both slopes showed a monotonic increasing trend in the prevalence of MCCs.

The slope of the MCC prevalence by gender in Figure 6 appeared to be very similar to the overall slope in Figure 5. The slopes for males and females in the NIS data were parallel with 1% to 2% difference for males and females and ultimately converged at approximately 84% by 2013. Pearson correlation coefficient showed a strong positive correlation between male and female older adults with an *R* value of .9966. The lines for OneFlorida data for males and females were nearly coincident and appeared to continue the slope of the NIS data.

In Figure 7, the prevalence of MCC by age group is presented for NIS and OneFlorida data. The NIS slopes for the 3 age groups were parallel through 2013. Pearson correlation coefficient showed a strong positive correlation among the 3 age groups—the *R* value between NIS 65 to 74 years age group and 75 to 84 years age group was .9972; the *R* value between NIS 65 to 74 years age group and NIS over 85 years age group was .9961. Nevertheless, there was about an 8-percentage point difference between the youngest age group (65-74 years) and the middle age group (75-84 years). The oldest age group (over 85 years) appeared to be about 4 percentage points higher than the 75 to 84 years age group throughout the time range. Similar differences were seen between the parallel slopes for OneFlorida data, although the 85 years and over group was trending somewhat higher as compared with the same age group in the NIS.

Finally, in Figure 8, we present the prevalence of MCCs by racial-ethnic groups. The general trend was the same as seen in Figures 5-7. The non-Hispanic black and non-Hispanic white groups ran parallel with the black group averaging about 2 percentage points higher. The Pearson correlation coefficient showed a strong positive correlation between non-Hispanic black and non-Hispanic white groups with an *R* value of .9959. The Hispanic group and the Asian and Pacific Islander group, both averaged a bit lower than the non-Hispanic white population, but there was more volatility probably due to smaller sample size. This was particularly true for the OneFlorida data.

Table 4. Prevalence of the 10 most common pairs of co-occurring chronic conditions. HCUP NIS: Healthcare Cost and Utilization Project National Inpatient Sample.

Condition A	Condition B	Number of patients, n (%)		
		OneFlorida 2013 (N=40,087)	OneFlorida 2012-2016 (N=147,900)	HCUP NIS 2013 (N=2,447,640)
Hyperlipidemia	Hypertension	19,452 (48.52)	73,732 (49.85)	960,388 (39.23)
Hypertension	Ischemic heart disease	16,672 (41.58)	56,945 (38.50)	745,865 (30.47)
Diabetes	Hypertension	13,410 (33.45)	49,079 (33.18)	698,256 (28.53)
Chronic kidney disease	Hypertension	11,727 (29.25)	45,482 (30.75)	671,397 (27.43)
Anemia	Hypertension	12,898 (32.17)	49,566 (33.51)	631,247 (25.79)
Hyperlipidemia	Ischemic heart disease	12,041 (30.04)	43,100 (29.14)	530,768 (21.68)
Heart failure	Hypertension	7452 (18.59)	28,625 (19.35)	490,243 (20.03)
Atrial fibrillation	Hypertension	9109 (22.72)	33,851 (22.89)	486,609 (19.88)
Diabetes	Hyperlipidemia	10,738 (26.79)	38,023 (25.71)	449,597 (18.37)
Hypertension	Chronic obstructive pulmonary disease and bronchiectasis	7180 (17.91)	27,285 (18.45)	414,983 (16.95)

Figure 5. Trends of multiple chronic conditions in older adults for the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS; 2003-2013) and OneFlorida (2012-2016). The denominator in the prevalence is the total number of older adults with at least one of the 25 conditions in each year.

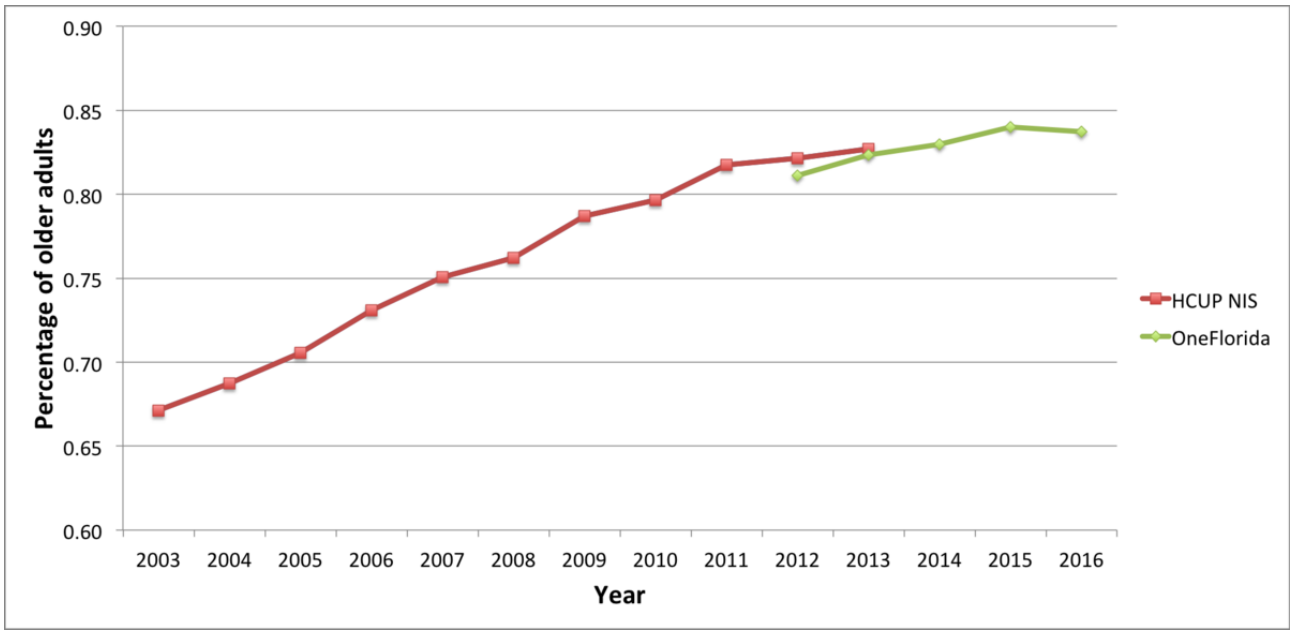


Figure 6. Trends of multiple chronic conditions in older adults by gender for the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS) (2003-2013) and OneFlorida (2012-2016). The denominator in the prevalence is the total number of older male or female with at least one of the 25 conditions in each year.

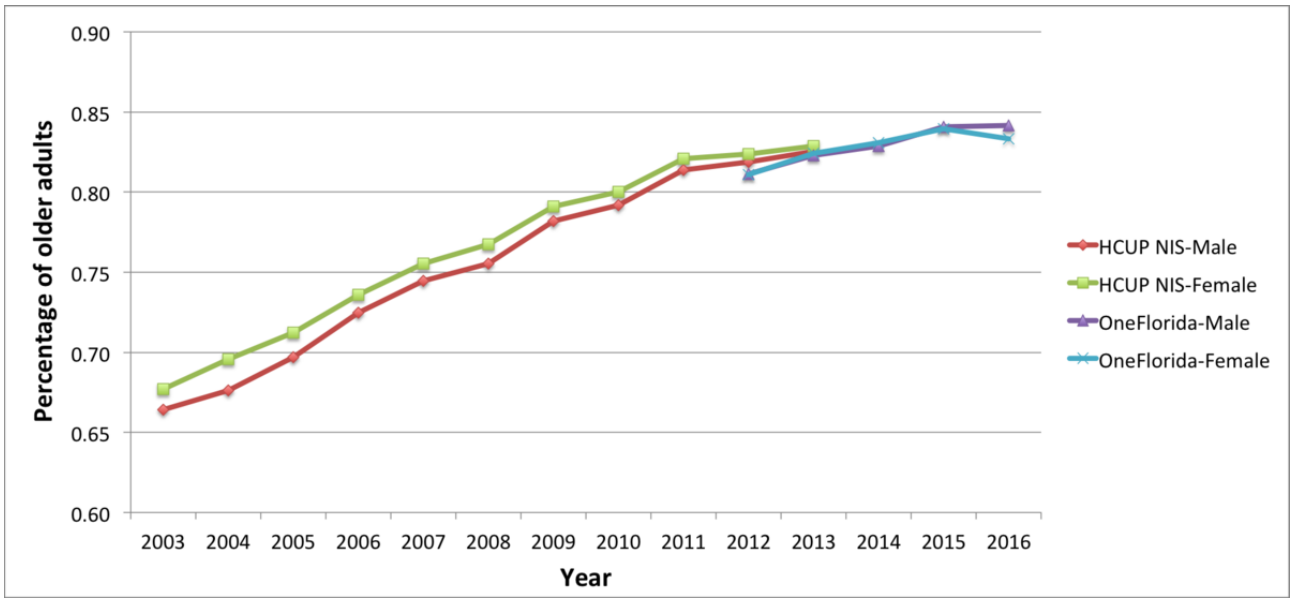


Figure 7. Trends of multiple chronic conditions in older adults by age group for the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS) (2003-2013) and OneFlorida (2012-2016). The denominator in the prevalence is the total number of older adults in each age group with at least one of the 25 conditions in each year.

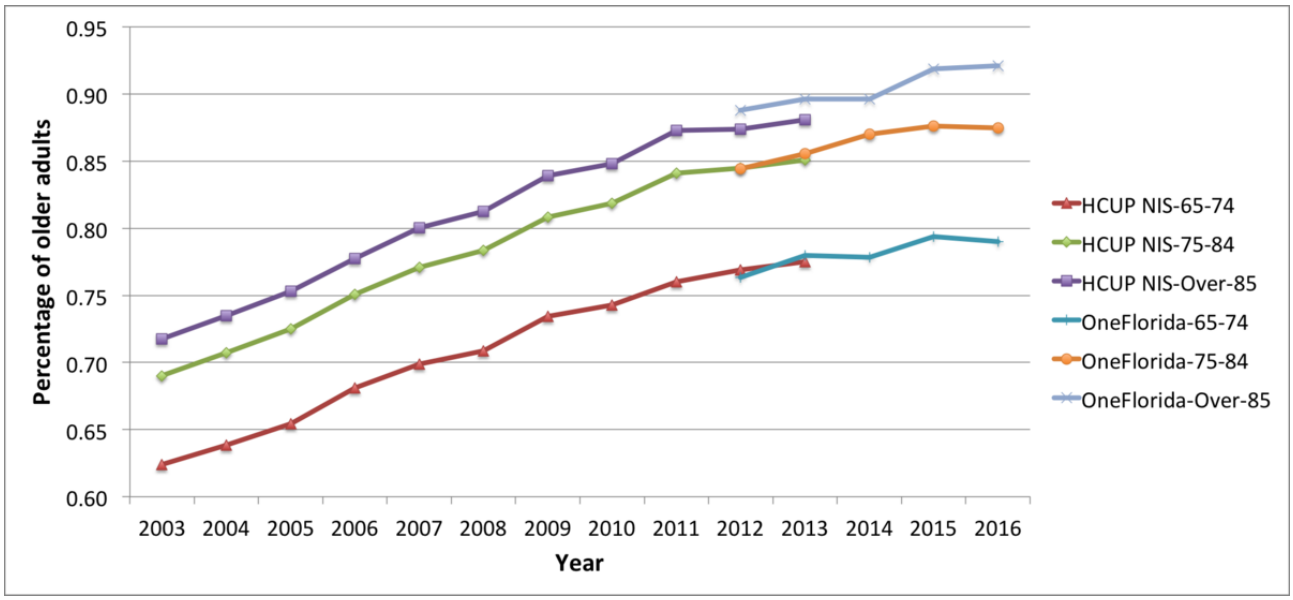
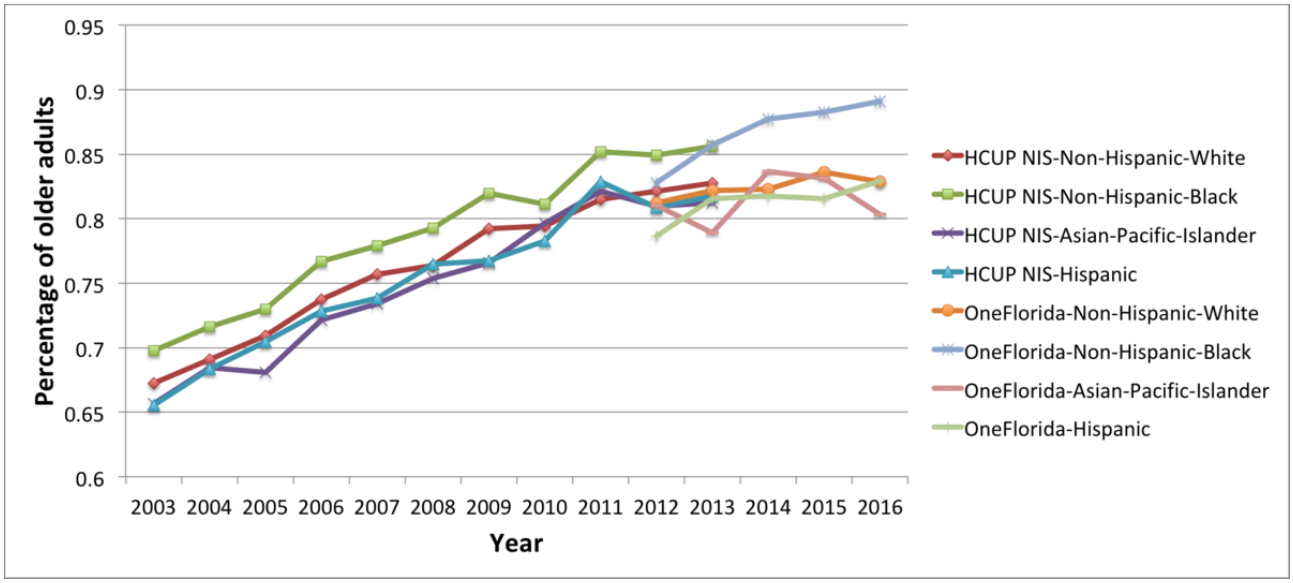


Figure 8. Trends of multiple chronic conditions in older adults by racial-ethnic groups for the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS; 2003-2013) and OneFlorida (2012-2016). The denominator in the prevalence is the total number of older adults in each racial-ethnic group with at least one of the 25 conditions in each year.



Discussion

Principal Findings

The main objective of our study was to compare the prevalence of common CCs and MCCs in older adults in the Florida and US national population using the OneFlorida Data Trust and the NIS of HCUP. The results showed that CCs and MCCs were prevalent in older adults, both nationally and in the Florida population. The most prevalent CCs were the same for older adults in the OneFlorida Data Trust and HCUP NIS. For hypertension, the largest differences in the estimates between the 2013 OneFlorida Data Trust and NIS were about merely 3% for non-Hispanic white females 85 years and older and males 65 to 74 years old. Regarding the number of MCCs, OneFlorida Data Trust and NIS did not exhibit any notable difference with respect to gender and race or ethnicity. The most prevalent MCCs of 2 CCs were also the same for OneFlorida 2013, NIS 2013, and OneFlorida 2012 to 2016. With regard to the MCC trends, the slopes of the increasing trend in the number of discharges reflecting 2 or more conditions appeared quite similar in both data sources. With respect to age group, the oldest age group (over 85 years of age) appeared to be about 4 percentage points higher than the 75 to 84 years age group and 12 percentage points higher than the 65 to 75 years age group throughout the time range. Even though slight differences were observed, similar estimates of prevalence of CCs and MCCs across OneFlorida Data Trust and NIS showed that large clinical research networks such as OneFlorida provide rich data resources for conducting large-scale secondary data analyses.

Although the MCC prevalence presented in this study is generalizable to the older US adults in the noninstitutionalized national population, the use of OneFlorida Data Trust and the HCUP NIS has limitations. OneFlorida and NIS both only captured the conditions that were confirmed by a doctor or health professionals in inpatient settings, potentially leading to the underrepresentation of conditions that remain undiagnosed or were not recorded in the inpatient care (eg, arthritis [3]). Many uninsured adults would not get into these databases until 65 years of age when they become eligible for Medicare. Undocumented immigrants would never make it into Medicare. For example, the prevalence of arthritis reported in a self-reported national survey almost doubles the prevalence of arthritis in the inpatient clinical data reported in Table 2. Of the conditions captured, we only used the CCW algorithm from the Centers for Medicare and Medicaid Services (CMS) and considered a single occurrence of the diagnosis code of a particular condition when identifying patients who had such a condition. There might be false positive cases included in the analysis. Furthermore, although the OneFlorida Clinical Research Consortium [41] covers care for approximately 48% Floridians, the consortium is missing representations from a few of the key health care markets in Florida, such as Tampa, and cities in the Florida panhandle area. Moreover, the prevalence of CCs might be overestimated for Florida, as there might be duplicated patient records across the different health care organizations in the OneFlorida consortium. For example, EHRs from health care providers and claims data from payers can have records for the same patient. In addition, the same patient can seek care in different health care organizations in the network. Thus, linking related data and resolving duplicates in a clinical research network is a significant task in improving the quality of a dataset. In our recent effort, we have linked and deduplicated patient records across 2 of the data sources in the OneFlorida consortium—University of Florida Health system and Florida Medicaid. We eliminated 430,106 duplicate patient records across these 2 sources, which is approximately 6.4% of the Florida Medicaid population.

Our study confirmed the previous literature [5] and showed the increasing trend in the prevalence of MCCs among the older US adults. We also showed that the characteristics of the patient population in these clinical research networks such as OneFlorida are comparable to national-level sample data. Furthermore, these clinical research networks have integrated fine-grained details of the patients (eg, encounters, procedures, diagnoses, medications, lab results, as well as patient-reported outcomes) from multiple health care organizations, which can provide a more complete picture of the patients' health traits. Enabled by clinical research networks such as OneFlorida, large-scale secondary data analyses can be conducted to discover novel findings in biomedical research, such as sophisticated relationships among diseases, medications, vital signs, adverse events, and outcomes.

Implication and Future Directions

The OneFlorida Data Trust is the informatics infrastructure that supports pragmatic trials, comparative effectiveness research, implementation science, and other research in the OneFlorida Clinical Research Consortium. The most key research functions supported by OneFlorida and PCORnet include cohort discovery and participant enrollment, recognizing the barriers in identifying and recruiting research participants for clinical research studies, especially for rare diseases. Furthermore, the population representativeness of clinical research has long been a concern [42]. Particularly, older adults are widely reported to be underrepresented in clinical studies across major medical conditions such as cardiovascular diseases [43,44], cancer [45,46], dementia [47], and diabetes [48,49]. Due to the lack of evidence in the clinical practice guideline in treating older adults with MCCs, it is imperative to generate such evidence by involving older adults with normal age-related organ impairment and comorbid conditions that may not interact with the treatment under study. However, older adults are often unfairly excluded by restrictive eligibility criteria in clinical studies [46,50]. Meanwhile, MCCs are most prevalent in the Medicare population. Persons with MCCs are at an increased risk of mortality, morbidity, hospitalization, high medical costs, and adverse events [51]. In order to understand how older adults with MCCs are represented in clinical trials, it is important to understand the prevalence of MCCs in older adults. In future work, we will use laboratory test results and medications to enhance the sensitivity and specificity of case assignment for some conditions. We will also compare the outpatient data of OneFlorida Data Trust with the national outpatient databases such as the Nationwide Emergency Department Sample.

Acknowledgments

This work was supported by the Institute for Successful Longevity of the Florida State University. The work was also supported in part by the OneFlorida Clinical Data Network, funded by PCORI #CDRN-1501-26692, in part by the OneFlorida Cancer Control Alliance, funded by the Florida Department of Health's James and Esther King Biomedical Research Program #4KB16, and in part by the University of Florida Clinical and Translational Science Institute, which is supported in part by the NIH National Center for Advancing Translational Sciences under award number UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of PCORI, its Board of Governors or Methodology, the OneFlorida Clinical Research Consortium, the University of Florida's Clinical and Translational Science Institute, the Florida Department of Health, or the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. Health Qual Life Outcomes 2009 Sep 08;7:82 [FREE Full text] [doi: 10.1186/1477-7525-7-82] [Medline: 19737412]
2. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med 2007 Dec;22(Suppl 3):391-395 [FREE Full text] [doi: 10.1007/s11606-007-0322-1] [Medline: 18026807]
3. Federal Interagency Forum on Aging-Related Statistics. Older Americans 2016 - Key Indicators of Well-Being Federal Interagency Forum on Aging-Related Statistics 2016 URL: https://agingstats.gov/docs/LatestReport/Older-Americans-2016-Key-Indicators-of-WellBeing.pdf [WebCite Cache ID 6tSAiX6FX]
4. Freid VM, Bernstein AB, Bush MA. Multiple chronic conditions among adults aged 45 and over: trends over the past 10 years. NCHS Data Brief 2012 Jul(100):1-8 [FREE Full text] [Medline: 23101759]
5. Ward BW, Schiller JS. Prevalence of multiple chronic conditions among US adults: estimates from the National Health Interview Survey, 2010. Prev Chronic Dis 2013 Apr 25;10:E65 [FREE Full text] [doi: 10.5888/pcd10.120203] [Medline: 23618545]
6. Ashman JJ, Beresovsky V. Multiple chronic conditions among US adults who visited physician offices: data from the National Ambulatory Medical Care Survey, 2009. Prev Chronic Dis 2013 Apr 25;10:E64 [FREE Full text] [doi: 10.5888/pcd10.120308] [Medline: 23618544]
7. He Z, Charness N, Bian J, Hogan WR. Assessing the comorbidity gap between clinical studies and prevalence in elderly patient populations. IEEE EMBS Int Conf Biomed Health Inform 2016 Feb;2016:136-139 [FREE Full text] [doi: 10.1109/BHI.2016.7455853] [Medline: 27738664]
8. Adler-Milstein J, Jha AK. HITECH Act drove large gains in hospital electronic health record adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]
9. Henry J, Pylypchuk Y, Searcy T, Patel V. The Office of the National Coordinator for Health Information Technology. 2016. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015 URL: https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php [WebCite Cache ID 6wcfwocuZ]
10. PracticeFusion. 2017. HR adoption rates: 20 must-see stats URL: https://www.practicefusion.com/blog/ehr-adoption-rates/ [WebCite Cache ID 6wcfzxBpe]
11. Paul MM, Greene CM, Newton-Dame R, Thorpe LE, Perlman SE, McVeigh KH, et al. The state of population health surveillance using electronic health records: a narrative review. Popul Health Manag 2015 Jun;18(3):209-216. [doi: 10.1089/pop.2014.0093] [Medline: 25608033]
12. Klompas M, Cocoros NM, Menchaca JT, Erani D, Hafer E, Herrick B, et al. State and local chronic disease surveillance using electronic health record systems. Am J Public Health 2017 Sep;107(9):1406-1412. [doi: 10.2105/AJPH.2017.303874] [Medline: 28727539]
13. Birkhead GS. Successes and continued challenges of electronic health records for chronic disease surveillance. Am J Public Health 2017 Sep;107(9):1365-1367. [doi: 10.2105/AJPH.2017.303938] [Medline: 28787206]
14. Perlman SE, McVeigh KH, Thorpe LE, Jacobson L, Greene CM, Gwynn RC. Innovations in population health surveillance: using electronic health records for chronic disease surveillance. Am J Public Health 2017 Jun;107(6):853-857. [doi: 10.2105/AJPH.2017.303813] [Medline: 28426302]
15. Weng C, Batres C, Borda T, Weiskopf NG, Wilcox AB, Bigger JT, et al. A real-time screening alert improves patient recruitment efficiency. AMIA Annu Symp Proc 2011;2011:1489-1498 [FREE Full text] [Medline: 22195213]
16. Boland MR, Shahn Z, Madigan D, Hripcsak G, Tatonetti N. Birth month affects lifetime disease risk: a phenome-wide method. J Am Med Inform Assoc 2015 Sep;22(5):1042-1053 [FREE Full text] [doi: 10.1093/jamia/ocv046] [Medline: 26041386]
17. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
18. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. Health Aff (Millwood) 2014 Jul;33(7):1178-1186. [doi: 10.1377/hlthaff.2014.0121] [Medline: 25006144]
19. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. J Am Med Inform Assoc 2014;21(4):576-577 [FREE Full text] [doi: 10.1136/amiajnl-2014-002864] [Medline: 24821744]
20. Corley DA, Feigelson HS, Lieu TA, McGlynn EA. Building data infrastructure to evaluate and improve quality: PCORnet. J Oncol Pract 2015 May;11(3):204-206 [FREE Full text] [doi: 10.1200/JOP.2014.003194] [Medline: 25980016]
21. Fleurence RL, Beal AC, Sheridan SE, Johnson LB, Selby JV. Patient-powered research networks aim to improve patient care and health research. Health Aff (Millwood) 2014 Jul;33(7):1212-1219. [doi: 10.1377/hlthaff.2014.0113] [Medline: 25006148]
22. pcornet. PCORnet, the National Patient-Centered Clinical Research Network URL: http://pcornet.org/ [accessed 2017-09-13] [WebCite Cache ID 6tSAuuOkt]

23. Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: shining light on a new research paradigm. Ann Intern Med 2015 Oct 20;163(8):635-636. [doi: 10.7326/M15-1460] [Medline: 26301537]
24. Lin KJ, Singer DE, Glynn RJ, Murphy SN, Lii J, Schneeweiss S. Identifying patients with high data completeness to improve validity of comparative effectiveness research in electronic health records data. Clin Pharmacol Ther 2018:-(forthcoming)(forthcoming). [doi: 10.1002/cpt.861] [Medline: 28865143]
25. Reams C, Powell M, Edwards R. State synergies and disease surveillance: creating an electronic health data communication model for cancer reporting and comparative effectiveness research in kentucky. EGEMS (Wash DC) 2014;2(2):1064 [FREE Full text] [doi: 10.13063/2327-9214.1064] [Medline: 25848604]
26. Fiks AG, Grundmeier RW, Steffes J, Adams WG, Kaelber DC, Pace WD, Comparative Effectiveness Research Through Collaborative Electronic Reporting (CER2) Consortium. Comparative effectiveness research through a collaborative electronic reporting consortium. Pediatrics 2015 Jul;136(1):e215-e224 [FREE Full text] [doi: 10.1542/peds.2015-0673] [Medline: 26101357]
27. Mo H, Thompson WK, Rasmussen LV, Pacheco JA, Jiang G, Kiefer R, et al. Desiderata for computable representations of electronic health records-driven phenotype algorithms. J Am Med Inform Assoc 2015 Nov;22(6):1220-1230 [FREE Full text] [doi: 10.1093/jamia/ocv112] [Medline: 26342218]
28. Vassy JL, Ho Y, Honerlaw J, Cho K, Gaziano JM, Wilson PW, et al. Yield and bias in defining a cohort study baseline from electronic health record data. J Biomed Inform 2018 Feb;78:54-59. [doi: 10.1016/j.jbi.2017.12.017] [Medline: 29305952]
29. Abrahão MT, Nobre MR, Gutierrez MA. A method for cohort selection of cardiovascular disease records from an electronic health record system. Int J Med Inform 2017 Dec;102:138-149. [doi: 10.1016/j.ijmedinf.2017.03.015] [Medline: 28495342]
30. Sidebottom AC, Johnson PJ, VanWormer JJ, Sillah A, Winden TJ, Boucher JL. Exploring electronic health records as a population health surveillance tool of cardiovascular disease risk factors. Popul Health Manag 2015 Apr;18(2):79-85. [doi: 10.1089/pop.2014.0058] [Medline: 25290223]
31. McVeigh KH, Newton-Dame R, Chan PY, Thorpe LE, Schreibstein L, Tatem KS, et al. Can electronic health records be used for population health surveillance? Validating population health metrics against established survey data. EGEMS (Wash DC) 2016;4(1):1267 [FREE Full text] [doi: 10.13063/2327-9214.1267] [Medline: 28154837]
32. OneFlorida. 2016. OneFlorida: Clinical Research Consortium URL: http://onefloridaconsortium.org [WebCite Cache ID 6tSB5DVH7]
33. PCORnet. 2016. PCORnet Common Data Model (CDM) URL: http://www.pcornet.org/pcornet-common-data-model [WebCite Cache ID 6tSBHS2Cg]
34. Agency for Healthcare Research and Quality. 2017. Overview of the National (Nationwide) Inpatient Sample (NIS) URL: https://www.hcup-us.ahrq.gov/nisoverview.jsp [WebCite Cache ID 6tSQzExbI]
35. Centers for Medicare & Medicaid Services. 2016. CCW Chronic Condition Algorithms in the Chronic Conditions Data Warehouse URL: https://www.ccwdata.org/web/guest/condition-categories [WebCite Cache ID 6tSQ2dLX6]
36. Ehrenpreis ED, Zhou Y. Hospital costs, length of stay and prevalence of hip and knee arthroplasty in patients with inflammatory bowel disease. World J Gastroenterol 2017 Jul 14;23(26):4752-4758 [FREE Full text] [doi: 10.3748/wjg.v23.i26.4752] [Medline: 28765696]
37. Wilson D, Jin DL, Wen T, Carmichael JD, Cen S, Mack WJ, et al. Demographic factors, outcomes, and patient access to transsphenoidal surgery for Cushing's disease: analysis of the Nationwide Inpatient Sample from 2002 to 2010. Neurosurg Focus 2015 Feb;38(2):E2. [doi: 10.3171/2014.11.FOCUS14694] [Medline: 25639320]
38. Khorgami Z, Aminian A, Shoar S, Andalib A, Saber AA, Schauer PR, et al. Cost of bariatric surgery and factors associated with increased cost: an analysis of national inpatient sample. Surg Obes Relat Dis 2017 Aug;13(8):1284-1289. [doi: 10.1016/j.soard.2017.04.010] [Medline: 28583812]
39. Spector WD, Limcangco R, Furukawa MF, Encinosa WE. The marginal costs of adverse drug events associated with exposures to anticoagulants and hypoglycemic agents during hospitalization. Med Care 2017 Sep;55(9):856-863. [doi: 10.1097/MLR.0000000000000780] [Medline: 28742544]
40. Balla A, Batista RG, Corradetti S, Balagué C, Fernández-Ananín S, Targarona EM. Outcomes after bariatric surgery according to large databases: a systematic review. Langenbecks Arch Surg 2017 Sep 01;402(6):885-899. [doi: 10.1007/s00423-017-1613-6] [Medline: 28780622]
41. OneFlorida. 2017. Partners of OneFlorida Clinical Research Consortium URL: https://www.ctsi.ufl.edu/research-initiatives/projects/oneflorida [WebCite Cache ID 6tSBRofgb]
42. Leaf C. The New York Times. 2013. Do Clinical Trials Work? URL: http://www.nytimes.com/2013/07/14/opinion/sunday/do-clinical-trials-work.html [WebCite Cache ID 6xuxP6i9X]
43. Lee PY, Alexander KP, Hammill BG, Pasquali SK, Peterson ED. Representation of elderly persons and women in published randomized trials of acute coronary syndromes. J Am Med Assoc 2001 Aug 08;286(6):708-713. [Medline: 11495621]
44. Sardar MR, Badri M, Prince CT, Seltzer J, Kowey PR. Underrepresentation of women, elderly patients, and racial minorities in the randomized trials used for cardiovascular guidelines. JAMA Intern Med 2014 Nov;174(11):1868-1870. [doi: 10.1001/jamainternmed.2014.4758] [Medline: 25264856]

45. Hutchins LF, Unger JM, Crowley JJ, Coltman CA, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. N Engl J Med 1999 Dec 30;341(27):2061-2067. [doi: 10.1056/NEJM199912303412706] [Medline: 10615079]
46. Lewis JH, Kilgore ML, Goldman DP, Trimble EL, Kaplan R, Montello MJ, et al. Participation of patients 65 years of age or older in cancer clinical trials. J Clin Oncol 2003 Apr 01;21(7):1383-1389. [doi: 10.1200/JCO.2003.08.010] [Medline: 12663731]
47. Schoenmaker N, Van Gool WA. The age gap between patients in clinical studies and in the general population: a pitfall for dementia research. Lancet Neurol 2004 Oct;3(10):627-630. [doi: 10.1016/S1474-4422(04)00884-1] [Medline: 15380160]
48. Cigolle CT, Blaum CS, Halter JB. Diabetes and cardiovascular disease prevention in older adults. Clin Geriatr Med 2009 Nov;25(4):607-41, vii. [doi: 10.1016/j.cger.2009.09.001] [Medline: 19944264]
49. He Z, Ryan P, Hoxha J, Wang S, Carini S, Sim I, et al. Multivariate analysis of the population representativeness of related clinical studies. J Biomed Inform 2016 Apr;60:66-76 [FREE Full text] [doi: 10.1016/j.jbi.2016.01.007] [Medline: 26820188]
50. Beers E, Moerkerken DC, Leufkens HG, Egberts TC, Jansen PA. Participation of older people in preauthorization trials of recently approved medicines. J Am Geriatr Soc 2014 Oct;62(10):1883-1890. [doi: 10.1111/jgs.13067] [Medline: 25283151]
51. Szlejf C, Farfel JM, Curiati JA, Couto Jr ED, Jacob-Filho W, Azevedo RS. Medical adverse events in elderly hospitalized patients: a prospective study. Clinics (Sao Paulo) 2012 Nov;67(11):1247-1252 [FREE Full text] [Medline: 23184198]

Abbreviations

CC: chronic condition
CCW: chronic condition warehouse
CDM: common data model
CDRN: Clinical Data Research Network
CMS: Center for Medicare and Medicaid Services
EHR: electronic health record
HCUP: Healthcare Cost and Utilization Project
HITECH: Health Information Technology for Economic and Clinical Health
ICD: International Classification of Diseases
MCC: multiple chronic conditions
NH: non-Hispanic
NIS: National Inpatient Sample
PCORnet: Patient-Centered Clinical Research Network
PCORI: Patient-Centered Outcomes Research Institute
RQ: research question

Edited by G Eysenbach; submitted 13.09.17; peer-reviewed by R Zhang, I de la Torre, S Wang, B Ward; comments to author 10.12.17; revised version received 20.01.18; accepted 15.02.18; published 12.04.18.

Please cite as:
He Z, Bian J, Carretta HJ, Lee J, Hogan WR, Shenkman E, Charness N
Prevalence of Multiple Chronic Conditions Among Older Adults in Florida and the United States: Comparative Analysis of the OneFlorida Data Trust and National Inpatient Sample
J Med Internet Res 2018;20(4):e137
URL: http://www.jmir.org/2018/4/e137/
doi:10.2196/jmir.8961
PMID:29650502

Original Paper

Identifying Unmet Treatment Needs for Patients With Osteoporotic Fracture: Feasibility Study for an Electronic Clinical Surveillance System

Fong-Ci Lin[1,2*], MIM; Chen-Yu Wang[3,4*], MS; Rung Ji Shang[5], PhD; Fei-Yuan Hsiao[3,4], PhD; Mei-Shu Lin[6], PhD; Kuan-Yu Hung[7,8], MD, PhD; Jui Wang[9], MS; Zhen-Fang Lin[2,3,4], PhD; Feipei Lai[1,10,11], PhD; Li-Jiuan Shen[2,3,4], PhD; Chih-Fen Huang[2,4], PhD

[1]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan
[2]Department of Pharmacy, National Taiwan University Hospital, Taipei, Taiwan
[3]Graduate Institute of Clinical Pharmacy, College of Medicine, National Taiwan University, Taipei, Taiwan
[4]School of Pharmacy, College of Medicine, National Taiwan University, Taipei, Taiwan
[5]Information Technology Office, National Taiwan University Hospital, Taipei, Taiwan
[6]Department of Development and Planning, National Taiwan University Hospital, Taipei, Taiwan
[7]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan
[8]Department of Internal Medicine, National Taiwan University Hospital Hsinchu Branch, Hsinchu, Taiwan
[9]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan
[10]Department of Computer Science & Information Engineering, National Taiwan University, Taipei, Taiwan
[11]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[*]these authors contributed equally

Corresponding Author:
Chih-Fen Huang, PhD
Department of Pharmacy
National Taiwan University Hospital
33 Linsen S. Rd
Taipei, 10050
Taiwan
Phone: 886 223562162
Fax: 886 223310930
Email: cfhuang1023@ntu.edu.tw

Abstract

Background: Traditional clinical surveillance relied on the results from clinical trials and observational studies of administrative databases. However, these studies not only required many valuable resources but also faced a very long time lag.

Objective: This study aimed to illustrate a practical application of the National Taiwan University Hospital Clinical Surveillance System (NCSS) in the identification of patients with an osteoporotic fracture and to provide a high reusability infrastructure for longitudinal clinical data.

Methods: The NCSS integrates electronic medical records in the National Taiwan University Hospital (NTUH) with a data warehouse and is equipped with a user-friendly interface. The NCSS was developed using professional insight from multidisciplinary experts, including clinical practitioners, epidemiologists, and biomedical engineers. The practical example identifying the unmet treatment needs for patients encountering major osteoporotic fractures described herein was mainly achieved by adopting the computerized workflow in the NCSS.

Results: We developed the infrastructure of the NCSS, including an integrated data warehouse and an automatic surveillance workflow. By applying the NCSS, we efficiently identified 2193 patients who were newly diagnosed with a hip or vertebral fracture between 2010 and 2014 at NTUH. By adopting the filter function, we identified 1808 (1808/2193, 82.44%) patients who continued their follow-up at NTUH, and 464 (464/2193, 21.16%) patients who were prescribed anti-osteoporosis medications, within 3 and 12 months post the index date of their fracture, respectively.

Conclusions: The NCSS systems can integrate the workflow of cohort identification to accelerate the survey process of clinically relevant problems and provide decision support in the daily practice of clinical physicians, thereby making the benefit of evidence-based medicine a reality.

(J Med Internet Res 2018;20(4):e142) doi:10.2196/jmir.9477

KEYWORDS

information systems; public health surveillance; osteoporotic fractures; pharmacovigilance; guideline adherence

Introduction

Clinical surveillance can provide information on disease prognosis and postmarketing medication safety, which helps researchers identify potential clinical issues [1,2]. We describe the development and implementation of a Web-based National Taiwan University Hospital Clinical Surveillance System (NCSS) that is an interoperable, reusable, and scalable platform for collaborative data sharing using the ASP.Net framework (Microsoft, WA, USA).

Traditional clinical surveillance relied on the results from clinical trials and observational studies of administrative databases. However, these studies not only required a lot of valuable resources but also faced a very long time lag. Several studies [3-15] have described the difficulty of reducing gaps between clinical research needs and proper data management technique. Therefore, we aimed to develop an automated system with the capability to use routinely collected electronic health care data to support clinical surveillance is an important issue.

The National Taiwan University Hospital launched a new 3-year strategic plan to build the NCSS in 2015. Therefore, we reviewed the literature to gain insights into the challenges of cross-networking, patient privacy, user-friendliness, and reusability (Table 1). To achieve data harmonization, we referenced the common data model from the Sentinel [4,5]. In addition, the PCORnet [3,9] developed a common data model derived from the Sentinel [4,5] to support the development of an analyzable research database that enhanced the performance of a cross-networking query. We now need to consider the system architectural and data management structures. The Research Electronic Data Capture [14] developed reusable tools for project-specific clinical and translational research data. We found that the design of the Stanford Translational Research Integrated Database Environment [15] for an anonymous patient cohort discovery tool provides a flexible research data management solution. In addition, the Integrating Biology and the Bedside [13] proposed a self-scaling, interoperable platform for collaborative data sharing. Finally, the Observational Health Data Sciences and Informatics is an international collaboration on open-source data analytic solutions [12].

Nonetheless, only a few successful efforts for high reusability and computerized workflow infrastructure have been accomplished in Taiwan to date. Therefore, we decided to implement a Web-based clinical surveillance system extensible to interdisciplinary collaboration and data sharing.

Our efforts are based on 2 highly diffusible, highly reusable, and thin client architectures for research network. The platform provides a high reusability infrastructure for a computerized workflow that captures relevant longitudinal clinical data and makes those data repositories reusable. Finally, the platform has been used for multi-municipality surveillance.

Table 1. Literature summary on electronic clinical surveillance, sorted by year of publication.

Author	Year	Main concept	Country
Brown et al [4,5]	2010	Data harmonization	United States
Obeid et al [14]	2013	Reusable tools for project-specific clinical and translational research data	International
Natter et al [13]	2013	A self-scaling, interoperable platform for collaborative data sharing	United States
Lowe et al [15]	2014	An anonymous patient cohort discovery tool and data management solution	United States
Waitman et al [3] and Fleurence et al [9]	2014	An analyzable research database that enhanced performance of a cross-networking query	United States
Hripcsak et al [12]	2015	An international collaboration on open-source data analytic solutions	International

Methods

Data Warehouse

The NCSS integrates a database of electronic medical records at National Taiwan University Hospital (NTUH), which is a medical center in Taiwan with over 2000 beds. The clinical data models were built using an Oracle 11g relational database; included the demographics, diagnosis, pharmacies, procedures, laboratories, and death records; and were implemented into the data warehouse process.

The data warehouse process is the collection of electronic medical records from the Integrated Medical Database (IMDB) through scheduling using an extraction, transformation, and loading tool. We refreshed the database during nonbusiness hours using 3 steps. First, the system extracted data changes by comparing the time difference with IMDB. Second, personal identifiable information was fully anonymized. Finally, the data

are synchronized back to the IMDB with a timestamp. This data warehouse provides a data access infrastructure for the NCSS.

Workflow

We aimed to present a Web-based NCSS for clinical surveillance in a secure, efficient, and interoperable platform. The NCSS used ASP.Net framework (version 4.5, Microsoft) for Web development and cloud batch process. The NCSS is configured to run using load balancing, including failover modes, to secure the system's availability and scalability. Firewalls were also installed to enhance the security of the NCSS. In addition, all queries were audited and logged, which assures compliance with institutional review board and other regulatory protections for subjects.

We designed thin client architecture so that the researchers can focus on the research design. After setting the parameters on the webpage, the researcher can leave the large computing delegate to cloud. Therefore, the researchers do not need to own a good performance of the computer and lower the computation of big data of barriers. The overall workflow of the NCSS is depicted in Figure 1.

Stage 1. Build a Template of Clinical Orders

The process supports the end user, typically a clinical researcher, to predefine a template of clinical orders using a Clinical Orders Navigator on the client side. The researcher can browse and search different dimensions, such as the diagnosis (International Classification of Diseases, ICD-9 or ICD-10), pharmacy (anatomical therapeutic chemical code), procedure, and laboratory in the integrated interface. The details of Clinical Orders Navigator are provided in Multimedia Appendix 1.

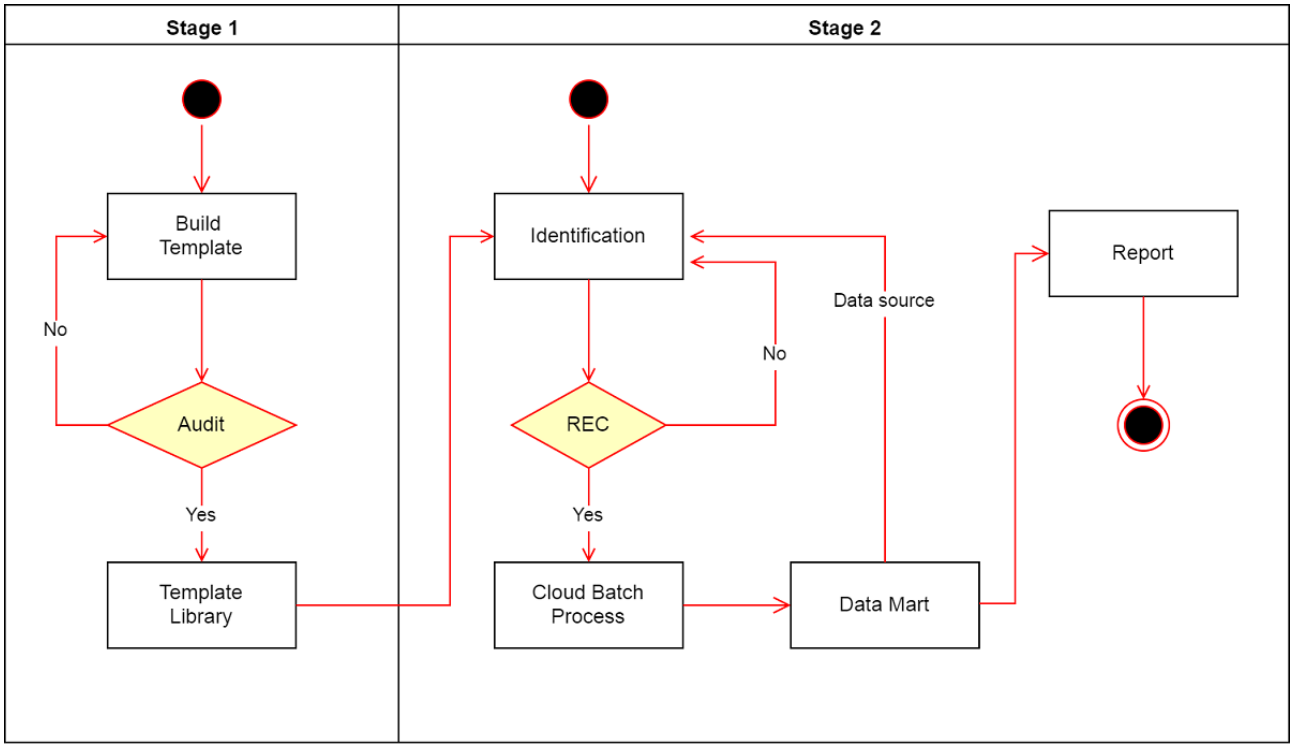
The stage can help clinical researchers build a protocol-based standardized process and save those clinical orders and specific guidelines to the database. The creator of the templates can choose to commit them to Template Library publicly. All templates submitted to the public Template Library will be reviewed by clinical professionals to ensure the quality and accuracy. All researchers can create their own template or use the open template applied to the Identification process.

Stage 2. Patient Identification

The stage of patient identification includes matching of the clinical needs to the optimal cohort study. The stage contains 5 processes, consisting of identification, REC (Research Institutional Ethics Committee) verify, cloud batch process, data mart of patient level, and report service.

1. Identification: We developed an electronic form to meet the cohort study flow that contains 2 setting as input as follows (when researchers finished the electronic form, the system will save that setting to the database and generate a universally unique identifier as case number of this setting.):
   - Basic setting: In this step, the researcher fills the surveillance topic, and the duration of observation from a particular data source, such as outpatient, admission, and emergency. In addition, this setting also supports the selection of specific patient list from data mart (patient level).
   - Order setting: In this step, the researcher chooses the order setting in Clinical Orders Navigator. The clinical orders can be in different dimensions, such as the diagnosis, pharmacy, procedure, and laboratory. For example, a disease can be defined using several diagnosis codes, such as hip or vertebral fractures, including ICD-9-codes 820, 805, and 806 in the Template Library. The researchers can reuse these templates as include or exclude criteria to design their study flow.
2. REC verify: All settings of Identification need to be been verified by REC. If the setting had been authorized, it will add a task to the Task Queue for cloud batch process in server side.
3. Cloud batch process: The cloud batch process maintains a Task Queue. Every task will be executed follow First-In-First-Out mechanism in batch server that read the eligible patients in IMDB. The batch process also records the cost of each query and the number of patients identified. After completion of the task, the patient list will be stored into the database and a mail notification is will be sent to the researchers.
4. Data mart (patient level): The data mart is a collection of the patient list that presents the result of every identification process. The researcher can adopt a patient list from data mart as their patient data source to query the next Identification process. Therefore, the identification process can support a hierarchical structure. This means that the process can generate a new study population based on a previous screening result. The researchers can reuse these patient lists to design a cohort study or case control study in fine-grained categorization.
5. Report service: The report service contains 3 views of summary statistics for patient list in data mart as follows:
   - View of characteristics: The view is a demographic summary that helps clarify the characteristics for the patient list, for example, the descriptive statistics of age, gender, body mass index (BMI), and income level.
   - View of longitudinal incidence trend: The view presents incidence trend by time series chart and provides a real-time interactive query by time interval, including monthly, quarterly, or yearly.
   - View of source record: The view presents the number of included/excluded patients in every Identification process. If the patient list contains a hierarchical structure that runs the process of identification more than once, the report service can track all results of identification in the aggregation table.

Figure 1. System workflow of the National Taiwan University Hospital Clinical Surveillance System (NCSS). REC: Research Institutional Ethics Committee.



A Practical Example of a Clinical Application of the National Taiwan University Hospital Clinical Surveillance System in the Identification of Osteoporotic Fracture Patients

For a practical example of a clinical application of the NCSS, we looked at the identification of osteoporotic fracture patients and their utilization in pharmacological therapy. Osteoporotic fractures, a major consequence of osteoporosis, are associated with a high mortality rate, increased risk of re-fracture, and poor quality of life, and incur heavy economic burden on the society.

According to a report from the Global Burden of Disease Study project, the global burden of osteoporosis-related problems has doubled in the past two decades and has shown a continuous increase in recent years. In 2016, 441,230 documented deaths could be attributed to osteoporosis-related problems [16,17].

In the United States, the direct economic burden of osteoporotic fractures was approximately US $17 billion in 2005, and is projected to increase 50% by 2025 [18]. Fortunately, the safety and efficacy of anti-osteoporosis medications (AOMs) used by patients with established osteoporotic fractures have been ascertained [19-23]. However, despite the readily available and effective treatment for osteoporosis, a care gap between established osteoporotic fractures and the pharmacological prevention of subsequent fractures is still being discussed worldwide [24-27]. We aimed to identify the unmet treatment needs for patients encountering major osteoporotic fractures with the NTUH-based clinical surveillance system.

Using NCSS as our data source, we identified patients newly diagnosed with a hip (ICD-9-code 820) or vertebral fracture (ICD-9-codes 805 and 806) between 2010 and 2014 as our study subjects, and defined them as patients requiring treatment. The initial diagnosis date of a hip or vertebral fracture was defined as the index date of the study subject. Patients under 50 years of age; with a diagnosis of malignant neoplasm (ICD-9-code 140-208), osteoporotic fracture (ICD-9-codes 820, 805, and 806), or Paget disease (ICD-9-code 731.0); or who had been prescribed AOMs within 1 year before the index date were excluded. Among them, we investigated the prescription pattern of the AOMs by distinguishing patients into those who began taking an AOM within 1 year after the index date, and those who did not. The AOMs evaluated in this study included alendronate, zolendronate, ibandronate, denosumab, raloxifene, teriparatide, estradiol valerate, conjugated estrogens, and calcitonin. The system setting of the identification process used by the NCSS, the demographics of the study population, and the treatment pattern of the study population were presented quarterly.

Results

The NCSS design uses a protocol-based standardized process of incremental development, testing, and deployment to meet specific clinical needs. We demonstrated a practical example of identifying the unmet treatment needs for patients encountering major osteoporotic fractures and implemented the hierarchical study population using the NCSS, as depicted in Figure 2.

We initially selected older patients diagnosed with a hip or vertebral fracture between 2010 and 2014. By adopting the identification and filter function of the NCSS, patients with a history of malignant neoplasm (n=557), or osteoporotic vertebral and hip fracture (n=1769) within 1 year before the index date, were excluded. In addition, to identify a new AOM user, we excluded 179 patients with an AOM prescription before the

index date. We identified 2193 incidence cases for hip or vertebral fractures within the period of 2010-2014. These patients were defined as "patients requiring treatment" according to the current treatment guidelines. In addition, in the Report Service, we used visualization tools for displaying the summary statistics for each patient list. This is depicted in Figure 3.

To ensure those participants are having a continuous follow-up, we excluded 385 patients who had not visited the hospital within 3 months after the index date. We enrolled in the study 1808 patients who had continued to follow-up at NTUH within 3 months post the index date of their fracture. To investigate the prescription pattern of the AOMs, we established 2 groups that were classified based on their AOM prescription date, and by adopting a filter function, we identified 1808 (1808/2193, 82.44%) patients who continued to follow-up at NTUH, and 464 (464/2193, 21.16%) patients prescribed with an AOM, within 3 and 12 months post the index date of their fracture, respectively.

The NCSS provided a summary of the baseline characteristics of the study population including gender, age, BMI, socioeconomic status, occupation types, and marital status, as shown in Figure 4. For example, among the patients who began taking an AOM within 1 year after the index date, their mean age was 76.47 (SD 10.10) and their mean BMI was 22.95 (SD 3.86) kg/m^2; in addition, the proportion of females was 82.7% (384/464). This population showed a high proportion of married patients 71.98% (334/464) with a normal income level 99.8% (463/464). NCSS provided information regarding the drug utilization of AOMs for the study population.

The longitudinal trends of patients newly diagnosed with an osteoporotic fracture and those who began taking an AOM within 1 year after the index date of their fracture are illustrated in Figure 5. For example, taking the information from Figure 5, we found that there were approximately 130 newly diagnosed osteoporotic fracture patients continuing their follow-up at NTUH in Quarter 4 2014, and among them, 42 (42/130, 32.3%) began taking an AOM within 1 year post the index date of their fracture.

Furthermore, the NCSS provides the choice of information stratification based on gender, index date, fracture type, and other types of information, which can be provided monthly, quarterly, and yearly, thereby increasing the flexibility of the clinical interpretation. For example, as shown in part (b) of Figure 5, there were 87 female and 43 male fracture patients in 2014 Q4, and among them, 30 (30/87, 35%) females and 12 (28%, 12/43) males began taking an AOM within 1 year post the index date of their fracture. Information on patients with different fracture types can be seen in part (c) of Figure 5.

Figure 2. Study flowchart implemented by National Taiwan University Hospital Clinical Surveillance System (NCSS). Each identification process had been assigned a universally unique identifier with a case number (marked by the blue background, such as 73, F124, F133, F134, F135, F136, and F137). The case number with an F as a prefix stands for its' own hierarchical structure. NTUH: National Taiwan University Hospital; AOMs: anti-osteoporosis medications.

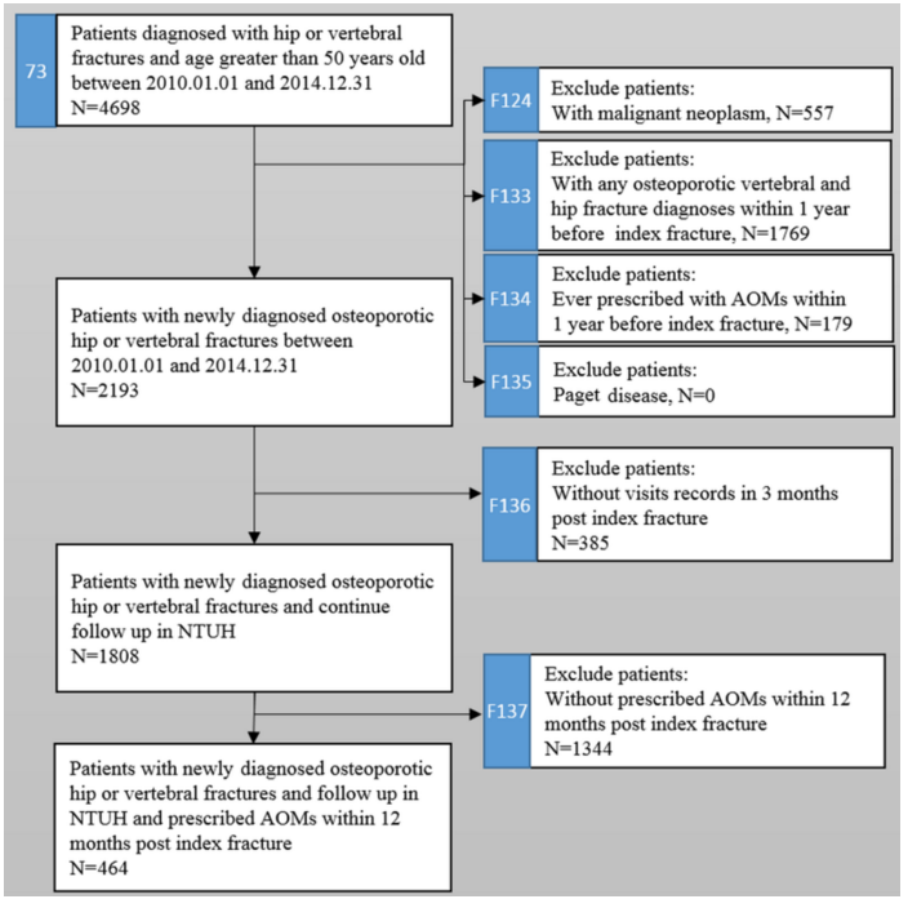
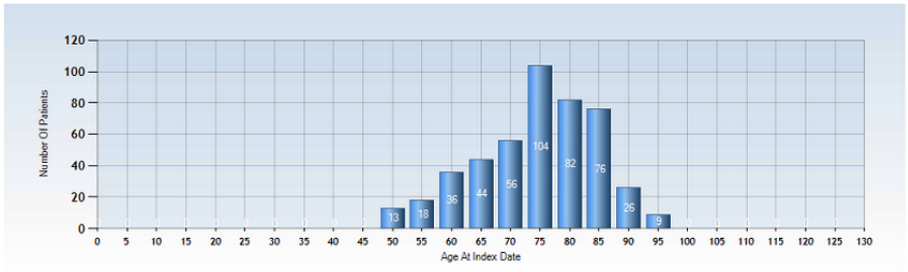


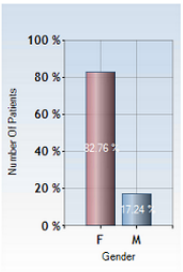
Figure 3. Snapshot of a source record in National Taiwan University Hospital Clinical Surveillance System (NCSS). Through the source record, we can get the number of included (and excluded) patients and the data source (patient list).

Characteristics | Longitudinal Incidence Trend | Source Record

Sync Record	ID	Patient List	Label	Number Of Included Patients	Number of Excluded Patients
View	73		[Osteoporotic Fracture] Incident Case	4,698	0
View	F124	73	[Osteoporotic Fracture] exclude malignant neoplasm	4,141	557
View	F133	F124	[Osteoporotic Fracture] exclude osteoporotic fracture within 1 year prior to index	2,372	1,769
View	F134	F133	[Osteoporotic Fracture] exclude AOMs	2,193	179
View	F135	F134	[[Osteoporotic Fracture] exclude Paget's Disease	2,193	0
View	F136	F135	[Osteoporotic Fracture] visit within 3 month post index	1,808	385
View	F137	F136	[Osteoporotic Fracture] initiated AOMs within 1 year post index	464	1,344

Figure 4. Snapshot of Report of characteristics in the National Taiwan University Hospital Clinical Surveillance System (NCSS).





Age

Total	464
Min	51
Max	98
Mean	76.47
Standard Deviation	10.10
Variance	102.02
Q1	70
Q2	77
Q3	84

Patient Source

Patient Class	Number of Patients	Percent
Admission	14	3.02 %
Outpatient	445	95.91 %
Emergency	5	1.08 %

Marital Status

Marital status	Number of Patients	Percent
Living together	1	0.22 %
Divorced	2	0.44 %
Widowed	29	6.44 %
Unmarried	84	18.67 %
Married	334	74.22 %

Income Level

Stat	Number of Patients	Percent
Low Income	1	0.22 %
Normal Income	463	99.78 %

Occupation

Occupation	Number of Patients	Percent
Housekeeping	121	29.16 %
Industry	19	4.58 %
Self-employment	9	2.17 %
Unemployed	88	21.20 %
Business	111	26.75 %
Agriculture	2	0.48 %
Fishery	1	0.24 %
Other	47	11.33 %
Service Industry	3	0.72 %
Government Employees	10	2.41 %
Education	4	0.96 %

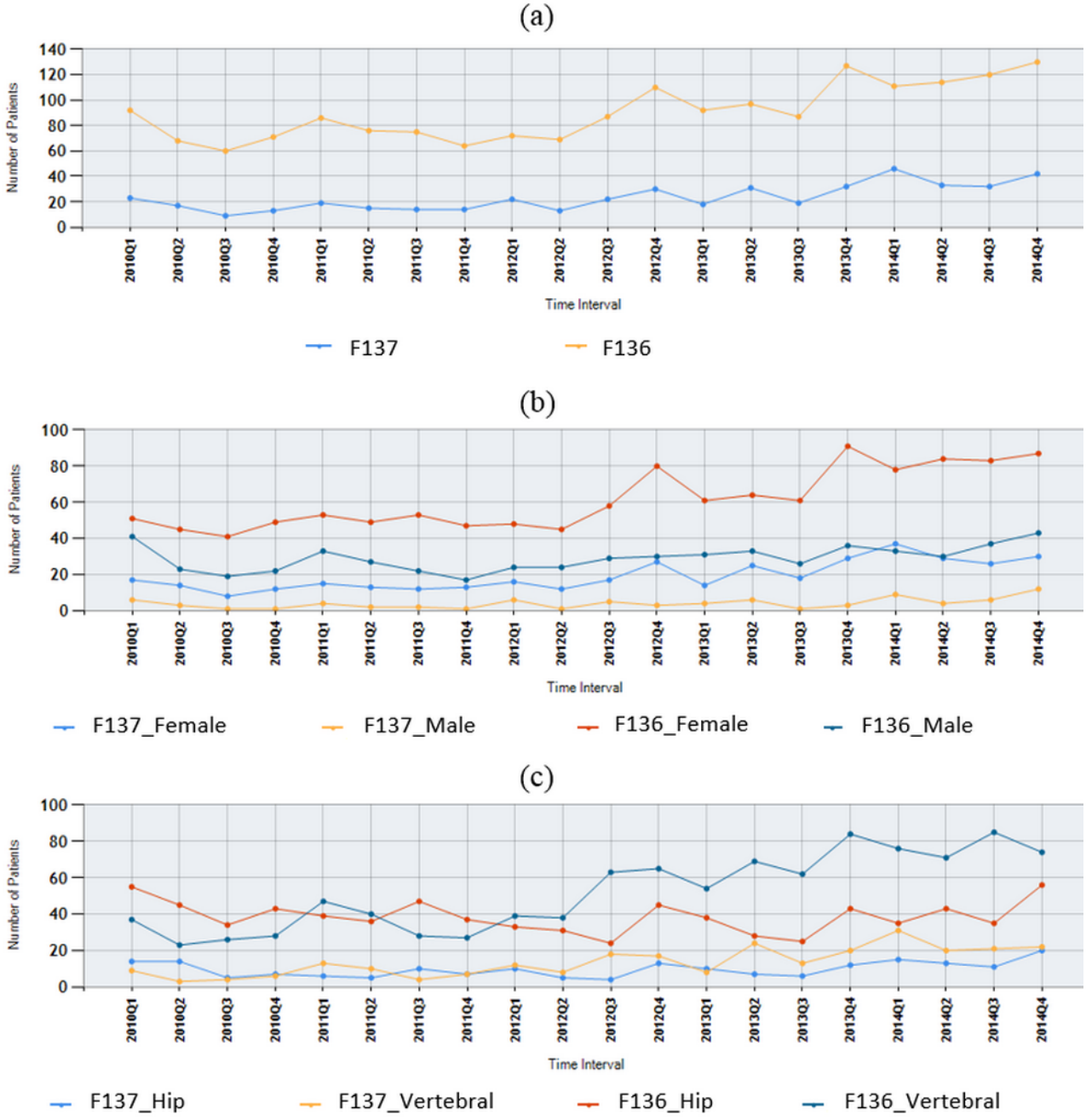
BMI

Total	464
Observation	351
Min	13.84
Max	36.85
Mean	22.95
Standard Deviation	3.86
Variance	14.88
Q1	20.20
Q2	22.64
Q3	25.3750

Gender

Total	464
Male	80
Female	384

Figure 5. Snapshot of the quarterly report for incidence trend of drug utilization of study cohorts.



Discussion

Principal Findings

To the best of our knowledge, NCSS is a pioneering electronic clinical surveillance system in its attempt to organize decision-making activities and facilitate the standard-based process for clinical needs in Taiwan. In this study, we demonstrated a practical example of an NCSS application. We can efficiently and correctly gather information on those patients requiring disease treatment and understand their treatment patterns, as well as identify any unmet treatment requirements in the hospital.

Through this practical example, we found that the pharmacological treatment rate of patients with an osteoporotic fracture is suboptimal at NTUH. On average, only 35% (30/87) of female and 28% (12/43) of male osteoporotic fracture patients initiated AOM therapy to prevent a subsequent fracture. More effort is warranted to improve the quality of care for these patients.

Limitations

Studies from a hospital-wide database require less time for data collection and can achieve a faster response to meet the needs of clinical research. However, there is a certain limitation to this study, that is, bias in the estimates may exist because it is difficult to determine whether the patient in a hospital-wide database is suffering from such injury for the first time.

Comparison With Prior Work

Clinical surveillance systems have been widely implemented [13,15,28], and studies have demonstrated the use of various algorithms to identify potential research patient cohorts. There are several important core concepts to our research, including access to clinical data for research purposes, the design of a flexible research data management solution, and clarifying the characteristics of each study population. However, previous studies [8,9,13,15] have not proposed how to integrate a protocol-based process in the building of a cohort discovery. The clinical surveillance systems should also have the capability to bring guidelines to the clinical practitioners. The NCSS integrates the workflow of cohort identification to accelerate the survey process based on certain guidelines. In particular, for quality assurance, clinical researchers or medical policy makers need to monitor specific quality indicators to ensure the quality of patient care. They can focus on clinical needs to achieve a continuous process integration using standardized NCSS templates. The different clinical contexts can be refined using a new scaffold to meet clinical needs based on the original standardized templates. In fact, different methodologies of capturing cases of patients would result in disparate estimates of incidence or mortality. For example, in the case on sepsis identification, there are four methods available, [29-32] including Angus et al, Martin et al, Dombrovskiy et al, and Wang et al. We believe this clinical knowledge should be preserved and converted into a shareable template for a collaborative research network. These templates can be quickly searched and reused for inclusion/exclusion criteria of patient identification in the Template Library. Therefore, the researchers can embrace change courageously because they only need to focus on any existing differences. More importantly, these processes should be conceptualized as continuous organizational efforts to lead to self-organizing innovation of the NCSS.

The hierarchical structure of the system design has not been proposed for system reusability. We designed a mechanism for reusing the patient list in data mart. We believe this identified list of patients can be reused by other researchers, especially for the design of a subgroup study, a case control study, or a similar context of research. To better interpret the process of identification of each patient list, we designed a hierarchical structure where each patient list could be traced back to its patient data source and researchers could compare every patient data source with characteristics and longitudinal incidence trend in the Report Service. This innovative method allows dynamically generating a clinical data mart and reduces the computation overhead through the reuse of the same patient list. This design can inspire researchers and allow them to focus on the research design rather than data processing.

Until 2017, the NCSS has been well constructed and continuously improving. The active users of NCSS are now in the form of research teams. These teams consist of multidisciplinary specialties such as a clinical doctor, pharmacist, epidemiologist, bio-medical engineer, and research assistants. Several research teams have been adopting the NCSS to improve the problem-solving process of their relevant clinical issue. This includes cardiovascular, diabetic, renal, liver, neurologic, and orthopedic medical teams. Each team consists of 5-7 people. These research teams regularly hold monthly meetings to discuss the problems related to the application of NCSS. In the period between 2015 and 2017, dozens of cohorts were created by each research team. Moreover, a large sum of meaningful feedback has been received for not only the problem related to using the NCSS but also recommendations for improving the NCSS. For example, the cardiovascular research team raised the problem of using different brands of drugs during different periods. Therefore, we retrieved the relevant drug products available and used in 2014 as well as 2004. This modification allows the research to more precisely capture and study the drug exposure across a population. The orthopedic research team also revealed the problems regarding the inconsistent days listed for a prescription and the pharmacological duration of the prescription. Some medication has a 1 year of effect duration; however, the prescription was recorded for only 1 day. Therefore, the NCSS modified the duration according to each drug property to correct the measurement of exposure. Finally, we accomplished the assistant resources such as the user handbook and Web-based video tutorials. This helps the research team readily assess the NCSS. These application experiences and associated feedback help to improve the NCSS efficiency and quality of clinical research at NTUH.

Future Work

The identification of the problem is the first step to solving and improving the clinical outcome of the patient. By applying computerized patient identification derived from the NCSS, we can create the infrastructure of an informatics system at NTUH. Furthermore, we can provide decision support in daily practice, thereby making the benefit of evidence-based medicine a reality. In fact, we believe that the best way to promote medical care is to provide relevant evidence to assist doctors in their decision-making process during their clinical daily practice. For example, an evaluation of the longitudinal trends of health care utilization can help create a baseline, track progress over time, and generate real-world evidence. Besides providing clinical support for physicians, the next step will be providing integrated real-time, interactive, and personalized support to individual patients [33,34]. This will be focused on in a future study. Finally, we strongly advocate developing a consistent strategy, as well as celebrating success and continuously sharing different experiences. The continual reduction in the gap between evidence and practice is an ongoing journey and not an end that can be simply reached shortly after the NCSS is implemented.

Conclusions

The results of this study confirm that the NCSS is an efficient electronic clinical surveillance system that can integrate the workflow of cohort identification to accelerate the survey process of disease and medication prescription patterns. The NCSS can serve the critical role of gathering data and making associations between evidence and practice in a rapid fashion, and can be a support system for researchers who wish to confirm certain clinical issues as well as for those requiring a computerized system to complete their studies. Finally, the NCSS can provide decision support in the daily practice of

clinical physicians and make the benefit of evidence-based medicine a reality.

Acknowledgments

This research received an Asia number one plan grant from National Taiwan University Hospital. This work was supported by National Taiwan University Hospital grant numbers 104-A129, 105-A129, and 106-A129.

Authors' Contributions

FL built the clinical surveillance system and was part of the final draft writing. CW designed the study, interpreted the results, and wrote the draft. CH and FH conceptualized and designed the study, interpreted the data, and critically revised the manuscript. All authors helped to critically review and revise the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A real-world use - case deployed at National Taiwan University Hospital Clinical Surveillance System.

[PDF File (Adobe PDF File), 1MB - jmir_v20i4e142_app1.pdf]

References

1. Linder JA, Ma J, Bates DW, Middleton B, Stafford RS. Electronic health record use and the quality of ambulatory care in the United States. Arch Intern Med 2007 Jul 09;167(13):1400-1405. [doi: 10.1001/archinte.167.13.1400] [Medline: 17620534]
2. Classen DC, Bates DW. Finding the meaning in meaningful use. N Engl J Med 2011 Sep 01;365(9):855-858. [doi: 10.1056/NEJMsb1103659] [Medline: 21879906]
3. Waitman LR, Aaronson LS, Nadkarni PM, Connolly DW, Campbell JR. The Greater Plains Collaborative: a PCORnet Clinical Research Data Network. J Am Med Inform Assoc 2014 Jul;21(4):637-641 [FREE Full text] [doi: 10.1136/amiajnl-2014-002756] [Medline: 24778202]
4. Brown J, Syat B, Lane K, Platt R. Effective Health Care AHRQ. 2010 Jun 23. Blueprint for a distributed research network to conduct population studies and safety surveillance URL: https://effectivehealthcare.ahrq.gov/topics/distributed-network-blueprint/research/ [accessed 2018-03-27] [WebCite Cache ID 6yDx5qtY0]
5. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. Med Care 2010 Jun;48(6 Suppl):S45-S51. [doi: 10.1097/MLR.0b013e3181d9919f] [Medline: 20473204]
6. Curtis LH, Weiner MG, Boudreau DM, Cooper WO, Daniel GW, Nair VP, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidemiol Drug Saf 2012 Jan;21(Suppl 1):23-31. [doi: 10.1002/pds.2336] [Medline: 22262590]
7. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System--a national resource for evidence development. N Engl J Med 2011 Feb 10;364(6):498-499. [doi: 10.1056/NEJMp1014427] [Medline: 21226658]
8. Ferranti JM, Gilbert W, McCall J, Shang H, Barros T, Horvath MM. The design and implementation of an open-source, data-driven cohort recruitment system: the Duke Integrated Subject Cohort and Enrollment Research Network (DISCERN). J Am Med Inform Assoc 2012 Jun;19(e1):e68-e75 [FREE Full text] [doi: 10.1136/amiajnl-2011-000115] [Medline: 21946237]
9. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014 Jul;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
10. Zerhouni EA. US biomedical research: basic, translational, and clinical sciences. JAMA 2005 Sep 21;294(11):1352-1358. [doi: 10.1001/jama.294.11.1352] [Medline: 16174693]
11. Sung NS, Crowley Jr WF, Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA 2003 Mar 12;289(10):1278-1287. [Medline: 12633190]
12. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]
13. Natter MD, Quan J, Ortiz DM, Bousvaros A, Ilowite NT, Inman CJ, et al. An i2b2-based, generalizable, open source, self-scaling chronic disease registry. J Am Med Inform Assoc 2013 Jan 01;20(1):172-179 [FREE Full text] [doi: 10.1136/amiajnl-2012-001042] [Medline: 22733975]

14. Obeid JS, McGraw CA, Minor BL, Conde JG, Pawluk R, Lin M, et al. Procurement of shared data instruments for Research Electronic Data Capture (REDCap). J Biomed Inform 2013 Apr;46(2):259-265 [FREE Full text] [doi: 10.1016/j.jbi.2012.10.006] [Medline: 23149159]
15. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE--An integrated standards-based translational research informatics platform. AMIA Annu Symp Proc 2009 Nov 14;2009:391-395 [FREE Full text] [Medline: 20351886]
16. Sànchez-Riera L, Carnahan E, Vos T, Veerman L, Norman R, Lim SS, et al. The global burden attributable to low bone mineral density. Ann Rheum Dis 2014 Sep;73(9):1635-1645. [doi: 10.1136/annrheumdis-2013-204320] [Medline: 24692584]
17. GBD 2016 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1345-1422 [FREE Full text] [doi: 10.1016/S0140-6736(17)32366-8] [Medline: 28919119]
18. Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. J Bone Miner Res 2007 Mar;22(3):465-475 [FREE Full text] [doi: 10.1359/jbmr.061113] [Medline: 17144789]
19. Cosman F, de Beur SJ, LeBoff MS, Lewiecki EM, Tanner B, Randall S, National Osteoporosis Foundation. Clinician's Guide to Prevention and Treatment of Osteoporosis. Osteoporos Int 2014 Oct;25(10):2359-2381 [FREE Full text] [doi: 10.1007/s00198-014-2794-2] [Medline: 25182228]
20. Compston J, Cooper A, Cooper C, Francis R, Kanis JA, Marsh D, National Osteoporosis Guideline Group (NOGG). Guidelines for the diagnosis and management of osteoporosis in postmenopausal women and men from the age of 50 years in the UK. Maturitas 2009 Feb 20;62(2):105-108. [doi: 10.1016/j.maturitas.2008.11.022] [Medline: 19135323]
21. Watts NB, Bilezikian JP, Camacho PM, Greenspan SL, Harris ST, Hodgson SF, AACE Osteoporosis Task Force. American Association of Clinical Endocrinologists Medical Guidelines for Clinical Practice for the diagnosis and treatment of postmenopausal osteoporosis. Endocr Pract 2010;16(Suppl 3):1-37 [FREE Full text] [Medline: 21224201]
22. Orimo H, Nakamura T, Hosoi T, Iki M, Uenishi K, Endo N, et al. Japanese 2011 guidelines for prevention and treatment of osteoporosis--executive summary. Arch Osteoporos 2012;7:3-20 [FREE Full text] [doi: 10.1007/s11657-012-0109-9] [Medline: 23203733]
23. Hwang JS, Chan DC, Chen JF, Cheng TT, Wu CH, Soong YK, et al. Clinical practice guidelines for the prevention and treatment of osteoporosis in Taiwan: summary. J Bone Miner Metab 2014 Jan;32(1):10-16. [doi: 10.1007/s00774-013-0495-0] [Medline: 24068612]
24. Formiga F, Rivera A, Nolla JM, Coscujuela A, Sole A, Pujol R. Failure to treat osteoporosis and the risk of subsequent fractures in elderly patients with previous hip fracture: a five-year retrospective study. Aging Clin Exp Res 2005 Apr;17(2):96-99. [Medline: 15977456]
25. Rabenda V, Vanoverloop J, Fabri V, Mertens R, Sumkay F, Vannecke C, et al. Low incidence of anti-osteoporosis treatment after hip fracture. J Bone Joint Surg Am 2008 Oct;90(10):2142-2148. [doi: 10.2106/JBJS.G.00864] [Medline: 18829912]
26. Wilk A, Sajjan S, Modi A, Fan CP, Mavros P. Post-fracture pharmacotherapy for women with osteoporotic fracture: analysis of a managed care population in the USA. Osteoporos Int 2014 Dec;25(12):2777-2786 [FREE Full text] [doi: 10.1007/s00198-014-2827-x] [Medline: 25112720]
27. Klop C, Gibson-Smith D, Elders PJ, Welsing PM, Leufkens HG, Harvey NC, et al. Anti-osteoporosis drug prescribing after hip fracture in the UK: 2000-2010. Osteoporos Int 2015 Jul;26(7):1919-1928 [FREE Full text] [doi: 10.1007/s00198-015-3098-x] [Medline: 25963232]
28. Yoo S, Kim S, Lee KH, Jeong CW, Youn SW, Park KU, et al. Electronically implemented clinical indicators based on a data warehouse in a tertiary hospital: its clinical benefit and effectiveness. Int J Med Inform 2014 Jul;83(7):507-516. [doi: 10.1016/j.ijmedinf.2014.04.001] [Medline: 24867195]
29. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med 2003 Apr 17;348(16):1546-1554. [doi: 10.1056/NEJMoa022139] [Medline: 12700374]
30. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med 2001 Jul;29(7):1303-1310. [doi: 10.1097/00003246-200107000-00002]
31. Wang HE, Shapiro NI, Angus DC, Yealy DM. National estimates of severe sepsis in United States emergency departments. Crit Care Med 2007 Aug;35(8):1928-1936. [doi: 10.1097/01.CCM.0000277043.85378.C1] [Medline: 17581480]
32. 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]
33. Duane S, Tandan M, Murphy AW, Vellinga A. Using Mobile Phones to Collect Patient Data: Lessons Learned From the SIMPle Study. JMIR Res Protoc 2017 Apr 25;6(4):e61 [FREE Full text] [doi: 10.2196/resprot.6389] [Medline: 28442451]
34. Fukuoka Y, Kamitani E, Bonnet K, Lindgren T. Real-time social support through a mobile virtual community to improve healthy behavior in overweight and sedentary adults: a focus group analysis. J Med Internet Res 2011 Jul 14;13(3):e49 [FREE Full text] [doi: 10.2196/jmir.1770] [Medline: 21752785]

Abbreviations

AOMs: anti-osteoporosis medications
BMI: body mass index
ICD: International Classification of Diseases
IMDB: Integrated Medical Database
NCSS: National Taiwan University Hospital Clinical Surveillance System
NTUH: National Taiwan University Hospital
REC: Research Institutional Ethics Committee

Edited by G Eysenbach; submitted 26.11.17; peer-reviewed by ZP Ho, N Zarinabad; comments to author 21.12.17; revised version received 23.01.18; accepted 26.01.18; published 24.04.18.

Please cite as:
Lin FC, Wang CY, Shang RJ, Hsiao FY, Lin MS, Hung KY, Wang J, Lin ZF, Lai F, Shen LJ, Huang CF
Identifying Unmet Treatment Needs for Patients With Osteoporotic Fracture: Feasibility Study for an Electronic Clinical Surveillance System
J Med Internet Res 2018;20(4):e142
URL: http://www.jmir.org/2018/4/e142/
doi:10.2196/jmir.9477
PMID:29691201

Original Paper

Psychosocial Factors of Health Professionals' Intention to Use a Decision Aid for Down Syndrome Screening: Cross-Sectional Quantitative Study

Samira Abbasgholizadeh Rahimi[1,2,3], Eng, PhD; Johanie Lépine[4], PhD; Jordie Croteau[2,5], MSc; Hubert Robitaille[2,3], PhD; Anik MC Giguere[1,3,6], PhD; Brenda J Wilson[7], MBChB, MSc, MRCP (UK), FFPH; François Rousseau[8], MSc, MD, FRCPC; Isabelle Lévesque[9], MD; France Légaré[1,2,3], MD, PhD, CCMF, FCMF

[1]Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada

[2]Canada Research Chair in Shared Decision Making and Knowledge Translation, Quebec, QC, Canada

[3]Université Laval Primary Care Research Centre (CERSSPL-UL), Université Laval, Quebec, QC, Canada

[4]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

[5]Quebec Support for People and Patient-Oriented Research and Trials Units, Université Laval, Quebec, QC, Canada

[6]Quebec Centre of Excellence on Aging, Laval University Research Centre on Primary Healthcare and Services, Quebec, QC, Canada

[7]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

[8]Department of Molecular Biology, Medical Biochemistry and Pathology, Faculty of Medicine, Université Laval, Quebec, QC, Canada

[9]Centre hospitalier universitaire de Québec, Quebec, QC, Canada

Corresponding Author:
France Légaré, MD, PhD, CCMF, FCMF
Canada Research Chair in Shared Decision Making and Knowledge Translation
2525, chemin de la Canardière
Pavillon Landry-Poulin, bureau A-4574, Entrée A-1-2
Quebec, QC, G1J 0A4
Canada
Phone: 1 418 525 4437
Fax: 1 418 525 4194
Email: France.Legare@mfa.ulaval.ca

Abstract

Background: Decisions about prenatal screening for Down syndrome are difficult for women, as they entail risk, potential loss, and regret. Shared decision making increases women's knowledge of their choices and better aligns decisions with their values. Patient decision aids foster shared decision making but are rarely used in this context.

Objective: One of the most promising strategies for implementing shared decision making is distribution of decision aids by health professionals. We aimed to identify factors influencing their intention to use a DA during prenatal visit for decisions about Down syndrome screening.

Methods: We conducted a cross-sectional quantitative study. Using a Web panel, we conducted a theory-based survey of health professionals in Quebec province (Canada). Eligibility criteria were as follows: (1) family physicians, midwives, obstetrician-gynecologists, or trainees in these professions; (2) involved in prenatal care; and (3) working in Quebec province. Participants watched a video depicting a health professional using a decision aid during a prenatal consultation with a woman and her partner, and then answered a questionnaire based on an extended version of the theory of planned behavior, including some of the constructs of the theoretical domains framework. The questionnaire assessed 8 psychosocial constructs (attitude, anticipated regret, subjective norm, self-identity, moral norm, descriptive norm, self-efficacy, and perceived control), 7 related sets of behavioral beliefs (advantages, disadvantages, emotions, sources of encouragement or discouragement, incentives, facilitators, and barriers), and sociodemographic data. We performed descriptive, bivariate, and multiple linear regression analyses to identify factors influencing health professionals' intention to use a decision aid.

Results: Among 330 health professionals who completed the survey, 310 met the inclusion criteria: family physicians, 55.2% (171/310); obstetrician-gynecologists, 33.8% (105/310); and midwives, 11.0% (34/310). Of these, 80.9% were female (251/310). Mean age was 39.6 (SD 11.5) years. Less than half were aware of any decision aids at all. In decreasing order of importance,

factors influencing their intention to use a decision aid for Down syndrome prenatal screening were as follows: self-identity (beta=.325, *P*<.001), attitude (beta=.297, *P*<.001), moral norm (beta=.288, *P*<.001), descriptive norm (beta=.166, *P*<.001), and anticipated regret (beta=.099, *P*=.003). Underlying behavioral beliefs significantly related to intention were that the use of a decision aid would promote decision making (beta=.117, 95% CI 0.043-0.190), would reassure health professionals (beta=.100, 95% CI 0.024-0.175), and might require more time than planned for the consultation (beta=−.077, 95% CI −0.124 to −0.031).

Conclusions: We identified psychosocial factors that could influence health professionals' intention to use a decision aid about Down syndrome screening. Strategies should remind them of the following: (1) using a decision aid for this purpose should be a common practice, (2) it would be expected of someone in their societal role, (3) the experience of using it will be satisfying and reassuring, and (4) it is likely to be compatible with their moral values.

(J Med Internet Res 2018;20(4):e114) doi:10.2196/jmir.9036

KEYWORDS

decision support techniques; Down syndrome; decision making; behavior; intention; physicians; midwifery; surveys; prenatal diagnosis

Introduction

Background

The decision about whether or not to take a prenatal screening test can be challenging for women [1]. First, the medical information is complex and results are sometimes uncertain [2-4]. Second, to make a good decision, women need to thoroughly understand probabilistic data and the characteristics of the various screening tests such as detection rates [5]. Third, prenatal testing for Down syndrome involves the woman's personal values, and she may find it difficult to articulate them [1]. Fourth, not knowing the outcome and then finding out the results of the screening may be stressful [6-8] and lead to more difficult decisions. If the results are positive, the pregnant woman will have to decide whether to undergo more invasive diagnostic testing (such as amniocentesis) or not. If these results, too, are positive, she will have to choose between terminating the pregnancy and preparing for a child who will have special needs throughout his or her life [9]. Finally, health professionals' communication of results is often insufficient, misleading, or negative and they may misjudge women's values and preferences [5,10-12]. Without sufficient decision support, women in this position may experience decisional conflict, defined as the uncertainty caused by a decision that involves risk, loss, or a challenge to their personal values [13]. Uncertainty may lead to decision regret, which in turn can lead to deteriorating health and perhaps litigation.

In Quebec, Canada, in family medicine groups for example, a first prenatal appointment with a nurse is usually scheduled for a pregnant woman starting from the fifth and sixth week of her pregnancy. The nurse enters information in the patient's obstetrical file and gives her an information kit that includes the Quebec Ministry of Health and Social Service's information leaflet on the prenatal test for Down syndrome. Starting from eighth to ninth week, the woman meets the family physician, and the doctor asks her if she wants to take the test.

At this stage, with the help of the physician, the woman should make an informed decision, that is, one that takes into account relevant information about the advantages and disadvantages of all the options and that is in keeping with her values and preferences [14]. However, in reality, studies have shown that in spite of some printed information they are given, few women are making informed decisions about prenatal screening, and health professionals are not engaging pregnant women in shared decision making (SDM) about this decision [12,15].

Shared Decision Making and Decision Aids

SDM is an interpersonal and interdependent process [16] in which the health professional and patient work together to make informed value-congruent decisions about the patient's health [17,18]. SDM about this and other health-related decisions not only improves care experiences but also is increasingly recognized as an ethical imperative in health policy and legislation [19]. Patient decision aids (DAs) are decision support tools that help the decision-making process by providing information about treatment or testing options, associated benefits, disadvantages, probabilities and uncertainties, as well as raising the question of values and preferences [20]. Studies have shown that by improving knowledge and allowing patients to make choices that are informed by evidence and by their values, DAs can reduce decisional conflict and indecision [20]. Despite the evidence in favor of SDM and DAs, they are rarely used in practice [21] and even less for prenatal screening decisions.

To increase the use of DAs in making informed decisions about Down syndrome testing, health professionals who are offering women prenatal testing should be targeted as distribution of DAs by health professionals has been shown to be the most promising implementation strategy of SDM to date [22]. According to the theory of planned behavior (TPB) and other sociocognitive theories and models (eg, theoretical domains framework [TDF]), the adoption of a behavior, such as using a DA, is mainly determined by the level of the person's intention to perform the behavior [23,24]. Some studies have suggested that health professionals' intention to use a DA would depend on their specialty [25], experience [26], and cultural beliefs [27]. However, these factors are not easily modifiable, whereas according to the TPB, the sociocognitive factors that influence health professionals' intentions can be modified. Knowledge of these modifiable factors would thus be useful in designing an intervention such as a care plan to foster health professionals' use of a DA for helping pregnant women make informed decisions about prenatal testing for Down syndrome.

Objectives

Therefore, we aimed to identify factors influencing health professionals' intention to use a DA during prenatal visit for decisions about Down syndrome screening.

Methods

Study Design and Context

We conducted a Web-based survey of health professionals in the province of Quebec (Canada) using a Web panel and used Checklist for Reporting Results of Internet E-Survey (CHERRIES) to guide reporting of results [28] (see Multimedia Appendix 1). This study was embedded in a larger research initiative called the PEGASUS project (Personalized Genomics for Prenatal Aneuploidy Screening Using Maternal Blood) aiming to validate the performance and utility of noninvasive prenatal testing in the general population. In this larger initiative, our overarching aim was to produce a DA to foster SDM in the context of prenatal screening for Down syndrome. Our study complements a similar survey of the intentions of pregnant women to use a DA for decisions about prenatal screening for Down syndrome [29]. We obtained ethics approval from the research ethics boards of the Centre de Santé et de Services Sociaux de la Vieille-Capitale (#2013-2014-29) and the CHU de Quebec (#B14-02-1929).

Participants and Recruitment

Prenatal care in the province of Quebec is provided by obstetrician-gynecologists (about 51% of pregnancies), family physicians (about 46%), and midwives (about 3%). These three types of health professionals were eligible to participate, and we expected to recruit a similar proportion of each type as is found in Quebec overall [30]. Recruitment took place from December 18, 2015, to October 4, 2016. Eligible health professionals were as follows: (1) family physicians, midwives, obstetrician-gynecologists, or trainees in these professions; (2) involved in prenatal care; and (3) working in the province of Quebec. We excluded health professionals who were on parental or sick leave and who had participated in a previous phase of the project.

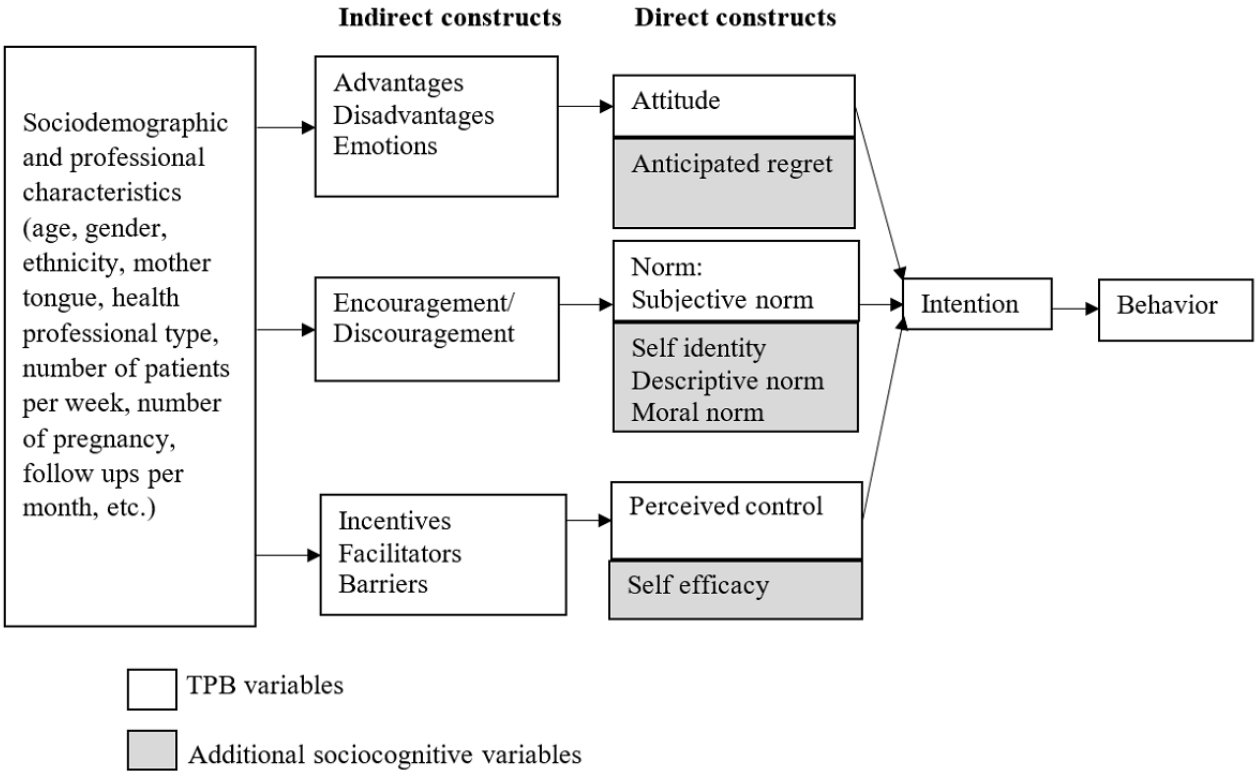
We mandated 2 private firms specialized in polling to program our Web survey and to recruit eligible health professionals in the province of Quebec (Canada). Canada's health care system consists of 13 (10 provincial and 3 territorial) independent health care systems. In this study, we focused on the province of Quebec, which is the second most populous Canadian province. Once the survey was programmed, we emailed invitations that included an open link to the survey as well as other relevant information. This included: (1) study context, (2) study aim, (3) survey content, (4) ethical approval, (5) funding information, (6) information about researchers, (7) time the survey would take (10 minutes to watch the video and 15 minutes for the questionnaire), (8) honoraria offered to eligible participants (50 Canadian dollars), and (9) coordinator contacts. Invited participants were asked to fill out the questionnaire as soon as possible since the link to the survey would be deactivated when the desired number of respondents was reached. We recruited additional eligible participants by the following means: (1) asking family physicians, midwives, and obstetrician-gynecologists known to the team to forward our invitation (snowball method) and (2) asking 3 relevant provincial health professional associations, which agreed to forward the invitation to their members. By clicking on the survey link, interested participants were directed to the open survey. Once eligibility criteria were confirmed (11 filter questions), eligible health professionals started the voluntary survey. No randomization of items or questionnaires was performed.

Data Collection

Clear preliminary statements provided information and instructions for the study and enabled participants to confirm their consent. Participating health professionals completed the Web-based survey through 22 Web pages that each included up to 6 items, appearing in the same order for all participants (see Multimedia Appendix 2). Participants were not expected to have experience in the use of a DA, so to have them understand the behavior of interest (Action: *use*; Target: *a DA for prenatal screening for DS*; Context: *during prenatal care visits*), they were invited to watch a 10-minute video that depicted 2 consecutive prenatal care follow-ups during which a pregnant woman, her partner, and a health professional use a DA to decide whether the woman will undergo prenatal screening for Down syndrome. It presented a health professional in a clinic in a scenario that would be relevant to any type of prenatal care provider. The DA is available in Multimedia Appendix 3. The production of the video and the DA followed validated processes used successfully in previous work [31]. After watching the video, eligible health professionals were directed to the survey.

The Web-based questionnaire was developed using constructs that extended the TPB, including some TDF constructs (Figure 1). We used the 3 direct constructs of the TPB: (1) attitude (perceived advantages or disadvantages of adopting the behavior), (2) subjective norm (the perceived social pressure from significant others to perform the behavior), and (3) perceived behavioral control (perceived control over performing the targeted behavior) to assess health professionals' intention to use a DA for prenatal screening for Down syndrome. We supplemented these with 5 more constructs from theories shown to improve the predictive capacity of the TPB: (1) anticipated regret or potential regret about not having adopted the target behavior; (2) self-identity or one's image of oneself, reflecting the extent to which a person sees himself or herself as fulfilling the criteria for any societal role; (3) moral norm or one's perceived moral duties; (4) descriptive norm or the perceived prevalence of the practice; and (5) self-efficacy or perceived ability to enact the behavior [32-34]. We also measured 7 related sets of behavioral beliefs as elicited through semistructured interviews in previous qualitative studies by our team [35], namely, perceived advantages and disadvantages of using the DA, predicted emotions following its use (beliefs underlying *attitude*), encouragement or discouragement (beliefs underlying *subjective norm*), perceived incentives, facilitators and barriers to using the DA (beliefs underlying *perceived control*; Figure 1).

Figure 1. Extended model of behavior change. TPB: theory of planned behavior.



Questionnaire and Measures

The questionnaire, which was developed both in English and French, included 68 closed items (67 scored on a 7-point Likert-type scale and 1 on a 5-point scale). The questionnaire was pilot tested (2-week test-retest). The aim of the questionnaire was to assess the theory-based factors influencing the use of a DA to decide about prenatal screening [29]. Attitude was measured with 5 items using bipolar adjectives assessing health professionals' cognitive and affective attitudes (ie, very difficult to very easy, useless to very useful). All other direct constructs were assessed with 3 or 4 items, except anticipated regret with 2 items, and all used opposing outcomes (ie, strongly disagree to strongly agree, very unlikely to very likely). Once data collection was completed, the 2 contracted companies sent us the data anonymously, which were then stored on our secure network (password-protected).

Sample Size

On the basis of power and sample size determination for linear models [36], we estimated that a sample size of 350 health professionals would be sufficient to detect a partial correlation of .15 between the intention and a model construct, with a power of 80% and a significance level of 5% for each group. To consider the missing data and ensure that our sample was large enough to perform subgroup analyses, we aimed to recruit 380 health professionals (175 gynecologist-obstetricians, 175 family physicians, and 30 midwives, reflecting the proportions of these health professionals who are involved in prenatal care in Quebec province).

Data Analysis

We used simple descriptive statistics (means, SDs and percentages) to summarize sociodemographics, professional characteristics and sociocognitive variables. For each sociocognitive direct construct and intention, we verified internal consistency by calculating Cronbach alphas. For each sociocognitive construct and continuous sociodemographic variable, we obtained Pearson correlations to assess the strength of their association with intention. For the categorical and dichotomous sociodemographic variables, we performed analysis of variance analyses and *t* tests. We also performed an initial multiple linear regression that included only the TPB variables (attitude, subjective norm and perceived control). Next, we compared the extended TPB model, including the additional variables of anticipated regret, self-identity, moral norm, descriptive norm, and self-efficacy, with the preceding TPB conventional model. We then added the external variables (health professional type and sociodemographics) to the model and used a backward approach in an attempt to obtain an adjusted model with better goodness-of-fit.

To identify significant underlying beliefs, we replaced constructs that significantly determined health professionals' intention with their associated underlying beliefs (eg, attitude was replaced by its underlying beliefs), and performed the regression model also including all other significant direct constructs. Following a backward approach, we kept significant beliefs ($P<.05$) while keeping the direct constructs in the model.

As our study participants included 3 types of health professionals (family physicians, obstetricians or gynecologists, and midwives), all bivariate analyses involving sociocognitive constructs were conducted on each of the 3 subsamples of health

professionals. Due to the small samples and the non-normality of the variables, we used Spearman correlation coefficients to explore the relationship between intention and the sociocognitive constructs per subsample. The multiple regression analysis, including the direct constructs of the extended TPB, was also reproduced in each of these subsamples.

For some of the multiple regression models described earlier, we used deviance and *F* tests to compare nested models and thus identify which one was best. In all regression models, the normality of residuals was satisfying, but their variance was not homogeneous. So we obtained heteroscedasticity-consistent standard errors [37] and reported the corresponding *t* tests for all inferred beta coefficients. All analyses were performed using SAS 9.4 (SAS Institute Inc, Cary, NC).

Results

Participant Characteristics

Figure 2 illustrates the flow of participants. Following CHERRIES [28], we considered as unique visitors all eligible participants who clicked on the personalized link to take the survey. The completion rate (ratio of users who finished the survey to those who agreed to participate) was 84.8% (330/389). The completion time was not kept for analysis as participating health professionals could stop and restart the survey later, and no data were missing as the Web-based survey did not allow participants to proceed beyond unanswered items. The respondents were not able to review and change their answers. Table 1 shows participant characteristics. Briefly, the mean age of the 310 participating health professionals included in the analyses was 39.6 (SD 11.5) years; 81.0% (251/310) were female; and 92.6% (287/310) were French-speaking. Of the 310 health professionals, 55.2% (171/310) were family physicians, 33.9% (105/310) were obstetricians or gynecologists, and 11% (34/310) were midwives. Surprisingly, although 23.6% (73/310) health professionals reported that they were already aware of a DA for decision making about prenatal screening for Down syndrome, only 40.97% (127/310) of health professionals reported that they knew of any DA for diagnostic or treatment decisions. We found no significant association between health professionals' knowledge of DAs (in prenatal screening or other contexts) and their intention to use one in the prenatal screening context.

Figure 2. The flowchart of the participants.

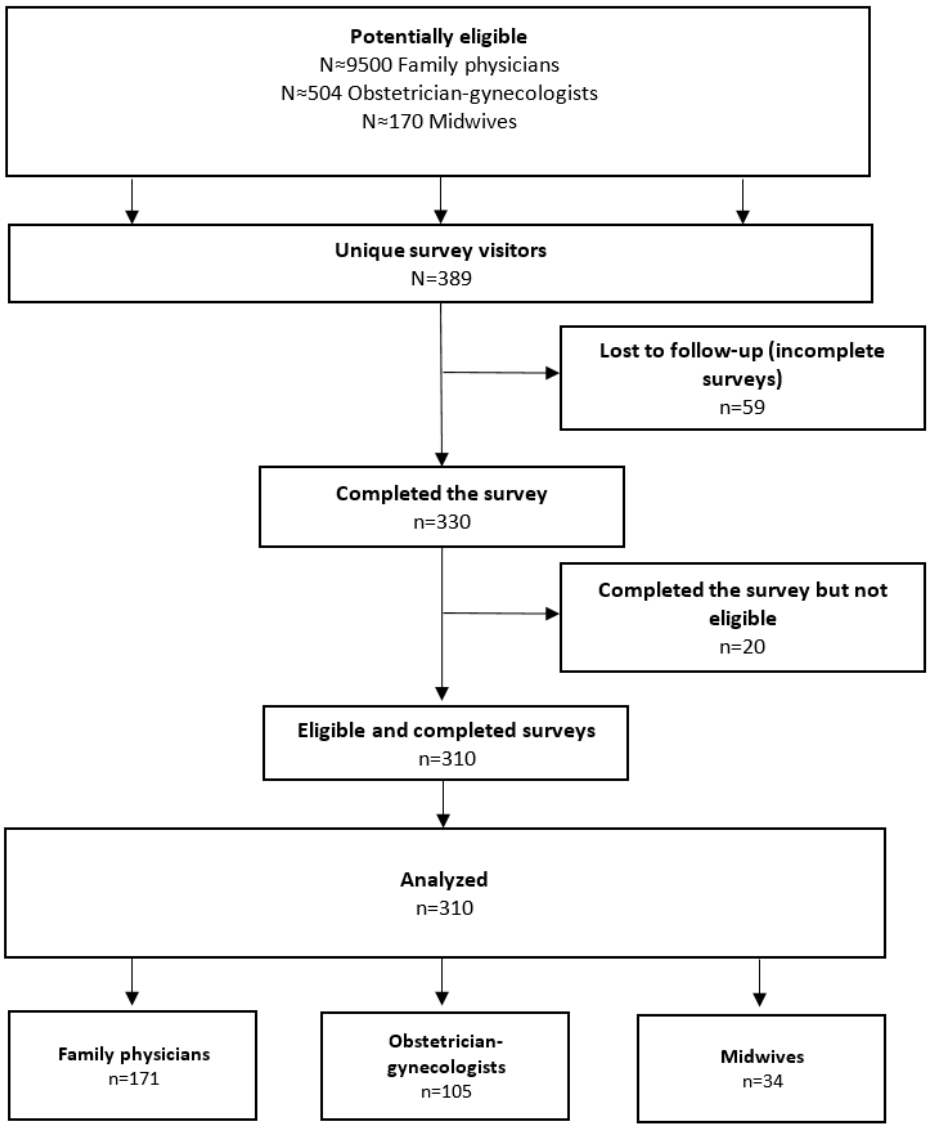


Table 1. Health professionals' characteristics, N=310.

Characteristic	Value
Age in years, mean (SD)	39.6 (11.5)
Number of patients per week, mean (SD)	59.5 (43.1)
Number of pregnancy follow-ups per month, mean (SD)	39.4 (54.1)
Sex, n (%)	
Male	59 (19.0)
Female	251 (81.0)
Type of health professional, n (%)	
Family doctor or family physician or general practitioner	171 (55.2)
Obstetrician/gynecologist	105 (33.8)
Midwife	34 (11.0)
Mother tongue, n (%)	
French	287 (92.6)
English	8 (2.6)
Other	15 (4.8)
Ethnicity, n (%)[a]	
White or Caucasian	289 (93.2)
African or African American	2 (0.7)
Latin American	3 (1.0)
Arab	7 (2.3)
South Asian	2 (0.7)
Southeast Asian	3 (1.0)
Chinese	2 (0.7)
Other	1 (0.3)
I would rather not answer	5 (1.6)
Prior knowledge about a decision aid for prenatal screening for Down syndrome, n (%)	
Yes	73 (23.5)
No	237 (76.5)
Prior knowledge about a decision aid regarding another issue, n (%)	
Yes	127 (41.0)
No	183 (59.0)

[a]Percentages add up to 101.5% because participants were allowed to select more than one ethnicity.

Descriptive and Bivariate Analyses

First, intention and all the 8 direct constructs analyzed showed adequate internal consistency (Cronbach alphas from .75 to .97). Except for anticipated regret (mean score of 3.57 out of 7), mean scores of each construct were higher than 4 out of 7 (4.70 to 5.91) (Table 2), that is, few participants thought they would regret not having used a DA in the context of prenatal screening for Down syndrome.

Bivariate analysis showed that all sociocognitive factors were significantly associated with intention (P<.01 to P<.001; Table 2). All constructs were also significantly associated with each other (P<.05 to P<.001) except for descriptive norm and anticipated regret in association with perceived control. In exploring associations between intention and sociocognitive constructs, because the sample was stratified by health professional type we found that, except for perceived control in the midwife sample, all constructs were associated with intention (P<.05 to P<.001). In order of magnitude, the mean intention scores among health professionals were as follows: midwives 5.78 (SD 0.84), family physicians 5.35 (SD 1.42), and obstetricians or gynecologists 4.91 (SD 1.67).

Table 2. Internal consistency of psychosocial constructs and descriptive analyses.

Constructs	Number of items	Cronbach alpha	Mean (SD)
Intention	3	.97	5.25 (1.48)
Attitude	5	.87	5.05 (0.96)
Subjective norm	3	.87	4.74 (1.11)
Perceived control	3	.75	5.27 (1.22)
Self-efficacy	4	.75	5.81 (0.85)
Self-identity	4	.91	5.06 (1.30)
Descriptive norm	3	.91	4.70 (1.44)
Moral norm	4	.90	5.91 (1.01)
Anticipated regret	2	.77	3.57 (1.54)

Multivariate Analyses

We identified the most significant factors in health professionals' intention to use the DA. In the first multivariate model, including only TPB variables, attitude (beta= 1.104, 95% CI 0.953-1.255) and subjective norm (beta=.157, 95% CI 0.028-0.286) were significant factors of health professionals' intention to use a DA in the context of prenatal Down syndrome screening (Table 3). No sociodemographic variable was added to the model.

In the second multivariate model, based on the extended TPB and thus including the additional variables of self-efficacy, self-identity, descriptive norms, moral norms, and anticipated regret, we found that self-identity (beta=.325, 95% CI 0.186-0.465), attitude (beta=.297, 95% CI 0.168-0.426), moral norm (beta=.288, 95% CI 0.153-0.422), descriptive norm (beta=.166, 95% CI 0.084-0.248), and anticipated regret (beta=.099, 95% CI 0.035-0.164) were significant factors of health professionals' intention to use a DA in the context of prenatal Down syndrome screening (Table 3). In the TPB-only model, the proportion of explained variance was .64, and in the extended model, it was .80. The increase of 16% in the extended model and the comparison of model deviance shows that health professionals' intention was better explained when additional sociocognitive variables were included (Δ deviance $F_{5,301}$=48.34; P<.001); see Table 3).

Table 3. Significant factors in health professionals' intention. TPB: theory of planned behavior; N/A: not applicable.

Construct	TPB, beta (95% CI)	Extended TPB, beta (95% CI)			
	Full sample[a] (N=310)	Full sample[b] (N=310)	Obstetricians/ gynecologists[c] (n=105)	Family physicians/ general practitioners[d] (n=171)	Midwives[e] (n=34)
Attitude	*1.104* (0.953-1.255)[f]	*.297* (0.168-0.426)[f]	*.487* (0.237-0.738)[f]	*.197* (0.035-0.359)[g]	*.297* (0.044-0.549)[g]
Subjective norm	*.157* (0.028-0.286)[g]	−.0041 (−0.098 to 0.089)	−0.061 (−0.223 to 0.100)	.018 (−0.118 to 0.154)	−.013 (−0.108 to 0.082)
Perceived control	.044 (−0.054 to 0.142)	.045 (−0.033 to 0.123)	.031 (−0.081 to 0.143)	.041 (−0.079 to 0.161)	−.018 (−0.134 to 0.097)
Self-efficacy	N/A	.193 (−0.01 to 0.395)	−.048 (−0.346 to 0.251)	*.353* (0.078-0.628)[g]	.061 (−0.175 to 0.297)
Self-identity	N/A	*.325* (0.186-0.465)[f]	*.308* (0.055 to 0.56)[g]	*.275* (0.113-0.438)[h]	*.334* (0.146-0.521)[h]
Descriptive norm	N/A	*.166* (0.084-0.248)[f]	*.201* (0.088 to 0.313)[f]	*.148* (0.030-0.266)[g]	.056 (−0.047 to 0.158)
Moral norm	N/A	*.288* (0.153-0.422)[f]	*.393* (0.162 to 0.624)[h]	*.224* (0.051-0.396)[g]	.233 (−0.051 to 0.517)
Anticipated regret	N/A	*.099* (0.035, 0.164)[h]	.072 (−0.028 to 0.172)	*.155* (0.071-0.238)[f]	−.02 (−0.134 to 0.093)

[a]R^2=.644; deviance=241.80.

[b]R^2=.803; deviance=134.12; $F_{5,301}$=48.34, P<.001 for comparison with TPB model deviance.

[c]R^2=.869; deviance=37.85.

[d]R^2=.766; deviance=80.37.

[e]R^2=.758; deviance=5.70.

[f]P<.001.

[g]P<.05.

[h]P<.01.

Table 4. Significant beliefs of health professionals. N/A: not applicable.

Construct	Underlying belief	Beta (SE)	*P* value
Attitude	Advantages: Using a decision aid would promote decision making	.117 (.037)	.002
	Emotions: Using a decision aid would reassure me	.100 (.038)	.01
	Disadvantages: Using a decision aid might require more time than planned for the consultation	−.077 (.024)	.001
Self-identity	N/A	.366 (.070)	<.001
Moral norms	N/A	.319 (.064)	<.001
Descriptive norms	N/A	.209 (.043)	<.001
Anticipated regret	N/A	.094 (.031)	.003

Analyses of the extended model within the strata of health professional type suggested some differences in the size effects of the factors, but due to collinearity issues we were not able to assess the statistical significance of these differences. For the obstetrician or gynecologist subgroup (N=105), we found that, in order of importance, attitude (beta=.487, 95% CI 0.237-0.738), moral norm (beta=.393, 95% CI 0.162-0.624), self-identity (beta=.308, 95% CI 0.055-0.560), and descriptive norm (beta=.201, 95% CI 0.088-0.313) were significant factors of intention. For the family physician subgroup, we found that self-efficacy (beta=.353, 95% CI 0.078-0.628), self-identity (beta=.275, 95% CI 0.113-0.438), moral norm (beta=.224, 95% CI 0.051-0.396), attitude (beta=.197, 95% CI 0.035-0.359), anticipated regret (beta=.155, 95% CI 0.071-0.238), and descriptive norm (beta=.148, 95% CI 0.030-0.266), were significant factors of intention. Finally, for the midwife subgroup, we found that self-identity (beta=.334, 95% CI 0.146-0.521) and attitude (beta=.297, 95% CI 0.044-0.549) were significant factors of intention (Table 3).

Attitude was the only significant construct among TPB variables. Thus, to identify significant underlying beliefs, we performed an additional multiple regression model where attitude was replaced with its underlying beliefs (Table 4). From this, we found 3 significant beliefs related specifically to the attitude construct, namely, that the use of a DA (1) would promote decision making (beta=.117, 95% CI 0.043-0.190), (2) would reassure health professionals (beta=.100, 95% CI 0.024-0.175), and (3) might require more time than planned for the consultation (beta=−.077, 95% CI −0.124 to −0.031).

Discussion

Principal Findings

With the aim of helping health professionals to support women to make informed, value-congruent decisions about prenatal testing, we identified psychosocial factors influencing the intentions of midwives, family physicians, and obstetricians or gynecologists to use a DA during a prenatal visit for decisions about Down syndrome screening. We found the following: (1) less than half of the health professionals were aware of DAs for contexts other than prenatal screening, and few of them knew of any DA for prenatal screening for Down syndrome; (2) all psychosocial measures except for anticipated regret scored high; (3) overall intention was high among health professionals but varied across the type of health professional, and attitude, self-identity, descriptive norm, moral norm and anticipated regret were all associated with intention to use a DA for prenatal screening among all types of health professionals; and (4) 3 significant beliefs related to attitude in all groups were that the use of a DA would promote decision making, would reassure health professionals, and might require more time than planned for the consultation. These results lead us to make the following four main observations.

First, health professionals do not know enough about DAs in any context, including in the prenatal screening context. Studies show the important role of health professionals in the delivery of DAs [22,38]. Our results, showing that more than half of all health professionals surveyed had never come across any DAs, indicate that more needs to be done to distribute DAs in health care systems and make health professionals aware of them. This concurs with results of qualitative studies on health professionals' attitudes to DAs, suggesting that their lack of awareness of the existence of DAs was a major barrier to their use [39,40]. In addition, we observed that the video was needed for showing health professionals what a DA was, as well as how it can be used during a consultation, to ensure that they had a clear idea of what they were being surveyed about. This result also suggests that another criterion could be added to the International Patient Decision Aids Standard, that is, that the purpose and potential use of the DA is comprehensible to health professionals. Further studies assessing factors influencing the use of DA should start with asking about target participants' awareness of DAs and make sure they know what a DA is and how they can be used in clinical settings.

Second, few participants thought they would regret not having used the DA. Regret is "a comparison-based emotion of self-blame, experienced when people realize or imagine that their present situation would have been better had they decided differently in the past" [41]. One of the conditions that determines anticipated regret is when the preferred option is not necessarily superior to other options [41,42]. It could be that health professionals feel they are already engaging their patients adequately in the decision about prenatal testing, and that the option of using the DA is not greatly superior to their own efforts. This is supported by studies showing that health professionals tend to think they are engaging in SDM more than their patients think they are [43]. Further investigation is required to identify the main reason for health care professionals reporting they do not anticipate regret for not using a DA in this context.

Third, intention to use a DA was high overall among health professionals. These results are congruent with previous studies showing high levels of intention among health professionals to engage in SDM in clinical contexts (including prenatal screening) [44]. Interestingly, some of our previous work has shown that pregnant women also have high levels of intention to use a DA for Down syndrome screening decisions [29,45]. Together, these results suggest that both health professionals and pregnant women seem inclined to use a DA in Down syndrome prenatal screening, and that lack of intention among either health professionals or pregnant women is not the cause of failure to implement DAs in this context. Although similar factors influenced their behavior (attitude, moral norm, descriptive norm, and anticipated regret), future interventions will need to be tailored to each member of the dyad.

However, although overall intention was high, we also observed that health professionals' intention to use a DA for prenatal Down syndrome screening may differ by type: midwives had the highest intention and obstetrician or gynecologists the lowest. Our results are the first to document that this intention varies across types of health professional. The variation could be due to obstetricians or gynecologists feeling that they have less time to enter into a lengthy discussion about Down syndrome prenatal screening. One study highlighted the importance of using DAs in a flexible manner, that is, adapted to timing appropriate to the needs of different types of health professionals [40]. The variation in intention could also be explained by their different views regarding their role and responsibilities [46-48]. For instance, although midwives see their role more as one of "providing information" and letting the patient decide, physicians more often consider their role as that of an "advisor" or an "educator" and feel the decision is their responsibility [48]. These differences reflect differences in training, philosophy, professional culture, and practice among the 3 types of health care professionals [46,47]. Researchers, curriculum developers, and providers of continuing education should adapt SDM training to the different types of health professionals.

Fourth, our findings indicate that health professionals' intention to use a DA in this context is determined by, in order of importance, their image of themselves as fulfilling a societal role (*self-identity*); the consequent advantages, disadvantages, and emotions they perceive (*attitude*); its compatibility with their moral values (*moral norm*); their perception of how much other health professionals use DAs (*descriptive norm*); and the regret they perceive they might feel if they do not use it (*anticipated regret*). These results align with our earlier qualitative results regarding these factors [35]. Interventions to foster the use of a DA for Down syndrome prenatal screening by health professionals should address these factors, for example, by introducing the advantages of using the DA (attitude), spreading the culture of using DAs through social media (moral norms, self-identity), presenting the use of DAs as a desirable practice (descriptive norm), and suggesting to health professionals that they might regret not using it (anticipated regret).

In addition, three significant underlying beliefs were identified. One was the belief that it might require more time than planned for the consultation. In a 2017 Cochrane systematic review of DAs, 8 out of 10 studies that measured consultation length reported no significant difference for the DA group compared with the control group [20]. Further studies are required to investigate if the use of DA takes more time or not. Key statements regarding these 3 salient beliefs could be added to SDM training materials (eg, continuing professional development course material, videos) to increase health professionals' intention to use DAs in clinical practice.

Limitations

This study has a few limitations. First, those who agreed to participate may have been more inclined to use DAs than those who did not. Second, the study targeted health professionals in Quebec, that is, in one health care system, so we cannot infer that our results are applicable to other health care systems including those in other Canadian provinces and territories. Although our results can inspire other efforts, interventions need to be adapted to each prenatal care pathway. Third, the invitation to complete the survey was sent to health professionals' organizations and personal email lists, and so calculation of a precise view rate (ratio of unique survey visitors to unique receivers of survey invitation) was not possible. Finally, we focused on 3 types of health professionals who are directly involved in the decision-making process with couples (family physicians, obstetrician or gynecologists, and midwives), although other health professionals, such as nurses and geneticists, are also likely to be involved at some stage of prenatal follow up and in prenatal screening decisions. Further studies will be needed to elucidate their specific roles and beliefs.

Conclusions

On the basis of a theoretical approach to behavior change and following best practices for conducting Web-based surveys [28], this study identifies psychosocial factors that could influence health professionals' intention to use a DA for helping pregnant women make informed decisions about Down syndrome screening, and suggests which factors will need to be addressed in an intervention to increase their intention. An earlier study investigating factors influencing intention to use such a DA among pregnant women observed high levels of intention, and in this study too, in general, all types of health professional showed high intention. These combined results, as well as our new detailed information on what behavioral factors to address, lead us to suggest that the time is ripe for implementing an intervention to foster DA use in this context.

Acknowledgments

For their collaboration in the recruitment of health professionals, the authors acknowledge the following 3 organizations: the Association des omnipraticiens en périnatalité du Québec [Quebec association of perinatal physicians], the Association des

obstétriciens et gynécologues du Québec [Quebec association of obstetrician-gynecologists], and the Regroupement des sages-femmes de Québec [Quebec association of midwives]. The authors acknowledge the following 3 physicians: Patrick Archambault, Emmanuel Bujold, and Caroline Fiset. The authors also thank the editor and two peer reviewers for their constructive comments, as well as Louisa Blair for editing the paper.

The authors are investigators in a larger research project (PEGASUS) funded under the auspices of Génome Québec, Genome Canada, the Government of Canada, the Ministère de l'Enseignement Supérieur de la Recherche de la Science et de la Technologie du Québec [the Quebec ministry of higher education, research, science and technology], and the Canadian Institutes for Health Research, all nonprofit government-funded organizations. As requested by these organizations, the project received additional funding from private partners that provide commercial genomics-based NIPT (Ariosa Diagnostics Inc, San Jose, CA) and reagents and equipment that can be used to perform genomics-based non-invasive prenatal assays (Thermo Fisher Scientific, Waltham, MA; Illumina, Inc, San Diego, CA; QIAGEN, Hilden, Germany; Perkin Elmer, Inc,Waltham, MA). Private funding was at arm's length. None of the funders had any role in the design of the study, interpretation of the results, or approval of the manuscript.

Authors' Contributions

FL, HR, JL, AG, BW, and FR conceived the study and participated in its design, method, and coordination. IL helped in recruiting health professionals and in data collection. JC performed analysis. SAR, FL, JC, and HR contributed to data interpretation. SAR, FL, and HR drafted the manuscript. All authors contributed ideas, read, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Checklist for Reporting Results of Internet E-Surveys (CHERRIES).

[PDF File (Adobe PDF File), 68KB - jmir_v20i4e114_app1.pdf]

Multimedia Appendix 2

Online questionnaire.

[PDF File (Adobe PDF File), 1MB - jmir_v20i4e114_app2.pdf]

Multimedia Appendix 3

The decision aid for prenatal screening for Down syndrome.

[PDF File (Adobe PDF File), 587KB - jmir_v20i4e114_app3.pdf]

References

1. St-Jacques S, Grenier S, Charland M, Forest J, Rousseau F, Légaré F. Decisional needs assessment regarding Down syndrome prenatal testing: a systematic review of the perceptions of women, their partners and health professionals. Prenat Diagn 2008 Dec;28(13):1183-1203. [doi: 10.1002/pd.2059] [Medline: 19097031]
2. Woolf SH, Chan ECY, Harris R, Sheridan SL, Braddock CH, Kaplan RM, et al. Promoting informed choice: transforming health care to dispense knowledge for decision making. Ann Intern Med 2005 Aug 16;143(4):293-300. [Medline: 16103473]
3. O'Connor AM, Drake ER, Wells GA, Tugwell P, Laupacis A, Elmslie T. A survey of the decision-making needs of Canadians faced with complex health decisions. Health Expect 2003 Jun;6(2):97-109 [FREE Full text] [Medline: 12752738]
4. Légaré F, O'Connor A, Graham I, Saucier D, Côté L, Cauchon M, et al. Supporting patients facing difficult health care decisions: use of the Ottawa Decision Support Framework. Can Fam Physician 2006 Apr;52:476-477 [FREE Full text] [Medline: 17327891]
5. Asplin N, Wessel H, Marions L, Georgsson Öhman S. Pregnant women's experiences, needs, and preferences regarding information about malformations detected by ultrasound scan. Sex Reprod Healthc 2012 Jun;3(2):73-78. [doi: 10.1016/j.srhc.2011.12.002] [Medline: 22578754]
6. Green J, Hewison J, Bekker H, Bryant L, Cuckle HS. Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review. Health Technol Assess 2004 Aug;8(33):G-124 [FREE Full text] [Medline: 15298822]
7. Lou S, Mikkelsen L, Hvidman L, Petersen OB, Nielsen CP. Does screening for Down's syndrome cause anxiety in pregnant women? A systematic review. Acta Obstet Gynecol Scand 2015 Jan;94(1):15-27. [doi: 10.1111/aogs.12482] [Medline: 25146713]
8. Georgsson Öhman S, Sissel S, Ulla W, Charlotta G, Sonja OL. Pregnant women's responses to information about an increased risk of carrying a baby with Down syndrome. Birth 2006 Mar;33(1):64-73. [Medline: 16499533]

9. van Gameren-Oosterom HM, Fekkes M, Oudesluys-Murphy AM, van Wouwe JP. [Young people with Down syndrome: independence and social functioning]. Ned Tijdschr Geneeskd 2014;158:A7983. [Medline: 25406816]
10. Shojai R, Boubli L, d'Ercole C. [Prognosis work-up in prenatal medicine: the example of Down's syndrome]. Gynecol Obstet Fertil 2005;33(7-8):514-519. [doi: 10.1016/j.gyobfe.2005.05.015] [Medline: 16005658]
11. Farrell R, Nutter B, Agatisa PK. Patient-centered prenatal counseling: aligning obstetric healthcare professionals with needs of pregnant women. Women Health 2015;55(3):280-296. [doi: 10.1080/03630242.2014.996724] [Medline: 25794055]
12. Gagnon S, Labrecque M, Njoya M, Rousseau F, St-Jacques S, Légaré F. How much do family physicians involve pregnant women in decisions about prenatal screening for Down syndrome? Prenat Diagn 2010 Feb;30(2):115-121. [doi: 10.1002/pd.2421] [Medline: 20013876]
13. O'Connor AM. Validation of a decisional conflict scale. Med Decis Making 1995;15(1):25-30. [doi: 10.1177/0272989X9501500105] [Medline: 7898294]
14. Marteau TM, Dormandy E, Michie S. A measure of informed choice. Health Expect 2001 Jun;4(2):99-108 [FREE Full text] [Medline: 11359540]
15. Chiang HH, Chao YM, Yuh YS. Informed choice of pregnant women in prenatal screening tests for Down's syndrome. J Med Ethics 2006 May;32(5):273-277 [FREE Full text] [doi: 10.1136/jme.2005.012385] [Medline: 16648277]
16. Légaré F, Witteman H. Shared decision making: examining key elements and barriers to adoption into routine clinical practice. Health Aff (Millwood) 2013 Feb;32(2):276-284. [doi: 10.1377/hlthaff.2012.1078] [Medline: 23381520]
17. Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. Patient Educ Couns 2006 Mar;60(3):301-312. [doi: 10.1016/j.pec.2005.06.010] [Medline: 16051459]
18. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango). Social Science & Medicine 1997 Mar;44(5):681-692. [doi: 10.1016/S0277-9536(96)00221-3]
19. The Secretary of State for Health. Equity and Excellence: Liberating the NHS. 2010. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf [accessed 2018-04-16] [WebCite Cache ID 6yjCgI84h]
20. 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 Dec 12;4:CD001431. [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
21. Légaré F, Ratté S, Gravel K, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: update of a systematic review of health professionals' perceptions. Patient Educ Couns 2008 Dec;73(3):526-535. [doi: 10.1016/j.pec.2008.07.018] [Medline: 18752915]
22. Hsu C, Liss D, Frosch D, Westbrook EO, Arterburn D. Exploring Provider Reactions to Decision Aid Distribution and Shared Decision Making: Lessons from Two Specialties. Med Decis Making 2017 Dec;37(1):113-126. [doi: 10.1177/0272989X16671933] [Medline: 27899745]
23. Godin G. Les comportements dans le domaine de la santé; Comprendre pour mieux intervenir. Montréal: Les Presses de l'Université de Montréal; 2012.
24. Fishbein M, Triandis H, Kanfer FH, Becker M, Middlestadt SE, Eichler A. Factors influencing behavior and behavior change. In: Baum A, Revenson T, Singer J, editors. Handbook of Health Psychology. New Jersey: Lawrence Erlbaum Associates; 2001:3-17.
25. Legare F, Godin G, Ringa V, Dodin S, Turcot L, Norton J. Variation in the psychosocial determinants of the intention to prescribe hormone therapy prior to the release of the Women's Health Initiative trial: a survey of general practitioners and gynaecologists in France and Quebec. BMC Med Inform Decis Mak 2005 Sep 08;5:31 [FREE Full text] [doi: 10.1186/1472-6947-5-31] [Medline: 16150149]
26. van der Horst K, Giger M, Siegrist M. Attitudes toward shared decision-making and risk communication practices in residents and their teachers. Med Teach 2011;33(7):e358-e363. [doi: 10.3109/0142159X.2011.577465] [Medline: 21696268]
27. Rahimi SA, Alizadeh M, Légaré F. Shared decision making in Iran: Current and future trends. Z Evid Fortbild Qual Gesundhwes 2017 Jun;123-124:52-55. [doi: 10.1016/j.zefq.2017.05.018] [Medline: 28549749]
28. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Dec 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
29. Delanoë A, Lépine J, Turcotte S, Leiva Portocarrero ME, Robitaille H, Giguère AM, et al. Role of Psychosocial Factors and Health Literacy in Pregnant Women's Intention to Use a Decision Aid for Down Syndrome Screening: A Theory-Based Web Survey. J Med Internet Res 2016 Oct 28;18(10):e283 [FREE Full text] [doi: 10.2196/jmir.6362] [Medline: 27793792]
30. Association pour la santé publique du Québec (ASPQ). Suivi de grossesse: quel professionnel choisir? Protégez-vous October. 2012. URL: https://www.protegez-vous.ca/association-pour-la-sante-publique-du-quebec/suivi-de-grossesse-quel-professionnel-choisir.html [accessed 2018-04-16] [WebCite Cache ID 6yjEwsuYG]
31. Stacey D, Brière N, Robitaille H, Fraser K, Desroches S, Légaré F. A systematic process for creating and appraising clinical vignettes to illustrate interprofessional shared decision making. J Interprof Care 2014 Sep;28(5):453-459. [doi: 10.3109/13561820.2014.911157] [Medline: 24766619]
32. Conner M, Armitage CJ. Extending the Theory of Planned Behavior: A Review and Avenues for Further Research. J Appl Social Pyschol 1998 Aug;28(15):1429-1464. [doi: 10.1111/j.1559-1816.1998.tb01685.x]

33. Rivis A, Sheeran P. Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. Curr Psychol 2003 Sep;22(3):218-233. [doi: 10.1007/s12144-003-1018-2]
34. Sandberg T, Conner M. Anticipated regret as an additional predictor in the theory of planned behaviour: a meta-analysis. Br J Soc Psychol 2008 Dec;47(Pt 4):589-606. [doi: 10.1348/014466607X258704] [Medline: 18039428]
35. Lépine J, Leiva Portocarrero ME, Delanoë A, Robitaille H, Lévesque I, Rousseau F, et al. What factors influence health professionals to use decision aids for Down syndrome prenatal screening? BMC Pregnancy Childbirth 2016 Dec 05;16:262 [FREE Full text] [doi: 10.1186/s12884-016-1053-2] [Medline: 27596573]
36. Castelloe JM, O'Brien RG. Power and Sample Size Determination for Linear Models. 2001 Presented at: Twenty-sixth Annual SAS Users Group International Conference; 2001; Cary, NC: SAS Institute Inc p. 240-226.
37. Scott Long J, Ervin H. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. The American Statistician 2000 Aug;54(3):217-224. [doi: 10.1080/00031305.2000.10474549]
38. Graham ID, Logan J, Bennett CL, Presseau J, O'Connor AM, Mitchell SL, et al. Physicians' intentions and use of three patient decision aids. BMC Med Inform Decis Mak 2007 Jul 06;7:20 [FREE Full text] [doi: 10.1186/1472-6947-7-20] [Medline: 17617908]
39. O'Brien MA, Charles C, Lovrics P, Wright FC, Whelan T, Simunovic M, et al. Enablers and barriers to using patient decision aids in early stage breast cancer consultations: a qualitative study of surgeons' views. Implement Sci 2014 Nov 29;9:174 [FREE Full text] [doi: 10.1186/s13012-014-0174-0] [Medline: 25433370]
40. Rees KM, Shaw AR, Bennert K, Emmett CL, Montgomery AA. Healthcare professionals' views on two computer-based decision aids for women choosing mode of delivery after previous caesarean section: a qualitative study. BJOG 2009 Jun;116(7):906-914 [FREE Full text] [doi: 10.1111/j.1471-0528.2009.02121.x] [Medline: 19522794]
41. Zeelenberg M, Pieters R. A Theory of Regret Regulation. Journal of Consumer Psychology 2007 Jan 01;17(1):3-18. [doi: 10.1207/s15327663jcp1701_3]
42. Janis IL, Mann L. Decision making: a psychological analysis of conflict, choice, and commitment. New York, NY: Free Press; 1977.
43. Elwyn G, Edwards A, Kinnersley P, Grol R. Shared decision making and the concept of equipoise: the competences of involving patients in healthcare choices. Br J Gen Pract 2000 Nov;50(460):892-899 [FREE Full text] [Medline: 11141876]
44. Thompson-Leduc P, Clayman ML, Turcotte S, Légaré F. Shared decision-making behaviours in health professionals: a systematic review of studies based on the Theory of Planned Behaviour. Health Expect 2015 Oct;18(5):754-774 [FREE Full text] [doi: 10.1111/hex.12176] [Medline: 24528502]
45. Légaré F, St-Jacques S, Gagnon S, Njoya M, Brisson M, Frémont P, et al. Prenatal screening for Down syndrome: a survey of willingness in women and family physicians to engage in shared decision-making. Prenat Diagn 2011 Apr;31(4):319-326. [doi: 10.1002/pd.2624] [Medline: 21268046]
46. Sutcliffe K, Caird J, Kavanagh J, Rees R, Oliver K, Dickson K, et al. Comparing midwife-led and doctor-led maternity care: a systematic review of reviews. J Adv Nurs 2012 Nov;68(11):2376-2386. [doi: 10.1111/j.1365-2648.2012.05998.x] [Medline: 22489571]
47. Sandall J, Devane D, Soltani H, Hatem M, Gates S. Improving quality and safety in maternity care: the contribution of midwife-led care. J Midwifery Womens Health 2010;55(3):255-261. [doi: 10.1016/j.jmwh.2010.02.002] [Medline: 20434086]
48. Dubé E, Vivion M, Sauvageau C, Gagneur A, Gagnon R, Guay M. How do Midwives and Physicians Discuss Childhood Vaccination with Parents? J Clin Med 2013 Nov 04;2(4):242-259 [FREE Full text] [doi: 10.3390/jcm2040242] [Medline: 26237146]

Abbreviations

SDM: shared decision making
DA: patient decision aid
TPB: theory of planned behavior
TDF: theoretical domains framework

Edited by G Eysenbach; submitted 25.09.17; peer-reviewed by M Skjøth, T Irizarry; comments to author 30.11.17; revised version received 22.12.17; accepted 14.01.18; published 25.04.18.

Please cite as:
Abbasgholizadeh Rahimi S, Lépine J, Croteau J, Robitaille H, Giguere AMC, Wilson BJ, Rousseau F, Lévesque I, Légaré F
Psychosocial Factors of Health Professionals' Intention to Use a Decision Aid for Down Syndrome Screening: Cross-Sectional Quantitative Study
J Med Internet Res 2018;20(4):e114
URL: http://www.jmir.org/2018/4/e114/
doi:10.2196/jmir.9036
PMID:29695369

Original Paper

The Effectiveness of Health Care Information Technologies: Evaluation of Trust, Security Beliefs, and Privacy as Determinants of Health Care Outcomes

Victoria Kisekka[1*], PhD; Justin Scott Giboney[2*], PhD

[1]Information Security and Digital Forensics, School of Business, University at Albany, State University of New York, Albany, NY, United States
[2]Information Technology Department, Brigham Young University, Provo, UT, United States
[*]all authors contributed equally

Corresponding Author:
Victoria Kisekka, PhD
Information Security and Digital Forensics
School of Business
University at Albany, State University of New York
1400 Washington Ave,
Massry Center for Business (BB) 371
Albany, NY, 12222
United States
Phone: 1 518 956 8361
Email: vkisekka@albany.edu

Abstract

Background: The diffusion of health information technologies (HITs) within the health care sector continues to grow. However, there is no theory explaining how success of HITs influences patient care outcomes. With the increase in data breaches, HITs' success now hinges on the effectiveness of data protection solutions. Still, empirical research has only addressed privacy concerns, with little regard for other factors of information assurance.

Objective: The objective of this study was to study the effectiveness of HITs using the DeLone and McLean Information Systems Success Model (DMISSM). We examined the role of information assurance constructs (ie, the role of information security beliefs, privacy concerns, and trust in health information) as measures of HIT effectiveness. We also investigated the relationships between information assurance and three aspects of system success: attitude toward health information exchange (HIE), patient access to health records, and perceived patient care quality.

Methods: Using structural equation modeling, we analyzed the data from a sample of 3677 cancer patients from a public dataset. We used R software (R Project for Statistical Computing) and the Lavaan package to test the hypothesized relationships.

Results: Our extension of the DMISSM to health care was supported. We found that increased privacy concerns reduce the frequency of patient access to health records use, positive attitudes toward HIE, and perceptions of patient care quality. Also, belief in the effectiveness of information security increases the frequency of patient access to health records and positive attitude toward HIE. Trust in health information had a positive association with attitudes toward HIE and perceived patient care quality. Trust in health information had no direct effect on patient access to health records; however, it had an indirect relationship through privacy concerns.

Conclusions: Trust in health information and belief in the effectiveness of information security safeguards increases perceptions of patient care quality. Privacy concerns reduce patients' frequency of accessing health records, patients' positive attitudes toward HIE exchange, and overall perceived patient care quality. Health care organizations are encouraged to implement security safeguards to increase trust, the frequency of health record use, and reduce privacy concerns, consequently increasing patient care quality.

(J Med Internet Res 2018;20(4):e107) doi:10.2196/jmir.9014

KEYWORDS

medical informatics; privacy; quality of health care; trust

Introduction

Background

Today, the health care industry primarily relies on health information technologies (HITs) such as electronic medical record (EMR) systems, patient health record (PHR) systems, and technical devices to deliver patient care services. Despite the continued diffusion of HITs within the health care sector, there is no theory explaining how HIT success influences perceived patient care quality.

Substantial strides have been made to study the success of HITs and their impact on patient care outcomes such as care quality, patient satisfaction, patient empowerment, and increased likelihood of adherence to medications [1-5]. But, with the increase in data breaches and privacy concerns [6], the success of HITs is now also contingent on how well the privacy of patient medical data is secured. There is a scarcity of empirical research evaluating the success of HITs from the perspective of information assurance. *Information assurance* is the protection of information and information systems, the detection of threats, and the reaction to threats [7]. Existing research has not answered the question: how does the success of the information assurance attributes of HITs influence perceived patient care quality?

Existing work misses 3 critical components of information assurance namely—information security beliefs, privacy concerns, and trust in health information. *Information security beliefs* are the perception of the user that data provided to the organization will be accurate and available. *Privacy concerns* are the perceived lack of confidentiality of personal information provided to the organization. Trust is the perception of the user that health information provided by the organization is reliable [8]. From an organization standpoint, organizations increase security beliefs by enacting security and privacy controls, undertaking tasks that ensure data accuracy and availability, and developing controls to protect the confidentiality of user data [8]. This research advances our understanding of the role of information security beliefs, privacy concerns, and trust in health information in a success model of a health care system. We extend existing research by going beyond the influence of privacy concerns. We included 2 new determinants of patient care quality—information security beliefs and trust in health information.

The objective of this research is threefold; first, we seek to examine the role of information assurance constructs (ie, information security beliefs, privacy concerns, and trust in health information) as measures of HIT effectiveness. Second, we seek to empirically investigate the critical yet unknown relationship between information assurance constructs and three aspects of system success: attitude toward health information exchange (HIE), patient access to health records, and perceived patient care quality. Third, our research extends current literature by extending the DeLone and McLean Information Systems Success Model (DMISSM) to the information assurance area in a health care context. We added a new variable to the model namely—attitude toward HIE.

Prior Literature

Ever since the call for improving patient care quality outcomes by Institute of Medicine [9]; a growing stream of health care research has explored the relationship between HIT use and numerous aspects of patient satisfaction with the health care organization. The correlates of HIT use include improved care coordination, enhanced communication between providers and patients, and increased effectiveness in various measures of quality outcomes and provider performance [10-16]. Han et al [12] showed that meeting the objectives of HIT use resulted in increased patient adherence to recommended diabetes tests and reduced hospital utilization. Similar findings related to better medication management and adherence was reported in recent studies [17,18].

As health care is a service, it is important to understand users' perceptions of HIT and service quality. Setia et al [19] theoretically and empirically demonstrate the link between information quality and service capabilities and performance. Their findings inform us that improving information quality enhances the effectiveness of service quality efforts.

A major perception of an HIT by users is information assurance. Brown et al [20] explain the trade-off between information privacy controls and patients' access to electronic health records. They show that obtaining an optimal balance between information privacy controls and access to patients' information primarily requires a clear understanding of the patient. Efforts for optimizing patient care quality outcomes demand health care providers to accurately identify patients' preferences for privacy to determine the acceptable levels of access to sensitive health information without violating patients' privacy. Brown et al's [20] work also underscores the importance of privacy controls in achieving positive patient care outcomes.

In the context of privacy concerns, the existing literature suggests that privacy concerns influence not only patients' perceptions of patient care quality but also behavioral intentions of HIT usage [21-25]. These findings suggest that patient satisfaction and confidence in using provider-managed technologies such as EMR and PHR technologies and nonprovider-managed HITs such as health social networks and other Web-based health information resources are obtained through strong perceptions of the effectiveness of security and privacy controls [26]. In fact, patients often express a preference for security features in Web-based patient portals [27]. We extend this body of literature by empirically testing the influence of information security beliefs, privacy concerns, and trust in health information; the role of these factors as measures of HIT effectiveness is currently not well understood. We theoretically extend existing literature using the DMISSM.

The DeLone and McLean Information Systems Success Model in Health Care

In this study, we examined the effect of HITs on perceived health care service quality. Health care is made of HITs and services that interact together to deliver patient care [28,29]. Because this research investigates the effectiveness of HITs within health care, it is necessary to adopt a theoretical framework that can explain how the components interact and

lead to perceived patient care quality. Thus, we employ the DMISSM presented in Figure 1 [30]. The DMISSM was developed to help organizations understand the benefits of information systems (IS) and how the effectiveness of IS impacts users and organizations. The model has been widely adopted to better understand IS success in different contexts [30]. We limit our discussion of DMISSM to its application in the health care discipline. The remainder of this section explains how the DMISSM has been used in previous HIT research.

As shown in Figure 1, the dimensions of success in the model are information quality, system quality, service quality, system use or intention to use, and user satisfaction [30,31]. In the context of patient safety, *information quality* refers to the completeness, relevance, accuracy, and timeliness of medical information; *system quality* refers to the usability, compatibility, reliability, and response time of the HIT; and *service quality* refers to the technical support and assurance (availability, integrity, and authenticity) of the HIT as well as the quality of service received [31]. Patients and health care providers alike, evaluate HITs in terms of information quality, system quality, and service quality [32,33], all of which are components of the DMISSM [30,31]. Beyond just HIT features, a patient's decision to use HIT also depends on nontechnical success factors and facilitating conditions such as behavioral controls and work processes [34], HIT cost, the individual's technical background and skill set, health conditions (eg, visual impairment), and information privacy concerns [35]. These constructs are commonly studied along with success measures of perceptions about the system such as the ease of use, usefulness, and enjoyment [36]. A large part of the purpose of an HIT is protecting and improving patient safety [31].

A significant component of HIT is how both HIT functionality and HIT use by patients influence patient care quality. Consistent with the DMISSM, HIT has 3 dimensions of quality: information, system, and service [30,37]. In turn, the perceptions of the dimensions of quality should positively influence the intention and decision to use the HIT, which consequently impacts the perceived benefit to the user. The perceived benefit to the user in our study is patient care quality. To our knowledge health care and IS research scholars have not yet made the critical link between patients' experiences with and perceptions of HITs and patient care quality outcomes. Existing research focuses on the relationship between HIT implementation and health outcomes from a macro level (ie, are there societal benefits?) [38] but not from a micro level (ie, are there individual benefits?). Another current gap relates to studying information assurance variables as dimensions of IS success. The rife of data breaches and patients' concern for privacy has created the need to understand HIT effectiveness from the perspective of their security. This research addresses the limitation by studying system quality and service quality in terms of the effectiveness of information security controls and information privacy safeguards.

In summary, many papers study the adoption of HITs but stop short of investigating how patients' use of HITs affects the ultimate goal of health care organizations: quality of care (eg, [39]). Other papers have studied health outcomes in the context of patient HIT success without statistical analysis [20]. Still, others have studied the continued use of patient HIT as the ultimate dependent variable and focus solely on the technology-of-interest [21]. Papers have also studied the adoption of HIT by health care providers [40,41]. We build on this research by empirically testing the effect of patients' perceptions of HIT quality measures and the impact of patient care quality.

Contextualized Hypotheses

This research studies the technical and social elements of a health care system. In the next paragraphs, we develop hypotheses for the model (see Figure 2).

Patient Care Quality

The ultimate dependent variable and the raison d'être for a health care system is to increase the quality of life of patients. Closely akin to the actual quality of care is the perceived quality of care of the patient. If patients perceive to have received high-quality care, it is likely that they are actually receiving quality care. Therefore, *patient care quality* is a perception of one's belief that he or she is receiving the best possible care from the health care system.

Figure 1. Adapted from the DeLone and McLean Information Systems Success Model.

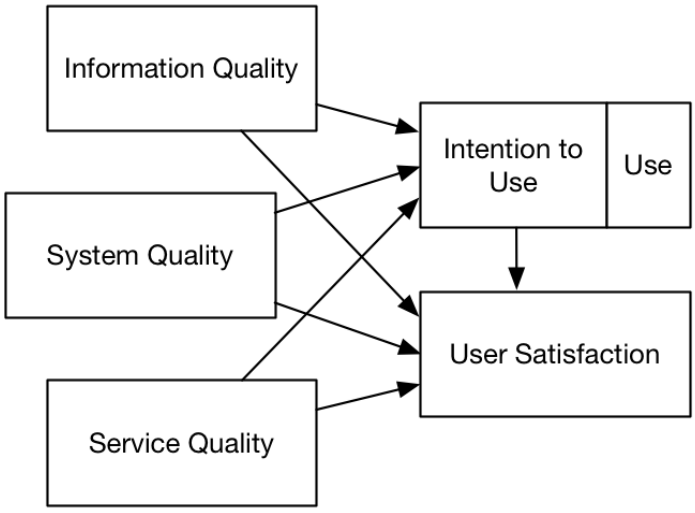
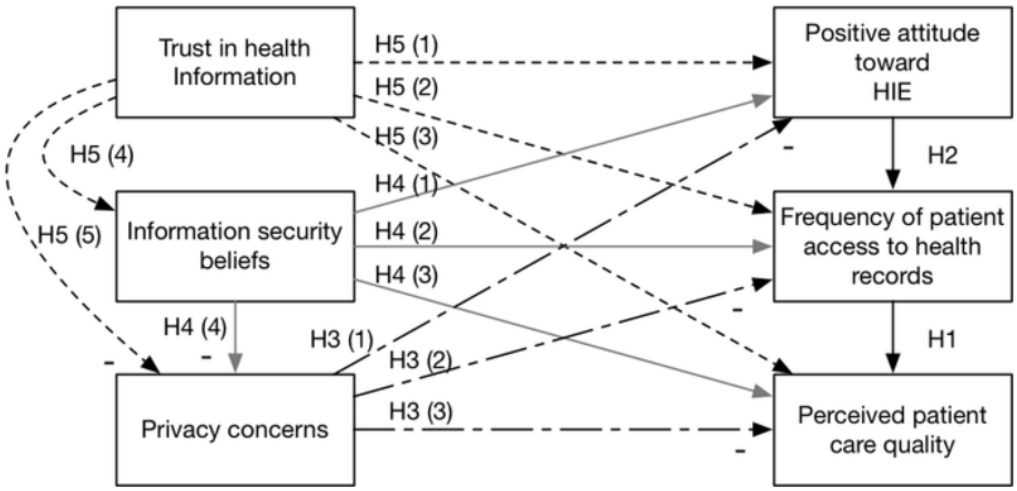


Figure 2. The health care system success model. HIE: health information exchange.



The Frequency of Patient Access to Health Records

This variable pertains to the frequency with which the patient uses the EMR to access his/her health records. Patient satisfaction has been shown to be positively associated with access to health records [42]. Under the Health Insurance Portability and Accountability Act regulation, patients have the legal right to access their medical records. Health care providers are increasingly adopting technologies, such as EMRs, to not only interact with their patients but also comply with government regulations [43]. According to DMISSM, when users engage with a system that helps them achieve their goals, they become more satisfied with the system [44]. It has also been previously established that patient access to medical records promotes communication between patients and physicians, consequently improving the quality of care that the patient receives [45,46]. Thus, we hypothesize the following:

> *H1. Patient access to health records increases the level of perceived patient care quality.*

Attitude Toward Health Information Exchange

This variable pertains to the patient's attitude toward information sharing among health care providers. A significant enabler of coordination in a health care system is the ability to share medical information electronically. Sharing of medical records facilitates timely delivery of care, which is a benefit for the patient. Sharing of medical information is, therefore, a valuable service for patients, as has been shown by researchers [47]. We argue that because of the value associated with the technical capabilities of information sharing, patients who support provider use of EMRs for information exchange may increase their own access to EMR. Patients are more accepting of HITs when they perceive the technologies to be beneficial to care delivery [48]. This acceptance includes patients' actual use of HITs and support and endorsement of providers' use of HITs [48]. Also, DMISSM explains that information quality and the service quality of the system increase the intention to use a system and actual use of the system [44]. Patients who desire that their health care provider exchange information using HITs are showing an intention to use EMR and participate in the health care organization as a whole. Thus, we hypothesize the following:

> *H2. Positive attitudes toward HIE increase the patient's access to health records.*

Privacy Concerns

There is an increasing amount of research related to the privacy of one's information [49]. Privacy is especially important with regard to health care information [35]. People often desire to keep their EMR out of the public domain. Privacy concerns in the health care system are part of service quality. According to DMISSM, service quality increases intention to use and use of a system [44]. This is because users perceive the system to be more reliable. Because privacy concerns are a negative measure of service quality, as privacy concerns increase, use of EMR, positive attitude toward HIE, and ultimately, perceived quality of care will decrease. Thus, we hypothesize the following:

> *H3. Privacy concerns will*
> 1. *decrease patients' positive attitudes toward HIE,*
> 2. *decrease the frequency of patient access to health records, and*
> 3. *decrease the level of perceived patient care quality.*

Information Security Beliefs

Information security beliefs are related to privacy concerns in that both deal with the assurance (integrity, availability, and authenticity) of health information. Information security beliefs are distinct from privacy concerns because information security beliefs are the idea that the system is protecting health information, and privacy concerns are the worry that health information will not be confidential. Information security beliefs will decrease privacy concerns. Also, information security beliefs are part of service quality of the health care system. Just as privacy concerns influence intention to use and use of the system, an increase in information security beliefs will increase use of EMR, positive attitude toward HIE, and ultimately, perceived quality of care. Thus, we hypothesize:

> *H4. Information security beliefs will*
> 1. *increase patients' positive attitudes toward HIE,*

2. *increase the frequency of patient access to health records,*
3. *increase the level of perceived patient care quality, and*
4. *decrease privacy concerns.*

Trust in Health Information

As a component of information quality, trust in internet health information increases a user's expected improved decision making and positive outcomes [44]. As a user's expected improved decision making and positive outcomes increase, the likelihood that the user will continue to use the system increases. Thus, as health care system users believe that they are receiving quality information from the internet, they will continue to participate in the health care system (as explained by the DMISSM). Specifically, as trust in internet health information increases, use of EMR, positive attitudes toward HIE, and ultimately, perceived quality of care will also increase. Furthermore, the concept that health information will be tampered with will decrease. As trust in internet health information increases, a person's information security beliefs will also increase and their privacy concerns will decrease. Thus, we hypothesize:

H5. Trust in internet health information will
1. *increase patients' positive attitudes toward HIE,*
2. *increase the frequency of patient access to health records, and*
3. *increase the level of perceived patient care quality.*
4. *increase information security beliefs and*
5. *decrease privacy concerns.*

Methods

This study employed a structural equation modeling (SEM) of the Health Information National Trends Survey (HINTS) of cancer patients to test our hypotheses [50]. SEM was chosen as the appropriate method because we are simultaneously analyzing multiple paths and multiple dependent variables, and we have a large enough sample that use of partial least squares is unnecessary. We used the HINTS 4 Cycle 4 dataset [50]. The HINTS datasets are publically available responses to surveys about health-related topics. Since the dataset is anonymized public data, there was no need for IRB approval, however, all data were kept confidential. For full details regarding the method of survey collection, see National Trends Survey [50]. The HINTS 4 Cycle 4 survey contained questions related to our phenomena of interest. Table 1 contains the questions from the survey with their corresponding construct.

The survey targeted known minority and nonminority populations. The survey targeted 1 adult per household in selected areas of the United States. In total, 3677 of 13,996 surveys (26.27%) were completed. Of the participants, 467 (12.70%) were in the age group of 18 to 34 years, 743 (20.21%) were aged between 35 and 49 years, 1220 (33.18%) were aged between 50 and 64 years, 637 (17.32%) were in the age group of 65 to 74 years, 428 (11.64%) were older than 75 years, and 182 (4.95%) did not specify. Of the participants, 2184 (59.40%) were female, 1424 (38.72%) were male, and 69 (1.88%) did not specify. Of the participants, 90 (2.45%) had completed less than 8 years of school, 218 (5.93%) had completed 8 to 11 years of school, 670 (18.22%) had completed 12 years or high school, 806 (21.92%) had some college, 284 (7.72%) had post-high school training other than college, 889 (24.18%) were college graduates, 569 (15.48%) had postgraduate schooling, and 151 (4.11%) did not specify.

Table 1. Survey questions.

Construct	Survey Question
Trust in internet health information	In general, how much would you trust information about cancer from [the internet]? (*Not at all, A little, Some, A lot*)
Information security beliefs	How confident are you that safeguards (including the use of technology) are in place to protect your medical records from being seen by people who aren't permitted to see them?
	Having safeguards (including the use of technology) in place has to do with the security of your medical records. (*Very confident, Somewhat confident, Not confident*)
Privacy concerns	If your medical information is sent electronically from one health care provider to another, how concerned are you that an unauthorized person would see it? Electronically means from computer to computer, instead of by telephone, mail, or fax machine. (*Very concerned, Somewhat concerned, Not concerned*)
Support for electronic medical record	Please indicate how important it is that [Doctors and other healthcare providers should be able to share your medical information with each other electronically]. (*Very important, Somewhat important, Not at all important*)
Patient access to health records	How many times did you access your personal health information online through a secure website or app in the last 12 months? (*None, 1 to 2 times, 3 to 5 times, 6 to 9 times, 10 or more times*)
Patient care quality	Overall, how would you rate the quality of healthcare you received in the past 12 months? (*Excellent, Very good, Good, Fair, Poor*)

Results

We tested all the hypotheses in one model (as shown in Figure 2). We used R software [51] and the Lavaan package [52] to create the SEM to analyze the hypotheses. We performed checks of goodness-of-fit by checking for a nonsignificant chi-square and a low root mean square error of approximation (RMSEA)–a measure of model fit. Our chi-square test was significant (<.001) indicating other relationships among our constructs not modeled. The RMSEA was low (.099), indicating that our model fits the data. We also verified that there were no multicollinearity issues in two ways. First, we verified that all correlations were below .80 (they ranged from −0.28 to 0.35) [53]. Second, we ran a regression of all the constructs to predict perceived care quality and checked the variance inflation factors (VIF) as well [54,55]. All VIF values were close to 1 (with a range 1.03-1.12) showing no signs of multicollinearity.

Overall, our extension of the DMISSM to the health care system was supported. Figures 3-6 diagram the tested relationships a portion at a time to make interpreting the results easier. Most of the hypotheses were supported (see Table 2).

Figure 3. Results of H1 and H2. HIE: health information exchange.

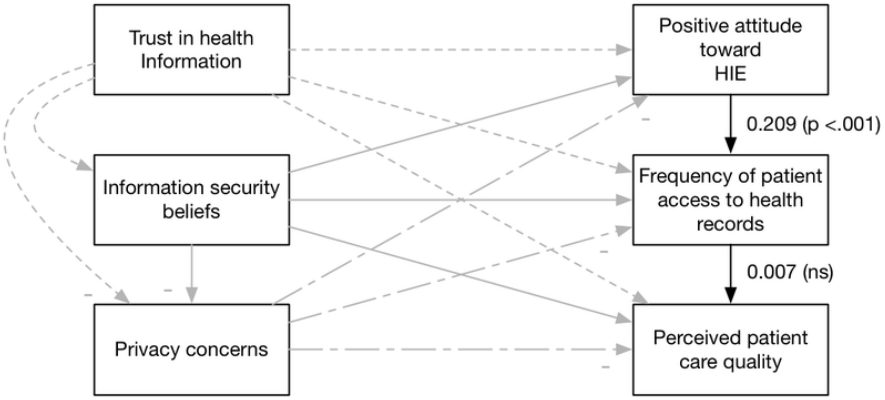


Figure 4. Results of H3. HIE: health information exchange.

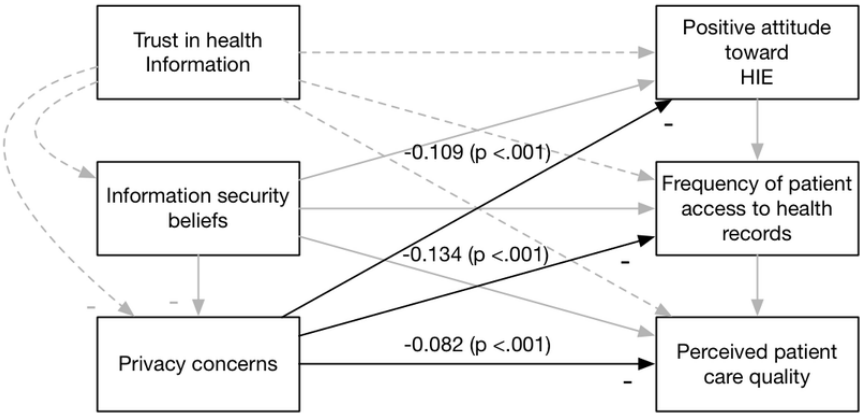


Figure 5. Results of H4. HIE: health information exchange.

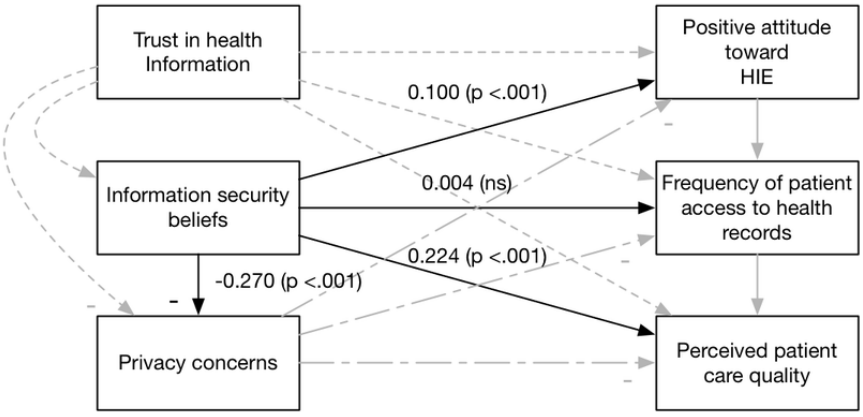


Figure 6. Results of H5. HIE: health information exchange.

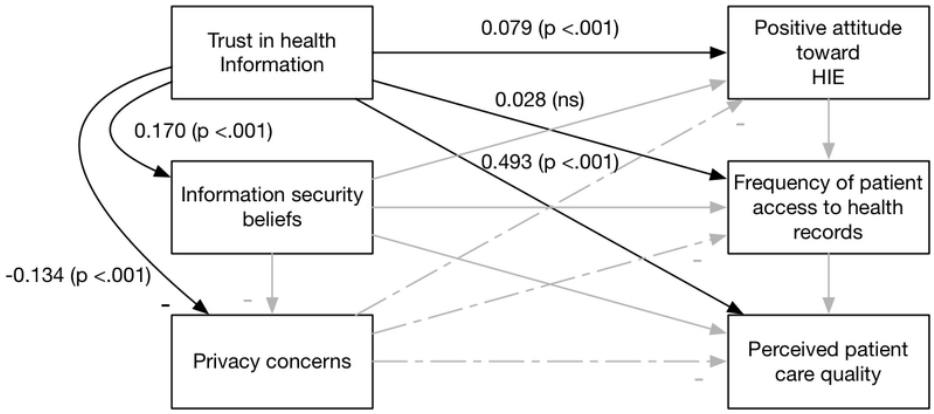


Table 2. Summary of hypothesis testing. HIE: health information exchange.

Hypothesis	Supported?
H1. Patient access to health records increases the level of perceived patient care quality.	No
H2. Positive attitudes toward HIE increase the patient's access to health records.	Yes
H3. Privacy concerns will	
(1) decrease support for electronic medical records	Yes
(2) decrease use of patient access to health records	Yes
(3) decrease the level of perceived patient care quality	Yes
H4. Information security beliefs will	
(1) increase support for electronic medical records	Yes
(2) increase patient access to health records	No
(3) increase the level of perceived patient care quality.	Yes
(4) decrease privacy concerns	Yes
H5. Trust in internet health information will	
(1) increase support for electronic medical records	Yes
(2) increase patient access to health records	No
(3) increase the level of perceived patient care quality	Yes
(4) increase information security beliefs	Yes
(5) decrease privacy concerns	Yes

Discussion

Principal Findings

The results indicate that increased privacy concerns reduce the frequency of EMR use, positive attitudes toward HIE, and ultimately, perceptions of patient care quality. These findings confirm and extend previous reports of privacy concerns deterring patients' adoption of HITs [6]. We also found that information security beliefs increase positive attitudes toward HIE and perceptions of patient care quality. There is, however, an indirect relationship between information security beliefs and frequency of EMR use through decreasing privacy concerns. Likewise, trust in health information has a positive association with positive attitudes toward HIE and patient care quality. Finally, the results show that patients' positive attitudes toward HIE have a positive relationship with patient care quality.

These findings have several practical implications for health care providers and policy makers. First, it was shown that while patients' privacy concerns impede their use of HIS and increase the negative attitude toward HIE and care quality, the perceived effectiveness of security controls lessens privacy concerns. As such, health care providers can mitigate privacy and security concerns by developing more secure privacy safeguards to prevent security attacks and unauthorized access to information. Providers must be transparent with patients [56] by providing clear information about how the security and privacy of patient data are preserved, under what circumstances data is shared, and with whom. This level of transparency, combined with adequate communication may reduce patients' privacy concerns and reluctance to share information, consequently increasing HIT adoption and patient care quality. Second, because the trust of health information is of critical importance to the success of HIT [8], more attention needs to be paid to solutions for

increasing patients' trust. Incorrect information not only limits the efficacy of HIT and reduces patients' trust but also negatively impacts medical decisions [57]. More research is needed to identify the antecedents of patient trust of health information in the context of HIT quality. Third, the results suggest that patients consider information sharing among health care providers to be a valuable capability in care delivery. In other words, patients believe that they benefit when their doctors have the capability to easily exchange information [58]. Entities within a health care organization are therefore encouraged to pursue solutions for effective health information exchange. A key finding is that trust in health information had no direct effect on EMR use; however, it has an indirect relationship through information privacy concerns. This shows that trust in health information lowers patients' privacy concerns, which in turn, increases EMR use by patients.

Our research makes the following theoretical contributions: we extend the DMISSM to a health care system level, above and beyond looking at one particular technology. We also extend the model to the information assurance discipline by adding 2 new constructs of system quality: information trust and information security beliefs.

Limitations

There are several limitations that present an opportunity for future study. First, the hypothesized model was tested using secondary data from a national survey of cancer patients. This limits our findings because the opinions expressed may not reflect those of patients suffering from other diseases as suggested in Zhang et al, 2012 [59]. More research is needed to extend our findings to other patient populations. The second limitation relates to our measurement of patient care quality; we used patients' perception of patient care quality instead of an objective measure. While this is a limitation, the only way to get close to capturing care quality objectively is to look at readmittance data.

Conclusions

In today's health care organizations, information technologies have become critical to the provision of medical care. Meaningful use regulations, combined with the desire to improve care coordination, reduce costs, and improve patient engagement all depend on the success of health care technologies. Yet, there is currently limited understanding of the antecedents of HIS success. Using the Delone and Mclean Information Systems Mode l [37], we investigated the relationships between HIT quality measures and HIT use and the impact of HIT use on patient care quality. The HIT quality measures we studied included trust in health information, information security beliefs, and information privacy controls. We found strong support for the relationships between information trust and the positive attitudes toward HIE among providers, and information trust and frequency of EMR use by patients. The results also showed that trust in health information and information security beliefs have a positive relationship with patient care quality, while privacy concerns reduced patient care quality. The findings have several theoretical and practical implications. It informs health care providers and leaders of the critical importance of effective security controls. While it is known that patients care about privacy, this research shows that patients care about HIT security controls as well. Hospitals can, therefore, improve patient care by implementing effective security and privacy controls. In addition, health care providers must not only provide access to medical records but also encourage patients to check medical records frequently.

Conflicts of Interest

None declared.

References

1. Callen J, Paoloni R, Li J, Stewart M, Gibson K, Georgiou A, et al. Perceptions of the effect of information and communication technology on the quality of care delivered in emergency departments: a cross-site qualitative study. Ann Emerg Med 2013 Feb;61(2):131-144. [doi: 10.1016/j.annemergmed.2012.08.032] [Medline: 23083964]
2. Jarvis B, Johnson T, Butler P, O'Shaughnessy K, Fullam F, Tran L, et al. Assessing the impact of electronic health records as an enabler of hospital quality and patient satisfaction. Acad Med 2013 Oct;88(10):1471-1477. [doi: 10.1097/ACM.0b013e3182a36cab] [Medline: 23969372]
3. Nirel N, Rosen B, Sharon A, Blondheim O, Sherf M, Samuel H, et al. The impact of an integrated hospital-community medical information system on quality and service utilization in hospital departments. Int J Med Inform 2010 Sep;79(9):649-657. [doi: 10.1016/j.ijmedinf.2010.06.005] [Medline: 20655276]
4. Yang H, Guo X, Wu T. Exploring the influence of the online physician service delivery process on patient satisfaction. Decis Support Syst 2015;78:113-121. [doi: 10.1016/j.dss.2015.05.006]
5. Sherifali D, Nerenberg KA, Wilson S, Semeniuk K, Ali MU, Redman LM, et al. The effectiveness of eHealth technologies on weight management in pregnant and postpartum women: systematic review and meta-analysis. J Med Internet Res 2017 Oct 13;19(10):e337 [FREE Full text] [doi: 10.2196/jmir.8006] [Medline: 29030327]
6. Rainie L, Duggan M. Pewinternet. 2016 Jan 14. Privacy and information sharing URL: http://www.pewinternet.org/2016/01/14/privacy-and-information-sharing/ [accessed 2018-02-19] [WebCite Cache ID 6xM2yrlbW]
7. Hamill JT, Deckro RF, Kloeber JM. Evaluating information assurance strategies. Decis Support Syst 2005 May;39(3):463-484. [doi: 10.1016/j.dss.2003.11.004]

8. Maconachy WV, Schou CD, Ragsdale D, Welch D. A model for information assurance: an integrated approach. 2001 Presented at: Proceedings of the 2001 IEEE Workshop on Information Assurance and Security; June 5-6, 2001; United States Military Academy, West Point, NY URL: http://it210web.groups.et.byu.net/lectures/MSRW%20Paper.pdf
9. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
10. Appari A, Johnson EM, Anthony DL. Meaningful use of electronic health record systems and process quality of care: evidence from a panel data analysis of U.S. acute-care hospitals. Health Serv Res 2013 Apr;48(2 Pt 1):354-375 [FREE Full text] [doi: 10.1111/j.1475-6773.2012.01448.x] [Medline: 22816527]
11. Frimpong J, Jackson B, Stewart LM, Singh KP, Rivers PA, Bae S. Health information technology capacity at federally qualified health centers: a mechanism for improving quality of care. BMC Health Serv Res 2013 Jan 31;13(1):35 [FREE Full text] [doi: 10.1186/1472-6963-13-35] [Medline: 23363660]
12. Han W, Sharman R, Heider A, Maloney N, Yang M, Singh R. Impact of electronic diabetes registry 'Meaningful Use' on quality of care and hospital utilization. J Am Med Inform Assoc 2016 Mar;23(2):242-247. [doi: 10.1093/jamia/ocv040] [Medline: 26133895]
13. Kern LM, Barrón Y, Dhopeshwarkar RV, Edwards A, Kaushal R, HITEC Investigators. Electronic health records and ambulatory quality of care. J Gen Intern Med 2013 Apr;28(4):496-503 [FREE Full text] [doi: 10.1007/s11606-012-2237-8] [Medline: 23054927]
14. Lund S, Boas IM, Bedesa T, Fekede W, Nielsen HS, Sørensen BL. Association between the safe delivery app and quality of care and perinatal survival in Ethiopia: a randomized clinical trial. JAMA Pediatr 2016 Aug 01;170(8):765-771. [doi: 10.1001/jamapediatrics.2016.0687] [Medline: 27322089]
15. Pifer EA, Smith S, Keever GW. EMR to the rescue. an ambulatory care pilot project shows that data sharing equals cost shaving. Healthc Inform 2001 Feb;18(2):111-114. [Medline: 11225061]
16. 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]
17. Enriquez JR, de Lemos JA, Parikh SV, Simon DN, Thomas LE, Wang TY, et al. Modest associations between electronic health record use and acute myocardial infarction quality of care and outcomes: results from the national cardiovascular data registry. Circ Cardiovasc Qual Outcomes 2015 Nov;8(6):576-585 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.115.001837] [Medline: 26487739]
18. Khunlertkit A, Carayon P. Contributions of tele-intensive care unit (Tele-ICU) technology to quality of care and patient safety. J Crit Care 2013 Jun;28(3):315.e1-315.12. [doi: 10.1016/j.jcrc.2012.10.005] [Medline: 23159139]
19. Setia P, Venkatesh V, Joglekar S. Leveraging digital technologies: how information quality leads to localized capabilities and customer service performance. MIS Q 2013;37(2):565-590.
20. Brown SM, Aboumatar HJ, Francis L, Halamka J, Rozenblum R, Rubin E, Privacy, Access, and Engagement Task Force of the Libretto Consortium of the Gordon and Betty Moore Foundation. Balancing digital information-sharing and patient privacy when engaging families in the intensive care unit. J Am Med Inform Assoc 2016 Sep;23(5):995-1000. [doi: 10.1093/jamia/ocv182] [Medline: 26984048]
21. Aicken CR, Fuller SS, Sutcliffe LJ, Estcourt CS, Gkatzidou V, Oakeshott P, et al. Young people's perceptions of smartphone-enabled self-testing and online care for sexually transmitted infections: qualitative interview study. BMC Public Health 2016 Dec 13;16(1):974 [FREE Full text] [doi: 10.1186/s12889-016-3648-y] [Medline: 27624633]
22. Anderson M. Pewresearch. 2016. 8 conversations shaping technology URL: http://www.pewresearch.org/fact-tank/2016/03/10/8-conversations-shaping-technology/ [accessed 2017-11-14] [WebCite Cache ID 6uxHMNI6v]
23. Irizarry T, DeVito DA, 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]
24. Meingast M, Roosta T, Sastry S. Security and privacy issues with health care information technology. 2006 Presented at: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 31, 2006; New York City, New York.
25. Vodicka E, Mejilla R, Leveille SG, Ralston JD, Darer JD, Delbanco T, et al. Online access to doctors' notes: patient concerns about privacy. J Med Internet Res 2013 Sep 26;15(9):e208 [FREE Full text] [doi: 10.2196/jmir.2670] [Medline: 24072335]
26. Burkhard RJ, Schooley BL, Dawson J, Horan TA. Information systems and Healthcare XXXVII: when your employer provides your personal health record-Exploring employee perceptions of an employer-sponsored PHR system. Commun Assoc Inf Syst 2010;27(19):323-338.
27. Haun JN, Chavez M, Nazi K, Antinori N, Melillo C, Cotner BA, et al. Veterans' preferences for exchanging information using veterans affairs health information technologies: focus group results and modeling simulations. J Med Internet Res 2017 Oct 23;19(10):e359 [FREE Full text] [doi: 10.2196/jmir.8614] [Medline: 29061553]
28. Josell SD. Mandibular displacement during mastication. J Baltimore Coll Dent Surg 1981;34(2):13-16. [Medline: 7019199]
29. Burton-Jones A, McLean ER, Monod E. Theoretical perspectives in IS research: from variance and process to conceptual latitude and conceptual fit. Eur J Inf Syst 2017;24(6):664-679. [doi: 10.1057/ejis.2014.31]

30. DeLone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2003;19(4):9-30 [FREE Full text] [doi: 10.1080/07421222.2003.11045748]
31. Salahuddin L, Ismail Z. Classification of antecedents towards safety use of health information technology: a systematic review. Int J Med Inform 2015 Nov;84(11):877-891. [doi: 10.1016/j.ijmedinf.2015.07.004] [Medline: 26238706]
32. Cohen JF, Coleman E, Kangethe MJ. An importance-performance analysis of hospital information system attributes: a nurses' perspective. Int J Med Inform 2016 Feb;86:82-90. [doi: 10.1016/j.ijmedinf.2015.10.010] [Medline: 26564330]
33. Winterling J, Wiklander M, Obol CM, Lampic C, Eriksson LE, Pelters B, et al. Development of a self-help web-based intervention targeting young cancer patients with sexual problems and fertility distress in collaboration with patient research partners. JMIR Res Protoc 2016 Apr 12;5(2):e60 [FREE Full text] [doi: 10.2196/resprot.5499] [Medline: 27073057]
34. Sebetci O, Çetin M. Developing, applying and measuring an e-Prescription Information Systems Success Model from the perspectives of physicians and pharmacists. Health Policy Technol 2016 Mar;5(1):84-93. [doi: 10.1016/j.hlpt.2015.10.008]
35. Kapadia V, Ariani A, Li J, Ray PK. Emerging ICT implementation issues in aged care. Int J Med Inform 2015 Nov;84(11):892-900. [doi: 10.1016/j.ijmedinf.2015.07.002] [Medline: 26250987]
36. Fontaine G, Cossette S, Heppell S, Boyer L, Mailhot T, Simard M, et al. Evaluation of a web-based e-learning platform for brief motivational interviewing by nurses in cardiovascular care: a pilot study. J Med Internet Res 2016 Aug 18;18(8):e224 [FREE Full text] [doi: 10.2196/jmir.6298] [Medline: 27539960]
37. DeLone WH, McLean ER. Information systems success: the quest for the dependent variables. Inf Syst Res 1992;3(1):60-95.
38. Agha L. The effects of health information technology on the costs and quality of medical care. J Health Econ 2014 Mar;34:19-30 [FREE Full text] [doi: 10.1016/j.jhealeco.2013.12.005] [Medline: 24463141]
39. Hsieh PJ. An empirical investigation of patients' acceptance and resistance toward the health cloud: the dual factor perspective. Comput Human Behav 2016 Oct;63:959-969. [doi: 10.1016/j.chb.2016.06.029]
40. Hadji B, Degoulet P. Information system end-user satisfaction and continuance intention: a unified modeling approach. J Biomed Inform 2016 Jun;61:185-193 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.021] [Medline: 27033175]
41. Hadji B, Martin G, Dupuis I, Campoy E, Degoulet P. 14 Years longitudinal evaluation of clinical information systems acceptance: the HEGP case. Int J Med Inform 2016 Feb;86:20-29. [doi: 10.1016/j.ijmedinf.2015.11.016] [Medline: 26725691]
42. Blaya JA, Shin SS, Yagui MJ, Yale G, Suarez CZ, Asencios LL, et al. A web-based laboratory information system to improve quality of care of tuberculosis patients in Peru: functional requirements, implementation and usage statistics. BMC Med Inform Decis Mak 2007 Oct 28;7:33 [FREE Full text] [doi: 10.1186/1472-6947-7-33] [Medline: 17963522]
43. Restuccia JD, Cohen AB, Horwitt JN, Shwartz M. Hospital implementation of health information technology and quality of care: are they related? BMC Med Inform Decis Mak 2012 Sep 27;12:109 [FREE Full text] [doi: 10.1186/1472-6947-12-109] [Medline: 23016699]
44. Petter S, DeLone W, McLean ER. Information systems success: the quest for the independent variables. J Manag Inf Syst 2013;29(4):7-62. [doi: 10.2753/MIS0742-1222290401]
45. Ferreira A, Correia A, Silva A, Corte A, Pinto A, Saavedra A, et al. Why facilitate patient access to medical records. Stud Health Technol Inform 2007;127:77-90. [Medline: 17901601]
46. Ross SE, Lin CT. The effects of promoting patient access to medical records: a review. J Am Med Inform Assoc 2003;10(2):129-138 [FREE Full text] [Medline: 12595402]
47. Baird A, Raghu TS. Associating consumer perceived value with business models for digital services. Eur J Inf Syst 2013;24(1):4-22.
48. Esmaeilzadeh P, Sambasivan M. Patients' support for health information exchange: a literature review and classification of key factors. BMC Med Inform Decis Mak 2017 Dec 04;17(1):33 [FREE Full text] [doi: 10.1186/s12911-017-0436-2] [Medline: 28376785]
49. Bélanger F, Crossler RE. Privacy in the digital age: a review of information privacy research in information systems. MIS Q 2011;35(4):1017-1042.
50. Hints.cancer. 2015. Health information national trends survey 4 (HINTS 4): cycle 4 methodology report URL: https://hints.cancer.gov/docs/HINTS_4_Cycle2_Methods_Report.pdf [WebCite Cache ID 6uxHTvwv5]
51. R Core Team. Gbif. 2013. R: a language and environment for statistical computing URL: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing [accessed 2018-02-22] [WebCite Cache ID 6xQIbulK1]
52. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw 2012;48(2):1-36 [FREE Full text]
53. Field A, Miles J, Field Z. Discovering statistics Using R. Thousand Oaks, California: SAGE Publications Inc; 2012.
54. Cenfetelli RT, Bassellier G. Interpretation of formative measurement in Information Systems research. MIS Q 2009;33(4):689-707.
55. Petter S, Straub D, Rai A. Specifying formative constructs in information systems research. MIS Q 2007;31(4):623-656.
56. Palermo TM, Janicke DM, McQuaid EL, Mullins LL, Robins PM, Wu YP. Recommendations for training in pediatric psychology: defining core competencies across training levels. J Pediatr Psychol 2014 Oct;39(9):965-984 [FREE Full text] [doi: 10.1093/jpepsy/jsu015] [Medline: 24719239]
57. Beldad AD, de Jong MD, Steehouder MF. How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. Comput Human Behav 2010;26(5):857-869. [doi: 10.1016/j.chb.2010.03.013]

58. Tharmalingam S, Hagens S, Zelmer J. The value of connected health information: perceptions of electronic health record users in Canada. BMC Med Inform Decis Mak 2016 Jul 16;16:93 [FREE Full text] [doi: 10.1186/s12911-016-0330-3] [Medline: 27422571]
59. Zhang Y, Lauche R, Sibbritt D, Olaniran B, Cook R, Adams J. Comparison of health information technology use between American adults with and without chronic health conditions: findings from the National Health Interview Survey 2012. J Med Internet Res 2017 Oct 05;19(10):e335 [FREE Full text] [doi: 10.2196/jmir.6989] [Medline: 28982644]

Abbreviations

DMISSM: DeLone and McLean Information Systems Success Model
EMR: electronic medical record
HIE: health information exchange
HINTS: Health Information National Trends Survey
HIT: health information technology
PHR: patient health record
RMSEA: root mean square error of approximation
SEM: structural equation modeling

Edited by G Eysenbach; submitted 21.09.17; peer-reviewed by R Gupta, R Pankomera; comments to author 26.10.17; revised version received 06.12.17; accepted 02.01.18; published 11.04.18.

Please cite as:
Kisekka V, Giboney JS
The Effectiveness of Health Care Information Technologies: Evaluation of Trust, Security Beliefs, and Privacy as Determinants of Health Care Outcomes
J Med Internet Res 2018;20(4):e107
URL: http://www.jmir.org/2018/4/e107/
doi:10.2196/jmir.9014
PMID:29643052

Original Paper

Perceived Drivers and Barriers to the Adoption of eMental Health by Psychologists: The Construction of the Levels of Adoption of eMental Health Model

Milou A Feijt[1], MSc; Yvonne AW de Kort[1], Prof Dr; Inge MB Bongers[2], Prof Dr; Wijnand A IJsselsteijn[1], Prof Dr

[1]Human-Technology Interaction, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, Netherlands

[2]Tranzo, Tilburg School of Social and Behavioural Sciences, Tilburg University, Tilburg, Netherlands

Corresponding Author:
Milou A Feijt, MSc
Human-Technology Interaction
Department of Industrial Engineering & Innovation Sciences
Eindhoven University of Technology
IPO Building, 2nd Floor
P.O. Box 513
Eindhoven, 5600 MB
Netherlands
Phone: 31 40 247 5391
Email: m.a.feijt@tue.nl

Abstract

Background: The internet offers major opportunities in supporting mental health care, and a variety of technology-mediated mental and behavioral health services have been developed. Yet, despite growing evidence for the effectiveness of these services, their acceptance and use in clinical practice remains low. So far, the current literature still lacks a structured insight into the experienced drivers and barriers to the adoption of electronic mental health (eMental health) from the perspective of clinical psychologists.

Objective: The aim of this study was to gain an in-depth and comprehensive understanding of the drivers and barriers for psychologists in adopting eMental health tools, adding to previous work by also assessing drivers and analyzing relationships among these factors, and subsequently by developing a structured representation of the obtained findings.

Methods: The study adopted a qualitative descriptive approach consisting of in-depth semistructured interviews with clinical psychologists working in the Netherlands (N=12). On the basis of the findings, a model was constructed that was then examined through a communicative validation.

Results: In general, a key driver for psychologists to adopt eMental health is the belief and experience that it can be beneficial to them or their clients. Perceived advantages that are novel to literature include the acceleration of the treatment process, increased intimacy of the therapeutic relationship, and new treatment possibilities due to eMental health. More importantly, a relation was found between the extent to which psychologists have adopted eMental health and the particular drivers and barriers they experience. This differentiation is incorporated in the Levels of Adoption of eMental Health (LAMH) model that was developed during this study to provide a structured representation of the factors that influence the adoption of eMental health.

Conclusions: The study identified both barriers and drivers, several of which are new to the literature and found a relationship between the nature and importance of the various drivers and barriers perceived by psychologists and the extent to which they have adopted eMental health. These findings were structured in a conceptual model to further enhance the current understanding. The LAMH model facilitates further research on the process of adopting eMental health, which will subsequently enable targeted recommendations with respect to technology, training, and clinical practice to ensure that mental health care professionals as well as their clients will benefit optimally from the current (and future) range of available eMental health options.

(J Med Internet Res 2018;20(4):e153) doi:10.2196/jmir.9485

KEYWORDS

eHealth; mental health; psychology, clinical; diffusion of innovation; technology

Introduction

The internet offers major opportunities in supporting mental health interventions [1]. A variety of technology-mediated mental and behavioral health services are available [2,3], with a growing body of evidence supporting their efficacy (eg, [4-7]). Over the past decades, a mix of terms and definitions has been used to describe eletronic mental health (eMental health; eg, [8-10]). This study will use the term eMental health to refer to "any delivery of mental and behavioral health services, including but not limited to therapy, consultation and psycho-education, by a licensed practitioner to a client in a non-face-to-face setting through distance communication technologies such as the telephone, asynchronous email, synchronous chat, and videoconferencing [8]."

Unique benefits of eMental health tools include increased access to psychological treatment, convenience, as well as enhanced self-reflection and increased emotional disinhibition of the client [9,11,12]. These positive findings, however, are in contrast to the low adoption (ie, acceptance, uptake and use) of eMental health by psychologists. Although exact numbers on a national or international level are scarce, the World Health Organization reported that in 2015 only a third of its member states indicated to have at least one program for technology-mediated mental health services [13]. Moreover, the report shows that most of these programs have a small scale and consist primarily of pilots or informal projects.

Multiple studies have been conducted to investigate therapists' attitudes toward eMental health, and several possible impeding or facilitating factors for adoption have been identified. A barrier frequently reported by therapists pertains to the lack of the full range of nonverbal cues during mediated communication, as they feel this heightens the risk of misunderstanding and does not allow for the development of a strong therapeutic relationship [14]. It has to be noted though that systematic studies that investigate these concerns are lacking [15]. A major concern often reported by therapists is how to deal with crisis situations online (eg, when a client expresses suicidal thoughts) [16]. Technology-mediated modalities allow clients to disconnect at any time, without the therapist knowing whether this is due to technology failure or because the client is in some kind of crisis, and the therapist is not in the same space to ensure their safety [17]. Another reason found for therapists' reluctance is the risk of clients misrepresenting themselves, as it is harder to verify an individual's identity when interacting remotely [17,18]. Therapists also mention more practical concerns such as costs of setting up and maintaining the infrastructure, licensure and jurisdiction constraints, lack of clear ethical guidelines for practice and confidentiality, patient privacy, and the potentially detrimental effects of technology failure [9,14,16,17]. In addition, some studies emphasize the importance of contextual factors of daily clinical practice such as the level of knowledge and training, available time and resources, perceived social norms, forces within the current care system, and the design and usability of the technological tools [14,16,19-25]. These factors vary between different mental health care institutions (eg, whether or not the management of a mental health care institution has allocated time during working hours to invest in eMental health), and in this way the institutional context of a psychologist also has a significant influence.

Despite these efforts to clarify therapists' perceptions toward eMental health, the exact nature of therapists' reluctance to its adoption has remained hard to grasp, and as a result, attempts to increase the uptake and use have not been very successful [26]. eMental health comprises a new way of working for psychologists, as they have to integrate new tools into their existing clinical practice. This requires psychologists to change their current behavior and adopt new behaviors. There is a vast body of literature on behavior change and the adoption of innovations, and multiple theories and models have been developed in an attempt to understand or explain influencing factors. Some prominent ones are the diffusion of innovation theory (DIT) [27], the theory of planned behavior [28], the transtheoretical model [29], and the technology acceptance model (TAM) [30]. For extensive reviews of behavior change and implementation theories, models, and frameworks, see Davis et al [31] and Nilsen [32].

Some studies have applied these theories to the topic of eMental health (eg, [14,20,22]), whereas other studies took a more exploratory approach (eg, [21,23,24,33,34]). However, this research has mainly resulted in lists of factors that impact the adoption of eMental health without structuring their relative weights or considering the influence of individual differences in practitioners' willingness and experience to explore and use technology-mediated therapeutic tools. Moreover, these studies have mostly employed written questionnaires for data collection, which afford a large sample size but restrict in-depth understanding of therapists' experiences. Another limitation in studies on therapists' adoption of eMental health pertains to the prevalent focus on barriers within those studies. This probably reflects the relatively large proportion of psychologists who have not adopted eMental health. Thus, a random sample drawn from this population, typical for survey studies, includes only a small percentage of active users and a much larger percentage of nonusers (eg, [22,34-37]). Although nonusers may still see advantages in eMental health, it is fair to say that nonusers will be more focused on barriers to adoption than active users. As yet, relatively little attention has been given to identifying perceived drivers, whereas research shows that perceived value is an important factor in reducing resistance to use a new technology [38], and some studies even suggest that perceived benefits are rated as more important than perceived barriers in the decision to use a novel technology [39,40]. Moreover, research shows that perceiving a new technology as advantageous compared with the existing practices is a key factor in the adoption of an innovation [27]. Hence, this research employed a stratified sampling strategy that specifically allows for *both* potential drivers as well as barriers in adopting eMental health to emerge.

In addition to its emphasis on barriers, the current literature tends to present both drivers and barriers as relatively undifferentiated lists of factors. How these factors combine or relate to each other and to the level of an individual's acceptance and use of technology-mediated therapeutic tools has been relatively unstudied. A conceptual model describing the

interrelationship between these factors could lead to a more structured understanding of the process of technology acceptance and use in relation to eMental health tools in psychologists' clinical practice. In turn, these structured insights may help address and prioritize selected drivers and barriers over others, thus potentially informing processes of technology development, interface design and evaluation, professional training and coaching, targeted clinical use, and organizational embedding of eMental health tools.

The aim of this study, therefore, was to gain an in-depth and comprehensive understanding of clinical psychologists' perspectives on the adoption of eMental health tools. To reach this objective, the authors will identify both drivers and barriers and analyze possible relationships among the involved factors and strive to structure the obtained findings. This will subsequently enable targeted recommendations with respect to technology, training, and clinical practice to ensure that mental health care professionals as well as their clients will benefit optimally from the current (and future) range of available eMental health options in mental health care.

Methods

Design

This study adopted a qualitative descriptive approach consisting of in-depth semistructured interviews with clinical psychologists. The study consisted of 3 phases. First, a qualitative data collection and analysis phase was aimed at gathering in-depth information about the drivers and barriers to adoption of eMental health from the perspective of clinical psychologists. Second, based on these qualitative findings, a model was constructed that captures different levels of adoption of eMental health and the drivers and barriers related to each level. Third, the model was validated through the process of communicative validation, that is, a second round of interviews to examine whether the model matched the perceptions and experiences of the participants.

Sampling and Recruitment

The sample consisted of practicing clinical psychologists working in the Netherlands. A total of 17 individuals were approached via emails through referrals from contacts in the health community, of which 12 agreed to participate in the first part of the study and 11 also participated in the communicative validation. Ethical approval was granted by the Eindhoven University of Technology Research Ethics Committee (ID: 581), and each participant was offered a €12 gift as a small token of acknowledgment.

The strategy of theoretical sampling was used—a process in which data are simultaneously collected and analyzed to determine who to approach next to yield (most) new insights [41]. The eventual sample size was determined by a saturation criterion, which is generally defined as the point where no new themes, findings, concepts, or problems emerged from the data [42]. This procedure involves the specification of a minimum sample size for initial analysis and a stopping criterion, that is, how many more interviews will be conducted without new information emerging before it is concluded that the point of saturation has been reached. On the basis of earlier findings, this study employed an initial sample size of 10 participants, and a stopping criterion of 2 participants [43]. As no new themes emerged during the last interviews, the data collection was terminated after interviewing 12 participants. In contrast to earlier questionnaire studies, which mostly used convenience samples, this study used stratified sampling to ensure the inclusion of clinical psychologists with different levels of use and experience with eMental health, and to represent a mix of age, gender, job position, and type of mental health care institution. Table 1 shows the distribution of these characteristics.

Procedure of the Interviews

Before the start of the interviews, participants were informed about the purpose and content of the study and signed an informed consent form. The interviewer followed a semistructured interview guide containing a mix of both open-ended and closed questions. The interview guide consisted of 29 questions. The topics covered via these questions were based on findings from previous research described earlier [9,11,12,14-16,18-25] and pertained to participants' knowledge, experience and attitudes toward the use of eMental health tools, covering current use, perceived advantages and disadvantages, and influences of their working environment. Each interview lasted between 45 and 60 min. Most interviews were held at the offices of the participating clinical psychologists or otherwise in a quiet public space. If preferred by the participant, the interview was conducted via telephone or video call (2 interviews). With participants' consent, interviews were audio-recorded to allow for transcription and subsequent analysis. All files were stored in a secured location accessible only to the interviewer.

In the service of the communicative validation, a second interview was scheduled with the same participants (N=11). Only one participant could not participate in this second round due to restricted availability during the time period of the study. However, because of the diversity of the sample's characteristics, the authors believe that this has not compromised the validity of the results. At the start, all participants were given a print of the constructed model, a short summary of the results, and a document with statements about the different levels that characterized them. The interviewer followed a semistructured interview guide with questions focusing on the general impression of the model and whether it matched their perceptions and experiences. These interviews lasted between 20 and 40 min.

Table 1. Demographic details of the participants (n=12).

Characteristic	n (%)
Age (years)	
20-30	2 (17)
31-40	3 (25)
41-50	3 (25)
51-60	3 (25)
60+	1 (8)
Gender	
Male	4 (33)
Female	8 (67)
Mental health care setting	
Public	6 (50)
Private institution	3 (25)
Private practice	3 (25)
Level of use	
No use	2 (17)
Minimal use	2 (17)
Passive use	2 (17)
Active use	4 (33)
Innovative use	2 (17)

Analysis

The interviews were transcribed and analyzed using QSR International's NVivo 11 software. The researchers employed a thematic analysis approach to derive themes in participants' perceptions of barriers and drivers to accept eMental health. The transcripts were systematically analyzed using the procedure outlined by Boeije [44], consisting of 3 phases: open coding, axial coding, and selective coding, resulting in a list of codes, categories, and main themes. Commonly used indicators for the quality of qualitative research are internal validity or credibility, reliability/dependability, objectivity/confirmability, and external validity/transferability [45]. This study addressed these criteria by applying strategies described by Miles and Huberman [45] and Wester and Peters [46]. Credibility was ensured by communicative validation. Dependability was established by performing a coding check on part of the data by an independent scholar. The interrater reliability was determined at a Cohen kappa of .78, which is considered substantial [47] and acceptable for exploratory research [48]. The dependability criterion was further supported by peer debriefing, which consisted of discussing findings with peers and colleagues on various moments during the process. This also enhanced confirmability, as was providing clear examples of key themes. Finally, to improve transferability, theoretical sampling was used with the inclusion of psychologists in various job positions, mental health care institutions, as well as age and level of adoption of eMental health. In addition, connecting the results to previous theories further helps in establishing this criterion.

Results

Qualitative Data Collection and Analysis

The results of the interviews are structured along 4 main themes that emerged from the thematic analysis: general characteristics of eMental health, drivers for the adoption of eMental health, barriers for the adoption of eMental health, and contextual factors of daily clinical practice. Within these themes, several subthemes were identified that provide more detailed information.

General Characteristics of eMental Health

All participants clearly expressed the indispensability of face-to-face contact for the delivery of their treatment, because they felt mediated forms of communication lack the subtle signs in facial expressions, posture, and appearance they believe to be crucial for an accurate understanding of their client. In line with this, all participants stated they only wanted to use eMental health in combination with face-to-face sessions, as a complement to their treatment. One participant said:

> *I do not think it can truly be a replacement. Because the way that people behave and look is indispensable… I particularly find it a very nice complement to my treatment.* [P3]

Another general point emerging from all the interviews is that not every kind of eMental health works for every client; it strongly depends on the client's specific needs, capabilities, and preferences, which is clear from the following statement by one of the participants:

> *You have to make eHealth really adapted, customized; what does that client need, and what fits the specific situation.* [P3]

Factors often mentioned to have an important influence on the specific utilization of eHealth were the nature and complexity of the psychological disorder, the client's age, level of computer skills, intelligence, and the devices available to the client. This was evident by what one of the participants expressed:

> *Not everyone is equally skilled with computers, that does make a difference too. Then just using the discussion feature [of the eMental health platform] is already an accomplishment.* [P7]

Besides characteristics of clients, participants also reported characteristics of therapists that made them more or less suitable such as the level of computer skills, affinity with technology, age, and therapeutic approach:

> *But I have to admit, I do not even use...I am very bad with computers, so I do not have any experience with all those online telephone things. So I don't know how that would feel in my day-to-day work.* [P8]

Perceived Drivers to the Adoption of eMental Health

All participants agreed that one needs to be convinced of the benefits to adopt eMental health:

> *If I would know that it would bring something positive to my clients, then I would definitely be much more motivated.* [P8]

Furthermore, it became clear that beyond being or becoming aware of the benefits, experiencing them is crucial in decreasing resistance and developing intrinsic motivation to continue using eMental health. Reiterating this, one of the participants said:

> *I become more and more aware of the benefits, definitely. Especially with my target group, addiction, I experience that it truly is an addition [to the treatment].* [P1]

A frequently mentioned benefit was that mediated contact in-between regular sessions affords a more intimate and personal therapeutic relationship, because it increases the frequency of contact between the therapist and client, and in this way enhances a sense of continuity:

> *People are making a stronger link with you, like "hey, you also think about me outside of that room." And that is very beneficial to your relationship with people, they really feel I still exist for them.* [P6]

Most participants using eMental health reported that this increased frequency of contact intensifies the treatment. Moreover, it stimulates clients to engage in a higher level of therapeutic activity at home, thus accelerating the therapeutic process:

> *eHealth gives you something of an intensifier. People can work on something every day.* [P11]

Several therapists recognized the benefit that both the technology-mediated interactions between sessions and heightened client activity at home allow for more consolidated progress. In addition, the higher frequency of interactions allows for the introduction of new therapeutic elements during the actual face-to-face sessions, rather than the repetition of earlier steps in the process. In the words of one participant:

> *I notice that they really are much more active at home and also return to the next session with more to discuss. Or that they have already thought about that, whereas with others you have to make them reflect at that moment itself, and then it is hard sometimes, then time passes by much quicker and you can make less progress, whereas when you make them work at home then you find that you can do much more in the sessions too.* [P9]

Moreover, in some cases eMental health allows for a better satisfaction of client needs, for example, when it is very burdensome for clients to travel to the therapist's office, when they are abroad, or in cases of illness, pregnancy, or other limiting circumstances. One participant stated:

> *I regularly have videoconferences with people with young children who cannot leave their house, or who are too ill to come. I also had several people living abroad, then it is also pretty convenient.* [P3]

Although less prominent, several therapists mentioned practical personal benefits of eMental health, such as increased efficiency in administrative tasks. Finally, some expressed that their enthusiasm toward eHealth was mostly due to the new treatment possibilities offered by eMental health, such as interventions with virtual reality and biofeedback, enabling them to treat their clients in ways that were not possible before.

Perceived Barriers to the Adoption of eMental Health

Perhaps the most important barrier reported by participants was a lack of knowledge and experience on their side with respect to various aspects of eMental health such as how to integrate eHealth into their treatments and the possibilities of available tools. This last point seems to be partly due to the relatively large—and fast growing—number of available tools, combined with the lack of a comprehensive overview of the availability, relevance, and efficacy of technology-mediated therapeutic tools in their clinical context. This was reflected in the statement made by one participant:

> *Sometimes I hear something from a colleague that I say "oh, is that available? I did not know." And if I am not aware of something, then I will not use it.* [P7]

Related to this, an important barrier was the strong professional obligation most participants feel to be an expert in eMental health before they can apply it in their daily practice, as is evident from the following statement:

> *I feel like it does not come across professionally to set up such an eHealth module with someone if I would not know exactly how it works. I find that unacceptable.* [P12]

Several participants reported that it would be helpful if they would receive training and have the opportunity to practice more with an eMental health platform and try out various tools:

> *It would make a big difference if I could practice more with the system. Because then it becomes familiar, whereas now it is not.* [P12]

As a counterpoint to the advantages of increased availability through eMental health tools described above, participants mentioned that being more accessible to clients also comes at a cost. It may put a higher demand on therapists by increasing the communication channels they must keep track of and the times at which they have to make themselves available to their clients. This is echoed by the following statements by 2 participants:

> *Now I not only have to watch my email, not only my phone, but then I also have to watch that Whatsapp.* [P10]

> *In a way eHealth makes the job more burdensome. You have to juggle multiple tasks as a mental health professional.* [P4]

Increased accessibility sometimes also increases feelings of responsibility, especially when dealing with crisis situations. This issue seems to be complicated by the lack of clear ethical guidelines and instructions on how to handle these kinds of situations:

> *If I am home at night and I read the message that someone is suicidal, what do I do? So that aggravates matters, because you feel like you always have to be available and you also wonder "where do my responsibilities lie?"* [P6]

Both increased availability and sense of moral responsibility beyond office hours are likely to add to the stress and workload that therapists experience and may negatively affect the balance between a therapist's work and private life.

Furthermore, all participants who were using eMental health tools experienced some technological issues. Most frequently mentioned were usability problems such as having troubles with logging into the system, and functionality issues of the eMental health platform, for example, the lack of an adequate search tool to explore the available content. In addition, complaints with respect to the quality of the videoconferencing technology were reported frequently.

Contextual Factors of Daily Clinical Practice

An experience that was reported in every interview was a perceived pressure from managers and health insurance companies to use eMental health. Some participants expressed not to be bothered much by this, but for others this caused a general sense of distrust against management, feeling that their management's interest in eMental health was solely driven by the goal to save money and not because psychologists or their clients would benefit from it. This elicited strong feelings of resistance. As one of the participants stated:

> *That is how eHealth is often looked at, like it is just a way to claim a two percent or higher hourly rate from the insurance companies.* [P4]

Another common experience was a high pressure on productivity during work and a lack of organizational support in terms of time and resources. Participants argued that not having sufficient time to invest in eHealth during working hours significantly hinders the implementation of eMental health:

> *There is nothing that exploring [eHealth] can be registered as, so then it reduces my productivity rate and then I think "no, I will not do it. That will only cost me and does not benefit me at all."* [P12]

Several participants also mentioned the low visibility and awareness of eMental health during daily practice as a prominent obstacle to the adoption of eMental health. Some attributed this to the topic of eMental health not being discussed much among colleagues or in team meetings:

> *I just barely encounter it, so then it also fades to the background more easily.* [P9]

> *eHealth is primarily a topic that is just not that much discussed.* [P8]

As a solution, participants suggested that external triggers and standard procedures such as automatic reminders would be helpful to facilitate the adoption of eMental health. In addition, they mentioned the importance of keeping an open dialogue on the topic to trigger awareness and reflection upon one's opinion about eMental health.

Construction of the Levels of Adoption of eMental Health Model

The previous sections described sets of drivers and barriers, as well as contextual conditions that affect the adoption of eMental health tools. An important insight that emerged across the various interviews is that although some experiences and attitudes are common among the entire sample, a strong differentiation could be seen in the nature and importance of the various drivers and barriers perceived across the participant sample; that is, the list of drivers and barriers was not uniformly or randomly distributed across the interviewees but rather seemed to be related quite strongly to the actual exposure and hands-on experiences that people had had regarding eMental health tools. This is in line with the notion, discussed in the Introduction, that the adoption of eMental health tools, similar to the adoption of any technological innovation, requires psychologists to adopt new behaviors.

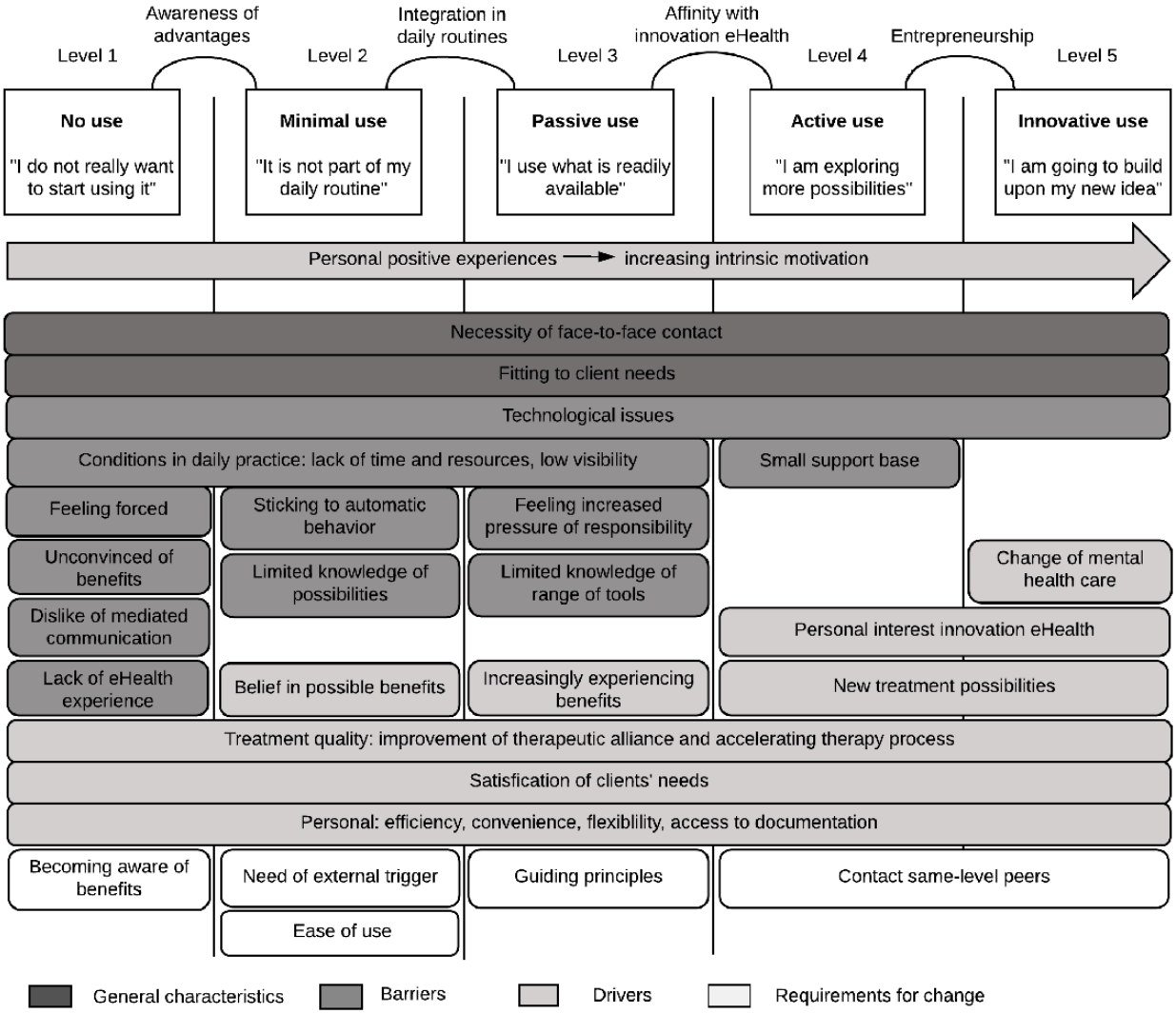
A model that seems to be particularly well suited to describe the process of behavior change to adopt innovative technology is Rogers' well-established diffusion of innovations theory (DIT) [27]. The DIT states that people differ in the time they need for adoption because of several individual characteristics and Rogers has distinguished 5 discrete adopter categories from fast to slow adoption based on these differences. This conception seems to be applicable to this study, as differences were found within the group of participants in the extent to which they had adopted eMental health. On the basis of similarities between the characteristics of Rogers' adopter categories and characteristics of the participants regarding their attitude and use of eHealth tools, the authors determined 5 levels of adoption of eMental health (see Table 2).

Table 2. Adopter categories from Rogers' diffusion of innovations theory (DIT) and their corresponding levels of adoption of electronic mental health (eMental health).

Adopter category by Rogers	Level of adoption of eMental health
Laggards	No use
Late majority	Minimal use
Early majority	Passive use
Early adopters	Active use
Innovators	Pioneer or innovative use

When applying these levels of adoption—that is, the extent to which clinical psychologists had already adopted eMental health in their clinical practice—clear differences can be found in the types of perceived drivers and barriers that characterize the psychologists at the different levels of adoption, as well as differences in the types of processes required for psychologists to move from one level of adoption to the next. Developing this notion, the authors propose the Levels of Adoption of eMental Health (LAMH) model. This model incorporates the 5 levels of adoption identified by the authors and links them to the general characteristics, drivers, barriers, and requirements for change that were found relevant for each level. These factors are directly derived from the main themes extracted during the thematic analysis of the in-depth interviews and hence are entirely based on the perceptions expressed by the participants. The final LAMH model is shown in Figure 1.

Figure 1. The Levels of Adoption of eMental Health (LAMH) model.



The 5 top rectangles in Figure 1 present the 5 levels of adoption, including a characterizing phrase that the authors formulated based on the interviews, exemplifying the attitude of a typical user at each level. The text above the connecting lines between levels describes the factor that was found to distinguish these levels. The bars below the rectangles represent factors that constitute general characteristics, drivers, barriers, and requirements for change. When a factor is located under a particular level, this means that it is most important for clinical psychologists at that level.

Clinical psychologists at level 1 (No use) are generally not convinced and even skeptical about the advantages that eMental health can have and are therefore averse to using it. When they experience pressure by management to do this, it results in a strong feeling of resistance. Psychologists in this group might also feel that eMental health does not suit their profession and show an aversion for computer-mediated communication in general. Psychologists in this category are further characterized by a relatively low level of computer literacy and lack of exposure to eMental health. In accordance with this, use of eMental health tools is nearly or entirely absent, possibly with the exception of telephone, and email for administrative purposes (ie, to schedule an appointment).

At level 2 (Minimal use), clinical psychologists are becoming convinced that eMental health may carry some advantages, which is the most important distinction between levels 1 and 2. However, they are generally unsure how to implement it into their daily practice. Because their intrinsic motivation is fairly low, participants do not want to spend a lot of time and effort in learning to use eMental health. Therefore, there tends to be a lack of knowledge about the possible ways in which eHealth can be applied in practice, and the use of eMental health is likely to be restricted to familiar and easy-to-access tools such as email, telephone, and instant messaging. Functionality limitations and usability issues have a relatively large influence in this group and ease of use of the eMental health tools is a major requirement. Because applying eMental health tools is not integrated into their daily practice, psychologists at this level are prone to maintain their existing way of working and as a result do not gain the positive experiences that could increase their intrinsic motivation.

Clinical psychologists at the third level (Passive use) are using eMental health tools as part of their daily routines. In a positive situation, this daily use leads to an increased conviction of the added value of eMental health, as they gather more and more positive experiences. However, regular use also might confront them with the challenges and limitations of eMental health such as the perceived lack of nonverbal communication or concerns about the pressures and responsibilities associated with being much more easily accessible. Although members of this group are generally motivated to use eMental health, they tend to stick to the applications that are readily available and are not inclined to actively search for other possibilities, which results in a limited overview of the entire range of eMental health tools and limited in-depth knowledge of specific tools. Their use of tools mostly consists of familiar applications such as email, telephone, and WhatsApp, or other mobile phone apps that are relatively well-known, for example consisting of mindfulness exercises. When easily available, an eHealth platform is regularly used.

Compared with practitioners at level 3, clinical psychologists at level 4 show a high level of personal interest in the developments of eMental health and hence have a higher intrinsic motivation to actively keep track of new developments in the field of eMental health. The new treatment possibilities technology-mediated tools offer can act as an additional driver for this group. They make use of a broad range of tools and are eager to try newly available options such as virtual environments or digital games. Most psychologists in this group function as experts of eHealth within their working environment and might be one of the few in their team actively using eMental health. The lack of interest from colleagues may lead to frustration, and frequent contact with same-level peers is important to support their positive attitude toward eMental health.

The characteristics of level 5 are largely similar to those of level 4. The most pronounced difference can be described as entrepreneurship, that is, the initiation of projects to develop and test new eMental health tools or apply the existing tools in novel contexts. In addition, the participants in the highest level might also have a clear vision about major and largely positive changes they expect eMental health will bring to the field of mental health care and even society in general, such as the enabling of personalized care and the opportunity to use gathered data for preventive measures.

Communicative Validation

In the second round of interviews, a communicative validation of the LAMH model was performed to evaluate whether the model matched the perceptions and experiences of the participants and refined it. When the LAMH model was presented first without any introduction, all participants reported that they considered the model a clear representation and felt that they recognized the displayed differences in adoption of eMental health from their experiences in daily clinical practice. When asked to which level they would classify themselves, all participants were able to categorize themselves to 1 level or in between 2 consecutive levels. This self-classification rarely diverged from the classification a priori made by the interviewer (ie, never more than 1 level higher or lower), based on the interview data. In the few cases (N=3) where there was a small divergence between interviewer classification and self-classification, the self-classification was always in the direction of higher acceptance of technology—that is, toward a higher level in the model. In general, participants agreed with the factors in the model, in particular those that were shared by all groups or those related to their own level:

> *I think these levels reflect the current situation very aptly, the different phases that you can be in and the barriers you experience.* [P12]

On the basis of the results, a few relatively minor adaptations were made to the LAMH model; some elements were added or removed, and some of the elements were rephrased to better match participants' experiences. With respect to the layout of the model, some of the elements were reordered and lines were

added to make it easier to understand. Figure 1 shows the final LAMH model.

Discussion

Principal Findings

Technology-mediated therapeutic tools have great potential in supporting the clinical practice of psychologists, and there is a substantial and growing evidence base in support of the efficacy of eMental health. However, at present, such tools are being underused in clinical practice, with only a minority of mental health care professionals and organizations employing a strategy to implement and use these new technologies. The discrepancy between the promise of eMental health tools and the documented reality of their use raises the following questions: why have these tools not been embraced more fully, what are the underlying barriers that hinder the adoption of eMental health tools, what are the perceived drivers that would help increase the acceptance and use of such tools, and How can these factors be structured in a way that improves systematic understanding? This study was aimed to elucidate these questions. In this context, research on technology acceptance provided a useful lens through which the drivers and barriers that clinical psychologists report could be analyzed and structured. On the basis of this, the authors developed a conceptual model for understanding the adoption of eMental health tools.

To arrive at these insights, this study utilized a qualitative descriptive approach consisting of in-depth interviews with clinical psychologists. The results indicated, first, that all participants consider a minimum amount of face-to-face contact vital to the quality of their treatment and stress the importance of basing their choice of eMental health tool on the needs and capabilities of the specific client. A major driver to the adoption of eMental health for psychologists is the belief and personal experience that eMental health can be beneficial for them or their practice, as it increases intrinsic motivation to use eMental health. Perceived benefits consist of the improvement of the therapeutic relationship, acceleration of the treatment process, increased satisfaction of client needs, personal benefits for therapists, and new treatment possibilities. Barriers that are most frequently reported are as follows: lack of knowledge and experience, increased demands due to increased accessibility, and technological issues. Furthermore, several contextual factors in daily clinical practice emerged as impeding factors to the adoption of eMental health, most notably lack of time and resources, feeling forced to use eMental health, and low visibility and awareness of eMental health.

Many of our findings are in line with results from previous research on the adoption of eMental health. Earlier studies also found lack of knowledge and experience, limited time and resources, and technological issues to be the most important impeding factors (eg, [9,14,37]. A barrier not highlighted in previous research pertains to the perceived increase of work demands caused by increased accessibility and a sense of moral responsibility beyond working hours implicated by the use of eMental health tools. However, such increased expectations and responsibilities associated with the use of new communication technologies have been identified outside the realm of mental health care, for instance in work focusing on social awareness systems within domestic settings [49,50]. In contrast to earlier work reporting that psychologists are influenced by social pressure from colleagues to reject eMental health [14,19], this study found that the general attitude in the workplace seems to be rather more disinterested than judgmental, as the topic of eMental health is rarely being discussed. In other words, in the context of day-to-day work pressures, exploring eMental health tools has a relatively low priority and visibility.

Compared with prior work in this area, this study puts a greater emphasis on the identification of drivers to the adoption of eMental health. Practitioners' belief in the beneficial outcomes of eMental health is a key driver of its adoption. Such a benefit-driven approach is not uncommon in the motivated acceptance and use of new communication technologies and resonates with earlier studies pointing to the importance of users perceiving benefits of an innovation as a precondition of use [39,40]. The results suggest that clinical psychologists perceive the acceleration of the treatment process as a primary advantage. One plausible factor causing this acceleration, suggested by various participants in this study, is that mediated contact in-between regular sessions intensifies the treatment and affords a more intimate and personal therapeutic relationship. This explanation is in line with research suggesting that higher session frequencies are related to faster clinically significant gains in recovering from psychological distress [51].

In addition, although previous studies on therapists' attitudes regarding eMental health have mainly resulted in an undifferentiated list of barriers and drivers, this study is the first to recognize a systematic relationship between the extent to which psychologists have adopted eMental health and the particular drivers and barriers they experience. On the basis of this insight, the LAMH model was developed that incorporates these differences. The model distinguishes 5 levels of adoption of eMental health and the corresponding drivers and barriers perceived by psychologists, determined based on several characteristics regarding their attitude and use of eHealth. At the heart of the LAMH model is the proposition that the willingness and ability of clinical psychologists to use technology-mediated therapeutic tools within their clinical context is contingent on the extent to which specific informational, motivational, technical, and organizational barriers at different stages are overcome and that specific drivers are present that will motivate and support the adoption process. Each level of adoption has a set of associated drivers, barriers, and requirements for change specific to that level. Although some barriers may be detrimental to adoption across all levels of the LAMH model—for example critical functionality limitations or severe usability problems associated with the eMental health tools—the model nevertheless assumes that lower level drivers and barriers will not be as relevant to higher levels of adoption, and vice versa. For example, at the levels 4 (active use) or 5 (innovative use) of the LAMH model, one would not expect that professionals experience a lack of knowledge about eMental health—a barrier that is typically found at lower adoption levels. Similarly, at level 1 (no use), one would not expect professionals to entertain a positive future vision on the role of eHealth in mental health care—a driver

that does characterize professionals at level 5 (innovative use). The model also reveals that psychologists having lower levels of adoption experience relatively more barriers than drivers, whereas this balance shifts to the opposite for higher levels, which in turn could result in an increase of intrinsic motivation to continue or even expand the use of eMental health tools.

At any point in time, the LAMH model could be applied to determine the momentary level of adoption of eMental health by an individual mental health care professional, for example, to gauge individuals' readiness to accept new technology-mediated therapeutic tools or to support organizational decision making when choosing a strategy to implement these innovative technologies. From the LAMH model, it can be inferred that interventions to increase adoption should be tailored to the practitioners' individual level of adoption of eMental health. Even stronger, the model can be practically applied by informing how this level can be influenced on a much more specific level than was previously possible by enabling the development of interventions that are tailored to one's level of adoption, hence targeting the specific barriers and drivers that are experienced. In addition, the model can be used to facilitate discussions among mental health care professionals about eMental health by forming a recognizable starting point and providing a shared language.

However, the model can also be conceptualized as a dynamic description of the process and stages of change. By distinguishing different levels of adoption and corresponding relevant drivers and barriers, the LAMH model enables examining the transitions between the levels, how these factors vary for different transitions, and their influence relative to each other. Besides relating specific drivers, barriers, and requirements for change to one's level of adoption, the results can also be structured into various dimensions that are at play in the adoption of eMental health such as compatibility with the current way of working, personal innovativeness, usability of technology, and organizational support. Moreover, the results suggest that these dimensions play a different role for different levels of adoption. Clustering the themes in this way thus provides another view on the factors that influence this process. The gained insights could be further expanded by investigating how the levels of adoption of eMental health defined in this study relate to the existing models describing processes of behavior change and implementation that use similar dimensions such as the DIT [27] and the TAM [30].

Following the construction of the LAMH model, a questionnaire is currently under development that will allow fast and easy assessment of the level of adoption of eMental health by mental health care professionals. Such an instrument will be a powerful tool in conducting research on larger samples of clinical professionals and relating levels of adoption to, among others, specific therapeutic approaches, technological innovations, and organizational contexts.

Limitations

A study of this kind has a number of limitations. Even though the participants in this study were carefully selected to comprise a representative sample of clinical psychologists, a limitation of this study is that it consisted of in-depth qualitative research with a small sample compared with most of the conducted studies with a quantitative approach. Conducting studies with larger sample sizes (enabled by the use of the assessment instrument that is being developed) will allow for further validation of the LAMH model. Moreover, although many of our insights are likely to hold true for the larger population of mental health care professionals, there may be subtle differences in experienced drivers and barriers across different mental health care occupations, which include psychotherapists, clinical psychologists, psychiatrists, and psychiatric nurses. This will be a subject for future investigations. Finally, although during the communicative validation phase of the study most participants agreed to their LAMH classification, few small divergences could be observed. These divergences pointed in the direction of a higher technology acceptance level, which is suggestive of a potential small social desirability bias; that is, it may be more desirable to come across as technology-savvy and open to technological innovations in mental health care. Although only a minor potential effect in our study, it is an issue that researchers need to be aware of when investigating technology acceptance among highly trained professionals.

Conclusions

By studying the adoption of eMental health tools by clinical psychologists, this study contributes to their effective use, supporting the availability of timely, high-quality, well-integrated mental health care. The mediated nature of these tools has many well-documented advantages that serve this purpose, which makes the low utilization of eMental health a challenge that urgently needs to be addressed and resolved. This study addresses this need as it aimed to produce a more comprehensive and systematic understanding of this problem space. This goal is achieved by clarifying both barriers and drivers, several of which are new to the literature. In addition, the authors identified a relationship between the actual level of adoption and level-specific drivers and barriers and subsequently structured the obtained findings through the construction of the LAMH model. In this way, the authors hope to advance the existing insights into the specific drivers and barriers relevant for specific groups of mental health care professionals. In turn, these insights will inform processes of technology development, interface design, professional training, clinical use, and organizational embedding of eMental health tools that enable reaching the full potential of eMental health. Eventually, this is likely to bring significant improvements to the quality and efficiency of mental health care practice, from which both professionals and clients will benefit.

Acknowledgments

The authors thank the clinical psychologists for their participation. This work was supported by NWO grant number 055.16.141, project title Psychosocial Games Supporting Mental Health Professionals in Stress Reduction and Empathic Interactions in Remote/online Psychotherapy, under the Serious Games for Professional Skills Program.

Authors' Contributions

WIJ and MF conceived the study. MF conducted the interviews, data analysis and interpretation of data, and drafted the manuscript. WIJ, YK, and IB contributed to critical revision of the paper. Each author listed on the manuscript has seen and approved the submission of this version of the manuscript and takes full responsibility for the manuscript.

Conflicts of Interest

None declared.

References

1. Suler J. Cybertherapeutic theory and techniques. In: Barak A, editor. Psychological Aspects of Cyberspace: Theory, Research, Applications. Cambridge, UK: Cambridge University Press; 2008:102-128.
2. 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]
3. Mucic D, Hilty DM. E-Mental Health. Cham, Switzerland: Springer; 2016.
4. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
5. 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]
6. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
7. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
8. Mallen MJ, Vogel DL. Online counseling: reviewing the literature from a counseling psychology framework. Couns Psychol 2005;33(6):819-871. [doi: 10.1177/0011000005280182]
9. Rochlen AB, Zack JS, Speyer C. Online therapy: review of relevant definitions, debates, and current empirical support. J Clin Psychol 2004 Apr;60(3):269-283. [doi: 10.1002/jclp.10263] [Medline: 14981791]
10. Manhal-Baugus M. E-therapy: practical, ethical, and legal issues. Cyberpsychol Behav 2001 Oct;4(5):551-563. [doi: 10.1089/109493101753235142] [Medline: 11725648]
11. Abney PC, Maddux CD. Counseling and technology: some thoughts about the controversy. J Technol Hum Serv 2004;22(3):1-24. [doi: 10.1300/J017v22n03]
12. Chester A, Glass CA. Online counselling: a descriptive analysis of therapy services on the Internet. Br J Guid Counc 2006 May;34(2):145-160. [doi: 10.1080/03069880600583170]
13. World Health Organization. Geneva, CH: World Health Organization Document Production Services; 2016. Global diffusion of eHealth: making universal health coverage achievable. Report of the third global survey on eHealth URL: http://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf?sequence=1&isAllowed=y
14. Lovejoy TI, Demireva PD, Grayson JL, McNamara JR. Advancing the practice of online psychotherapy: an application of Rogers' diffusion of innovations theory. Psychotherapy (Chic) 2009 Mar;46(1):112-124. [doi: 10.1037/a0015153] [Medline: 22122574]
15. Simpson S. Psychotherapy via videoconferencing: a review. Br J Guid Couns 2009 Aug;37(3):271-286. [doi: 10.1080/03069880902957007]
16. Perle JG, Langsam LC, Nierenberg B. Controversy clarified: an updated review of clinical psychology and tele-health. Clin Psychol Rev 2011 Dec;31(8):1247-1258. [doi: 10.1016/j.cpr.2011.08.003] [Medline: 21963670]
17. Rummell CM, Joyce NR. “So wat do u want to wrk on 2day?”: The ethical implications of online counseling. Ethics Behav 2010 Dec 13;20(6):482-496. [doi: 10.1080/10508422.2010.521450]
18. Alleman JR. Online counseling: The Internet and mental health treatment. Psychother Theory, Res Pract Train 2002;39(2):199-209. [doi: 10.1037//0033-3204.39.2.199]
19. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]

20. Carper MM, McHugh RK, Barlow DH. The dissemination of computer-based psychological treatment: a preliminary analysis of patient and clinician perceptions. Adm Policy Ment Health 2013 Mar;40(2):87-95. [doi: 10.1007/s10488-011-0377-5] [Medline: 22001968]
21. Gun SY, Titov N, Andrews G. Acceptability of Internet treatment of anxiety and depression. Australas Psychiatry 2011 Jun;19(3):259-264. [doi: 10.3109/10398562.2011.562295] [Medline: 21682626]
22. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
23. Kivi M, Eriksson MC, Hange D, Petersson EL, Björkelund C, Johansson B. Experiences and attitudes of primary care therapists in the implementation and use of internet-based treatment in Swedish primary care settings. Internet Interv 2015 Sep;2(3):248-256. [doi: 10.1016/j.invent.2015.06.001]
24. Sinclair C, Holloway K, Riley G, Auret K. Online mental health resources in rural Australia: clinician perceptions of acceptability. J Med Internet Res 2013;15(9):e193 [FREE Full text] [doi: 10.2196/jmir.2772] [Medline: 24007949]
25. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, et al. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9. [doi: 10.1016/j.invent.2017.01.001]
26. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark S, et al. Implementing and up-scaling evidence-based eMental health in Europe: the study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
27. Rogers EM. Diffusion of Innovations. 5th Ed. New York, NY: The Free Press; 2003.
28. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
29. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [Medline: 6863699]
30. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
31. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015 Sep;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
32. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015;10(53):1-13 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
33. Wilhelmsen M, Høifødt RS, Kolstrup N, Waterloo K, Eisemann M, Chenhall R, et al. Norwegian general practitioners' perspectives on implementation of a guided web-based cognitive behavioral therapy for depression: a qualitative study. J Med Internet Res 2014;16(9):e208 [FREE Full text] [doi: 10.2196/jmir.3556] [Medline: 25208886]
34. Vigerland S, Ljótsson B, Bergdahl Gustafsson F, Hagert S, Thulin U, Andersson G, et al. Attitudes towards the use of computerized cognitive behavior therapy (cCBT) with children and adolescents: a survey among Swedish mental health professionals. Internet Interv 2014 Jul;1(3):111-117 [FREE Full text] [doi: 10.1016/j.invent.2014.06.002]
35. van der Vaart R, Atema V, Evers AW. Guided online self-management interventions in primary care: a survey on use, facilitators, and barriers. BMC Fam Pract 2016;17(27):1-9 [FREE Full text] [doi: 10.1186/s12875-016-0424-0] [Medline: 26961547]
36. Mora L, Nevid J, Chaplin W. Psychologist treatment recommendations for Internet-based therapeutic interventions. Comput Human Behav 2008 Sep;24(6):3052-3062. [doi: 10.1016/j.chb.2008.05.011]
37. Perle JG, Langsam LC, Randel A, Lutchman S, Levine AB, Odland AP, et al. Attitudes toward psychological telehealth: current and future clinical psychologists' opinions of internet-based interventions. J Clin Psychol 2013 Jan;69(1):100-113. [doi: 10.1002/jclp.21912] [Medline: 22975897]
38. Kim HW, Kankanhalli A. Investigating user resistance to information systems implementation: a status quo bias perspective. MIS Q 2009;33(3):567-582. [doi: 10.2307/20650309]
39. Melenhorst AS, Rogers WA, Bouwhuis DG. Older adults' motivated choice for technological innovation: evidence for benefit-driven selectivity. Psychol Aging 2006 Mar;21(1):190-195. [doi: 10.1037/0882-7974.21.1.190] [Medline: 16594804]
40. Sharit J, Czaja SJ, Perdomo D, Lee CC. A cost-benefit analysis methodology for assessing product adoption by older user populations. Appl Ergon 2004 Mar;35(2):81-92. [doi: 10.1016/j.apergo.2003.12.003] [Medline: 15105069]
41. Flick U. An Introduction To Qualitative Research (5th Edition). London, UK: SAGE; 2010.
42. Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychol Health 2010 Dec;25(10):1229-1245. [doi: 10.1080/08870440903194015] [Medline: 20204937]
43. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
44. Boeije H. Analyseren in kwalitatief onderzoek. Denken en doen [Analyzing in qualitative research. Think and do.] (2nd Edition). The Hague, Netherlands: Boom Lemma Publishers; 2012.

45. Miles MB, Huberman AM. Qualitative Data Analysis: A Sourcebook (2nd Edition). Thousand Oaks, CA: Sage Publications; 1994.
46. Wester F, Peters V. Kwalitatieve analyse: uitgangspunten en procedures [Qualitative analysis: starting points and procedures]. Bussum, NL: Coutinho; 2004.
47. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310]
48. Neuendorf KA. The Content Analysis Guidebook. Thousand Oaks, CA: Sage Publications; 2002.
49. IJsselsteijn W, van Baren J, Markopoulos P, Romero N, de Ruyter B. Measuring affective benefits and costs of mediated awareness: development and validation of the ABC-Questionnaire. In: Markopoulos P, de Ruyter B, Mackay W, editors. Awareness Systems. Human-Computer Interaction Series. London, UK: Springer; 2009:473-488.
50. Yarosh S, Markopoulos P, Abowd GD. Towards a questionnaire for measuring affective benefits and costs of communication technologies. : ACM; 2014 Presented at: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing; February 15-19, 2014; Baltimore, MD p. 84-96. [doi: 10.1145/2531602.2531634]
51. Erekson DM, Lambert MJ, Eggett DL. The relationship between session frequency and psychotherapy outcome in a naturalistic setting. J Consult Clin Psychol 2015 Dec;83(6):1097-1107. [doi: 10.1037/a0039774] [Medline: 26436645]

Abbreviations

DIT: diffusion of innovations theory
eMental health: electronic mental health
LAMH: Levels of Adoption of eMental Health
TAM: technology acceptance model

Edited by G Eysenbach; submitted 23.11.17; peer-reviewed by K Weisel, T Antunes, M Beutel; comments to author 11.01.18; revised version received 20.02.18; accepted 23.02.18; published 24.04.18.

Please cite as:
Feijt MA, de Kort YAW, Bongers IMB, 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;20(4):e153
URL: http://www.jmir.org/2018/4/e153/
doi:10.2196/jmir.9485
PMID:29691215

Review

The Use of Virtual Reality in Patients with Eating Disorders: Systematic Review

Damien Clus[1], MD; Mark Erik Larsen[2], PhD; Christophe Lemey[3], MD; Sofian Berrouiguet[4,5], MD

[1]Department of Mental Health, University Hospital of Brest, Brest, France

[2]Black Dog Institute, University of New South Wales, Sydney, Australia

[3]Department of Mental Health, University Hospital of Brest, Université de Bretagne Occidentale, Brest, France

[4]UMR CNRS 6285 Lab-STICC, Institut Mines Télécom Atlantique, Université Bretagne Loire F-29238 Brest, Brest, France

[5]EA 7479 Soins Primaires, Santé Publique et Registre des cancers de Bretagne Occidentale, Department of Mental Health, Université de Bretagne Occidentale, Brest, France

Corresponding Author:
Sofian Berrouiguet, MD
UMR CNRS 6285 Lab-STICC
Institut Mines Télécom Atlantique
Université Bretagne Loire F-29238 Brest
Brest Medical University Hospital at Bohars, Adult Psychiatry
Route de Ploudalmezeau
Brest, 29200
France
Phone: 33 668204178
Email: sofian.berrouiguet@gmail.com

Abstract

Background: Patients with eating disorders are characterized by pathological eating habits and a tendency to overestimate their weight and body shape. Virtual reality shows promise for the evaluation and management of patients with eating disorders. This technology, when accepted by this population, allows immersion in virtual environments, assessment, and therapeutic approaches, by exposing users to high-calorie foods or changes in body shape.

Objective: To better understand the value of virtual reality, we conducted a review of the literature, including clinical studies proposing the use of virtual reality for the evaluation and management of patients with eating disorders.

Methods: We searched PubMed, PsycINFO, ScienceDirect, the Cochrane Library, Scopus, and Web of Science up to April 2017. We created the list of keywords based on two domains: virtual reality and eating disorders. We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses to identify, select, and critically appraise relevant research while minimizing bias.

Results: The initial database searches identified 311 articles, 149 of which we removed as duplicates. We analyzed the resulting set of 26 unique studies that met the inclusion criteria. Of these, 8 studies were randomized controlled trials, 13 were nonrandomized studies, and 5 were clinical trials with only 1 participant. Most articles focused on clinical populations (19/26, 73%), with the remainder reporting case-control studies (7/26, 27%). Most of the studies used visual immersive equipment (16/26, 62%) with a head-mounted display (15/16, 94%). Two main areas of interest emerged from these studies: virtual work on patients' body image (7/26, 27%) and exposure to virtual food stimuli (10/26, 38%).

Conclusions: We conducted a broad analysis of studies on the use of virtual reality in patients with eating disorders. This review of the literature showed that virtual reality is an acceptable and promising therapeutic tool for patients with eating disorders.

(J Med Internet Res 2018;20(4):e157) doi:10.2196/jmir.7898

KEYWORDS

virtual reality exposure therapy; feeding and eating disorders; binge-eating disorder; anorexia nervosa; bulimia nervosa

Introduction

Patients with eating disorders are characterized by pathological eating habits and a tendency to overestimate their weight and body shape, according to the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (DSM-5) [1] and the *International Classification of Diseases, Tenth Revision* [2]. Eating disorders are associated with severe medical and psychological outcomes [3], a high risk of death [4,5], and high public health costs [6]. Prevention programs, such as media literacy and psychoeducation, have been put in place to limit these consequences [7]. A large American retrospective survey [8] found prevalences of 0.9% for anorexia nervosa, 1.5% for bulimia nervosa, and 3.5% for binge eating disorder among women and prevalences of 0.3%, 0.5%, and 2%, respectively, among men. These rates were based on the fourth edition of the DSM and would be higher with the application of DSM-5 criteria [9].

International guidelines from the UK National Institute for Health and Care Excellence [10] and the American Psychiatric Association [11] recommend the psychological management of patients with eating disorders, mainly cognitive behavioral therapy (CBT) approaches. The choice of psychotherapy is made according to the patient's preference; his or her social support, age, and motivation; and the stage of the disease. According to the recommendations of the World Federation of Societies of Biological Psychiatry [12], there is no specific drug therapy for eating disorders. The perceptual component of a person's body image is addressed during the psychotherapeutic follow-up, with the possible use of photography or video; however, there is a lack of consensus on the media that should be used. Virtual reality (VR) is one possibility.

VR can be defined as a computer technology that reproduces a real or imaginary environment and that simulates the user's presence in that physical environment, with which the user can interact through engagement of his or her senses (sight, touch, hearing, and smell). VR has been assessed in mental health conditions [13-15] for the rehabilitation of patients with schizophrenia [16], for the treatment of posttraumatic stress disorder symptomatology [17], or for the management of phobic disorders [18]. VR exposes users to interactive 3-dimensional environments that simulate a specific situation [19] and, through guided imagination, overcomes the disadvantages of exposure to a real-life situation, including a possible lack of control over participants' thoughts and imaginative difficulties [20]. Exposing patients to VR allows delivery of therapy in a form that they may find more acceptable [21]. The virtual environment makes it possible to control the unexpected and to be exposed in a safe environment to certain fears that may be difficult to reproduce in real situations [22], and it affords a greater degree of confidentiality [18]. The main limitations of this technology are the insufficient number of therapists trained in its use, side effects such as "simulator sickness" [23,24], and the high cost of equipment.

Technological advances have made it possible to improve VR techniques, which has led to an increasing number of studies on this subject. However, the latest literature reviews on the use of VR for eating disorders were conducted in 2012 [25-27]. A more recent review of the literature on the use of VR in general psychological treatment of mental illness [13] did not include keywords or derivatives of "eating disorders."

To better understand the value of this technology, we reviewed the literature including clinical studies proposing the use of VR in patients with eating disorders. The objective of this study was to provide a thorough review of the applications of VR in patients with eating disorders.

Methods

We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [28] to identify, select, and critically appraise relevant research while minimizing bias.

Search Strategy

We extensively searched PubMed, PsycINFO, Science Direct, the Cochrane Library, Scopus, and Web of Science up to April 2017. We based the list of keywords on two domains: virtual reality and eating disorders. A search strategy was constructed using the Boolean operators AND and OR and applied in the Medical Subject Heading, title, and abstract fields. The keywords and search strategy we used were (virtual reality OR virtual environment OR immersive reality) AND (eating disorders OR anorexia OR bulimia OR disordered eating OR binge eating disorder) in any language, but referenced in the selected databases and meeting the inclusion criteria. To limit the risk of selection bias, we applied no restrictions in terms of article type or clinical population. We did not include serious games, defined as video games with a pedagogical, diagnostic, or therapeutic interest, in the search terms because these do not necessarily rely on immersive VR technology. We excluded all articles not written in English.

Study Selection

We included randomized controlled trials and nonrandomized studies. The primary end point was the collection of studies using VR for eating disorders, specifically articles that addressed the use of VR techniques in an eating disorder sample only, in comparison with either another treatment condition or a control group. We excluded studies without a clinical population and theoretical articles presenting an application of VR without final results. We also excluded studies exploring the use of serious games, without the use of VR, for these disorders.

Data Extraction

We screened each study and extracted data independently using standard forms. The following information was extracted from each study: first author's last name, publication year, and country; study design; population analyzed; number of patients; evaluation scales used; study objectives; study protocol; and study results.

Results

Figure 1 shows the PRISMA flowchart summarizing the stages of the review. The initial database searches identified 311 articles, 149 of which we removed as duplicates. Following

review of the titles and abstracts, we excluded 119 articles. We downloaded the remaining 43 articles for full-text review, following which we excluded an additional 17 articles: 7 articles presented a protocol without results; 3 studies evaluated healthy participants; 2 articles were not entirely in English; 2 articles reported studies already included; 2 studies reported preliminary data; and 1 study described a serious game without VR. We analyzed the resulting set of 26 unique studies that met the inclusion criteria.

Multimedia Appendix 1 [29-54] presents the results of this review in a descriptive way and in the form of a table.

Study Authors, Year of Publication, and Country of Origin

The team of Riva et al [29-36] produced approximately one-third of the included articles (8/26, 31%). The last literature reviews on this subject [25-27] included articles that were published up to 2010. Since then, 12 (46%) articles included in our selection were published, suggesting an increase in interest in using these new technologies for patients with eating disorders (Figure 2).

All the selected articles were from European teams and 2 countries are the most represented in this literature review: Italy (14/26, 54%) and Spain (11/26, 42%); the remaining study was conducted in the United Kingdom.

Study Design, Population, and Sample Size

Concerning the study methodologies, 8 studies were randomized controlled trials, 13 were nonrandomized studies, and 5 were clinical trials with only 1 participant. Most articles focused on clinical populations (19/26, 73%), with the remainder reporting case-control studies (7/26, 27%). Most participants were female (24/26, 92%); only 2 studies included mixed samples (male and female). The clinical populations observed in this literature review were heterogeneous from a diagnostic point of view in 15 studies (58%). The other studies targeted specific disordered eating: obese patients without psychiatric comorbidities (n=6, 23%), anorexia nervosa (n=3, 12%), and binge eating disorder (n=3, 12%). The numbers of patients varied from single case studies (5/26, 19%) to larger samples of more than 100 patients (8/26, 31%).

Figure 1. Flow diagram of study selection. MeSH: medical subject headings.

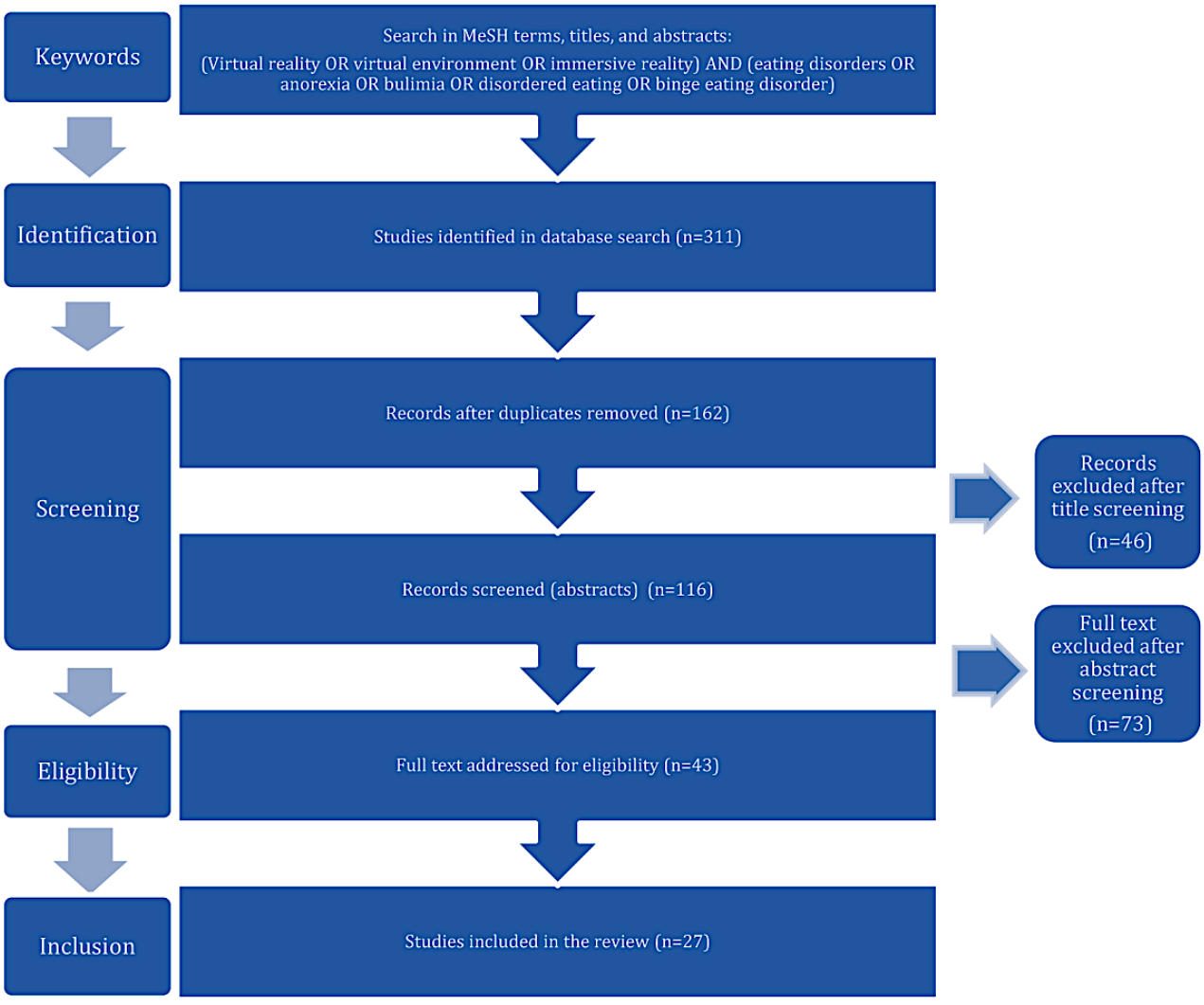
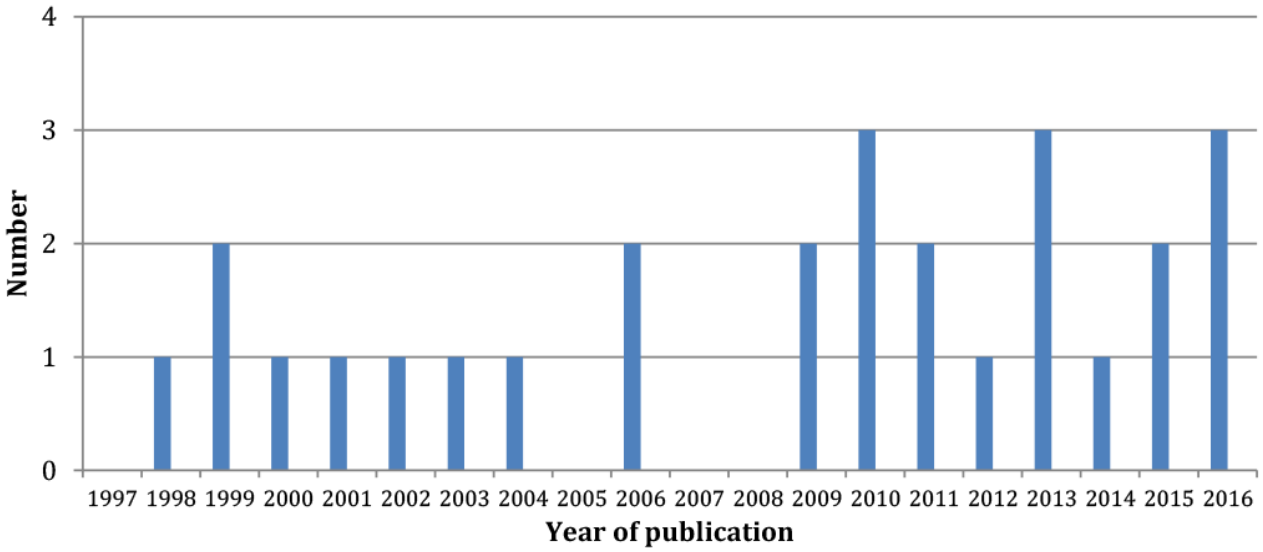


Figure 2. Number of articles included in the review by year published.



Technology Used

Most of the studies used visual immersive equipment (16/26, 62%) with a head-mounted display (15/16, 94%). One study used a cubic immersive room with the projection of a virtual environment on the walls (cave system). A total of 7 studies used audio stimuli in combination with visual kinematics to allow user–virtual world interaction and gradual exposure to high-calorie food stimuli associated with chewing sounds or comments on ingested foods. Also, only 1 study used visual-tactile stimulation complemented with immersive material, which included an obese patient to test the effect on her assessment of her body measurements.

Objectives of the Studies

Two main areas of interest emerged from these studies: virtual work on patients' body image (7/26, 27%) and exposure to virtual food stimuli (10/26, 38%). Some study protocols analyzed both of these fields (6/26, 23%). A total of 15 (58%) studies had a primary therapeutic objective; 11 (42%) studies evaluated the users' tolerance of the protocol and their emotional reactions during the VR immersion. Multimedia Appendix 1 presents the main results.

Discussion

Principal Findings

We conducted a broad analysis of studies on the use of VR in patients with eating disorders. This review of the literature showed an increase in interest in using these new technologies for patients with eating disorders, with 26 articles. Since the last reviews of the literature dealing with this topic, 12 articles have been published, testifying to the always-present interest in this technology. Commercialized VR technology is increasingly accessible to the general public, and medical research, especially in mental health, is increasingly using this new tool, with diagnostic, therapeutic, and preventive aims. The use of VR in the evaluation or treatment of patients with eating disorders is being led by European teams.

It is difficult to interpret the results of this research because of the heterogeneity of the populations studied, the studies' objectives, and the content of the VR protocols. Many studies drew conclusions without differentiating the subtypes of disordered eating, with low sample sizes.

This review of the literature nevertheless allows us to discuss certain aspects of this field of investigation.

Virtual Reality, an Acceptable and Effective Therapeutic Tool

Some studies with medium- and long-term follow-up showed less loss to follow-up with a VR protocol compared with other groups [38,54]. This can be explained by the attractiveness of new technologies and exposure to stimuli in a virtual environment in the presence of a therapist. Starting therapy could help foster a therapeutic alliance, an active participation in a process of change. Several studies observed patients' increased motivation for change [31,32,37,55].

The use of a VR module in addition to CBT showed greater efficacy in the main variables analyzed in comparison with control groups [36,52] or CBT alone [29,30,35,38,43,44,54]. European [56] and American [11] guidelines recommend the use of CBT in the psychological care of bulimia nervosa and binge eating disorder. CBT is considered one of the solutions for psychological interventions in patients with anorexia nervosa [11,56,57]. CBT approaches are recommended for patients with obesity [58]. Thus, CBTs are valid therapies in for patients with eating disorders and are viable for a comparative judgment of the effectiveness of a new approach.

These results are, however, to be qualified by certain limits cited by this research. Samples of the studies were small due to recruitment difficulties presented by the low frequency of these disorders. The difficulties that certain populations, such as those with anorexia nervosa, have in engaging with care may also explain this result. Most of this research evaluated a female population, which is explained by the female predominance of eating disorders [3]. Also, the lack of controlled randomized clinical trials leads us to be cautious in interpreting these results.

All of the studies were developed by European teams, notably by Riva's team. It is surprising that there were no studies of North American origin due to the socioeconomic impact of these pathologies on this continent [6]. Similarly, most of the technologies used to enable immersion in the virtual environment have been developed in the United States. The predominance of Riva's research team may also be related to an interpretation bias in the final analysis of the articles.

Theory of Maintenance of the Perturbations of Body Image

Some studies [37,42] advanced assumptions on the maintenance of pathological disorders. The hypothesis developed to explain the perturbations of body image in patients with eating disorders is the allocentric lock hypothesis [59,60]. Body image disorders in patients with eating disorders are related to a deficiency in their ability to update their negative body image stored in their memory (allocentric function) with sensory motor and proprioceptive inputs in real time (egocentric function). In patients with anorexia nervosa, this deficit is related to an involvement of the lower parietal lobe and precuneus, the gyrus of the inner surface of the parietal lobe of the cerebral cortex [61]. This hypothesis could also apply in patients with obesity [62]. The use of VR could make it possible to unblock this transmission [63]. The theory of objectification as a specific cognitive process is cited to understand these perturbations: a person internalizes an objectified self-image when they use a reference allocentric frame (observer mode) to recall the events in which they evaluate themselves on the basis of body appearance [64].

Virtual Environments for Patients' Disorders

The validity of using a virtual environment in a population can be judged by the individual's emotional reactions. Unique exposure to virtual food stimuli increases anxiety and changes mood [45,47,49-51,53], reproducing physiological reactions to a real situation. Similar emotional reactions were found between exposures to real and virtual foods [49]. The repetition of VR sessions with modules of exposure to food stimuli and silhouettes of the patient or mannequin reduces negative emotions [29,34,46,54], by progressive attenuation of the anxiogenic response.

Nevertheless, one may question the relevance of the results of certain studies [50,51,53] whose objective was to evaluate emotional reactions in a virtual environment. The use of an evaluation tool such as the Barcelona Depression Questionnaire, a self-questionnaire of low scientific validity measuring variations in depressed mood, warrants interpreting these conclusions cautiously.

Limitations

This analysis of the literature had some limitations. It excluded studies with a nonclinical population—these typically described the use of VR with objectives of acceptability and tolerance, to be reproduced later with clinical samples. We excluded from the analysis those articles proposing a VR study model on eating disorders without final results, as well as meeting abstracts and a protocol using a serious game, without immersive material, to evaluate impulsivity in bulimic patients [65].

In the bibliographic search, we did not use the term obesity. Indeed, according to international classifications recognized in psychiatry [1], obesity without psychiatric comorbidities is not part of the spectrum of eating disorders. Certain psychiatric pathologies belonging to this spectrum (binge eating disorder) can result in weight gain up to obesity [3]. The lifetime prevalence of obesity in patients with eating disorders is estimated at 28.8% [66]. In the final analysis, because of this frequent association, we included articles with obese participants.

Conclusion

VR is increasingly being applied in the evaluation and management of patients with eating disorders, with a recent increase in articles being published. This technology, when accepted by this population, allows patients to be immersed in virtual environments that are adapted to their psychological state, causing reactionary emotions as in real life, in a safe environment, under the supervision of a therapist. Upcoming technical improvements of VR will also provide a better sense of presence. Overall, VR techniques enable the evaluation of pathological eating behaviors and body image distortions. In addition to CBT, use of VR techniques by patients with eating disorders decreased their negative emotional responses to virtual food stimuli or exposure to their body shape.

Acknowledgments

This research was supported by the Fondation de l'Avenir.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of the reviewed articles.

[PDF File (Adobe PDF File), 93KB - jmir_v20i4e157_app1.pdf]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. Washington, DC: American Psychiatric Association Publishing; 2013.
2. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision. Geneva, Switzerland: WHO; 1992.
3. Treasure J. Eating disorders. Medicine 2012 Nov;40(11):607-612. [doi: 10.1016/j.mpmed.2012.08.005]
4. Fichter MM, Quadflieg N. Mortality in eating disorders - results of a large prospective clinical longitudinal study. Int J Eat Disord 2016 Apr;49(4):391-401. [doi: 10.1002/eat.22501] [Medline: 26767344]
5. Arcelus J, Mitchell AJ, Wales J, Nielsen S. Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. Arch Gen Psychiatry 2011 Jul;68(7):724-731. [doi: 10.1001/archgenpsychiatry.2011.74] [Medline: 21727255]
6. Samnaliev M, Noh HL, Sonneville KR, Austin SB. The economic burden of eating disorders and related mental health comorbidities: an exploratory analysis using the U.S. Medical Expenditures Panel Survey. Prev Med Rep 2015;2:32-34 [FREE Full text] [doi: 10.1016/j.pmedr.2014.12.002] [Medline: 26844048]
7. Watson HJ, Joyce T, French E, Willan V, Kane RT, Tanner-Smith EE, et al. Prevention of eating disorders: a systematic review of randomized, controlled trials. Int J Eat Disord 2016 Sep;49(9):833-862. [doi: 10.1002/eat.22577] [Medline: 27425572]
8. Hudson JI, Hiripi E, Pope HG, Kessler RC. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. Biol Psychiatry 2007 Feb 01;61(3):348-358 [FREE Full text] [doi: 10.1016/j.biopsych.2006.03.040] [Medline: 16815322]
9. Mustelin L, Silén Y, Raevuori A, Hoek HW, Kaprio J, Keski-Rahkonen A. The DSM-5 diagnostic criteria for anorexia nervosa may change its population prevalence and prognostic value. J Psychiatr Res 2016 Jun;77:85-91. [doi: 10.1016/j.jpsychires.2016.03.003] [Medline: 27014849]
10. National Institute for Health and Care Excellence. Eating Disorders in Over 8s: Management. Clinical Guideline CG9. London, UK: NICE; 2004.
11. Yager J, Devlin M, Halmi K, Herzog D, Mitchell IJ, Powers P, et al. Practice guideline for the treatment of patients with eating disorders. 3rd edition. Washington, DC: American Psychiatric Association; 2010. URL: https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/eatingdisorders.pdf [accessed 2018-03-21] [WebCite Cache ID 6y5MOSygN]
12. Aigner M, Treasure J, Kaye W, Kasper S, WFSBP Task Force On Eating Disorders. World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the pharmacological treatment of eating disorders. World J Biol Psychiatry 2011 Sep;12(6):400-443. [doi: 10.3109/15622975.2011.602720] [Medline: 21961502]
13. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: a systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]
14. Malbos E, Boyer L, Lançon C. [Virtual reality in the treatment of mental disorders]. Presse Med 2013 Nov;42(11):1442-1452. [doi: 10.1016/j.lpm.2013.01.065] [Medline: 23702202]
15. Clough BA, Casey LM. Technological adjuncts to enhance current psychotherapy practices: a review. Clin Psychol Rev 2011 Apr;31(3):279-292. [doi: 10.1016/j.cpr.2010.12.008] [Medline: 21382535]
16. Smith MJ, Fleming MF, Wright MA, Roberts AG, Humm LB, Olsen D, et al. Virtual reality job interview training and 6-month employment outcomes for individuals with schizophrenia seeking employment. Schizophr Res 2015 Aug;166(1-3):86-91 [FREE Full text] [doi: 10.1016/j.schres.2015.05.022] [Medline: 26032567]
17. McLay R, Ram V, Murphy J, Spira J, Wood DP, Wiederhold MD, et al. Effect of virtual reality PTSD treatment on mood and neurocognitive outcomes. Cyberpsychol Behav Soc Netw 2014 Jul;17(7):439-446. [doi: 10.1089/cyber.2013.0383] [Medline: 24635120]
18. Lambrey S, Jouvent R, Allilaire J, Pélissolo A. Les thérapies utilisant la réalité virtuelle dans les troubles phobiques. Ann Méd Psychol Revue Psychiatr 2010 Feb;168(1):44-46. [doi: 10.1016/j.amp.2009.10.003]
19. Rothbaum BO, Anderson P, Zimand E, Hodges L, Lang D, Wilson J. Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying. Behav Ther 2006 Mar;37(1):80-90. [doi: 10.1016/j.beth.2005.04.004] [Medline: 16942963]
20. Vincelli F. From imagination to virtual reality: the future of clinical psychology. Cyberpsychol Behav 1999;2(3):241-248. [doi: 10.1089/cpb.1999.2.241] [Medline: 19178241]
21. Garcia-Palacios A, Botella C, Hoffman H, Fabregat S. Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. Cyberpsychol Behav 2007 Oct;10(5):722-724. [doi: 10.1089/cpb.2007.9962] [Medline: 17927544]
22. Garcia-Palacios A, Hoffman HG, See SK, Tsai A, Botella C. Redefining therapeutic success with virtual reality exposure therapy. Cyberpsychol Behav 2001 Jun;4(3):341-348. [doi: 10.1089/109493101300210231] [Medline: 11710258]
23. Keshavarz B, Stelzmann D, Paillard A, Hecht H. Visually induced motion sickness can be alleviated by pleasant odors. Exp Brain Res 2015 May;233(5):1353-1364. [doi: 10.1007/s00221-015-4209-9] [Medline: 25633319]

24. Sharples S, Cobb S, Moody A, Wilson JR. Virtual reality induced symptoms and effects (VRISE): comparison of head mounted display (HMD), desktop and projection display systems. Displays 2008 Mar;29(2):58-69. [doi: 10.1016/j.displa.2007.09.005]
25. Koskina A, Campbell IC, Schmidt U. Exposure therapy in eating disorders revisited. Neurosci Biobehav Rev 2013 Feb;37(2):193-208. [doi: 10.1016/j.neubiorev.2012.11.010] [Medline: 23201859]
26. Ferrer-García M, Gutiérrez-Maldonado J. The use of virtual reality in the treatment of eating disorders. Stud Health Technol Inform 2012;181:17-21. [Medline: 22954820]
27. Ferrer-García M, Gutiérrez-Maldonado J. The use of virtual reality in the study, assessment, and treatment of body image in eating disorders and nonclinical samples: a review of the literature. Body Image 2012 Jan;9(1):1-11. [doi: 10.1016/j.bodyim.2011.10.001] [Medline: 22119329]
28. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
29. Riva G, Bacchetta M, Cesa G, Conti S, Molinari E. The use of VR in the treatment of eating disorders. Stud Health Technol Inform 2004;99:121-163. [Medline: 15295149]
30. Riva G, Bacchetta M, Cesa G, Conti S, Molinari E. Six-month follow-up of in-patient experiential cognitive therapy for binge eating disorders. Cyberpsychol Behav 2003 Jun;6(3):251-258. [doi: 10.1089/109493103322011533] [Medline: 12855080]
31. Riva G, Bacchetta M, Baruffi M, Rinaldi S, Molinari E. Experiential cognitive therapy: a VR based approach for the assessment and treatment of eating disorders. Stud Health Technol Inform 1998;58:120-135. [Medline: 10350912]
32. Riva G, Bacchetta M, Baruffi M, Rinaldi S, Molinari E. Virtual reality based experiential cognitive treatment of anorexia nervosa. J Behav Ther Exp Psychiatry 1999 Sep;30(3):221-230. [Medline: 10619546]
33. Riva G, Bacchetta M, Baruffi M, Rinaldi S, Vincelli F, Molinari E. Virtual reality-based experiential cognitive treatment of obesity and binge-eating disorders. Clin Psychol Psychother 2000 Jul 01;7(3):209-219. [doi: 10.1002/1099-0879(200007)7:3<209::AID-CPP232>3.0.CO;2-V]
34. Riva G, Cárdenas-López G, Duran X, Torres-Villalobos GM, Gaggioli A. Virtual reality in the treatment of body image disturbances after bariatric surgery: a clinical case. Stud Health Technol Inform 2012;181:278-282. [Medline: 22954871]
35. Riva G, Bacchetta M, Cesa G, Conti S, Castelnuovo G, Mantovani F, et al. Is severe obesity a form of addiction? Rationale, clinical approach, and controlled clinical trial. Cyberpsychol Behav 2006 Aug;9(4):457-479. [doi: 10.1089/cpb.2006.9.457] [Medline: 16901250]
36. Riva G, Bacchetta M, Baruffi M, Molinari E. Virtual reality-based multidimensional therapy for the treatment of body image disturbances in obesity: a controlled study. Cyberpsychol Behav 2001 Aug;4(4):511-526. [doi: 10.1089/109493101750527079] [Medline: 11708731]
37. Serino S, Scarpina F, Keizer A, Pedroli E, Dakanalis A, Castelnuovo G, et al. A novel technique for improving bodily experience in a non-operable super-super obesity case. Front Psychol 2016;7:837 [FREE Full text] [doi: 10.3389/fpsyg.2016.00837] [Medline: 27378965]
38. Manzoni GM, Cesa GL, Bacchetta M, Castelnuovo G, Conti S, Gaggioli A, et al. Virtual reality-enhanced cognitive-behavioral therapy for morbid obesity: a randomized controlled study with 1 year follow-up. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):134-140. [doi: 10.1089/cyber.2015.0208] [Medline: 26430819]
39. Perpiñá C, Roncero M. Similarities and differences between eating disorders and obese patients in a virtual environment for normalizing eating patterns. Compr Psychiatry 2016 May;67:39-45. [doi: 10.1016/j.comppsych.2016.02.012] [Medline: 27095333]
40. Pla-Sanjuanelo J, Ferrer-Garcia M, Gutiérrez-Maldonado J, Vilalta-Abella F, Andreu-Gracia A, Dakanalis A, et al. Trait and state craving as indicators of validity of VR-based software for binge eating treatment. Stud Health Technol Inform 2015;219:141-146. [Medline: 26799896]
41. Ferrer-Garcia M, Gutiérrez-Maldonado J, Pla-Sanjuanelo J, Vilalta-Abella F, Andreu-Gracia A, Dakanalis A, et al. External eating as a predictor of cue-reactivity to food-related virtual environments. Stud Health Technol Inform 2015;219:117-122. [Medline: 26799891]
42. Serino S, Dakanalis A, Gaudio S, Carrà G, Cipresso P, Clerici M, et al. Out of body, out of space: impaired reference frame processing in eating disorders. Psychiatry Res 2015 Dec 15;230(2):732-734. [doi: 10.1016/j.psychres.2015.10.025] [Medline: 26541204]
43. Cesa GL, Manzoni GM, Bacchetta M, Castelnuovo G, Conti S, Gaggioli A, et al. Virtual reality for enhancing the cognitive behavioral treatment of obesity with binge eating disorder: randomized controlled study with one-year follow-up. J Med Internet Res 2013;15(6):e113 [FREE Full text] [doi: 10.2196/jmir.2441] [Medline: 23759286]
44. Marco JH, Perpiñá C, Botella C. Effectiveness of cognitive behavioral therapy supported by virtual reality in the treatment of body image in eating disorders: one year follow-up. Psychiatry Res 2013 Oct 30;209(3):619-625. [doi: 10.1016/j.psychres.2013.02.023] [Medline: 23499231]

45. Perpiñá C, Roncero M, Fernández-Aranda F, Jiménez-Murcia S, Forcano L, Sánchez I. Clinical validation of a virtual environment for normalizing eating patterns in eating disorders. Compr Psychiatry 2013 Aug;54(6):680-686. [doi: 10.1016/j.comppsych.2013.01.007] [Medline: 23452907]
46. Cardi V, Krug I, Perpiñá C, Mataix-Cols D, Roncero M, Treasure J. The use of a nonimmersive virtual reality programme in anorexia nervosa: a single case-report. Eur Eat Disord Rev 2012 May;20(3):240-245. [doi: 10.1002/erv.1155] [Medline: 21932282]
47. Ferrer-Garcia M, Gutierrez-Maldonado J. Virtual reality exposure in patients with eating disorders: influence of symptom severity and presence. Stud Health Technol Inform 2011;167:80-85. [Medline: 21685646]
48. Gutiérrez-Maldonado J, Ferrer-García M, Caqueo-Urízar A, Moreno E. Body image in eating disorders: the influence of exposure to virtual-reality environments. Cyberpsychol Behav Soc Netw 2010 Oct;13(5):521-531. [doi: 10.1089/cyber.2009.0301] [Medline: 20950176]
49. Gorini A, Griez E, Petrova A, Riva G. Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. Ann Gen Psychiatry 2010 Jul 05;9:30 [FREE Full text] [doi: 10.1186/1744-859X-9-30] [Medline: 20602749]
50. Ferrer-García M, Gutiérrez-Maldonado J. Effect of the mood produced by virtual reality exposure on body image disturbances. Stud Health Technol Inform 2010;154:44-49. [Medline: 20543267]
51. Ferrer-García M, Gutiérrez-Maldonado J, Caqueo-Urízar A, Moreno E. The validity of virtual environments for eliciting emotional responses in patients with eating disorders and in controls. Behav Modif 2009 Nov;33(6):830-854. [doi: 10.1177/0145445509348056] [Medline: 19822737]
52. Manzoni GM, Pagnini F, Gorini A, Preziosa A, Castelnuovo G, Molinari E, et al. Can relaxation training reduce emotional eating in women with obesity? An exploratory study with 3 months of follow-up. J Am Diet Assoc 2009 Aug;109(8):1427-1432. [doi: 10.1016/j.jada.2009.05.004] [Medline: 19631051]
53. Gutiérrez-Maldonado J, Ferrer-García M, Caqueo-Urízar A, Letosa-Porta A. Assessment of emotional reactivity produced by exposure to virtual environments in patients with eating disorders. Cyberpsychol Behav 2006 Oct;9(5):507-513. [doi: 10.1089/cpb.2006.9.507] [Medline: 17034316]
54. Perpiñá C, Botella C, Baños R, Marco H, Alcañiz M, Quero S. Body image and virtual reality in eating disorders: is exposure to virtual reality more effective than the classical body image treatment? Cyberpsychol Behav 1999;2(2):149-155. [doi: 10.1089/cpb.1999.2.149] [Medline: 19178251]
55. Riva G, Bacchetta M, Baruffi M, Rinaldi S, Molinari E. Experiential cognitive therapy in anorexia nervosa. Eat Weight Disord 1998 Sep;3(3):141-150. [Medline: 10728164]
56. National Institute for Clinical Excellence. Eating disorders: core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders. London, UK: NICE URL: https://www.nice.org.uk/guidance/cg9/documents/eating-disorders-information-for-the-public-second-consultation2 [accessed 2018-03-21] [WebCite Cache ID 6y5N7bd7I]
57. Haute Autorité de Santé. Recommandation de bonne pratique: Anorexie mentale: prise en charge. 2010. URL: https://www.has-sante.fr/portail/jcms/c_985715/fr/anorexie-mentale-prise-en-charge [accessed 2018-03-21]
58. Haute Autorité de Santé. Recommandation de bonne pratique: Surpoids et obésité de l'adulte: prise en charge médicale de premiers recours. 2011. URL: https://www.has-sante.fr/portail/jcms/c_964938/fr/surpoids-et-obesite-de-l-adulte-prise-en-charge-medicale-de-premier-recours [accessed 2018-03-21]
59. Riva G. Out of my real body: cognitive neuroscience meets eating disorders. Front Hum Neurosci 2014;8:236 [FREE Full text] [doi: 10.3389/fnhum.2014.00236] [Medline: 24834042]
60. Riva G. Neuroscience and eating disorders: the allocentric lock hypothesis. Med Hypotheses 2012 Feb;78(2):254-257. [doi: 10.1016/j.mehy.2011.10.039] [Medline: 22100628]
61. Riva G, Gaudio S. Allocentric lock in anorexia nervosa: new evidences from neuroimaging studies. Med Hypotheses 2012 Jul;79(1):113-117. [doi: 10.1016/j.mehy.2012.03.036] [Medline: 22546757]
62. Riva G, Gaudio S, Dakanalis A. I'm in a virtual body: a locked allocentric memory may impair the experience of the body in both obesity and anorexia nervosa. Eat Weight Disord 2014 Mar;19(1):133-134. [doi: 10.1007/s40519-013-0066-3] [Medline: 24068525]
63. Riva G. The key to unlocking the virtual body: virtual reality in the treatment of obesity and eating disorders. J Diabetes Sci Technol 2011 Mar;5(2):283-292 [FREE Full text] [Medline: 21527095]
64. Riva G, Gaudio S, Dakanalis A. The neuropsychology of self-objectification. Eur Psychol 2015 Jan;20(1):34-43. [doi: 10.1027/1016-9040/a000190]
65. Giner-Bartolomé C, Fagundo AB, Sánchez I, Jiménez-Murcia S, Santamaría JJ, Ladouceur R, et al. Can an intervention based on a serious videogame prior to cognitive behavioral therapy be helpful in bulimia nervosa? A clinical case study. Front Psychol 2015;6:982 [FREE Full text] [doi: 10.3389/fpsyg.2015.00982] [Medline: 26236261]
66. Villarejo C, Fernández-Aranda F, Jiménez-Murcia S, Peñas-Lledó E, Granero R, Penelo E, et al. Lifetime obesity in patients with eating disorders: increasing prevalence, clinical and personality correlates. Eur Eat Disord Rev 2012 May;20(3):250-254 [FREE Full text] [doi: 10.1002/erv.2166] [Medline: 22383308]

Abbreviations

CBT: cognitive behavioral therapy
DSM: Diagnostic and Statistical Manual of Mental Disorders
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
VR: virtual reality

Edited by G Eysenbach; submitted 28.04.17; peer-reviewed by C Buche, Q Cheng, K Gurditta, S Tuo; comments to author 15.06.17; revised version received 28.09.17; accepted 16.11.17; published 27.04.18.

Please cite as:
Clus D, Larsen ME, Lemey C, Berrouiguet S
The Use of Virtual Reality in Patients with Eating Disorders: Systematic Review
J Med Internet Res 2018;20(4):e157
URL: http://www.jmir.org/2018/4/e157/
doi:10.2196/jmir.7898
PMID:29703715

Original Paper

The Korean eHealth Literacy Scale (K-eHEALS): Reliability and Validity Testing in Younger Adults Recruited Online

SeonYoon Chung[1], RN, PhD; Bu Kyung Park[2], RN, PhD; Eun-Shim Nahm[3], RN, PhD

[1]Mennonite College of Nursing, Illinois State University, Normal, IL, United States

[2]College of Nursing, Research Institute of Nursing Science, Kyungpook National University, Daegu, Republic Of Korea

[3]School of Nursing, University of Maryland, Baltimore, MD, United States

Corresponding Author:
Bu Kyung Park, RN, PhD
College of Nursing, Research Institute of Nursing Science
Kyungpook National University
640 Gukchaebosang-ro Jung-gu
Daegu, 41944
Republic Of Korea
Phone: 82 53 420 4929
Fax: 82 53 421 2758
Email: bukpark@knu.ac.kr

Abstract

Background: In this digital era, eHealth literacy is an essential skill set to leverage health information available online to promote health outcomes. South Korea has an advanced health information technology infrastructure, including widespread use of the internet and mobile phones. A few studies have explored eHealth literacy in South Korea using translated versions of the eHEALS; however, they were not fully validated. A unified reliable and valid assessment tool is critical to assess and enhance the eHealth literacy level across the population.

Objective: The aim was to develop a Korean version of eHealth Literacy Scale (K-eHEALS) and evaluate its reliability and validity employing healthy young adults in Korea.

Methods: The K-eHEALS was developed based on eHEALS, a widely used tool that measures eHealth literacy, and was validated using a sample of 500 young adults recruited from a pool of a Korean internet survey panel. Content validity was assessed using the content validity index (CVI) for individual items and for scale. Construct validity was examined using exploratory factor analysis and hypothesis testing. The Cronbach alpha coefficient was used to determine the internal consistency and the Pearson correlation coefficient was used to evaluable the stability of the measure (n=55).

Results: Both individual and scale CVIs were acceptable (individual CVIs>0.67; scale CVI=0.83). Single factors accounting for 50.3% of the variance in the scales were extracted revealing the unidimensional latent structure of K-eHEALS. Hypothesis testing showed significant association between eHealth literacy and hours of internet use per day, supporting the construct validity. Items of the K-eHEALS were internally consistent (Cronbach alpha=.88) and stable over a 1-month period (r=.754, P<.001).

Conclusions: The findings of this study suggest that K-eHEALS is a valid and reliable measure of eHealth literacy in Korean young adults. Additional studies are needed with more diverse groups of adults in Korea.

(J Med Internet Res 2018;20(4):e138) doi:10.2196/jmir.8759

KEYWORDS

eHEALS; eHealth; literacy; reliability; validity

Introduction

In this digital era, the internet and mobile devices are integral to our daily life and the majority of the population uses the internet and social media to find health information [1]. Literacy in eHealth is an essential skillset to leverage these resources and produce better outcomes [2]. The eHealth Literacy Scale (eHEALS) is a measure to assess eHealth literacy initially developed in Canada by Norman and Skinner [3]. It is based on the Lily model that outlines six core literacies: (1) traditional literacy, (2) health literacy, (3) information literacy, (4) scientific literacy, (5) media literacy, and (6) computer literacy. The

validity and reliability of eHEALS have been evaluated in numerous age groups and it has been translated globally [4-12]. However, the existing Korean versions of eHEALS (K-eHEALS) are limited in the number of items translated for use or have not been tested for their reliability and validity.

Korea has advanced health information technology infrastructure including widespread use of the internet and mobile phones [13,14]. In Korea, the concept of health literacy was introduced in the 2000s and a Korean health literacy assessment tool was first developed in 2005 [15]. This tool, Korean Health Literacy Assessment Tool, was developed by translating and modifying the Rapid Estimate of Adult Literacy in Medicine (REALM) [16] in the context of Korean culture [15]. More recently, the most frequently used tools were the Korean version of REALM, Test of Functional Health Literacy in Adults, and Newest Vital Sign [17]. These tools, however, are limited in assessing eHealth literacy.

A few studies have explored eHealth literacy in Korea. For example, Lee et al [18] examined how eHealth literacy affects communication between patients and doctors. Cho et al [19] examined the effects of cognitive factors including eHealth literacy on the health app use. More recently, eHealth literacy was assessed among nursing students in Korea [20] and relationships between eHealth literacy and health behaviors were examined in Korean adults [21]. These studies used a translated version of eHEALS [3] to measure eHealth literacy. The translated version of the eHEALS used in those studies, however, was not fully validated [18,20,21]. Not having a reliable and valid Korean version of eHealth literacy scale can significantly limit eHealth literacy in Korea that is necessary to optimize the development of interventions aimed at promoting eHealth literacy. This highlights the importance of having a unified and reliable assessment tool to compare eHealth literacy levels across the population. Thus, the purpose of this study was to develop a full eight-item K-eHEALS and evaluate its reliability and validity among healthy young adults in Korea.

Methods

Design

In this study, we translated eHEALS and tested the psychometric aspects of the measure using a survey. The initial data collection was conducted from September 5 to 16, 2016. To test the stability of the measure, another wave of the survey was conducted from October 3 to 10, 2016.

Participants

The study participants were recruited (N=500) from a pool of registrants of Banana Lab, a Korean internet survey panel service agency. The survey was conducted using SurveyMonkey (SurveyMonkey Inc, Palo Alto, CA, USA). Banana Lab is an exclusive survey panel service agency working with SurveyMonkey Korea that had approximately 437,511 voluntarily registered participants in 2016. The sample size was determined based on previous studies that used a subject-to-item ratio greater than or equal to two or a sample size greater than or equal to 100 to validate a scale [22]. Potential participants who met inclusion criteria (younger adults aged 20-39 years in South Korea) received an email asking to respond to an online survey. A proportioned quota sampling by sex and age was used to minimize selection bias. For equal distribution of the younger adults, we recruited equal sample size for four sample groups by sex and age (age 20-29 male: 25%; age 20-29 female: 25%; age 30-39 male: 25%; age 30-39 female: 25%). The survey panel service agency randomly selected the potential participants from a list of registered participants with demographic information. The survey was sent out via email to potential participants in accordance to the proportion for stratified sampling, and this continued until it met our target number of participants in each quota.

Among the participants who completed the K-eHEALS, a subset of participants was randomly selected and invited to complete the same survey for test-retest reliability testing after 1 month from the initial response. If randomly selected participants did not respond, the next randomly selected participant was invited until 55 participants completed the survey. After the targeted number of 55 participants voluntarily signed and completed the survey, the online survey was closed.

Measurement

Sociodemographic Attributes and Internet Use Behaviors

The survey included items on demographics and internet use behaviors, including hours of internet use per day, purpose of internet use, types of health information searched, level of trust for health information, and its usefulness. Previous studies found internet usage was related to eHealth literacy; thus, hours of internet use was measured to validate construct validity of K-eHEALS.

The eHEALS

The eHEALS is composed of eight items measuring eHealth literacy on a 5-point Likert scale (1=strongly disagree, 5=strongly agree), with a total score that ranges from 8 to 40, in which a higher score indicates higher literacy. The eight items measure perceived knowledge, skills, and confidence in locating, evaluating, and using electronic health information to make health decisions. The measure also includes two additional questions that are not included in the total score. These two questions assess the perception of the internet as a tool to assess health information and make decisions about health.

The K-eHEALS

The K-eHEALS was developed following the process of translation and adaptation of instrument proposed by the World Health Organization [23]. After we acquired permission from the original developers of the eHEALS (Dr Cameron and Dr Norman), two bilingual professionals in nursing (two of the authors) conducted forward translation independently and compared the translated instrument. Next, a bilingual expert panel, consisting of four faculty members of nursing schools in Korea and the United States and two professionals in computer and information technology, evaluated the translated instrument. All six members of the expert panel rated each item of the K-eHEALS in terms of its relevance to the underlying construct on a 4-point scale (1=not relevant, 2=somewhat relevant, 3=quite relevant, 4=highly relevant) [24]. In addition, the expert panel

commented on each item if they had any suggestions or questions. Through the expert panel discussion, the translated instrument was adjusted and the complete K-eHEALS was produced. Then, back-translation of the K-eHEALS was done by an independent translator and was compared against the original eHEALS. Multimedia Appendix 1 is the final version of the K-eHEALS.

Ethical Considerations

The study was reviewed and approved as an exempt study by the corresponding author's Institutional Review Boards (1041078-201608-HRSB-151-01). The data collection was conducted by a third party, Banana Lab, a Korean internet survey panel service agency; the researchers were not in direct contact with the potential participants and received deidentified survey data. Eligible individuals who were interested in the study received an email including a link to the online consent form, which explained the study and other information listed in the face-to-face consent form. On review of the form, those who agreed to participate in the survey typed their name and clicked on the "I agree" button to consent to participate in the study and proceed to the survey.

Data Analysis

Validity

Content Validity

To determine the content validity index (CVI) for individual items, six members of the expert panel rated each item in terms of its relevance to the underlying construct on a 4-point scale. Then, individual CVI was computed for each item as the number of experts giving a rating of either 3 or 4, divided by the number of experts (the proportion of agreement about relevance). The CVI for the scale was calculated as the mean of the individual CVIs for all items on the scale [25]. An individual CVI higher than 0.78 was considered excellent; a scale CVI higher than 0.80 was considered acceptable [26].

Construct Validity: Exploratory Factor Analysis

After content validity was confirmed, the Korean eHEALS was administered to young adults through an online survey. Exploratory factor analysis was conducted to ensure the construct validity [27]. Sufficiency of the sample size relative to the number of items was determined using the Kaiser-Meyer-Olkin value (>0.70) and factorability of the data were evaluated based on the Bartlett test of sphericity. A scree plot and the eigenvalue (>1) were used to determine the number of factors to be extracted [28].

Construct Validity: Hypothesis Testing

Construct validity was also assessed using a hypothesis-testing approach [27]. Based on prior studies [2,12,29], we hypothesized that young adults who used the internet for more hours would have higher eHealth literacy scores. Both analysis of variance and Tukey post hoc analyses were used to test the association between duration of internet use and eHealth literacy.

Reliability

Interitem Consistency

Interitem reliability was calculated by Cronbach alpha. A value of .70 or higher was considered acceptable [30].

Stability of the Measure

Test-retest reliability testing was conducting using Pearson correlation.

Results

Characteristics of the Young Adult Participants

The total number of young adults included in the study was 500. Half of the participants were male (n=250). Half of the participants were in their twenties, whereas the other half were in their thirties. More than two-thirds of the participants were single (72.6%, 363/500) and had at least some university education (437/500, 87.4%) (Table 1). The highest percentage of participants used the internet for 1 to 3 hours per day (personal computer; PC: 38.6%, 193/500; portable device: 49.2%, 246/500), followed by 4 to 7 hours per day (PC: 29.6%, 148/500; portable device: 23.6%, 118/500). More than half of the participants (54.6%, 273/500) used the internet to search for information. Types of health information searched for included healthy lifestyle (45.0%, 225/500), disease (32.8%, 164/500), and treatment and medicine (15.8%, 79/500). More than half (53.6%, 268/500) of the participants neither trusted nor distrusted health information online, but 64.8% (324/500) reported the internet is useful in making decisions about health and 64.4% (322/500) reported the internet is important to have access to health resources. The mean total score on the K-eHEALS was 28.06 (SD 4.80, range 8-40) (Table 1). The mean of items in the K-eHEALS was 3.51 ranging from 3.31 to 3.68 (item range 1-5) (Table 2).

Content Validity

The individual CVI were excellent, scoring higher than 0.78 for all items except for item 3 ("I know how to find helpful health resources on the internet") and item 6 ("I have the skills I need to evaluate the health resources I find on the internet"). Although scoring was lower than 0.78 for two items, the experts commented that the underlying construct for these two items were apparent but they gave them a low individual CVI score because of the low fluency of the translation. Therefore, the K-eHEALS was used after editing items 3 and 6 per the experts' suggestions. The scale CVI was acceptable, scoring higher than 0.80 (scale CVI=0.83) (Table 3).

Table 1. Characteristics of the young adult participants (N=500).

Characteristics	n (%)
Gender	
Male	250 (50.0)
Female	250 (50.0)
Age group (years)	
20-29	250 (50.0)
30-39	250 (50.0)
Education level	
High school diploma	166 (33.2)
University degree and above	334 (66.8)
Marital status	
Single	363 (72.6)
Married	137 (27.4)
Internet use on personal computer per day (hours)	
<1	55 (11.0)
1-3	193 (38.6)
4-7	148 (29.6)
>8	104 (20.8)
Internet use on phone per day (hours)	
l<1	77 (15.4)
1-3	246 (49.2)
4-7	118 (23.6)
>8	59 (11.8)
Purpose of internet use	
Social networking service (eg, Kakaotalk,[a] Instagram)	171 (34.2)
Searching information	273 (54.6)
Game	30 (6.0)
Others	26 (5.2)
Health information searched for	
Disease	164 (32.8)
Healthy lifestyle	225 (45.0)
Medicine	36 (7.2)
Treatment	43 (8.6)
Medical personnel	4 (0.8)
Others	28 (5.6)
Engine used to search health information	
Google	38 (7.6)
Naver	423 (84.6)
Daum	32 (6.4)
YouTube	5 (1.0)
Others	2 (0.4)
Level of trust for health information	
Strongly trust	11 (2.2)

Characteristics	n (%)
Quite trust	187 (37.4)
Neutral	268 (53.6)
Quite distrust	31 (6.2)
Never trust	3 (0.6)
How useful do you feel the internet is in helping you in making decisions about your health?	
Not useful at all	4 (0.8)
Not useful	29 (5.8)
Unsure	120 (24.0)
Useful	324 (64.8)
Very useful	23 (4.6)
How important is it for you to be able to access health resources on the internet?	
Not important at all	4 (0.8)
Not important	22 (4.4)
Unsure	105 (21.0)
Important	322 (64.4)
Very important	47 (9.4)

[a]Kakaotalk is one of the most popular messenger apps in South Korea.

Table 2. Total and item means for the K-eHEALS in young adult participants (N=500).

K-eHEALS items	Mean (SD)
1. I know what health resources are available on the internet	3.53 (0.76)
2. I know where to find helpful health resources on the internet	3.47 (0.80)
3. I know how to find helpful health resources on the internet	3.59 (0.80)
4. I know how to use the internet to answer my questions about health	3.68 (0.77)
5. I know how to use the health information I find on the internet to help me	3.62 (0.77)
6. I have the skills I need to evaluate the health resources I find on the internet	3.31 (0.85)
7. I can tell high quality from low quality health resources on the internet	3.41 (0.87)
8. I feel confident in using information from the internet to make health decisions	3.44 (0.81)
Total means	3.51

Construct Validity-Exploratory Factor Analysis

The results supported the validity of the K-eHEALS. The Bartlett test of sphericity was significant (χ^2_{28}=1859.0, P<.001) suggesting the factorability of the correlation matrix. The results of the Kaiser-Meyer-Olkin test (0.91) was high, showing adequate sampling relative to the number of items present. Based on the initial eigenvalue (4.52) and the scree plot that was suggestive of a unidimensional latent structure (Figure 1), a single factor was retained. In this single factor model, the sum of squared loadings of the eight items on the extracted factor based on maximum likelihood method was 4.02, explaining 50.3% of the variance in the scale (Table 4).

Construct Validity-Hypothesis Testing

The results from hypothesis testing further supported the construct validity of the K-eHEALS (Table 5). There was a significant association between eHealth literacy and the hours of internet use per day using PC ($F_{4,106.3}$=5.608, P<.001). Post hoc analysis showed that the difference was evident between adults using internet on PC for less than 1 hour per day compared to other groups that used more than 1 hour: 1 to 3 hours, 4 to 7 hours, or 8 to 11 hours. Similarly, there was significant association between eHealth literacy and hours of internet use per day using a portable device ($F_{4,98.0}$=4.610, P=.002). Post hoc analysis showed that the difference in mean eHealth literacy was evident between adults using internet on a portable device for less than 1 hour per day compared to other groups that used more than 1 hour: 1 to 3 hours, 4 to 7 hours, or more than 12 hours. The difference was also shown between those using the Web for 8 to 11 hours and more than 12 hours.

Table 3. Individual content validity index (CVI) and scale CVI scores for the Korean eHEALS (K-eHEALS).

K-eHEALS Items	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Individual CVI
1. I know what health resources are available on the internet	3	3	3	3	2	4	0.83
2. I know where to find helpful health resources on the internet	3	4	3	3	2	4	0.83
3. I know how to find helpful health resources on the internet	4	4	2	4	2	4	0.67
4. I know how to use the internet to answer my questions about health	4	4	4	3	2	4	0.83
5. I know how to use the health information I find on the internet to help me	3	4	4	3	4	4	1.00
6. I have the skills I need to evaluate the health resources I find on the internet	2	4	2	3	3	3	0.67
7. I can tell high quality from low quality health resources on the internet	3	4	2	4	4	4	0.83
8. I feel confident in using information from the internet to make health decisions	4	4	3	4	3	4	1.00
Scale CVI							0.83

Figure 1. Scree Plot of the K-eHEALS.

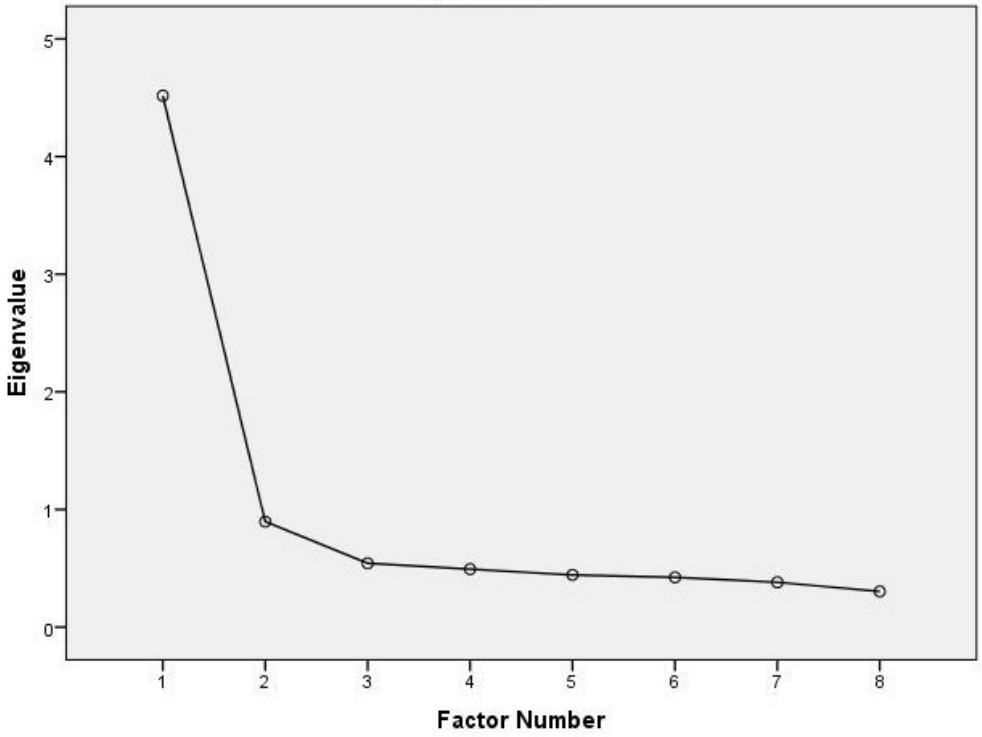


Table 4. Factor loadings and factor score coefficients for the K-eHEALS and single extracted factor (N=500).

K-eHEALS Items	Factor loadings[a]
1. I know what health resources are available on the internet	0.758
2. I know where to find helpful health resources on the internet	0.737
3. I know how to find helpful health resources on the internet	0.762
4. I know how to use the internet to answer my questions about health	0.787
5. I know how to use the health information I find on the internet to help me	0.705
6. I have the skills I need to evaluate the health resources I find on the internet	0.623
7. I can tell high quality from low quality health resources on the internet	0.616
8. I feel confident in using information from the internet to make health decisions	0.664
Sums of squared loadings	4.022

[a] Extraction method: maximum likelihood. One factor extracted, four iterations required.

Table 5. Participant eHealth literacy by hours of internet use (N=500).

Hours of internet use	N	Mean (SD)	*P* value	*F* (*df1*, *df2*)	*P* value
Internet use on personal computer (hours)	500	28.06 (4.81)		5.608 (4,106.30)	<.001
<1	55	25.60 (5.91)	—		
1-3	193	27.68 (5.30)	.032[a]		
4-7	148	28.63 (3.89)	.001[a]		
8-11	83	29.34 (3.41)	<.001[a]		
>12 hours	21	28.90 (5.09)	.050[a]		
Internet use on phone (hours)	500	28.06 (4.81)		4.610 (4,98.03)	.002
<1	77	26.30 (5.90)	—		
1-3	246	28.13 (4.07)	.026[b]		
4-7	118	28.86 (4.43)	.002[b]		
8-11	32	26.81 (6.90)	.986[b]		
>12	27	30.48 (4.52)	.001[b]		

[a]Versus <1 hour internet use on personal computer.

[b]Versus <1 hour internet use on phone.

Reliability

Regarding interitem reliability, the calculated Cronbach alpha coefficient was .88, suggesting that the K-eHEALS was internally consistent. The measure also showed stability over time as evidenced by high test-retest reliability (r=.754, P<.001).

Discussion

Principal Findings

This study is the first attempt to translate a full version of eHEALS into Korean and to test its psychometric aspects. Although additional psychometric testing is necessary to further establish validity of this measure in this population, results of this study are promising and support the validity and reliability of K-eHEALS. Content validity was acceptable (individual CVIs>0.67, scale CVI=0.83), and so was construct validity as supported by unidimensional latent structure of K-eHEALS and significant association between eHealth literacy and hours of internet use per day. In terms of reliability, the items of K-eHEALS were internally consistent (Cronbach alpha=.88) and stable over a 1-month period (r=.754, P<.001). Therefore, the results of this study reveal that the translated K-eHEALS is reliable.

The eHEALS is a measure that has been globally validated in multiple languages, including Japanese [10], Dutch [5], Spanish [11], Chinese [9], German [6], Italian [4], Iranian [8], Hebrew [12], and Turkish [7]. In Korea, several studies have used a translated version of eHEALS, but those studies did not provide psychometric evaluation of the eHEALS in Korean [18,20,21].

Content validity of K-eHEALS reported here was evaluated by six experts and it showed good scale CVI and individual CVI except for two items. However, previous studies on various

language versions of eHEALS did not conduct content validity or report the CVI score (all six). Only one study, an Iranian version of eHEALS [8], measured face validity by four experts, but did not evaluate the CVI score. Evaluation of content validity is proposed for future psychometric studies to enhance construct validity of all instruments [26,31].

Construct validity of the K-eHEALS was evaluated by exploratory factor analysis and hypothesis testing. The K-eHEALS showed a monofactorial unidimensional structure and explained 50.3% of the variance in the measure. This finding supports previous studies that yielded a single factor solution explaining from 52.6% (Spanish version) to 70.5% (Iranian version) of variance in the measure.

Internal consistency of the K-eHEALS was .88, which was comparable with previous findings ranging from .78 (Turkish version) to .93 (Japanese version). Test-retest reliability of the K-eHEALS showed stability (r=.754, P<.001), with the r coefficient within range of previous studies (r=.63 for Japanese version and r=.96 for Iranian version). Therefore, the K-eHEALS is a reliable and valid tool compared to other translated versions of eHEALS.

In South Korea, a higher percentage of Koreans use the internet to search and read health information (66.4%) compared to those who seek information from mass media (40.8%) or a health care provider (11.8%) [32]. Moreover, there is an increasing interest in the use of personal health records (PHRs). For young adults, eHealth literacy is essential for effective use of online resources and their PHR to assist in self-management and promote health conditions. Acknowledging the importance of eHealth literacy, the US Office of Disease Prevention and Health Promotion specified the need for increased heath literacy, access to the internet, and use of health information technologies to promote health of the public in *Healthy People 2020*, a 10-year national objective for improving public health in the United States [33]. The Health Information Technology for Economic and Clinical Health (HITECH) Act [34] also encourages health care institutions to use electronic health records (EHRs), PHRs tethered to EHRs, and related technology for meaningful use of EHRs to improve health [35]. However, there is no prominent national initiative available to promote eHealth literacy in South Korea. More research aimed at understanding of the current level of eHealth literacy in this population is needed to develop more effective eHealth interventions to promote the health of the public. The K-eHEALS, which is a reliable and valid measure, can significantly contribute to these efforts [18].

Limitations

A main limitation of our study is that the participants were recruited from a pool of registrants of an internet survey panel service agency who are likely active online users and this group sample may not be representative of the general young adult population. Moreover, the results cannot be generalized to older adults because this was studied in young adults. Further studies employing diverse age groups are needed to address this issue.

Conclusion

To promote eHealth literacy, researchers and health care providers should first understand the eHealth literacy of the individuals. The psychometric findings from this study suggest that K-eHEALS is a reliable and valid measure of eHealth literacy in young Korean adults who are active online users. We hope K-eHEALS can help Korean researchers who conduct studies in eHealth by providing a reliable and valid measure that can properly gauge participants' eHealth literacy and develop optimal interventions.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MISP) (No: NRF-2016R1C1B2013649).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Korean version of eHEALS (K-eHEALS).

[PDF File (Adobe PDF File), 224KB - jmir_v20i4e138_app1.pdf]

References

1. Norman CD. Social media and health promotion. Glob Health Promot 2012 Dec;19(4):3-6. [doi: 10.1177/1757975912464593] [Medline: 24803437]
2. 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]
3. 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]
4. De Caro W, Corvo E, Marucci AR, Mitello L, Lancia L, Sansoni J. eHealth Literacy Scale: a nursing analysis and Italian validation. Stud Health Technol Inform 2016;225:949. [Medline: 27332425]

5. 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]
6. Soellner R, Huber S, Reder M. The concept of eHealth literacy and its measurement. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
7. Coskun S, Bebis H. Psychometric evaluation of a Turkish version of the e-health literacy scale (e-heals) in adolescent. Gulhane Med J 2015;57(4):378. [doi: 10.5455/gulhane.157832]
8. Bazm S, Mirzaei M, Fallahzadeh H, Bazm R. Validity and reliability of Iranian version of eHealth Literacy Scale. J Commun Health Res 2016;5(2):121-130.
9. Koo M, Norman CD, Hsiao-Mei C. Psychometric Evaluation of a Chinese Version of the eHealth Literacy Scale in school age children. Int Electron J Health Educ 2012;15:29-36.
10. 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(5):361-371. [Medline: 21905612]
11. Paramio PG, Almagro BJ, Hernando GA, Aguaded GJI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish university students]. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
12. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
13. Poushter J. Pew Research Center. 2016 Feb 22. Smartphone ownership and Internet usage continues to climb in emerging economies URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-Internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2018-03-26] [WebCite Cache ID 6yDjbDFVW]
14. Akamai. 2016. State of the Internet Report 2016 URL: https://www.akamai.com/kr/ko/about/news/press/2016-press/akamai-first-quarter-2016-state-of-the-internet-connectivity-report.jsp [accessed 2018-03-26] [WebCite Cache ID 6yDjpxvl7]
15. Kim S, Kim S, Lee S. Health literacy: development of a Korean health literacy assessment tool. Korean J Heal Educ Promot 2005;22(4):215-227.
16. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]
17. Kim S, Oh J, Lee Y. Health literacy: an evolutionary concept analysis. J Korean Acad Soc Nurs Educ 2013;19(4):558-570 [FREE Full text]
18. Lee B, Byoun W, Lim J. The influence of individual's e-Health literacy on doctor-patient communication. J Cybercommunication Acad Soc 2010;27(3):89-125 [FREE Full text]
19. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014 May 09;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
20. Park H, Lee E. Self-reported eHealth literacy among undergraduate nursing students in South Korea: a pilot study. Nurse Educ Today 2015 Feb;35(2):408-413. [doi: 10.1016/j.nedt.2014.10.022] [Medline: 25466791]
21. Kim S, Son Y. Relationships between eHealth literacy and health behaviors in Korean adults. Comput Inform Nurs 2017 Feb;35(2):84-90. [doi: 10.1097/CIN.0000000000000255] [Medline: 27258808]
22. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin J. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes 2014 Dec 09;12:176 [FREE Full text] [doi: 10.1186/s12955-014-0176-2] [Medline: 25492701]
23. World Health Organization. 2016. Process of translation and adaptation of instruments URL: http://www.who.int/substance_abuse/research_tools/translation/en/ [accessed 2018-03-26] [WebCite Cache ID 6sRzhrOYT]
24. Davis LL. Instrument review: getting the most from a panel of experts. Appl Nurs Res 1992 Nov;5(4):194-197. [doi: 10.1016/S0897-1897(05)80008-4]
25. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
26. Polit D, Beck C. Nursing Research: Generating and Assessing Evidence for Nursing Practice. 10th ed. Philadelphia, PA: Wolters Kluwer; 2017.
27. Waltz C, Strickland O, Lenz E. Measurement in Nursing and Health Research. 4th ed. New York: Springer; 2010.
28. Tabachnick B, Fidell L. Using Multivariate Statistics. 6th ed. Harlow, UK: Pearson Education Limited; 2014.
29. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156. [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
30. Nunnally J, Bernstein I. Psychometric Theory. New York: McGraw-Hill; 1994.
31. US Department of Health and Human Services FDA Center for Drug Evaluation and Research, US Department of Health and Human Services FDA Center for Biologics Evaluation and Research, US Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical

product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006 Oct 11;4:79 [FREE Full text] [doi: 10.1186/1477-7525-4-79] [Medline: 17034633]
32. Cho KW, Kam S, Chae YM. Analysis of Internet usage patterns of health consumers for Internet health information assessment criteria. Korean J Heal Educ Promot 2007;24(2):15-28.
33. Office of Disease Prevention and Health Promotion. 2016. Healthy People 2020: health communication and health information technology URL: https://www.healthypeople.gov/2020/topics-objectives/topic/health-communication-and-health-information-technology/objectives [accessed 2018-03-26] [WebCite Cache ID 6sRzerPvw]
34. HealthIT.gov. HITECH programs & advisory committees URL: https://www.healthit.gov/policy-researchers-implementers/hitech-programs-advisory-committees [accessed 2018-03-26] [WebCite Cache ID 6sRzcxin4]
35. HealthIT.gov. Meaningful use definition and objectives URL: https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives [accessed 2018-03-26] [WebCite Cache ID 6sRzF2iC9]

Abbreviations

CVI: content validity index
eHEALS: eHealth Literacy Scale
EHR: electronic health record
K-eHEALS: Korean version of eHealth Literacy Scale
PHR: personal health record
REALM: Rapid Estimate of Adult Literacy in Medicine

Edited by G Eysenbach; submitted 17.08.17; peer-reviewed by E Neter, J Cho; comments to author 29.11.17; revised version received 03.01.18; accepted 24.02.18; published 20.04.18.

Please cite as:
Chung S, Park BK, Nahm ES
The Korean eHealth Literacy Scale (K-eHEALS): Reliability and Validity Testing in Younger Adults Recruited Online
J Med Internet Res 2018;20(4):e138
URL: http://www.jmir.org/2018/4/e138/
doi:10.2196/jmir.8759
PMID:29678800

Original Paper

Optimizing Electronic Consultation Between Primary Care Providers and Psychiatrists: Mixed-Methods Study

Jennifer M Hensel[1,2,3], MSc, MD; Rebecca Yang[1], MPH; Minnie Rai[1], BA; Valerie H Taylor[2,3], MD, PhD

[1]Women's College Hospital Institute for Healthcare Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Department of Psychiatry, Women's College Hospital, Toronto, ON, Canada

[3]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Jennifer M Hensel, MSc, MD
Women's College Hospital Institute for Healthcare Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 323 6230
Email: jennifer.hensel@wchospital.ca

Abstract

Background: The use of electronic consultation (e-consult) between primary care providers (PCPs) and psychiatrists has potential, given the high prevalence of mental health issues in primary care and problematic access to specialist care. Utilization and uptake, however, appears to be lower than would be expected.

Objective: This study aimed to examine actual utilization of e-consult between PCPs and psychiatrists and investigate the perceptions of PCPs about this form of psychiatric advice to inform how to optimize the utility and thereby the uptake of this service.

Methods: In this mixed-methods study, we conducted a chart review of psychiatry e-consults (N=37) over 2 platforms during early implementation in Ontario, Canada, as well as 3 group interviews and 1 individual interview with PCPs (N=10) with variable experience levels and from a range of practice settings. The chart review assessed response times and referral content including the type of request, referral attachments, and consultant responses. Interviews explored the perceptions of the PCPs about the uses and barriers of psychiatry e-consult. Thematic content analysis of interview data identified common themes as well as themes unique to different provider profiles (eg, experienced PCPs vs new PCPs and rural vs urban practice). On the basis of interpretation of the quantitative and qualitative findings, we developed recommendations for the optimization of psychiatry e-consultation services.

Results: During the study period, psychiatry e-consults comprised 3.66% (49/1339) of all e-consults submitted on the studied platforms. Among the e-consults reviewed, different psychiatric diagnoses were represented: 70% of requests (26/37) queried about medication safety or side effects, whereas 59% (22/37) asked about psychiatric symptom management. Moreover, 81% (30/37) of e-consults were answered within 24 hours, and 65% (24/37) were addressed in a single exchange. Themes from the interview data included psychiatry having a complexity that differentiates it from other specialties and may limit the utility of e-consult, other than for psychopharmacology advice. Variability in awareness exists in the way e-consultation could be used in psychiatry, with new PCPs feeling unsure about the appropriateness of a question. In general, new PCPs and PCPs practicing in rural areas were more receptive to psychiatry e-consult. PCPs viewed e-consult as an opportunity to collaborate and desired that it be integrated with other available services. Recommendations include the need for appropriate specialist staffing to address a wide range of requests, adequate education to referrers regarding the use of psychiatry e-consult, and the need to integrate psychiatry e-consult with other geographically relevant services, given the complexity of psychiatric issues.

Conclusions: E-consult is a viable and timely way for PCPs to get much-needed psychiatric advice. For optimizing its utility and uptake, e-consult needs to be integrated into reliable care pathways with adequate referrer and consultant preparation.

(J Med Internet Res 2018;20(4):e124) doi:10.2196/jmir.8943

KEYWORDS

eHealth; psychiatry; primary care; consultation; health services

Introduction

Specialist wait times are a major health care access barrier [1-3]. As a potential solution to specialist wait times, electronic consultations (e-consults) were used to facilitate rapid access to specialist advice via asynchronous written communication [4]. Data across a range of medical specialties support that 75% of e-consults receive a timely response within 3 days, and most take less than 10 min to complete, with up to one third of intended in-person referrals being avoided [5]. Primary care providers (PCPs) are extremely satisfied with e-consult due to its convenience and its contribution to their confidence in managing their patients [5,6].

E-consult between PCPs and psychiatrists has potential as some of the longest specialist wait times have been observed in psychiatry [7]. The studies, which simulated psychiatric referrals from primary care in Canadian and US cities, confirmed that referrals were frequently rejected altogether [8,9]. This would be highly problematic as mental illness is among the leading causes of disability, with economic implications projected to rise by five or six times in the next 30 years [10,11]. As such, rapid access to advice from a psychiatrist would be extremely beneficial for PCPs and their patients. Recently published studies have reported that PCPs feel increased support as a result of having access to psychiatry e-consult [12], and that psychiatry e-consults are predominantly used to address medication-related questions. Utilization of e-consult for psychiatry has been low, however, relative to other specialties [6,13] and the theoretical need that might exist based on mental health visit volumes to primary care [14,15]. The aim of this study was to quantitatively examine actual utilization and the content of e-consults between PCPs and psychiatrists and qualitatively investigate the perceptions of PCPs about this form of psychiatric advice, to inform how to optimize the utility and thereby the uptake of this service as a mechanism for consultation.

Methods

Study Design

We used a convergent parallel mixed-methods study design consisting of a retrospective chart review of completed e-consults and a series of interviews with PCPs. The overall approach to the study was a pragmatic one, seeking practical solutions for programmatic change that triangulated between the quantitative and qualitative results, as well as the existing literature [16]. Quantitative and qualitative analyses were completed separately and integrated, looking for both convergence and divergence of findings [17] to inform recommendations. Research ethics approval was obtained from the Women's College Hospital Research Ethics Board.

Chart Review

Setting and Participants

A sample of e-consults completed between January 1, 2015, and March 31, 2016, was obtained from 2 e-consult platforms (one private and one government-funded) available in Ontario, Canada. These platforms were implemented in target regions and have gradually expanded to be accessible to most of the province with voluntary participation. Platforms offer PCPs access to a range of specialists, with compensation for consultants and referrers. The operator of each platform provided the study team with utilization data for the study period and email addresses for psychiatrist consultants who were requested to provide their e-consultation for review. There are no restrictions for referrers regarding the types of questions that can be posed over these e-consult platforms, although psychiatrists may specify their areas of expertise.

Data Collection and Analysis

Among the 8 active psychiatrists, 5 provided all of their de-identified e-consult reports to the study team by secure fax or encrypted email. All the consultations were reviewed to capture the content using a data collection form, which was modified iteratively during the chart review until a set of nonexclusive categories were established. Data collected included referrer details and patient's age, gender, and psychiatric diagnosis. We classified the content of the referral question, types of referral attachments, and the components of the consultant's response including any attachments provided back to the referrer, and calculated time to consultant response and the number of exchanges. Data were inputted into a Microsoft Access database and extracted into Microsoft Excel for analysis. From the data, we generated variable counts and means, where applicable.

Qualitative Interviews

Setting and Participants

We conducted 4 interviews with a total of 10 PCPs—3 were small group interviews and 1 was an individual interview. We used purposive and maximal variation sampling to recruit participants of different experience levels, such as PCPs from urban and rural areas, and those conducting group and solo practice. Details of the interviewed participants are presented in Table 1.

Data Collection and Analysis

Interviews were conducted by one of the 2 researchers who had no pre-existing relationship with the participants. Participants were asked about their previous experience with e-consult and 3 questions specifically about e-consult for psychiatry:

1. Do you think e-consult is useful for psychiatry?
2. What are the limitations or barriers of using e-consult for psychiatry?
3. Is there anything that would make e-consult more useful for psychiatry?

Table 1. Focus group participants. N/A: not applicable.

Interview	Participants	Practice location	Type of practice	Years in practice, mean (range)	Type of interview	Length of interview (min)
1	5 family physicians	Urban	Group-based[a]	17.8 (10-36)	In-person (at a clinic meeting)	30
2	1 family physician	Urban	Solo practice	3 (N/A)	Teleconference	10
3	2 family physicians	Urban	1 solo practice; 1 group-based	2.5 (2-3)	Teleconference	15
4	1 family physician and 1 nurse practitioner	Rural	Solo practice	6.5 (5-8)	Teleconference	20

[a]Refers to a group-based practice of family physicians and nurse practitioners with access to multidisciplinary support including social work and colocated psychiatry with wait times till on the order of months.

Open-ended questions were used intentionally to generate discussions between the participants. Interviews varied in length from 10 to 30 min, and were audio-recorded and transcribed verbatim. Subsequently, 2 reviewers independently openly coded each transcript, and then convened to compare codes. From the first transcript of a group interview, preliminary themes were developed, and a constant comparative analysis was applied for subsequent transcripts. The coders reviewed for common themes overall, as well as common and contrasting themes from different participant profiles (eg, experienced vs novice and urban vs rural). The reviewers also applied intersubjectivity during coding and thematic development [16], drawing from both the objective data and their subjective personal experience, as one was an e-consult psychiatrist (JH) and the other had been involved in local e-consult implementation (RY). Once a final set of themes was agreed upon, transcripts were independently rereviewed, and any disagreements were discussed until consensus was reached in all cases.

Results

Chart Review

Between January 1, 2015, and March 31, 2016, among all e-consults submitted across all specialty areas, 3.66% (49/1339) e-consults were submitted for psychiatry on the 2 platforms: 4.2% (38/887) on the government-funded platform and 2.4% (11/452) on the private platform. Out of the psychiatric e-consults, 78% (38/49) were obtained from the psychiatrists for review. Of them, one e-consult had been declined and resubmitted to another psychiatrist; thus, it was counted only once. The e-consults represented 37 different patients from 30 unique PCP referrers (ranging from 1 to 5 e-consults each). The 5 psychiatrists answered between 1 and 25 e-consults each. The consultant knew the patient from a previous encounter in only one of the cases. Table 2 displays referral characteristics. Most requests (86%, 32/37) were submitted by family physicians, with 14% (5/37) coming from nurse practitioners. The years of practice of the referrers ranged from 1 to 42 years. Patients were mostly female (65%, 24/37), ranging in age from 15 to 90 years. A diverse range of diagnoses were represented, with more than half (57%, 21/37) of referrals having two or more diagnoses documented. The most common referral question was about medication side effects or safety (70%, 26/37), followed by psychiatric symptom management (59%, 22/37; Table 3). Most questions about medication side effects or safety pertained to antidepressants, followed by antipsychotics, lithium, stimulants, and benzodiazepines. Questions about side effects or safety can be clustered into 3 main groups: (1) medical complications (eg, weight gain, hypothyroidism, sexual dysfunction), (2) safety in special populations (eg, pregnant, elderly, pediatric), and (3) psychiatric complications (eg, manic switch, suicidal ideation). Referrers provided various attachments for the specialist to review (Table 3).

Most e-consults (81%, 30/37) were responded to within 24 hours and nearly two-thirds (65%, 24/37) were answered with a single exchange between referrer and consultant (Table 3). Among the 8 cases where the consultant suggested a referral to a specialist for assessment, 5 were for psychiatry, 2 for cognitive behavioral therapy, and 1 for endocrinology. Consultants frequently suggested or attached provider and patient resources, which often included e-resources (Table 3).

Qualitative Interviews

Previous experience with e-consult among interviewed PCPs varied from minimal to frequent use. Participants described preferred general conditions for the use of e-consult such as “quick,” “simple” questions to a consultant who is reliable in responding promptly. They also discussed the preference for e-consult to be integrated with electronic medical records for ease of use, and a desire for short, unstructured forms. Several themes specific to perceptions regarding e-consult for psychiatry were identified.

Psychiatry is Perceived as More “Nuanced” Than Other Specialties Limiting Electronic Consultation Uses Other Than Psychopharmacology Advice

PCPs felt that e-consults were more applicable in other medical specialties where consultants could advise on objective assessments, compared with psychiatry, where the questions were pertinent to the entire person and often too “nuanced” for e-consult:

> *Most times when I use e-consults in general, it’s usually for an abnormal result, an abnormal lab, an abnormal ultrasound, that I’m debating how urgent it is, what’s the next step, is this a formal consultation. That’s really not the situation in psychiatry, where anything that I would get abnormal, related to*

psychiatry, I can manage...It's really the nuances in the interaction... [Interview 1]

Table 2. Referrer and patient characteristics from e-consults (N=37).

Referrer or patient characteristic	Value
Referrer years of practice, mean (range)	14.9 (1-42)
Referrer discipline, n (%)	
Family physician	32 (86)
Nurse practitioner	5 (14)
Patient gender, n (%)	
Male	13 (35)
Female	24 (65)
Patient age, mean (range)	39.7 (15-90)
Patient diagnosis, n (%)	
Depression	17 (46)
Anxiety	16 (43)
Bipolar disorder	6 (16)
Posttraumatic stress disorder	4 (11)
Substance use disorder	4 (11)
Attention deficit hyperactivity disorder	4 (11)
Sleep disorder	3 (8)
Psychotic disorder	2 (5)
Obsessive compulsive and related disorders	2 (5)
Intellectual and developmental disability	1 (3)
Other	3 (8)
Not stated	2 (5)
Two or more diagnoses	21 (57)

A participant stated, "psych(iatry) is so messy" (Interview 3). Moreover, as PCPs frequently provide mental health care, they are "pretty comfortable dealing with frontline psychiatric issues" (Interview 1). In these cases, they feel "it's rare, in fact, that [they] need to consult with psychiatry at all" (Interview 1). During the first group interview, comprising experienced PCPs, a strong group consensus was observed on this perspective. When the PCPs felt the need to consult, they expressed a preference for shared assessment and management of these complex cases:

> *Often when I need to refer somebody to psych(iatry), it's a pretty complicated issue where I feel like they actually need to see the person and it's not something that I can convey via [e-consult].* [Interview 3]

However, although all participants referred to the nuances of the psychiatric patient, they also identified a particular use of e-consult for medication advice. In this context, a participant stated:

> *But I just think...you need to see them,...Unless it's something like medication optimization and things like that.* [Interview 3]

Examples of using e-consult included "ensuring (medication) safety in pregnancy" (Interview 1) and getting assistance with the application of treatment guidelines in unique situations:

> *There's a lot of guideline support that's out there, which certainly I can rely on guideline support and not use the e-consult. But typically, the guideline support doesn't come with the vast experience that e-consult psychiatrists do.* [Interview 4]

A dedicated psychopharmacology e-consult service was suggested rather than an unspecified service to ensure that the PCP seeks relevant advice from a knowledgeable specialist.

A Lack of Awareness on the Range of Uses for Psychiatric Electronic Consultation Impacts Its Utilization and Perceived Utility

The first interview exposed variable awareness regarding the use of e-consult. A PCP mentioned using an e-consult to help with resource navigation for a patient, whereas another participant found the suggestion given via e-consult to be very compelling, and referred to it several times throughout the interview.

Moreover, another PCP described the shift in perspective after understanding the use of a psychiatry e-consult:

I think when it was first presented to me I was a little skeptical. When I saw some of the use cases it was definitely helpful. [Interview 2]

Table 3. Content and outcome of e-consult (N=37).

Referral domain	Value, n (%)
Content	
Reason for referral[a]	
Medication side effects or safety[b]	26 (70)
Psychiatric symptom management	22 (59)
Role of co-occurring medical illness	9 (24)
Seeking behavioral intervention strategies	1 (3)
Referral attachments	
Typed consult note	9 (24)
Cumulative patient profile	7 (19)
Previous consult reports	6 (16)
Laboratory results	4 (11)
Photos	1 (3)
Outcome	
Time to respond	
Within 24 hours	30 (81)
Between 24 and 72 hours	5 (14)
More than 72 hours	2 (5)
Number of exchanges[c]	
One	24 (65)
Two	12 (32)
More than two	1 (3)
Consultant response	
Requested clarification about question	9 (24)
Suggested referral	8 (22)
Attached provider resources (including e-resources)	14 (38)
Attached patient resources (including e-resources)	8 (22)

[a]Referrals may have contained questions that fit into more than one classification.

[b]Medication side effects/safety includes medical complications such as weight gain, hypothyroidism, and sexual dysfunction (n=10); safety in special populations, such as pregnant, pediatric, and elderly (n=9); psychiatric complications, such as antidepressant-induced mania or suicidal ideation (n=3); and other (n=6).

[c]An exchange includes one message from each of the referrer and consultant related to the referral question. If a reply only expressed gratitude for information provided, it was not counted.

New PCPs expressed more uncertainty regarding the appropriateness of a question to the specialist than more experienced PCPs:

I'm worried that I'm bothering the psychiatrist, or it's not an appropriate question, maybe a little bit more than for other specialties. [Interview 2]

because psych(iatry) can be so complex; I feel like it might be hard for the family doc[tor] to know when to refer or what's appropriate to refer. [Interview 3]

Applicability of Electronic Consultation Would Be Higher if Region-Specific Advice Was Provided and Integrated With Access to Other Services

Specifically, in the case of mental health care, PCPs suggested that e-consults should be provided by psychiatrists in the same jurisdiction or geographic area to ensure common knowledge about "the standard of care or the resources or the types of patients or just the way the system works..." (Interview 2). The rural PCPs also preferred the same, but they had been accustomed to accessing specialists through telemedicine. Many instances wherein e-consult could be integrated with other

services were mentioned, including phone consultation and in-person assessment. Experienced PCPs particularly expressed an interest in phone consultation:

> *I think, actually, an email saying, can we find a time to chat, could be very useful...* [Interview 1]

Similarly, face-to-face consultation was not considered to be well-integrated with e-consult. In one example, a psychiatrist e-consultant offered to see a rural patient in-person, but the patient had to travel for 6 hours to meet the consultant.

Rather than a consult, the PCPs viewed e-consult as a potential tool for shared care of patients by functioning as means of communication between providers to facilitate "collaboration" rather than "consultation" to bring providers "together with the best plan" (Interview 1). The participants discussed a preference to be consulting a psychiatrist who has knowledge of the patient and ideally has met them previously:

> *I think the main thing that is helpful around an e-consult, the way it helps me quickly, if it's somebody that they already know, and I don't have to go through the whole history, et cetera...* [Interview 1]

However, a barrier for this type of care is an appropriate compensation model. As one PCP described:

> *...although we're trying to advance care, and be more streamlined, and be more efficient, and work together in a more collaborative model, and meet patients where they're at, et cetera, we're not getting paid for this approach.* [Interview 1]

New Primary Care Providers and Primary Care Providers From Rural Areas May be More Ready to Adopt Electronic Consultation

Experienced PCPs clearly preferred verbal communication with consultants:

> *A practical issue for me is I think I'm more eloquent when I speak to somebody, than when I try to type.* [Interview 1]

Conversely, new PCPs were more open to using e-communication and were more optimistic about future applications of e-consults:

> *I think this has a good potential to be a light touch, umm...kind of intervention that can have a meaningful impact...* [Interview 2]

PCPs from rural areas were focused on gaining access and improving patient experience, and were already accustomed to using telemedicine for specialist care; thus, they were more acceptable of e-consult:

> *...for most of our patients, you can get a fairly timely response from e-consult, and certainly it provides the support that I need to help manage my patient without having them drive anywhere.* [Interview 4]

In fact, e-consult was seen as a necessary tool for access in the constrained mental health system:

> *I don't have a choice really to use the e-consult because a lot of time it takes just too long to see the psychiatrist.* [Interview 4]

Recommendations

For the most part, the quantitative and qualitative data converged and could be interpreted to yield 5 recommendations for future uptake and expansion of e-consult between PCPs and psychiatrists.

First, PCPs can be educated regarding the feasibility and promptness of e-consult in receiving psychiatric advice. They should also be given examples regarding the types of requests that can be addressed using e-consult.

Second, the primary use for psychiatric e-consult is mostly for seeking pharmacological advice. Unless specified before, this may cover any and all psychiatric disorders and medication classes. The e-consult service needs to be adequately staffed with questions appropriately directed to the most knowledgeable specialists.

Third, PCPs, who are new to practice and who are from rural areas, may be more receptive to psychiatric e-consult, but not exclusively, as long wait times for specialist care are universal.

Fourth, in mental health care, where community services and social determinants of health are important, e-consult psychiatrists ideally must have familiarity or relationships with the communities, and wherever possible with the providers, that their patients consult to.

Fifth, given the potential nuances of the psychiatric patient, e-consult should be integrated with other psychiatric services including telephone consultation and face-to-face assessment either in-person or by telemedicine, as well as with methods of communicating for ongoing collaboration.

Discussion

Summary of Findings

E-consult for psychiatry accounted for 4% of all e-consults on the 2 platforms that we studied, and represented nearly all psychiatric diagnostic categories. Most e-consults were addressed in a single exchange and completed within 24 hours. The quantitative and qualitative data converged to yield recommendations for the implementation, integration, and staffing of the psychiatry e-consult service to optimize utility and uptake.

Comparisons With Previous Work

Our finding that 4% of all e-consults were for psychiatry is strikingly similar to the findings in previous reports [6]. This percentage represents underutilization of e-consults, given the high proportion of individuals with mental illness who seek care in primary practice [14,15]. In our qualitative data, we found that PCPs perceived that e-consult had limited utility in psychiatric issues, which were considered more nuanced, and preference was given to face-to-face assessment of these patients. Similar to this finding, a US study reported a recommendation for an in-person assessment in a quarter of psychiatry e-consults [18], and another US study found a higher

rate of conversion to face-to-face visits for psychiatry e-consults than other medical specialties [13]. These findings support that e-consult is not a complete solution but could and should facilitate stepped care approaches to ensure access to the right care at the right time. We observed that the most common actual and suggested use for psychiatry e-consult was medication advice, consistent with the findings of another descriptive study of psychiatry e-consults [18]. We also found that PCPs who are from rural areas and those who are new to practice are more likely to be receptive to using psychiatry e-consult, reflecting differential adoption of technology and program change based on geography, age, and experience [19]. Ultimately, many of the recommendations that we arrived at are applicable to e-consult more broadly and have been described by others disciplines, such as expansion to be a part of the triage and referral pathway [5,20], and for comanagement [5].

Limitations

A limitation of this study was that data were collected too early during the implementation of both the e-consult platforms, which may not be generalizable to more established services or services that function in different ways. Our sample sizes were not large, but the volumes of e-consults were similar to the volumes that have been reported for other platforms during their early use [6,18]. Furthermore, for generating recommendations regarding future directions for psychiatry e-consult, we feel that the quantitative data were sufficiently informative, and that the themes from the qualitative interviews were saturated and triangulated well with the existing literature [6,13,19] and our personal experience with continued use of e-consult. By maximum variation sampling, we were able to identify pertinent provider differences amidst the common themes. Although we reported on the content of psychiatrist's response, we did not assess the quality of these responses or the actual rate at which recommendations were followed by the PCPs.

Conclusions

For health care technology, a key facilitator of adoption is the proof of utility [19]. We specifically undertook this study due to the issues of patient complexity and PCP skepticism that we encountered with respect to e-consultation for psychiatry. Empirical data that support the benefits of e-consult are emerging, and our study has identified some factors that could be optimized for improving the utility and the uptake of e-consults between PCPs and psychiatrists. Within current resource-constrained environments, new models of integrated care for mental illness are needed to improve quality of care for patients. Although overall access to psychiatry needs to be improved [21], the psychiatrist who specializes in complex mental health problems can typically offer advice with an e-consult to PCPs, and enhance their feeling of support while caring for their patients [12]. Additional study of the impact on patient outcomes and costs is required, along with established characteristics of a good e-consult psychiatrist and features of an effective e-consult. Resource considerations are essential as compensation and practice models are not often well aligned to incentivize novel methods of communication and collaboration.

Acknowledgments

This study was funded by a research grant from the Department of Psychiatry at the University of Toronto. The authors would like to acknowledge the psychiatrists who provided their e-consult records for review and the PCPs who participated in the qualitative interviews.

Authors' Contributions

JH was responsible for funding, study design, data analysis, and drafting of the manuscript. RY participated in recruitment, data collection and analysis, and drafting of the manuscript. MR participated in recruitment and data collection. VT was responsible for funding, study design, and drafting of the manuscript.

Conflicts of Interest

None declared.

References

1. Kreindler SA. Policy strategies to reduce waits for elective care: a synthesis of international evidence. Br Med Bull 2010;95:7-32. [doi: 10.1093/bmb/ldq014] [Medline: 20457662]
2. Berthelot J, Sanmartin C. : Statistics Canada; 2006. Access to health care services in Canada: Waiting times for specialized services (January to December 2005) Ottawa, Canada URL: http://www.statcan.gc.ca/pub/82-575-x/82-575-x2006002-eng.htm [accessed 2017-09-11] [WebCite Cache ID 6tOxtpGQY]
3. Barua B. Canada: The Fraser Institute; 2015. Waiting Your Turn: Wait Times for Health Care in Canada, 2015 Report URL: https://www.fraserinstitute.org/studies/waiting-your-turn-wait-times-for-health-care-in-canada-2015-report [accessed 2017-09-11] [WebCite Cache ID 6tOxzXPuE]
4. Vimalananda VG, Gupte G, Seraj SM, Orlander J, Berlowitz D, Fincke BG, et al. Electronic consultations (e-consults) to improve access to specialty care: a systematic review and narrative synthesis. J Telemed Telecare 2015 Sep;21(6):323-330 [FREE Full text] [doi: 10.1177/1357633X15582108] [Medline: 25995331]

5. Keely E, Liddy C, Afkham A. Utilization, benefits, and impact of an e-consultation service across diverse specialties and primary care providers. Telemed J E Health 2013 Oct;19(10):733-738 [FREE Full text] [doi: 10.1089/tmj.2013.0007] [Medline: 23980939]
6. 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;28(3):394-403 [FREE Full text] [doi: 10.3122/jabfm.2015.03.140255] [Medline: 25957372]
7. Jaakkimainen L, Glazier R, Barnsley J, Salkeld E, Lu H, Tu K. Waiting to see the specialist: patient and provider characteristics of wait times from primary to specialty care. BMC Fam Pract 2014 Jan 25;15:16 [FREE Full text] [doi: 10.1186/1471-2296-15-16] [Medline: 24460619]
8. Goldner EM, Jones W, Fang ML. Access to and waiting time for psychiatrist services in a Canadian urban area: a study in real time. Can J Psychiatry 2011 Aug;56(8):474-480. [doi: 10.1177/070674371105600805] [Medline: 21878158]
9. Malowney M, Keltz S, Fischer D, Boyd JW. Availability of outpatient care from psychiatrists: a simulated-patient study in three U.S. cities. Psychiatr Serv 2015 Jan 01;66(1):94-96. [doi: 10.1176/appi.ps.201400051] [Medline: 25322445]
10. Lim KL, Jacobs P, Ohinmaa A, Schopflocher D, Dewa CS. A new population-based measure of the economic burden of mental illness in Canada. Chronic Dis Can 2008;28(3):92-98 [FREE Full text] [Medline: 18341763]
11. Mental Health Commission of Canada. 2016. Making the case for investing in mental health in Canada URL: https://www.mentalhealthcommission.ca/sites/default/files/2016-06/Investing_in_Mental_Health_FINAL_Version_ENG.pdf; [accessed 2017-09-11] [WebCite Cache ID 6tOyNtTYJ]
12. Golberstein E, Kolvenbach S, Carruthers H, Druss B, Goering P. Effects of electronic psychiatric consultations on primary care provider perceptions of mental health care: survey results from a randomized evaluation. Healthc (Amst) 2017 Feb 02:pii: S2213-0764(16)30173-7. [doi: 10.1016/j.hjdsi.2017.01.002] [Medline: 28162990]
13. North F, Uthke LD, Tulledge-Scheitel SM. Internal e-consultations in an integrated multispecialty practice: a retrospective review of use, content, and outcomes. J Telemed Telecare 2015 Apr;21(3):151-159. [doi: 10.1177/1357633X15572204] [Medline: 25712108]
14. Lesage AD, Goering P, Lin E. Family physicians and the mental health system. Report from the Mental Health Supplement to the Ontario Health Survey. Can Fam Physician 1997 Feb;43:251-256 [FREE Full text] [Medline: 9040912]
15. Centers for Disease Prevention and Control. 2014. Percentage of Mental Health-Related Primary Care Office Visits, by Age Group - National Ambulatory Medical Care Survey, United States, 2010 URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6347a6.htm [accessed 2017-09-11] [WebCite Cache ID 6tOyR9Wpk]
16. Shannon-Baker P. Making paradigms meaningful in mixed methods research. J Mix Methods Res 2016;10(4):319-334 [FREE Full text] [doi: 10.1177/1558689815575861]
17. Creswell JW. United States: University of Nebraska- Lincoln; 2013. Steps in conducting a scholarly mixed methods study URL: http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1047&context=dberspeakers; [accessed 2017-09-09] [WebCite Cache ID 6tOzBapty]
18. Lowenstein M, Bamgbose O, Gleason N, Feldman MD. Psychiatric consultation at your fingertips: descriptive analysis of electronic consultation from primary care to psychiatry. J Med Internet Res 2017 Aug 04;19(8):e279 [FREE Full text] [doi: 10.2196/jmir.7921] [Medline: 28778852]
19. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
20. Patterson V, Humphreys J, Chua R. Email triage of new neurological outpatient referrals from general practice. J Neurol Neurosurg Psychiatry 2004 Apr;75(4):617-620 [FREE Full text] [Medline: 15026509]
21. Clatney L, Macdonald H, Shah SM. Mental health care in the primary care setting: family physicians' perspectives. Can Fam Physician 2008 Jun;54(6):884-889 [FREE Full text] [Medline: 18556498]

Abbreviations

e-consult: electronic consultation
PCP: primary care provider

Edited by CL Parra-Calderón; submitted 11.09.17; peer-reviewed by S Moor, J Apolinário-Hagen; comments to author 01.11.17; revised version received 06.12.17; accepted 27.12.17; published 06.04.18.

Please cite as:
Hensel JM, Yang R, Rai M, Taylor VH
Optimizing Electronic Consultation Between Primary Care Providers and Psychiatrists: Mixed-Methods Study
J Med Internet Res 2018;20(4):e124
URL: http://www.jmir.org/2018/4/e124/
doi:10.2196/jmir.8943
PMID:29625949

Original Paper

Impact of Low Back Pain Clinical Trials Measured by the Altmetric Score: Cross-Sectional Study

Amanda Costa Araujo[1], MSc; Dafne Port Nascimento[1], MSc; Gabrielle Zoldan Gonzalez[1], MSc; Christopher G Maher[2], PhD; Leonardo Oliveira Pena Costa[1], PhD

[1]Masters and Doctoral Programs in Physical Therapy, Universidade Cidade de São Paulo, São Paulo, Brazil

[2]Musculoskeletal Health Sydney, School of Public Health, The University of Sydney, Sydney, Australia

Corresponding Author:
Amanda Costa Araujo, MSc
Masters and Doctoral Programs in Physical Therapy
Universidade Cidade de São Paulo
Tatuapé
448 Cesario Galeno St
São Paulo, 03070-000
Brazil
Phone: 55 11 2476 5749
Email: mandaa_costa@hotmail.com

Abstract

Background: There is interest from authors and publishers in sharing the results of their studies over the Internet in order to increase their readership. In this way, articles tend to be discussed and the impact of these articles tends to be increased. In order to measure this type of impact, a new score (named Altmetric) was created. Altmetric aims to understand the individual impact of each article through the attention attracted online.

Objective: The primary objective of this study was to analyze potential factors related with the publishing journal and the publishing trial that could be associated with Altmetric scores on a random sample of low back pain randomized controlled trials (RCTs). The secondary objective of this study was to describe the characteristics of these trials and their Altmetric scores.

Methods: We searched for all low back pain RCTs indexed on the Physiotherapy Evidence Database (PEDro; www.pedro.org.au) published between 2010 and 2015. A total of 200 articles were randomly selected, and we extracted data related to the publishing trial, the publishing journal, methodological quality of the trials (measured by the 0-10 item PEDro scale), and total and individual scores of Altmetric mentioned and Altmetric reader. The study was a cross-sectional study, and multivariate regression models and descriptive statistics were used.

Results: A total of four variables were associated with Altmetric mentioned score: impact factor (β-coefficient=3.4 points), number of years since publication (β-coefficient=–4.9 points), number of citations divided by years since publication (β-coefficient=5.2 points), and descriptive title (β-coefficient=–29.4 points). Only one independent variable was associated with Altmetric reader score: number of citations divided by years since publication (β-coefficient=10.1 points, 95% CI 7.74-12.46). We also found that the majority of articles were published in English, with a descriptive title, and published in open access journals endorsing the Consolidated Standards of Reporting Trials (CONSORT) statement.

Conclusions: Researchers should preferably select high impact factor journals for submission and use declarative or interrogative titles, as these factors are likely to increase the visibility of their studies in social media.

(J Med Internet Res 2018;20(4):e86) doi:10.2196/jmir.9368

KEYWORDS

Altmetric; social impact; clinical trials; low back pain

Introduction

There is growing interest from both authors and publishers in sharing the results of their studies over the Internet in order to increase their readership [1,2]. Similarly, consumers of research (including clinicians and patients) also share articles that they found interesting, useful, or controversial with their peers over the Internet. One of the ways that articles can be disseminated

on a large scale is by sharing them on social media, such as Facebook, Twitter, Instagram, and others. Another way involves the use of reference manager websites such as Mendeley or Connotea [2]. Both methods offer new approaches to access, read, and discuss research articles; consequently, the dissemination of these articles increases [2-4].

The most traditional way of quantifying the scientific impact of an article or a journal is through the number of citations in peer-reviewed journals [5]. However, indices related to the number of journal citations do not necessarily reflect a greater dissemination of the content of articles to clinicians and patients. Until very recently, the impact of scientific articles on social media and reference managers was not quantified. In order to measure this type of impact a new score (named *Altmetric*) was created [2,6].

Altmetric is a tool developed by a group of British researchers [7]. *Altmetric* aims to understand the individual impact of each article through the attention attracted online (eg, on social media and reference managers) [6]. The *Altmetric* attention score is composed by two independent scoring systems: the *Altmetric mentioned* score and the *Altmetric reader* score [7]. The *Altmetric mentioned* score for an article reflects how widely an article is mentioned in a range of media, including social media (eg, Facebook, Twitter), newspapers, encyclopedias (eg, Wikipedia), online platforms (eg, Faculty1000 and Publication Peer-Reviews), videos on YouTube, sites on questions and answers (eg, Q&A stack overflow), and *policy documents* or PDF documents available over the Internet . Each of these mentions receives different weights to reflect the relative reach of each source, and contributes to the total score. For example, each mention on Facebook counts as 0.25 points while a mention on Twitter counts as 1.0 point. The *Altmetric reader* score can be visualized by clicking the *Altmetric* "donut" symbol (ie, a visual representation of the Altmetric score) and summing the number of readers. This score is hidden in the donut. The second score is the *Altmetric reader* score which measures the impact on online reference managers such as *Mendeley, CiteULike* and *Connotea* [7] . This score has identical weights for all reference managers (ie, 1.0 point for each mention). Readers can easily identify the *Altmetric mentioned* in the websites of most journals by clicking on the *Altmetric* "donut."

Most journal articles about *Altmetric* published to date are only introductory tutorials or editorials [2,4-6,8]. Smith et al [9] published a discussion paper about the importance of *Altmetric* in the field of health sciences that aimed to quantify the social impact of these articles. Patthi et al [10] published a systematic review that retrieved seven articles published between 2010 to 2016 in the dental area that aimed to analyze the correlations between journal citations measured by the Web of Science website and *Altmetric mentioned* scores. This review concluded that journal citations and *Altmetric* scores are positively correlated (between r=0.30 and 0.61) [10]. Finally, Rosenkrantz et al [11] also observed positive correlations between citations and *Altmetric* scores in radiology journals.

There are three articles [12-14] that have measured the correlation among *Altmetric*, Tweets, blogs, and Mendeley (a reference management software). These studies observed a large increase in the number of Tweets and blogs related to scientific journals, and these variables were correlated with the *Altmetric* score . Previously published articles [13,15,16] showed that the number of Tweets can predict citations within the first three days of article publication. These findings indicate that *Altmetric* scores are likely to be correlated with the journal's impact factor [8]. Rinald [17] published a tutorial about the benefits of open access journals with regards to visibility on social media. However, more research is needed to identify potential variables that might be associated with *Altmetric* scores, such as the journal impact factor, number of years since publication, study quality, and open access articles.

To our knowledge, there is no study describing the characteristics of randomized controlled trials (RCTs) and their *Altmetric* scores or predictive factors of *Altmetric* score. In this study, trials of nonpharmacological interventions for low back pain were chosen by the authors because back pain has the largest amount of evidence in the field of musculoskeletal health [18]. Additionally, back pain is extremely prevalent [19-21] and involves high costs [19,20,22]. According to a study that ranks the most disabling diseases in the world [21], low back pain has been one of the highest ranking musculoskeletal diseases since 1990 [21].

Therefore, the primary objective was to analyze potential factors related to the publishing journal (eg, online access) and the publishing trial (eg, trial quality) that could be associated with *Altmetric* scores in a random sample of low back pain RCTs. The secondary objective of this study was to describe the characteristics of these trials and their *Altmetric* scores.

Methods

Study Design

This is a cross-sectional study.

Search Strategies

We selected a random sample of 200 low back pain RCTs from the Physiotherapy Evidence Database (PEDro) [23]. We have chosen PEDro because this is the most comprehensive database of physiotherapy trials [24,25], and also because the PEDro scale has acceptably high reliability and validity [18]. In addition, all trials indexed on PEDro are rated for methodological quality using the 0-to-10-point PEDro scale [24-27]. The items are described below:

1. Eligibility criteria
2. Random allocation
3. Concealed allocation
4. Baseline comparability
5. Blinding of subjects
6. Blinding of therapists
7. Blinding of outcome assessors
8. Completeness of follow up
9. Intention to treat analysis
10. Between-group statistical comparisons
11. Presentation of point measures and measures of variability

Textbox 1. Data sources and details that were extracted.

- Downloaded from PEDro database: full title, authors' names, journal name, language of publication, year of publication, category of intervention according to the PEDro database (ie, exercise, manual therapy, behavioral modification, electrotherapy, acupuncture), and PEDro total and individual scores.
- Extracted from the full-text article: continent where the study was conducted, type of title categorized as declarative (title expressing the results of the trial), interrogative (title introducing the trial in the form of a question), or descriptive (title describing the aim, but does not reveal the main conclusions).
- Extracted from Web of Science: journal's impact factor, number of citations.
- Extracted from journal websites and *Directory of Open Access Journals*: if the paper was published as open access .
- Extracted from journal and Consolidated Standards of Reporting Trials (CONSORT) websites: if the journal endorses the CONSORT statement [28].
- Extracted scores from *Altmetric* website: total and individual scores of *Altmetric mentions* (individual scores [weights] from *Facebook [0.25 points], Twitter [1.0 point], Google+user [1.0 point], News Outlet [8.0 points], Blogs [5.0 points], Sina Weibo [1.0 point], Reddit [0.25 points], Linkedin [1.0 point], Highlight Platform [1.0 point], Pinterest [0.25 points], Wikipedia Page [3.0 points], Faculty1000 [1.0 point], Publication Peer-Reviews [1.0 point], YouTube [0.25 points], Q 0.25 points] &A [stack overflow; 0.25 points]* and *Policy Documents [3.0 points]*. We also collected total and individual scores related to *Altmetric reader* (individual scores from *Mendeley [1.0 point], CiteULike [1.0 point] and Connotea [1.0 point]).*

The total PEDro score is computed by summing *yes* responses to items 2-11. The first item does not count in the final score because this is related to external validity. All trials on PEDro are rated by at least two trained raters and, in cases of disagreement, a final arbitration is performed by a senior rater.

On February 1, 2016 we identified all low back pain trials indexed on PEDro that were published in the time period of 2010-2015 and selected a random sample of 40%. We excluded trial protocols, preliminary analyses of trials, and secondary analyses. The search strategy is described as follows:

> Strategy search: "2010 until 2015" [year of publication] and "low back pain" [part of body] and "pain" [problem] and "clinical trial" [method].

Data Extraction

Several pieces of data were extracted, as detailed in Textbox 1.

Data related to *Altmetric* scores and number of citations divided by years since publication were collected on May 10, 2016 for all articles because these scores are extremely dynamic.

Statistical Analyses

The number of years since publication of the article and the number of citations were determined as of May 10, 2016. The number of citations was normalized by the number of years since publication (number of citations divided by years since publication), as it is expected that older manuscripts are more likely to have a larger number of citations compared to newer ones. Descriptive statistics were used to present most of the data.

Separate multivariate regression models were built to predict (1) *Altmetric mentioned* score and (2) *Altmetric reader* score. The independent variables in both models were: impact factor, paper was published as open access (yes/no), total PEDro score, number of years since publication, normalized citation count, and type of title. These variables were chosen because it seems plausible that they would be associated with *Altmetric* scores. For example, we choose the variable number of years since publication and total PEDro score because they are related to the number of accesses on PEDro [18].

Initially, univariate regression analyses were performed and all variables that reached a P value of <0.20 were retained for inclusion in the multivariate model. Multivariate regression models were then built and the final model contained only variables that reached a statistical significance of $P<0.05$. The results were expressed as R^2 indexes (explained variability of the model) and the individual contribution of each variable was expressed through the presentation of β-coefficients and their respective 95% CIs. We used the Statistical Package for Social Sciences (SPSS) version 19 for the analyses.

Results

Selection of Eligible Articles

A total of 537 clinical trials were retrieved using the search strategy. Sixty-seven articles were excluded because they were related to conditions other than low back pain or were related to studies in progress. From the remaining 470 articles, 200 were randomly selected for analyses (Figure 1).

Descriptive Characteristics of Articles

Table 1 presents the characteristics of the trials. From the 200 articles, 186 had an *Altmetric* score with a mean *mentioned* score of 18.2 (SD 41.3) and a mean *reader* score of 34.9 (SD 41.6). Most of the articles were published in English, had a descriptive title (title describing the aim, but does not reveal the main conclusions) and were published as open access in journals that endorse the CONSORT statement. In addition, the mean impact factor of the journals publishing these trials was 2.1 (SD 2.6) with a mean total PEDro score of 5.8 points (SD 1.6; Table 1).

Predictive Factors

The univariate analysis for *Altmetric mentioned* score showed that being published in an open access journal was not independently associated with *Altmetric mentioned* score (Table 2). The final multivariate model is presented in Table 3. Four

variables were associated with *Altmetric mentioned* score: impact factor (β-coefficient=3.4 points), number of years since publication (β coefficient=–4.9), number of citations divided by years since publication (β-coefficient=5.2 points), and descriptive title (β-coefficient=–29.4 points). This model accounts for 28% of the explained variance. The interpretation of this model is that older articles and those with descriptive titles were associated with a lower *Altmetric mentioned* score, whereas articles from journals with a higher impact factor and with greater citations were associated with a higher *Altmetric mentioned* score.

The univariate analysis for *Altmetric reader* score showed that being published in an open access journal was not associated with *Altmetric reader* score (Table 4). The multivariate analysis showed that one independent variable was associated with *Altmetric reader* score: number of citations divided by years since publication (β-coefficient=10.1 points, 95% CI 7.74-12.46). This single variable accounted for 31% of the explained variance.

Figure 1. Flow diagram.

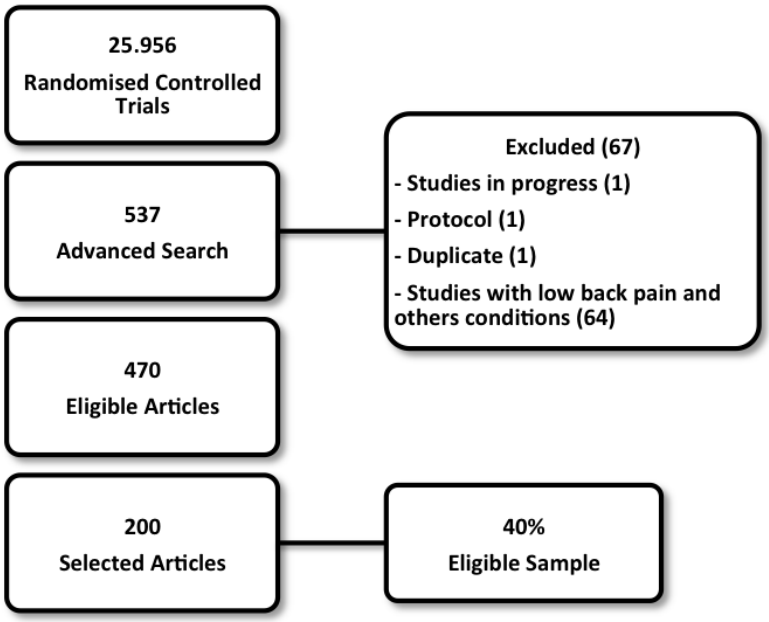


Table 1. Characteristics of the included trails (n=200), *Altmetric mentioned* and *Altmetric reader scores*. Categorical data were expressed as numbers (percentage). Continuous normal data were expressed as means (SD). PEDro: Physiotherapy Evidence Database.

Variables	All articles (n=200)	*Altmetric mentioned*[a] score (n=186)	*Altmetric reader* score (n=186)
Published in English (%)	198 (99.0)	18.4 (41.4)	35.3 (41.7)
Published in other languages (%)	2 (1.0)	0.0 (0.0)	0.0 (0.0)
Continent where the trial was conducted (%)			
Asia	65 (32.5)	3.9 (5.1)	18.8 (25.3)
Europe	70 (35.0)	13.6 (27.7)	40.0 (48.4)
America	46 (23.0)	35.9 (59.8)	40.3 (39.8)
Oceania	11 (5.5)	32.0 (35.9)	63.7 (49.3)
Africa	8 (4.0)	34.2 (91.2)	31.1 (42.7)
Category of interventions (%)			
Stretching, mobilization, manipulation and/or massage	88 (44.0)	28.5 (57.4)	42.8 (51.4)
Strength training	85 (42.5)	17.6 (44.8)	31.6 (36.6)
Behaviour modification	20 (10.0)	19.9 (43.0)	50.4 (64.3)
Neurodevelopmental therapy, neurofacilitation	1 (0.5)	0.0 (0.0)	0.0 (0.0)
Electrotherapies	24 (12.0)	11.3 (37.5)	10.0 (13.0)
Acupuncture	13 (6.5)	11.5 (12.8)	20.0 (15.9)
Skill training	41 (20.5)	26.6 (42.2)	51.5 (62.6)
Education	50 (25.0)	16.1 (31.5)	35.0 (36.0)
Fitness training	26 (13.0)	5.0 (14.1)	29.5 (35.5)
No appropriate value	2 (1.0)	12.5 (7.8)	17.0 (7.1)
Hidrotherapy, balneotherapy	8 (4.0)	3.4 (5.3)	17.1 (21.4)
Orthoses, taping, splinting	1 (0.5)	16.2 (32.3)	52.8 (55.7)
Type of title (%)			
Descriptive	180 (90.0)	15.2 (36.9)	32.4 (40.7)
Interrogative/Declarative	20 (10.0)	43.6 (62.9)	56.0 (44.6)
Open access (%)			
Yes	115 (57.5)	17.9 (42.8)	36.7 (44.0)
No	85 (42.5)	18.6 (39.0)	32.3 (38.0)
Journal endorses CONSORT statement (%)			
Yes	111 (55.5)	24.3 (50.1)	43.3 (47.5)
No	89 (44.5)	9.2 (19.6)	22.5 (26.7)
PEDro items (%)			
Eligibility criteria	166 (83.0)	19.4 (43.8)	37.7 (44.2)
Random allocation	197 (98.5)	18.4 (41.4)	35.2 (41.8)
Concealed allocation	96 (48.0)	25.3 (52.2)	42.3 (44.3)
Baseline comparability	172 (86.0)	19.0 (43.4)	36.6 (43.4)
Blinding of subjects	14 (7.0)	48.6 (76.3)	38.3 (39.2)
Blinding of therapists	3 (1.5)	10.0 (10.5)	34.7 (30.8)
Blinding of outcome assessors	78 (39.0)	27.3 (58.0)	42.9 (51.2)
Completeness of follow up	129 (64.5)	20.8 (47.4)	39.0 (46.4)
Intention to treat analysis	89 (44.5)	26.2 (46.7)	45.1 (42.0)

Variables	All articles (n=200)	*Altmetric mentioned*[a] score (n=186)	*Altmetric reader* score (n=186)
Between-group statistical comparisons	195 (97.5)	18.5 (41.6)	35.5 (41.9)
Presentation of point measures and measures of variability	187 (93.5)	17.9 (40.9)	35.2 (41.5)
Total PEDro score (mean [SD])	5.8 (1.6)		
Journal impact factor (mean [SD])	2.1 (2.6)		
Number of years since publication (mean [SD])	3.4 (1.7)		
Normalized citation count[b] (mean [SD])	2.3 (2.3)		
Score *Altmetric mentioned* (mean [SD])			
Tweeters	13.4 (31.0)		
Facebook pages	3.9 (13.3)		
Google+user	0.2 (1.0)		
News outlet	0.3 (1.6)		
Others	0.0 (0.0)		
Total	18.2 (41.3)		
Score *Altmetric reader* (mean [SD])			
Mendeley	34.4 (41.4)		
CiteULike	0.5 (4.8)		
Connotea	0.0 (0.2)		
Total	34.9 (41.6)		

[a]Others are *Altmetric mentioned* by: Wikipedia page, Blog, Weibo users, Highlight platform, Policy documents, Post-publication peer-reviews, Linkedin, Reddit, Faculty1000, Q&A (stack overflow), Youtube, Pinterest.

[b]Normalized citation count calculated by number of citations divided by years since publication. Years since publication calculated by current year–year of publication.

Table 2. Univariate model to predict characteristics that were associated with Altmetric mentioned score. PEDro: Physiotherapy Evidence Database.

Variable	Constant	β-coefficient	95% CI	*P* value
Journal characteristics				
Open Access	18.63	–0.65	–12.86 to 11.55	.92
Impact Factor	10.32	4.23	1.95 to 6.52	.00
Article characteristics				.01
Total PEDro score (/10)	–20.36	6.54	2.82 to 10.27	.00
Number of years since publication	28.41	–3.01	–6.53 to 0.50	.09
Normalized citation count (number of citations divided by years since publication)	3.61	6.11	3.84 to 8.38	.00
Descriptive title	43.65	–28.47	–47.35 to –9.60	.00

Table 3. Final multivariate model to predict characteristics that were associated with Altmetric mentioned score.

Variable	Constant	β-coefficient	95% CI	*P* value
Journal characteristics	43.02		22.42 to 63.63	.00
Impact Factor		3.42	0.98 to 5.86	.00
Article characteristics				
Number of years since publication		–4.99	–8.50 to –1.47	.00
Normalized citation count (number of citations divided by years since publication)		5.18	2.49 to 7.88	.00
Descriptive title		–29.36	–46.48 to –12.23	.00

Table 4. Univariate model to predict characteristics that were associated with Altmetric reader score. PEDro: Physiotherapy Evidence Database.

Variable	Constant	β-coefficient	95% CI	*P* value
Journal characteristics				
Open Access	32.33	4.34	–7.96 to 16.64	.49
Impact Factor	27.53	4.83	2.25 to 7.40	.00
Article characteristics				
Total PEDro score (/10)	–8.10	7.29	3.56 to 11.03	.00
Number of years since publication	23.59	3.36	–0.19 to 6.90	.06
Normalized citation count (number of citations divided by years since publication)	11.45	10.10	7.74 to 12.46	.00
Descriptive title	56.05	–23.67	–42.87 to –4.47	.02

Discussion

Principal Findings

The primary objective of this study was to analyze potential factors that could be associated with *Altmetric* score. The secondary objective was to describe the characteristics of low back pain RCTs and their *Altmetric* scores. We found that trials with interrogative/declarative titles, those published in higher impact factor journals, those published more recently, and those with a larger number of citations were associated with a higher *Altmetric mentioned* score. We observed that the number of citations was also associated with a higher *Altmetric reader* score. Finally, we found that the *Altmetric reader* score was higher than *Altmetric mentioned* score. Most of the articles were published in English, had descriptive titles, and were published as open access in journals that endorse the CONSORT statement.

There are three previous articles that have measured correlations between the number of citations and *Altmetric* scores in medical journals [10,11,29], and numerous others studying the relationship between tweets (the main *Altmetric* score component) and citations ([15] being the first). The conclusions of these articles are very similar to our study: there is an association between citations and *Altmetric* scores. This information confirms that conventional measures of scientific impact (based on citations) are associated with social impact (based on social media). The difference between our study and these three previous studies is that we used multivariate regression analyses rather than simple correlations. We believe that this approach allowed us to focus on the key independent variables, and the beta coefficients we provide are more interpretable than correlation coefficients. For example, we can predict that for every citation received, 5.2 and 10.1 points will be added to the *Altmetric mentioned* and *reader* scores, respectively. Scientific impact appears to follow the social impact in back pain trials.

We observed an association between *Altmetric mentioned* score and the journal's impact factor. The journal's impact factor is a measure that reflects the number of citations of scientific articles published in the journal divided for the two previous years [5]. We might infer that journals with higher impact factors have more credibility to a wider range of readers, health care providers, and media, which may reflect a large number of posts in social media. Many of these journals may have well-developed media strategies, such as preparation and distribution of releases to the media. This action encourages the promotion of their papers in newspapers, blogs, and social media more rapidly and efficiently than journals that do not do this. These journals also format the online versions of articles so that a reader can easily click an icon to post details about the study on social media, usually by embedding a key figure from the article in the post.

We also observed that trials with declarative/interrogative titles were associated with higher *Altmetric mentioned* scores than those with descriptive titles. Our study is, to our knowledge, the first that has investigated the effect of the format of the articles title on *Altmetric* scores. There is evidence that articles with shorter titles are more likely to be highly cited [30-32]. Jamali et al [33] concluded that articles with interrogative titles are also associated with a larger number of citations and downloads [33]. Therefore, authors should be aware that shorter and interrogative titles should be considered in order to attract a wider audience for their manuscripts. Finally, we observed that papers published more recently also have a higher *Altmetric* score; it seems that recent studies are more likely to be shared. This finding should be investigated further in future studies.

The strength of this study is the use of a representative sample of trials (N=200, or 40% of all trial reports indexed on PEDro and categorized as "low back pain"). A possible limitation of this study is related to external validity, as our dataset contains only articles about low back pain. It would be important to replicate our study in other health disciplines.

Conclusion

Our study brings new insights for authors on how to increase the visibility of their articles. First, researchers should preferably select high impact factor journals for submission and use declarative or interrogative titles, as these factors are likely to increase the visibility of their studies in social media. Furthermore, we suggest new studies that use different research designs (eg, systematic reviews and guidelines) in order to externally validate our findings.

Acknowledgments

We would like to thank Anne M Moseley for her insightful contribution on this project, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for funding ACA's PhD scholarship, and Australia's National Health and Medical Research Council for funding CGM's fellowship.

Conflicts of Interest

None declared.

References

1. Thelwall M, Haustein S, Larivière V, Sugimoto CR. Do altmetrics work? Twitter and ten other social web services. PLoS One 2013;8(5):e64841 [FREE Full text] [doi: 10.1371/journal.pone.0064841] [Medline: 23724101]
2. Melero R. Altmetrics - a complement to conventional metrics. Biochem Med (Zagreb) 2015;25(2):152-160 [FREE Full text] [doi: 10.11613/BM.2015.016] [Medline: 26110028]
3. Liu CL, Xu YQ, Wu H, Chen SS, Guo JJ. Correlation and interaction visualization of altmetric indicators extracted from scholarly social network activities: dimensions and structure. J Med Internet Res 2013 Nov 25;15(11):e259 [FREE Full text] [doi: 10.2196/jmir.2707] [Medline: 24275693]
4. Fabry G, Fischer MR. Beyond the impact factor - what do alternative metrics have to offer? GMS J Med Educ 2017;34(2):Doc27 [FREE Full text] [doi: 10.3205/zma001104] [Medline: 28584875]
5. Dinsmore A, Allen L, Dolby K. Alternative perspectives on impact: the potential of ALMs and altmetrics to inform funders about research impact. PLoS Biol 2014 Nov;12(11):e1002003 [FREE Full text] [doi: 10.1371/journal.pbio.1002003] [Medline: 25423184]
6. Brigham TJ. An introduction to Altmetrics. Med Ref Serv Q 2014;33(4):438-447. [doi: 10.1080/02763869.2014.957093] [Medline: 25316077]
7. Altmetric. Altmetric for Scopus Internet URL: https://help.altmetric.com/support/solutions/articles/6000062186-altmetric-for-scopus [accessed 2018-02-13] [WebCite Cache ID 6xCVHpDqD]
8. Rhee JS. High-impact articles-citations, downloads, and Altmetric score. JAMA Facial Plast Surg 2015;17(5):323-324. [doi: 10.1001/jamafacial.2015.0869] [Medline: 26226501]
9. Smith DR, Watson R. Career development tips for today's nursing academic: bibliometrics, altmetrics and social media. J Adv Nurs 2016 Nov;72(11):2654-2661. [doi: 10.1111/jan.13067] [Medline: 27399604]
10. Patthi B, Prasad M, Gupta R, Singla A, Kumar JK, Dhama K, et al. Altmetrics - a collated adjunct beyond citations for scholarly impact: a systematic review. J Clin Diagn Res 2017 Jun;11(6):ZE16-ZE20 [FREE Full text] [doi: 10.7860/JCDR/2017/26153.10078] [Medline: 28764311]
11. Rosenkrantz AB, Ayoola A, Singh K, Duszak R. Alternative metrics ("Altmetrics") for assessing article impact in popular general radiology journals. Acad Radiol 2017 Jul;24(7):891-897. [doi: 10.1016/j.acra.2016.11.019] [Medline: 28256440]
12. Maggio LA, Meyer HS, Artino AR. Beyond citation rates: a real-time impact analysis of health professions education research using Altmetrics. Acad Med 2017 Oct;92(10):1449-1455. [doi: 10.1097/ACM.0000000000001897] [Medline: 28817430]
13. Pinho-Costa L, Yakubu K, Hoedebecke K, Laranjo L, Reichel CP, Colon-Gonzalez MDC, et al. Healthcare hashtag index development: identifying global impact in social media. J Biomed Inform 2016 Oct;63:390-399 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.010] [Medline: 27645323]
14. Delli K, Livas C, Spijkervet FKL, Vissink A. Measuring the social impact of dental research: an insight into the most influential articles on the Web. Oral Dis 2017 Nov;23(8):1155-1161. [doi: 10.1111/odi.12714] [Medline: 28734099]
15. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
16. Finch T, O'Hanlon N, Dudley SP. Tweeting birds: online mentions predict future citations in ornithology. R Soc Open Sci 2017 Nov 01;4(11):171371. [doi: 10.1098/rsos.171371]
17. Rinaldi A. Spinning the web of open science: social networks for scientists and data sharing, together with open access, promise to change the way research is conducted and communicated. EMBO Rep 2014 Apr;15(4):342-346 [FREE Full text] [doi: 10.1002/embr.201438659] [Medline: 24652852]
18. Yamato TP, Arora M, Stevens ML, Elkins MR, Moseley AM. Quality, language, subdiscipline and promotion were associated with article accesses on Physiotherapy Evidence Database (PEDro). Physiotherapy 2017 Aug 12:1-7. [doi: 10.1016/j.physio.2017.08.003] [Medline: 28927671]
19. Murray CJL, Lopez AD. Measuring the global burden of disease. N Engl J Med 2013 Aug 01;369(5):448-457. [doi: 10.1056/NEJMra1201534] [Medline: 23902484]
20. Beattie PF, Silfies SP, Jordon M. The evolving role of physical therapists in the long-term management of chronic low back pain: longitudinal care using assisted self-management strategies. Braz J Phys Ther 2016;20(6):580-591 [FREE Full text] [doi: 10.1590/bjpt-rbf.2014.0180] [Medline: 28001268]

21. GBD 2015 DiseaseInjury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Dec 08;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
22. Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012 Jun;64(6):2028-2037 [FREE Full text] [doi: 10.1002/art.34347] [Medline: 22231424]
23. Physiotherapy Evidence Database. 2018 Feb 05. URL: http://www.pedro.org.au [accessed 2018-02-23] [WebCite Cache ID 6xSV2NA9S]
24. Moseley AM, Sherrington C, Elkins MR, Herbert RD, Maher CG. Indexing of randomised controlled trials of physiotherapy interventions: a comparison of AMED, CENTRAL, CINAHL, EMBASE, hooked on evidence, PEDro, PsycINFO and PubMed. Physiotherapy 2009 Sep;95(3):151-156. [doi: 10.1016/j.physio.2009.01.006] [Medline: 19635333]
25. Michaleff ZA, Costa LOP, Moseley AM, Maher CG, Elkins MR, Herbert RD, et al. CENTRAL, PEDro, PubMed, and EMBASE are the most comprehensive databases indexing randomized controlled trials of physical therapy interventions. Phys Ther 2011 Feb;91(2):190-197. [doi: 10.2522/ptj.20100116] [Medline: 21148259]
26. Maher CG, Moseley AM, Sherrington C, Elkins MR, Herbert RD. A description of the trials, reviews, and practice guidelines indexed in the PEDro database. Phys Ther 2008 Sep;88(9):1068-1077. [doi: 10.2522/ptj.20080002] [Medline: 18635670]
27. de Morton NA. The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. Aust J Physiother 2009;55(2):129-133 [FREE Full text] [Medline: 19463084]
28. Hopewel S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. [CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration]. Zhong Xi Yi Jie He Xue Bao 2008 Mar;6(3):221-232. [Medline: 18334138]
29. Amath A, Ambacher K, Leddy JJ, Wood TJ, Ramnanan CJ. Comparing alternative and traditional dissemination metrics in medical education. Med Educ 2017 Sep;51(9):935-941. [doi: 10.1111/medu.13359] [Medline: 28719136]
30. Letchford A, Moat HS, Preis T. The advantage of short paper titles. R Soc Open Sci 2015 Aug;2(8):150266 [FREE Full text] [doi: 10.1098/rsos.150266] [Medline: 26361556]
31. Deng B. Papers with shorter titles get more citations. Nature 2015 Aug 26;2(8):150266 [FREE Full text] [doi: 10.1038/nature.2015.18246]
32. Jacques TS, Sebire NJ. The impact of article titles on citation hits: an analysis of general and specialist medical journals. JRSM Short Rep 2010 Jun 30;1(1):2 [FREE Full text] [doi: 10.1258/shorts.2009.100020] [Medline: 21103094]
33. Jamali HR, Nikzad M. Article title type and its relation with the number of downloads and citations. Scientometrics 2011 May 19;88(2):653-661. [doi: 10.1007/s11192-011-0412-z]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
PEDro: Physiotherapy Evidence Database
RCT: randomized controlled trial
SD: standard deviation

Edited by G Eysenbach; submitted 08.11.17; peer-reviewed by C Liu, D Sanchez; comments to author 14.12.17; revised version received 19.12.17; accepted 02.01.18; published 05.04.18.

Please cite as:
Araujo AC, Nascimento DP, Gonzalez GZ, Maher CG, Costa LOP
Impact of Low Back Pain Clinical Trials Measured by the Altmetric Score: Cross-Sectional Study
J Med Internet Res 2018;20(4):e86
URL: http://www.jmir.org/2018/4/e86/
doi:10.2196/jmir.9368
PMID:29622526

Original Paper

Vaccine Images on Twitter: Analysis of What Images are Shared

Tao Chen[1], BE, PhD; Mark Dredze[1], BS, MA, PhD

Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Tao Chen, BE, PhD
Center for Language and Speech Processing
Johns Hopkins University
Hackerman Hall 226
3400 North Charles Street
Baltimore, MD, 21218
United States
Phone: 1 410 516 4237
Email: tchen77@jhu.edu

Abstract

Background: Visual imagery plays a key role in health communication; however, there is little understanding of what aspects of vaccine-related images make them effective communication aids. Twitter, a popular venue for discussions related to vaccination, provides numerous images that are shared with tweets.

Objective: The objectives of this study were to understand how images are used in vaccine-related tweets and provide guidance with respect to the characteristics of vaccine-related images that correlate with the higher likelihood of being retweeted.

Methods: We collected more than one million vaccine image messages from Twitter and characterized various properties of these images using automated image analytics. We fit a logistic regression model to predict whether or not a vaccine image tweet was retweeted, thus identifying characteristics that correlate with a higher likelihood of being shared. For comparison, we built similar models for the sharing of vaccine news on Facebook and for general image tweets.

Results: Most vaccine-related images are duplicates (125,916/237,478; 53.02%) or taken from other sources, not necessarily created by the author of the tweet. Almost half of the images contain embedded text, and many include images of people and syringes. The visual content is highly correlated with a tweet's textual topics. Vaccine image tweets are twice as likely to be shared as nonimage tweets. The sentiment of an image and the objects shown in the image were the predictive factors in determining whether an image was retweeted.

Conclusions: We are the first to study vaccine images on Twitter. Our findings suggest future directions for the study and use of vaccine imagery and may inform communication strategies around vaccination. Furthermore, our study demonstrates an effective study methodology for image analysis.

(J Med Internet Res 2018;20(4):e130) doi:10.2196/jmir.8221

KEYWORDS

vaccine; visual communication; image tweet; Twitter; retweet prediction; social media

Introduction

Background

Visual imagery plays an important role in communication in a range of domains. Their importance in health communication is widely acknowledged [1], and crafting effective health communication literature, materials, and campaigns involves creating visual content to extend the resonance of the message beyond the written word.

This is especially true in public health, where information awareness campaigns are one of the primary interventions available for changing public health behaviors; for example, the area of smoking cessation that has made extensive use of imagery in awareness campaigns. The *Tips from Former Smokers* campaign run by the Centers for Disease Control and Prevention used jarring imagery to encourage smoking cessation [2]. In addition, graphic warning labels on cigarette packages can dissuade people from smoking [3].

Images are especially important in communications related to vaccination, an area of public health with both proponents and

opponents of the advocated behavior. Vaccine supporters and skeptics rely on scientific arguments and logic and emotional resonance to convince people of their perspective. Images are not only effective tools for eliciting emotional reactions but also for conveying statistics and data in support of a position. These can be combined in "infographics"—visuals that blend relevant imagery and statistics.

Both vaccine skeptics and supporters have a large presence on social media in general and Twitter in particular, and use these platforms to advocate for their positions [4,5]. Although several studies have looked at these communities, the topic of vaccine-related images has received little attention. The few studies [6,7] that have considered vaccine-related images have focused on *Pinterest*, an image-based social media platform. Although *Pinterest* is growing increasingly popular, it has less than half the total number of monthly active users as Twitter [8]. In addition, these studies leveraged qualitative analysis (ie, manual annotation), and thus their analysis is limited to a few facets for very small datasets. Moreover, they did not consider how image characteristics correlate with message engagement. See the Prior Work section in the following for a detailed review of the literature in this area.

In this paper, we examine a large corpus of vaccine-related image tweets. We are the first to study vaccine images on Twitter and pose two research questions:

- Research question 1: What are the common characteristics of vaccine-related images shared on Twitter?
- Research question 2: What properties of these images are correlated with higher engagement with other users?

To answer these questions, we pose and analyze various image properties via automated image analytics, which allows us to scale up our analyses to a large image collection. In addition, we fit a logistic regression model to model whether an image tweet was retweeted as a means to identify characteristics of images and tweets that are correlated with engagement. Our goals were to understand how images are used in vaccine-related tweets and to identify characteristics of vaccine images correlated with a higher likelihood of being retweeted.

Prior Work

Images in General Public Health

Prior work on the effects of images in public health has focused on traditional media, such as brochures, advertisements, or magazines. Houts et al's [1] comprehensive literature review summarized the following four functions of images in health communication: increasing patients' *attention*, *comprehension*, *recall*, and *adherence* to health information. Images are most effective when they are closely linked to the written or spoken text and exhibit proper emotional stimuli. Effects can be more pronounced among low-literate people [9-11]. Houts et al [1] suggest the following seven guidelines for the effective use of images: (1) consider using images as visual aids; (2) prefer simple drawings or photographs to complex images; (3) simplify the accompanying text; (4) guide viewers toward an interpretation of the image; (5) be aware of the viewer's culture; (6) take an active role in creating images; and (7) evaluate effects of images systematically.

There have been several analyses of images for specific use cases. Chang [12] studied the effects of images in advertising surrounding four diseases (tinea pedis, periodontal disease, H1N1, and peptic ulcers) and found that images can be effective in risk communication (ie, increasing audience's perceptions of the severity of a disease) and educating the audience about prevention steps. In the same vein, pictorial warnings on cigarette packages have been shown to be a crucial factor for encouraging smokers to use cessation services and quit smoking and sustain their effects longer than text-only warnings [13-15].

Despite widespread acceptance that images are crucial in public health messaging, there is no visual theory that systematically guides the design and use of images, and characteristics of effective images in health communication remain unclear [16]. Furthermore, the community lacks standard tools to analyze health image content and estimate the effects on health behavior [16]. Our work aims to fill these gaps for vaccine images in social media. We identify the key features of vaccine images that correlate with image sharing through fitting a logistic regression model.

Images in Vaccination

To study the effectiveness of messages in a measles, mumps, and rubella (MMR) vaccine promotion (eg, textual information about the dangers of MMR diseases and images of sick children who have MMR diseases), Nyhan et al [17] conducted 2-wave Web-based survey experiments with 1759 parents who had children aged younger than 18 years. Unlike the positive effects of images in most public health studies, their experimental results showed that the image of a sick child had the opposite effect, that is, it increased parents' beliefs in serious vaccine side effects. This result highlights the necessity of carefully testing vaccination messages before their use in a campaign.

The other three works studied vaccine images in social media, including *Facebook* [18] and *Pinterest* (an image-oriented platform) [7,6]. Broniatowski et al [18] analyzed news articles related to vaccine and vaccine-preventable illnesses during the Disneyland measles outbreak and measured the extent of several factors (eg, whether the article included a story and the article contained an image) that influence sharing on Facebook. They found that the presence of images in the article increases the likelihood of sharing, but they did not conduct an analysis of the images themselves.

Guidry et al [7] collected 800 vaccine-related pins (ie, posts from *Pinterest* that consist of an image and a caption) via keyword search and conducted a quantitative analysis to characterize content and user behaviors. In terms of the stance, the authors found that most pins (74.0%) portray vaccinations in a negative light, and antivaccine pins use more narrative than statistical information, whereas provaccine pins are just the opposite. A total of 81.5% of pins have an external reference (ie, contain an external URL), but only 0.3% refer to a government website and 3.7% refer to an official medical website (eg, hospital). They also examined the distribution of 5 Health Belief Model constructs in the dataset. For example, only 16.5% of pins perceived vaccinations to be highly effective, whereas 59.8% of pins showed that barriers to vaccination are high. For user behaviors, the most popular user engagement

with a vaccine pin is repinning (a form of sharing), followed by "like" and comment.

Focusing on the images themselves, Milani [6] manually analyzed more than 1000 pins that clearly exhibited an antivaccination position. In all, 83.9% of the images were photos, 10.2% were charts and infographics, and the remaining 5.9% were drawings. In terms of the subject, syringes dominate (30.8% of pins), followed by children (19.8%), adults (14.6%), and the combination of children and syringes (11.7%). As such, syringes were the main semiotic sign of vaccination. They identified several stereotypes in the photos. Ethnically, all the physicians, paramedics, and 92.1% of babies appeared white. Emotionally, children often show neutral facial expressions when they are alone, smile when with family or in a group, and cry when taking a vaccination (with syringes) or portrayed as sick, whereas most adults are emotionless. Moreover, the author identified a common theme among the most repinned images, that is, rich in emotion but poor in information (ie, no textual information about vaccination). These posts use emotional appeal to try to persuade that vaccination is unnecessary and potentially harmful.

Although insightful, these previous studies have several limitations. Two of them [17,18] did not analyze the content of images, and the other two [7,6] conducted qualitative analysis of *Pinterest* images. As these works require manual labeling, their qualitative analysis is limited to a few facets for small datasets (less than a few thousand images). More importantly, none of them studied vaccine images on Twitter, a much more popular forum than *Pinterest*. In contrast, we rely on automated analytics to analyze millions of vaccine images on Twitter.

Vaccination on Twitter

Despite lack of prior work on vaccine-related images on Twitter, several studies have examined vaccine text tweets. A common thread is the study of attitudes and beliefs surrounding vaccination [4,5,19-21]. These studies typically leverage machine learning algorithms to automatically classify the sentiment (antivaccine, provaccine, or neutral) of vaccine tweets and then analyze the content based on this categorization. For instance, Dunn et al [19] found that users who were exposed to negative opinions about human papillomavirus (HPV) vaccine were more likely to subsequently post negative opinions, and Mitra et al [4] identified cohorts of users who persistently hold pro- or antivaccine attitudes via a longitudinal study.

Another line of research focuses on studying posts' topics, dissemination patterns, community structures, and user behaviors [20,22-24]. For example, Surian et al [22] first characterized tweets about HPV vaccine using topic modeling and then examined the alignment of topics and user community structure. Radzikowski et al [24] collected vaccination tweets in the aftermath of the 2015 measles outbreak and analyzed key terms, the connection among such terms, communication patterns, and geographical patterns. Others have developed novel machine learning algorithms to discover tweets with specific traits, including identifying pseudoscientific claims about Zika vaccine [20] and inferring intentions (received or intend to receive) toward flu vaccine [23].

General Twitter Images

A relatively large number of tweets contain images (17.2%, according to a recent study [25]). Because image tweets are more likely to be shared than tweets without images, users are incentivized to add images to their messages. Unlike other photo-sharing websites, such as Flickr, Twitter images are not limited to photographs but include figures, graphics, screenshots, and other images meant to convey information or advertise the underlying content. For example, many people tweet images of articles that are too long to fit in a tweet or photos meant to accompany a linked website or news story. As a result, numerous studies have examined a range of aspects regarding images on Twitter (eg, [26]). Aspects have included characterizing images [27,28], automatically identifying sentiment of images [29,30], predicting image tweet popularity [31,32], detecting multimedia events [33,34], identifying fake images [35,36], mining trends from images [37,38], and understanding users [25,39].

Methods

Overview

We will characterize the types of images used in a vaccine-related message on Twitter. Our analysis relies on automated analytics for images and tweet content, and we identify factors that influence the likelihood that an image will be shared (retweeted). This section will describe dataset creation, content and image analytics, and our retweet prediction task.

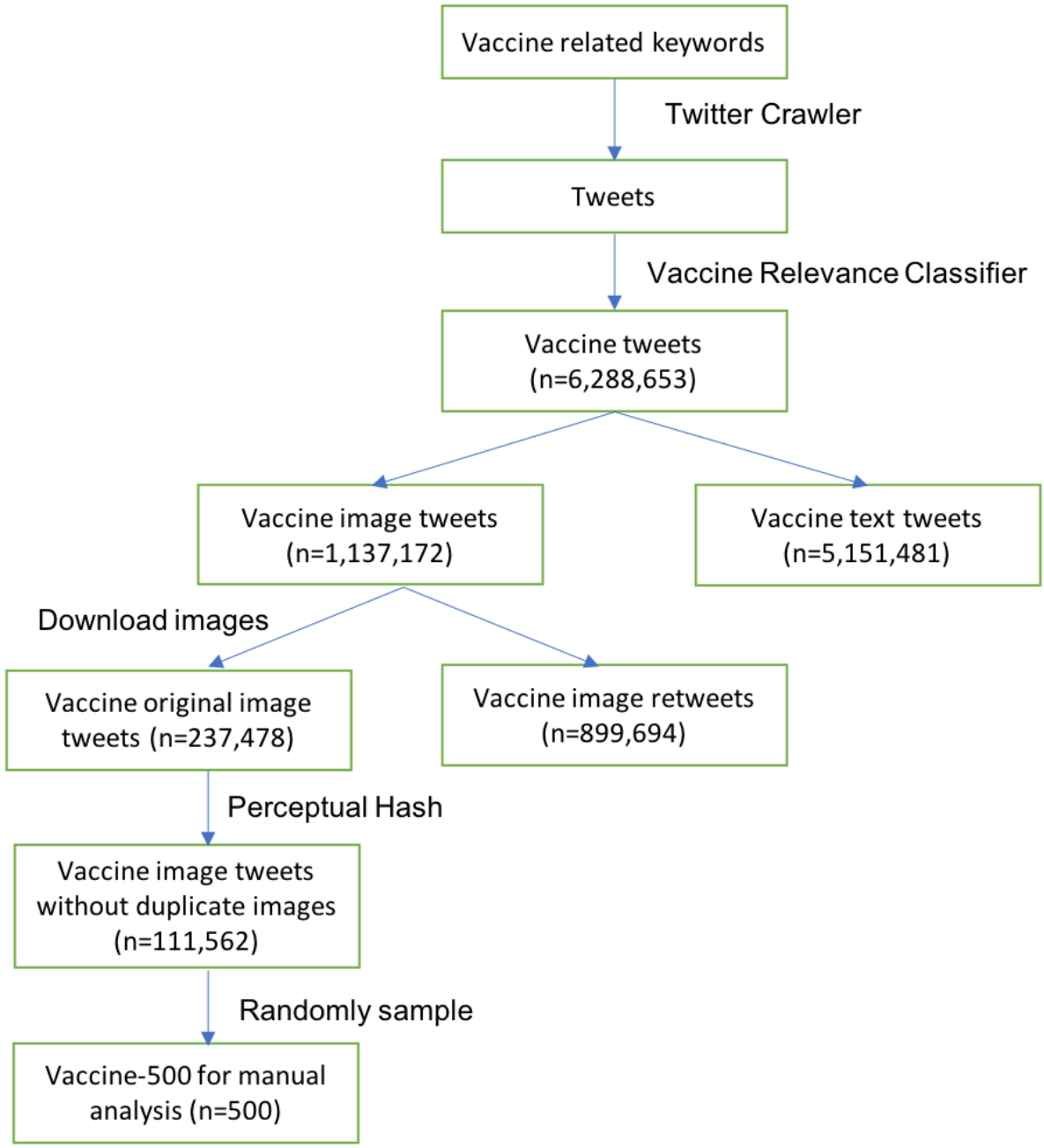
Dataset

Vaccine Twitter Data

We constructed a large corpus of tweets relevant to vaccinations. Figure 1 shows the flowchart of the data collection. We first collected tweets that contained a term from a set of vaccine-related keywords and vaccine-preventable diseases (detailed in Textbox 1) using the Twitter streaming application programming interface (API) from November 11, 2014 to August 8, 2016. This includes all public tweets containing these keywords subject to a rate limit of roughly 1% of the total Twitter volume at that time.

We next applied a statistical classifier [40] to identify tweets relevant to vaccination, as opposed to irrelevant tweets containing vaccine keywords. This support vector machine (SVM) classifier was trained on 1899 manually annotated tweets (released at [41]) and achieved good performance (precision=0.96, recall=0.91, F_1=0.93). After applying this classifier, we obtained a collection of 6,288,653 vaccine-related tweets (Table 1). A total of 18.08% (1,137,172/6,288,653) of the tweets contain an embedded image, a proportion similar to what has been previously reported (17.2%) for Twitter in general [25]. We obtained the number of times each original tweet was retweeted using the Twitter API [42] on December 11, 2016. We then downloaded all the images contained in the original tweets and referenced in the tweet metadata.

Figure 1. The flowchart of our vaccine Twitter data collection.



Textbox 1. The complete vaccine-related keywords and vaccine-preventable diseases used to query Twitter.

vaccine, vaccines, mmr, tdap, flushot, hpv, polio, rotavirus, chickenpox, smallpox, hepatitis, hepa, hepb, dtap, meningitis, shingles, vaccinate, vaccinated, vaccine, vaccines, vacine, vacines, tetanus, diptheria, pertussis, whoopingcough, dtp, dtwp, chickenpox, measles, mumps, rubella, varicella, diphtheria, haemophilus, papillomavirus, meningococcal, pneumococcal, rabies, tuberculosis, typhoid, yellowfever, immunizations, immunization, imunization, immune, imune, cholera, globulin, encephalitis, lyme, zika

Table 1. The demographics of our vaccine tweet dataset.

Medium of tweet	Original tweet, n	Retweet, n	Total, n
Image tweet	237,478	899,694	1,137,172
Text tweet	3,162,184	1,989,297	5,151,481
Total	3,399,662	2,888,991	6,288,653

In addition, we randomly sampled 500 images (without duplicates, see below) from our large vaccine image tweet dataset (hereafter *vaccine–500*). This small dataset was primarily used to conduct manual analysis, which is complementary to the automatic analysis on the large dataset (detailed in the following section).

General Twitter Data

As a baseline of comparison for the retweet prediction task, we obtained a corpus of 200,000 general image tweets sampled from the Twitter 1% public feed between January 1 and December 12, 2016. Retweet counts were obtained on December 29, 2016, and 77.80% (155,600/200,000) of the image tweets had been retweeted at least once. We did not remove duplicates as they were rare in general image tweets.

News Data

The second baseline of comparison for the retweet prediction task will be images included in news articles related to vaccines. Following the procedure described in the study by Broniatowski et al [18], we collected 144,867 vaccine news articles from November 18, 2014 to November 15, 2016, from Google and Bing news using three keywords (vaccine, vaccination, and measles) associated with vaccination. We extracted article contents from HTML using *Goose* [43], including the main text and image of an article and filtered out articles without central images (eg, excluded logos, menu bar graphics, etc). This resulted in a set of 43,664 articles which had an image that was still accessible online. We then used the *Facebook* sharing API [44] to obtain each article's share count, defined as the number of times the link has been shared on Facebook. We found that 51.51% (22,489/43,663) of the articles had been shared at least once. We did not remove duplicate images as they were rare.

Image Processing

Removing Duplicate Images

Many images appeared multiple times in our collection, either as exact copies of the same image file or different files with little change. We identified duplicate images using the Perceptual Hash [45], a popular algorithm for constructing fingerprints of images. Two photos will have similar hashes if they are nearly identical, for example, 2 images with identical content but different aspect ratios. To evaluate the performance of this algorithm, we manually checked 50 duplicate clusters (252 images in total) and found that it achieves 98.4% (248/252) accuracy. In our vaccine dataset, we found that 53.02% (125,916/237,478) of the images were near-duplicates. When we analyze a single tweet for an (duplicate) image that has multiple tweets, we select the tweet that is the most popular, as measured by retweet per follower (retweet count divided by follower count) [46]. This results in a set of 111,562 image tweets, of which 43.00% (47,972/111,562) have been retweeted.

User-Created Images

Users can tweet images that they create themselves, such as a picture taken by the user, but often times they distribute images that they obtain from other sources, such as an infographic or stock image. To differentiate between these images, we leveraged the Google Image Search query-by-image feature. We submitted each image in the *vaccine–500* set as a query and checked whether the image appeared on any other website.

Image Analytics

Extracting Text From Images

Many of the images contain text (see Figure 2 left for an example). Embedded text may be informative for interpreting the image [25]. We extract embedded text using *Tesseract* [47], an open source optical character recognition (OCR) toolkit. This tool is originally designed for printed text and thus works well for Twitter images similar to scanned text—detecting 89.5% of text-style images and 92.9% of screenshots that have text—and generally detects 68.4% of images with embedded text [25]. On the basis of the amount of text, we further categorize images containing embedded text into 3 groups: *primarily images* (no more than 10 words), *a mixture of an image and text* (between 10 and 30 words), and *primarily text* (more than 30 words).

Identifying Faces

Previous work found that many Twitter images contain pictures of people [27]. We identified and characterized human faces in images using *Face++* [48], an online face recognition tool. The tool identified faces and their estimated age, gender, and whether the person was smiling. *Face++* was reported to achieve 99.5% accuracy in face recognition [49], 83.0% accuracy in gender recognition, and have a mean absolute error at 11.0 for age estimation [50].

Figure 2. Three example vaccine images..

Object Recognition

We characterized the content of images using automated object recognition. We used *Clarifai* [51], a deep-learning powered commercial image recognition system. Clarifai provides textual tags to describe the content of an image and has a classification accuracy of 89.3% for the top 5 tags [52]. Consider the middle image in Figure 2 as an example. The top 5 tags from Clarifai are "syringe, injection, medicine, needle, and vaccination."

Topic Analysis

We analyzed the content of tweets using Latent Dirichlet Allocation (LDA) [53], an unsupervised topic model that identifies major themes in a corpus through co-occurring words. This model has been widely used in the topic analysis for traditional text documents [53,54] as well as social media posts [22,55]. A key parameter of LDA is the number of topics, which could be determined on a held-out set.

In our work, we trained two separate LDA models to analyze post's text and visual content. For the post's text, we used the corpus of vaccine-related tweets (both text and image tweets) and trained the LDA model on the text of the post with 50 topics (textual topics, hereafter). For the visual content, we used the vaccine image corpus and trained LDA on the images' tags provided by Clarifai with 40 topics (visual topics, hereafter). The number of textual or visual topics in both models was determined similarly by running an initial experiment with 20% of each dataset as a held-out set. LDA learns a topic distribution for each document and a word distribution for each topic. Following common practice [22,56], we used the topic with the largest probability as the document-level topic for a tweet or an image and manually assigned (by the first author of this paper) a label for each topic for clarity by looking at the top words in the word distribution. We then studied how the textual or visual topics correlate with the medium (image tweet or textual tweet), user engagement (retweeted or nonretweeted), and post's sentiment.

Manual Analysis

In addition to the automatic image feature analysis, we manually examined the images from the *vaccine–500* corpus (conducted by the first author of this paper). This analysis is complementary to the automatic analysis and aims to further characterize the image content and shed light on the functions of images in vaccine messaging.

Retweet Prediction

What makes a vaccine image tweet compelling or engaging? To answer this question, we consider a proxy task: What characteristics of the text and embedded image make it more likely to be shared? We frame this as a retweet prediction task, where we use binary classification to determine whether a tweet was retweeted or not (yes or no). To the best of our knowledge, we are the first to study retweet prediction for vaccine image tweets. Most prior studies concentrate on general text tweets [46,57-59], and only two have specifically considered general image tweets [31,32].

Our study uses a logistic regression, a generalized linear model, to estimate the probability of a binary response (in our case, retweeted or not). Logistic regression was used in previous work for the retweet prediction [57,58,46] and has a key advantage of interpreting the predictive power of features via computing odds ratio. To build a retweet classifier, the previous work exploited various features from tweet's textual content (eg, words, topics, and sentiment), contextual metadata (eg, posting time, the presence of URL), author's profile (eg, the number of followers and friends), and images (eg, color histograms, GIST descriptors, detected visual objects). In our study, we extracted a wide spectrum of features inspired by the previous work and additionally proposed several vaccine-specific features (described below).

To evaluate the model performance, we split our vaccine image tweet dataset (111,562 image tweets after removing duplicate images) into training and test set. To model the timeliness of tweets, we keep the most recent 20% tweets as test set and the rest 80% older tweets as training set. We report precision (true positives divided by true and false positives), recall (true positives divided by true positives and false negatives), and the F_1 score (harmonic mean of precision and recall). Following the same setting, we also build predictive models for the vaccine news and general image tweet dataset by using the same set of features (excluding vaccine-specific features).

Features Used for Predictive Model

Author Popularity Features

One of the most important predictors of whether a tweet will be retweeted is the popularity of the tweet's author [31]. To control for author popularity, we added two features to the model: (1) whether or not the user account is verified (Twitter's account verification establishes account authenticity for public figures or those at risk of impersonation) and (2) the follower count (log normalized). For vaccine news, we measured the author popularity by the Facebook share count of its URL domain (not the link to the story itself).

Metadata Features

We extracted three features from a tweet's metadata: (1) the number of included hashtags, (2) user mentions, and (3) images. We added features indicating the presence of a URL in the tweet, whether the link was to a government (*.gov*) or nongovernment page, and whether the page was no longer available. We similarly classified vaccine news as government or nongovernment page based on the URL.

Text Features

We considered three types of text features.

- Topic: We added features based on the inferred LDA topic model, where each feature is the probability of a topic in the document. Considering that the tweet's text is short, we followed previous work [25] to obtain an enriched textual representation for a tweet by combining (1) the tweet's text, (2) the text from webpages that were linked in a tweet, and (3) the embedded text from the images. We then trained the LDA topic model based on the enriched text. For vaccine news, we combined the OCR text with its article content. We then trained separate topic models for the three datasets.

- Vaccine names: We identified whether the tweet contained one of 25 vaccine names (eg, MMR, HIV) that could suggest the vaccination topic of the tweet. This feature is applied to vaccine image tweets and news but not for general image tweets.
- Sentiment: Previous work found that sentiment is a predictor of retweets [58]. To identify the sentiment of vaccine tweets, we built two SVM classifiers. We first trained a classifier with 447 labeled tweets to identify sentimental tweets from neutral tweets, and then we trained another SVM classifier with 153 labeled tweets to identify sentimental tweets as provaccine or antivaccine. These two classifiers obtain an F_1 score of 0.80 and 0.39 on the test set, respectively. To facilitate further research, we have released the labeled dataset at [41]. We included this sentiment label (neutral, provaccine, or antivaccine) as a feature. Due to the lack of proper tools, we did not extract this feature for general image tweets and vaccine news.

Image Features

We extracted features from the image that capture high-level semantics and low-level vision properties.

- Visual topics: We used the topics learned by fitting LDA to object recognition tags provided by Clarifai. We added topics in the same manner described previously for text topic features. Considering the images in general tweets differ significantly from vaccine images, we restrict these features only to vaccine image tweets and news.
- Face recognition: We extracted four face-related features: (1) number of faces, (2) gender (does the image have a male or female face), (3) age group (does the image have a face with an age falls in 0-2, 3-14, 15-24, 25-64, or over 65 years), and (4) smiling face.
- Image type: We identified images as *pure image*, *primarily image*, *a mixture of image and text*, and *primarily text* as defined previously.
- Visual sentiment: Low-level image features have been shown to be a simple but an effective way to capture emotions, sentiment, or effect of an image [28,60]. We extracted five sets of color-based features to capture visual sentiment: (1) saturation: the mean and standard deviation of saturations; (2) brightness: the mean and standard deviation of brightness; (3) hue: the mean and angular dispersion, with and without saturation weighted; (4) dominant color: the most prevalent of 11 basic colors (ie, black, blue, brown, green, gray, orange, pink, purple, red, white, and yellow) [61]; and (5) affectiveness: one set of three affective scores to measure the pleasure, arousal, and dominance [62].

Results

Analysis of Vaccine Image Tweets

Image Tweet Corpus Analysis

Although Twitter allows users to attach up to 4 images per tweet, most vaccine image tweets (1,089,411/1,137,172; 95.80%) have a single image. These images make the vaccine tweets more likely to be shared (72,906/237,478; 30.70% of image tweets were retweeted) than their text-only counterparts (483,815/3,162,184; 15.30%). While the text-only tweet retweeting rates are similar in the general tweet dataset (13.61%; 10,379/76,273 retweeted), a huge difference exists in image tweets, that is, general image tweets are 2.5 times (155,600/200,000; 77.80%) more likely to be shared than vaccine image tweets. This highlights the need to understand how images are used to discuss vaccines on Twitter and identify strategies that lead to effective images.

Most of the images were not user created but were instead prepared infographics, stock photos, or other imagery. In the *vaccine–500* corpus, 88.4% (442/500) of the images were found on other websites, suggesting that they were not user generated. Users tweeting about vaccines are sharing existing images, which may explain why so many images are reused by other users. In addition, a large number of vaccine images contain text; 39.90% (44,513/111,562) of vaccine images contained at least one embedded textual word. Of which, 42.90% (19,096/44,513) were *primarily images* (no more than 10 words), 30.60% (13,620/44,513) were *a mixture of an image and text* (between 10 and 30 words), and 26.50% (11,797/44,513) were *primarily text* (more than 30 words).

We also found that one-fourth (28,560/111,562; 25.60%) of vaccine images contain faces. The majority of these had a single face, and most of the remaining images (4798/28,560; 16.80%) had 2 faces. The large presence of faces agrees with the objects and concepts discovered by object recognition; four of the five most frequent Clarifai tags (“people, business, adult, woman, and man”) are explicitly about people.

Topic Analysis

Our topic model analysis (without retweets) showed that images were more likely to be associated with some textual topics. Figure 3 lists the manually assigned label for 25 of the topics that were found to be semantically coherent and had the highest deviation of the proportion of tweets containing an image from the mean. The vertical line indicates the mean across the corpus. The topics “poliovirus vaccine in Ethiopia,” “chimps used in hepatitis research,” and “the rate of taking or refusing vaccine” had the highest proportion of image tweets, whereas tweets about “toxic ingredients in vaccines” and “vaccine for soldier or veteran” had the lowest.

Figure 3. The distribution of vaccine image tweets and text tweets in the selected textual topics. On average, 18.08% (1,137,172/6,288,653) of vaccine tweets are image tweets (indicated by the vertical line).

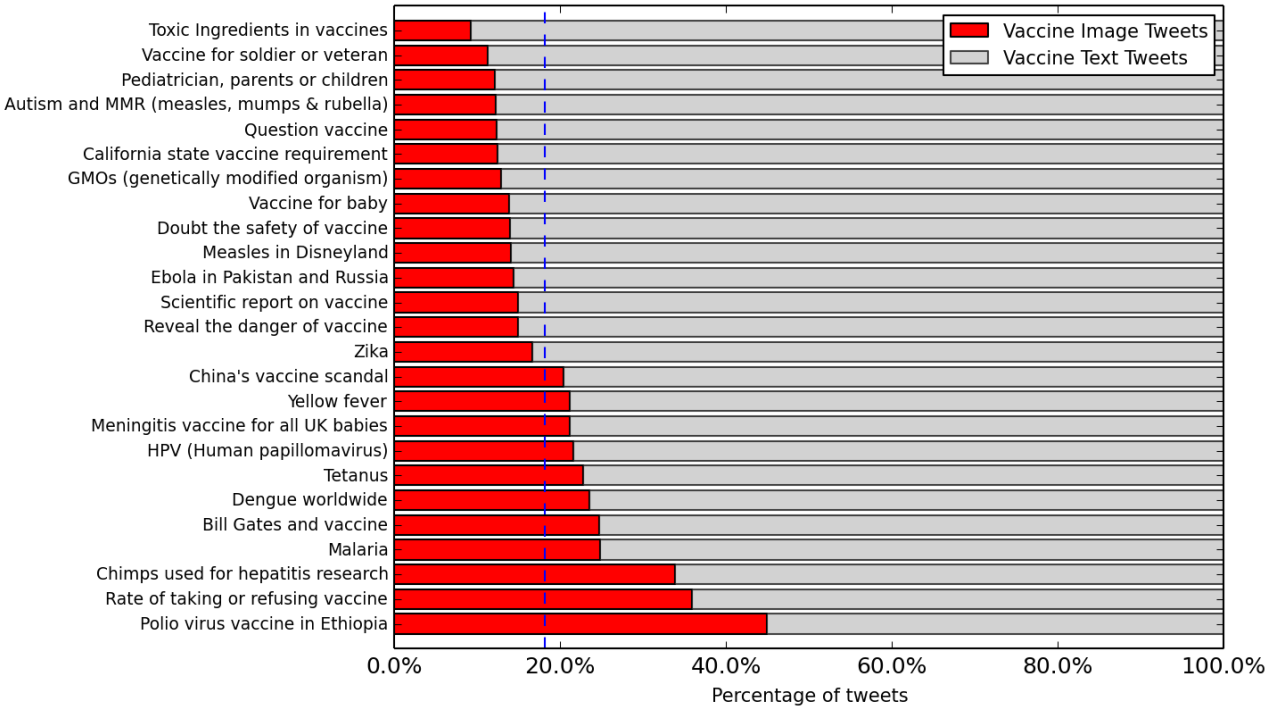


Figure 4. The topic distribution of vaccine image's visual tags.

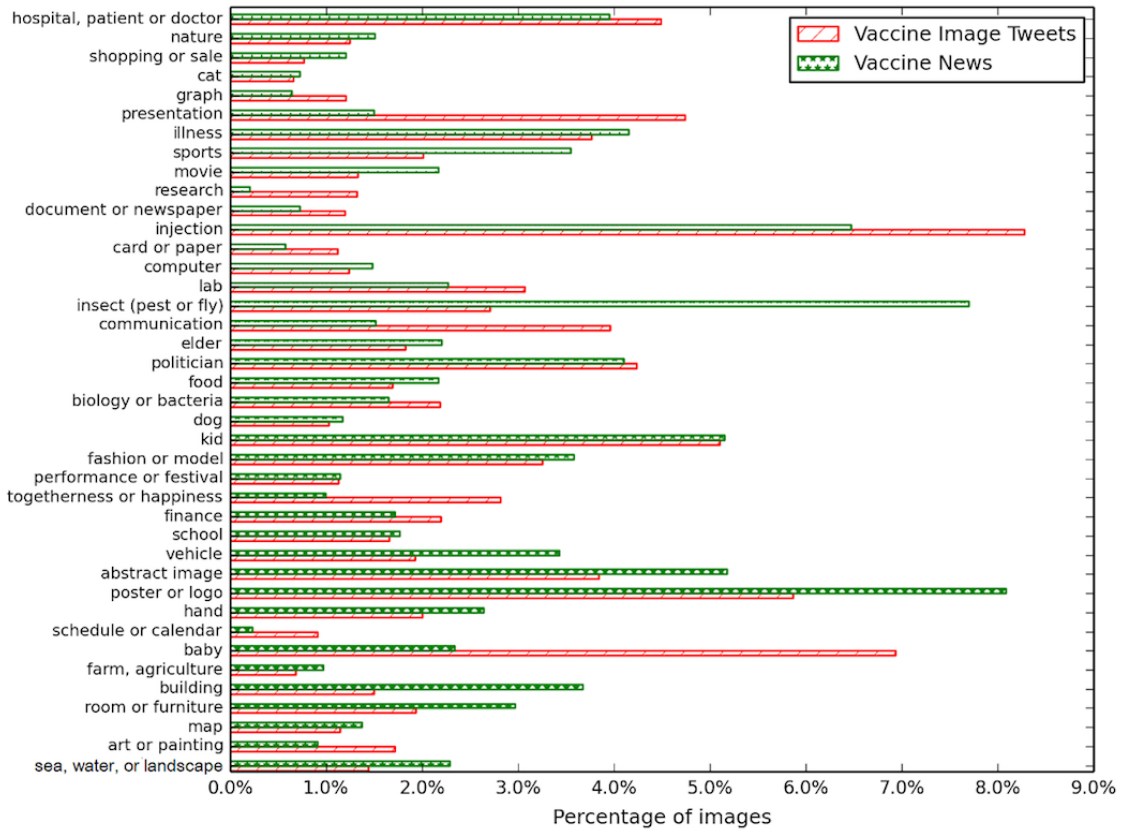


Table 2. The most dominant visual topic and its percentage in each textual topic.

Textual topic	Dominant visual topic	n (%)
Dengue worldwide	Insect (pest or fly)	334 (24.29)
Vaccine price	Poster or logo	231 (17.88)
Poliovirus vaccine in Ethiopia	Baby	3930 (91.43)
HPV[a]	Injection	376 (13.60)
H1N1, swine, or bird flu	Injection	275 (21.43)
Malaria	Insect (pest or fly)	342 (21.30)
Vaccines and fertility	Female	114 (71.5)
Debates of Republican candidates on vaccines and autism	Politician	852 (20.97)
Chimps used for hepatitis research	Nature	1021 (56.07)
Rabies	Dog	1585 (36.91)
Zika	Insect (pest or fly)	534 (23.56)
Meningitis vaccine for all UK babies	Baby	373 (15.84)
Autism and MMR[b]	Communication	410 (16.18)
Scientific report on vaccine	Biology or Bacteria	299 (14.06)
Genetically modified organism	Food	133 (7.53)
Ebola in Africa	Hospital, patient, or doctor	473 (17.81)
Anticancer vaccines	Abstract image (cell or cartoon)	182 (12.17)
HIV	Poster or logo	147 (10.55)
Vaccine for soldier or veteran	Vehicle	37 (6.9)
Promote vaccine	Baby	255 (10.36)

[a]HPV: human papillomavirus.

[b]MMR: measles, mumps, and rubella.

Turning to the topic model analysis of the Clarifai visual tags, the most common topics we found were injection (Figure 2, left) and baby (Figure 2, right). For comparison, we additionally applied the same procedures to the vaccine news images. Tweets and news exhibit differences in using images (see Figure 4). Compared with the visual tags for Twitter, poster or logo and insect are the most used images in news, which takes up 8.08% (3258/43,664) and 7.70% (3362/43,664) of news images, respectively. We then measured the correlation between the text topic and visual topic in the same tweet. Table 2 lists the most common visual topic for each of the 20 textual topics, with the proportion of images in a tweet and the text topic containing the visual topic. We observe strong semantic connections: posts about Dengue show images of insects, posts about polio show babies, and posts about HPV show injections.

Finally, we studied the correlation between the topic and sentiment. We plot the sentimental distribution within each textual and visual topic in Figures 5 and 6, respectively. From Figure 5, we see that many textual topics show skewed sentiment distribution. People primarily express provaccine attributes when they discuss the topic of "comparing vaccinated with unvaccinated people," "HPV," "Ebola in Africa," and "Meningitis vaccine for all UK babies," while expressing antivaccine sentiment in the discussion of "Chimps used for hepatitis research," "an antivaccine film," and "Autism and MMR." On the contrary, visual topics exhibit more balanced sentimental distribution, as most topics are close to the averaged sentimental distribution (Figure 6).

Manual Analysis

We found that most images (without the associated tweet's text) are indicative of the topic of vaccine or health (eg, Figure 2, middle). These images function as a *semiotic sign*, making the tweet differentiable from the huge amount of nonvaccine tweets on Twitter. For the people in the images, we noticed many of them were taking an injection, which further implies the tweet is discussing the vaccination for that specific group of people (Figure 2, right). We then turn to the images that contain text. Such images are an indispensable component of the tweets, either displaying long text as an image to overcome Twitter's text length restriction (up to 140 characters) or render key information of the tweet visually (eg, charts and figures). Aside from these, we identified the third use case of vaccine images that exhibit strong emotion (Figure 2, right), which is used to enforce the sentiment of the post. In summary, we identified three key functions of vaccine images: (1) expressing the topic visually, (2) supplementing information in the tweet, and (3) eliciting emotional responses.

Figure 5. The sentiment distribution of vaccine image tweets' textual topics. On average, 57.58% (71,417/124,029) of sentimental vaccine image tweets are provaccine (indicated by the vertical line). CDC: Centers for Disease Control and Prevention.

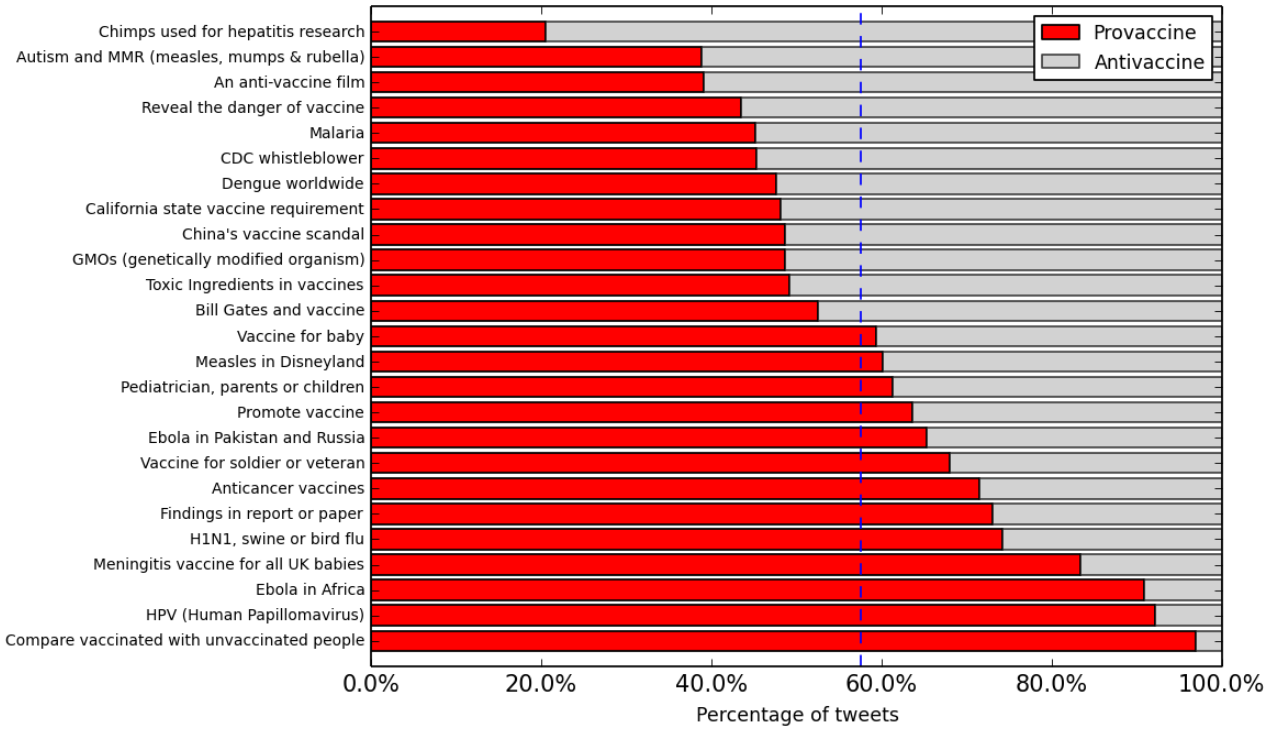


Figure 6. The sentiment distribution of vaccine image tweets' visual topics. On average, 57.58% (71,417/124,029) of sentimental vaccine image tweets are provaccine (indicated by the vertical line).

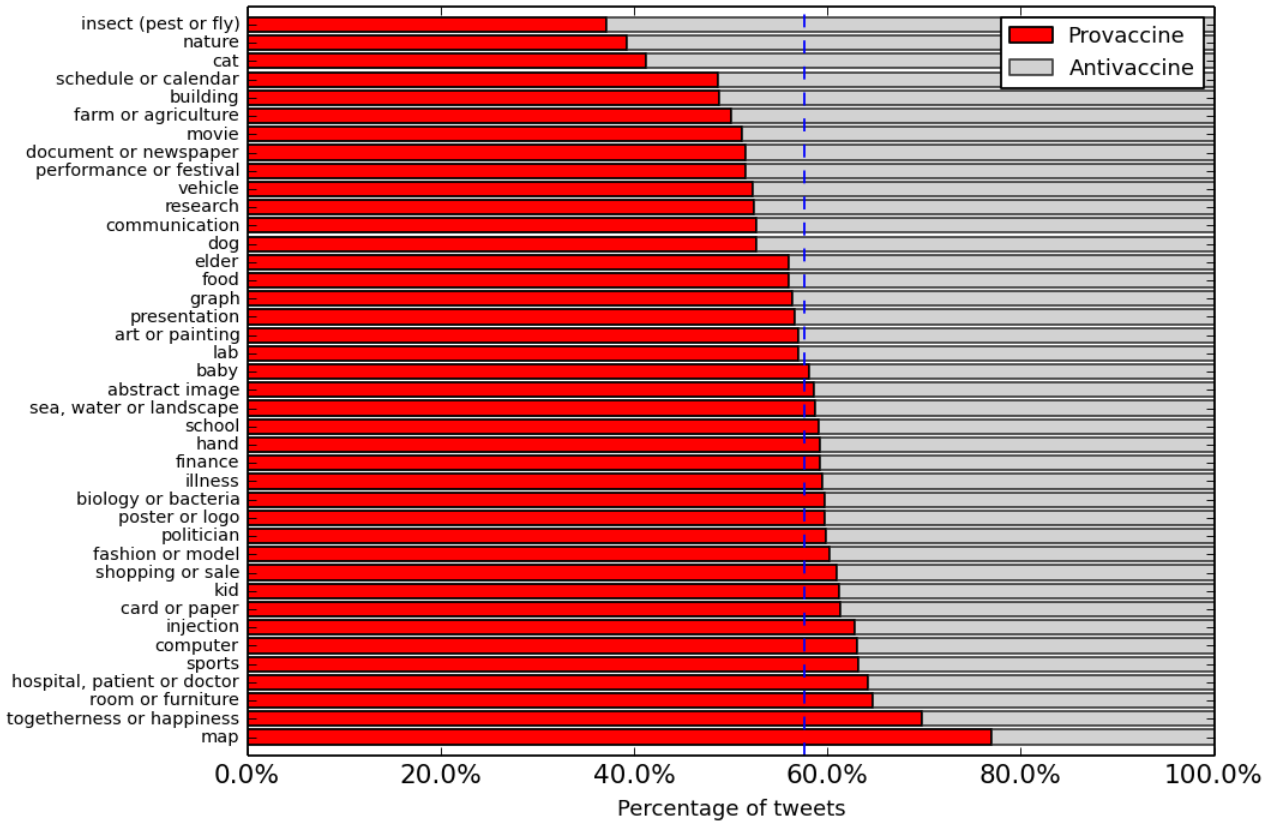


Table 3. Experimental results of sharing prediction. Large score denotes better performance.

Dataset	Precision	Recall	F_1
Vaccine image tweets	0.713	0.632	0.670
Vaccine news	0.750	0.596	0.683
General image tweets	0.856	0.947	0.899

Retweet Prediction

The logistic regression classifier obtained an F_1 score of 0.670 (precision=0.713, recall=0.632) when predicting if a vaccine tweet would be retweeted. For comparison, we conducted similar experiments for vaccine news and general image tweets. Table 3 shows that the predictive model achieves much better performance on general image tweets (F_1=0.899) than vaccine image tweets (F_1=0.670) and news (F_1=0.683). This implies retweet prediction in the focused vaccine domain is more challenging than the general domain.

We computed the odds ratio of the coefficients to determine the strength of a feature in predicting retweets. We compared the features among the three datasets when applicable. For both tweet datasets, the metadata features positively predict retweets, and information about the author (eg, their popularity) matters most (Table 4).

Table 5 shows the statistically significant text features for vaccine tweets and news. Mentioning some vaccines by name increases retweeting (HPV, pertussis, polio, and smallpox), whereas others decrease retweeting (Zika, flu, anthrax, and meningococcal). For news, six vaccine names exhibit strong indication for either sharing (meningococcal, HPV, and typhoid fever) or not sharing (shingles, Zika, and adenovirus). Expressing emotion (either pro- or antivaccine) in the text makes the vaccine tweet more retweetable. Although sharing URLs does not predict retweets, nongovernment and deleted URLs more negatively indicate retweets than government URLs. For news, the articles from government site are significantly more shared than other news. Finally, the textual topics play a key role in retweeting. Tweets that discussed topics such as "Scientific report on vaccine," "Yellow fever," "Trial," and "Vaccine for soldier or veteran" attract more retweeting, whereas other topics such as "Dengue worldwide," "Autism and MMR," "Poliovirus vaccine in Ethiopia," "Ebola in Pakistan and Russia," and "Chimps used for hepatitis research" are negative indicators for retweeting.

We finally turn to the features based on images (Table 5) and focus on the comparisons of vaccine tweets and news. In general, the high-level visual features are very predictive for both tweets and news. Figure 7 shows the proportion of retweeted and nonretweeted vaccine image tweets in each visual topic. Looking at the specific topics, the predictive topics are rather different on Twitter and Facebook, and some topics have contrary predictive power. For example, images of communication and sports increase retweeting but decrease news sharing on Facebook. Images with faces are not predictive in general, but the presence of a smiling face is a positive indicator for retweeting. Embedding some text in the image always increases retweeting regardless of the amount of text but does not have a significant impact news sharing. In general, visual sentiment features are predictive for both tweet and news, but the specific features and their impact often differ in the two datasets. Such feature discrepancies suggest that vaccine visual communication may differ between platforms (Twitter and Facebook) and mediums (tweet and news).

Table 4. The odds ratios of Twitter metadata features that are statistically significant.

Feature	Vaccine image tweets		General image tweets	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Follower count	2.55 (0.92 to 0.96)	<.001	4.78 (1.54 to 1.58)	<.001
Verified user	3.55 (1.19 to 1.34)	<.001	2.98 (0.97 to 1.20)	<.001
Mention count	1.33 (0.26 to 0.31)	<.001	1.39 (0.30 to 0.35)	<.001
Image count	1.15 (0.09 to 0.18)	.01	1.69 (0.50 to 0.56)	<.001
Hashtag count	1.24 (0.20 to 0.23)	<.001	1.05 (0.03 to 0.06)	<.001

[a]OR: odds ratio.

Table 5. The odds ratios of textual and image features that are statistically significant.

Feature	Vaccine image tweets			General image tweets		
	Feature	OR[a] (95% CI)	*P* value	Feature	OR (95% CI)	*P* value
Vaccine names	HPV[b]	1.24 (0.05 to 0.29)	<.001	HPV	1.20 (0.00 to 0.36)	.05
	Zika	0.76 (–0.43 to –0.11)	<.001	Zika	0.63 (–0.59 to –0.34)	<.001
	Meningococcal	0.56 (–0.72 to –0.45)	<.001	Meningococcal	1.28 (0.11 to 0.39)	<.001
	Pertussis	1.34 (0.14 to 0.45)	<.001	Shingles	0.71 (–0.63 to –0.06)	.02
	Polio	1.27 (0.15 to 0.33)	<.001	Adenovirus	0.55 (1.08 to –0.12)	.01
	Smallpox	1.65 (0.24 to 0.77)	<.001	Typhoid fever	1.70 (0.06 to 0.99)	.03
	Flu	0.85 (–0.23 to –0.09)	<.001			
	Anthrax	0.72 (–1.02 to 0.37)	.008			
Textual sentiment	Neutral	0.87 (–0.29 to –0.21)	<.001	N/A[c]		
URL reliability	Government URL	0.92 (-0.23 to 0.07)	<.001	Government URL	1.97 (0.71 to 1.79)	<.001
	Nongovernment URL	0.76 (-0.32 to -0.24)	<.001			
	Deleted URL	0.48 (-0.81 to -0.68)	<.001			
Textual topics	Autism and MMR[d]	0.76 (-0.73 to -0.41)	<.001	Omitted as a separate topic model was trained for news and the comparisons may not be meaningful.		
	Dengue worldwide	0.70 (–0.84 to –0.45)	<.001			
	Poliovirus vaccine in Ethiopia	0.12 (–2.61 to –2.14)	<.001			
	Vaccine for soldier or veteran	1.55 (0.12 to 0.42)	<.001			
	Scientific report on vaccine	1.60 (0.01 to 0.35)	.002			
	Ebola in Pakistan and Russia	0.73 (–0.80 to 0.41)	<.001			
	Chimps used for hepatitis research	0.42 (–1.39 to –0.95)	<.001			
	Yellow fever	1.41 (–0.12 to 0.23)	<.001			
	Trial	1.79 (0.06 to 0.52)	<.001			
Visual topics	Art or painting	1.21 (0.08 to 0.36)	<.001	Sea, water, or landscape	1.21 (0.03 to –0.48)	.03
	Communication	1.28 (0.14 to 0.40)	<.001	Communication	0.57 (–0.78 to –0.25)	<.001
	Sports	1.45 (0.25 to 0.53)	<.001	Sports	0.72 (–0.48 to –0.07)	.01
	School	1.35 (0.18 to 0.47)	<.001	Building	0.76 (0.40 to –0.02)	.03
	Schedule or calendar	1.38 (0.18 to 0.51)	<.001	Fashion, model, or female	1.31 (0.1 to –0.55)	<.001
	Togetherness or happiness	1.28 (0.13 to 0.41)	<.001	Biology or bacteria	1.60 (0.27 to 0.79)	<.001
	Performance or festival	0.76 (–0.47 to –0.16)	.03	Food	1.51 (0.24 to 0.70)	<.001
	Politician	1.38 (0.22 to 0.47)	<.001	Elder	1.29 (0.08 to 0.56)	.01
	Hand	0.77 (–0.40 to –0.08)	.003			
	Research	1.37 (0.20 to 0.48)	<.001			
	Presentation	1.26 (0.14 to 0.38)	<.001			
	Graph	1.46 (0.24 to 0.57)	<.001			
	Cat	0.34 (–1.20 to –0.92)	<.001			
Face	Smile	1.27 (0.16 to 0.32)	<.001			
	Primary image	1.24 (0.17 to 0.27)	<.001			
	Mixture	1.16 (0.09 to 0.20)	<.001			
	Primary text	1.34 (0.22 to 0.37)	<.001			

Feature	Vaccine image tweets			General image tweets		
	Feature	OR[a] (95% CI)	*P* value	Feature	OR (95% CI)	*P* value
Visual sentiment	Blue	0.84 (–0.83 to –0.27)	<.001	Blue	1.87 (0.20 to 1.21)	.006
	White	0.84 (–0.79 to –0.30)	<.001	Red	0.35 (–1.52 to –0.45)	<.001
	SD of Brightness	2.16 (0.49 to 1.05)	<.001	SD of Brightness	0.52 (–1.10 to –0.20)	.005
	Weighted mean of hue	5.65 (1.04 to 2.42)	<.001	Weighted mean of hue	0.08 (–3.70 to –1.41)	<.001
	Arousal	0.73 (–1.33 to –0.40)	<.001	Arousal	1.38 (0.15 to 1.65)	.02
				SD of Saturation	1.79 (0.17 to 1.00)	.006

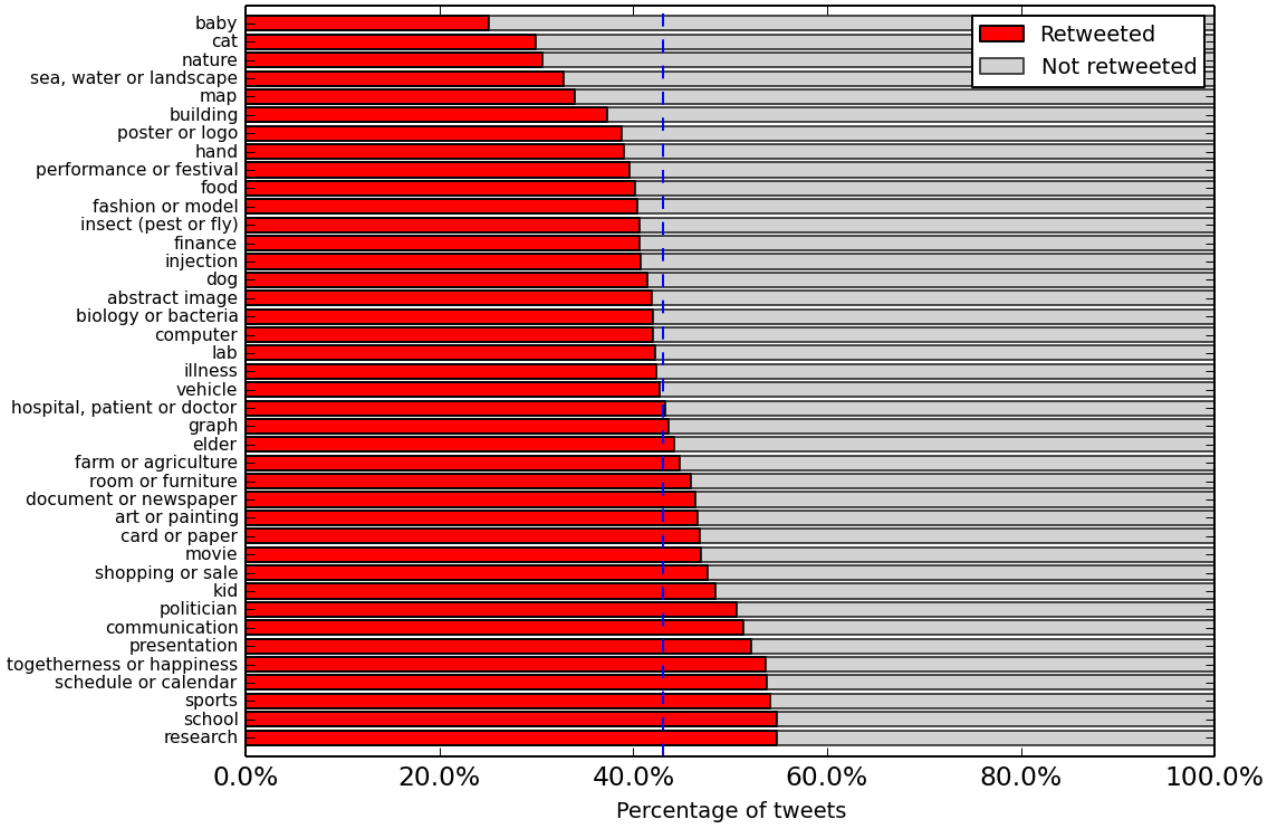
[a]OR: odds ratio.

[b]HPV: human papillomavirus.

[c]N/A: not applicable.

[d]MMR: measles, mumps, and rubella.

Figure 7. After image deduplication, the proportion of retweeted and nonretweeted vaccine image tweets in each visual topic. On average, 43.00% (47,972/111,562) of vaccine image tweets have been retweeted (indicated by the vertical line).



Discussion

Overview

Images have been largely used in vaccine tweets (1,137,172/6,288,653; 18.08% of vaccine tweets contain an image) but were neglected by the prior work. We are the first to study vaccine images on Twitter and particularly answer two research questions: (1) what are their characteristics? and (2) what are the kinds of traits that make them engaging? We summarize the key findings and their implications and highlight some future works in the following paragraphs.

Principal Findings

As with the previous work on general image tweets, an image makes a vaccine tweet more likely to be shared (72,906/237,478; 30.70% of image tweets were retweeted) than their text-only counterparts (483,815/3,162,184; 15.30%). This is a possible motivation for users to attach images to tweets [26,63]. The large number of vaccine images from Twitter that are duplicates (125,916/237,478; 53.02%) or found on other websites (442/500, 88.4%) suggests that most vaccine images are not user created (eg, a photo taken by a user); instead, they are selected from other sources by the user to help promote their vaccination message. Also, the much higher proportion of vaccine image tweets with external URLs (653,874/1,137,172; 57.50%)

compared with general image tweets (22.7% [25]) suggests that images play an important role in vaccine-related messaging. This makes Twitter an especially attractive platform for assessing the effectiveness of vaccine visual communication.

Furthermore, many of the vaccine images contain their own information beyond a visual supporting of the message in the tweet's text. Nearly 40% of images have embedded text, and embedded text is informative to interpret the overall message of the tweet. These images include screenshots, infographics, charts, and figures, which is consistent with vaccine images on *Pinterest* [6,7]. Focusing on the visual content, the two most recurring objects in vaccine images are syringes and people, which is consistent with the findings on *Pinterest* [6]. The visual content is also highly correlated with a tweet's textual topics. As such, the purpose of attaching an image to a tweet is to make it more attractive [27] and convey the topics of the tweet.

Vaccine tweets with images were twice as likely to be shared as nonimage tweets, which follows the trend of general tweets [26]. Our logistic regression identified the author as one of the most important factors for determining whether an image tweet would be shared, the same trend as in general tweets [46,57], and consistent with the findings in the study by Broniatowski et al [18]. Sentiment features, extracted from text and image, are also predictive of sharing behavior. Positive or negative sentiment vaccine image tweets are more likely to be retweeted than neutral tweets, which also matches the behavior of *Pinterest* [6]. One-fourth of vaccine images contain faces. Although previous work found that images with faces have a higher user engagement [64], we found that vaccine image tweets containing a face were equally likely to be retweeted as those without a face (25.5% with vs 25.7% without).

Comparison between retweet prediction for general image tweets and vaccine news shows that retweet prediction for vaccine image tweet is a much more difficult task. We found differing behaviors of features between vaccine tweets and vaccine news. For instance, a smiling face increased sharing for vaccine tweets, but not news, whereas pictures of landscape and nature contribute positively for news sharing but negatively for tweets. This suggests that different communication patterns exist in the two domains (tweet and news), or there could be a difference in how people decide to share content on the two social media platforms (Twitter and Facebook).

Implications

Our research has implications for public health researchers and practitioners.

We demonstrated that images are widely used in Twitter vaccine messages and characterized these images using several types of analyses. This should aid in understanding the information content of millions of vaccine tweets.

In addition, vaccine-related communication strategies could benefit from our analyses. Images boost the reach of a vaccine message. Our retweet predictive model could be used as a tool to assess the effectiveness of designed visual vaccine messages. From that model, we also identify a few key factors that correlate the retweeting of vaccine tweets. Although we have not established a causal relation, these factors could still guide message design.

Finally, our study demonstrates an effective methodology for image analysis studies. We found that Twitter is a productive platform for studying visual communication issues surrounding vaccines. This is an important finding because Twitter makes it relatively easy to collect large quantities of image data via the public Twitter API, compared with the lack of *Pinterest* APIs for creating large, unbiased datasets [65]. Unlike prior work that relied on human analysis of images, we used fully automated analytics to conduct a comprehensive analysis over a large dataset. Such techniques can be applied to analyze vaccine images from other sources and health-related images in general. To enable future studies, we have released the labeled datasets that were used to build our vaccine relevance and sentiment classifiers [41].

Future Directions

We see several avenues of future work. Although our study of Twitter adds to other work that has studied Pinterest, several large social media platforms, in which images are prevalent, have not been examined for vaccine content. These include Instagram and Facebook. Because effective messaging strategies need to be tailored for each platform, evidence of vaccine image effectiveness on these platforms would be welcome.

In addition, we are interested in understanding images beyond the analytics presented in this paper. For example, what images are most effective for different campaigns? How do images tie into existing narratives around vaccination? How are target populations of vaccination campaigns reflected in images? Finally, these questions can be applied broadly to public health awareness campaigns. We plan to extend our methodology to consider these questions in future work.

Acknowledgments

This work was supported in part by the National Institute of General Medical Sciences under grant number 5R01GM114771. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

MD has received consulting fees from Bloomberg LP and holds equity in Good Analytics Inc and Sickweather Inc. These organizations did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. There are no other conflicts to be declared.

References

1. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
2. Ayers JW, Althouse BM, Emery S. Changes in Internet searches associated with the "Tips from Former Smokers" campaign. Am J Prev Med 2015 Jun;48(6):e27-e29 [FREE Full text] [doi: 10.1016/j.amepre.2015.03.015] [Medline: 25998929]
3. Hammond D, Fong GT, McDonald PW, Brown KS, Cameron R. Graphic Canadian cigarette warning labels and adverse outcomes: evidence from Canadian smokers. Am J Public Health 2004 Aug;94(8):1442-1445. [Medline: 15284057]
4. Mitra T, Counts S, Pennebaker J. Understanding anti-vaccination attitudes in social media. Presented at: Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM); May 17–20, 2016; Cologne, Germany p. 269-278.
5. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying Multiple Data Collection Tools to Quantify Human Papillomavirus Vaccine Communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
6. Milani E. 2015. Images for advocacy in communication of risk URL: https://www.slideshare.net/memotak/images-for-advocacy-in-communication-of-risk [accessed 2018-03-09] [WebCite Cache ID 6wptcy3WK]
7. Guidry JP, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. Vaccine 2015 Sep 22;33(39):5051-5056. [doi: 10.1016/j.vaccine.2015.08.064] [Medline: 26319742]
8. Salman A. Omni Core Agency. 2018 Jan 1. Twitter by the Numbers: Stats, Demographics & Fun Facts URL: https://www.omnicoreagency.com/twitter-statistics/ [accessed 2018-03-09] [WebCite Cache ID 6xnNm4o4t]
9. Garcia-Retamero R, Dhami MK. Pictures speak louder than numbers: on communicating medical risks to immigrants with limited non-native language proficiency. Health Expect 2011 Mar;14 Suppl 1:46-57 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00670.x] [Medline: 21323820]
10. Delp C, Jones J. Communicating information to patients: the use of cartoon illustrations to improve comprehension of instructions. Acad Emerg Med 1996 Mar;3(3):264-270 [FREE Full text] [Medline: 8673784]
11. Ngoh LN, Shepherd MD. Design, development, and evaluation of visual aids for communicating prescription drug instructions to nonliterate patients in rural Cameroon. Patient Educ Couns 1997 Jul;31(3):243-261. [Medline: 9277247]
12. Chang C. Seeing is believing: the direct and contingent influence of pictures in health promotion advertising. Health Commun 2013 Feb 28;28(8):822-834. [doi: 10.1080/10410236.2012.726403] [Medline: 23448568]
13. Fong GT, Hammond D, Hitchman SC. The impact of pictures on the effectiveness of tobacco warnings. Bull World Health Organ 2009 Aug;87(8):640-643 [FREE Full text] [Medline: 19705020]
14. Hammond D, Fong GT, McNeill A, Borland R, Cummings KM. Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: findings from the International Tobacco Control (ITC) Four Country Survey. Tob Control 2006 Jun;15 Suppl 3:iii19-iii25 [FREE Full text] [doi: 10.1136/tc.2005.012294] [Medline: 16754942]
15. Hammond D. Health warning messages on tobacco products: a review. Tob Control 2011 Sep;20(5):327-337. [doi: 10.1136/tc.2010.037630] [Medline: 21606180]
16. McWhirter JE, Hoffman-Goetz L. A systematic review of visual image theory, assessment, and use in skin cancer and tanning research. J Health Commun 2014;19(6):738-757. [doi: 10.1080/10810730.2013.837562] [Medline: 24512597]
17. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: a randomized trial. Pediatrics 2014 Mar;133(4):e835-e842 [FREE Full text] [doi: 10.1542/peds.2013-2365] [Medline: 24590751]
18. Broniatowski DA, Hilyard KM, Dredze M. Effective vaccine communication during the Disneyland measles outbreak. Vaccine 2016 Jun 14;34(28):3225-3228 [FREE Full text] [doi: 10.1016/j.vaccine.2016.04.044] [Medline: 27179915]
19. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
20. Dredze M, Broniatowski DA, Hilyard KM. Zika vaccine misconceptions: a social media analysis. Vaccine 2016 Jun 24;34(30):3441-3442 [FREE Full text] [doi: 10.1016/j.vaccine.2016.05.008] [Medline: 27216759]
21. Smith M, Broniatowski DA, Dredze M. Using Twitter to examine social rationales for vaccine refusal. Presented at: International Engineering Systems Symposium (CESUN); June 27-29, 2016; Washington, DC, USA.
22. 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]

23. Huang X, Smith M, Paul M, Ryzhkov D, Quinn S, Broniatowski D, et al. Examining patterns of influenza vaccination in social media. Presented at: AAAI Joint Workshop on Health Intelligence (W3PHIAI); Feb 4–5, 2017; San Francisco, California, USA p. 542-546.
24. 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]
25. Chen T, He X, Kan MY. Context-aware image tweet modelling and recommendation. Presented at: Proceedings of the 2016 ACM on Multimedia Conference (MM); October 15-19, 2016; Amsterdam, The Netherlands p. 1018-1027. [doi: 10.1145/2964284.2964291]
26. Chen T. National University of Singapore. Analyzing Image Tweets in Microblogs URL: http://scholarbank.nus.sg/handle/10635/124157 [accessed 2018-03-09] [WebCite Cache ID 6wpuzT3RY]
27. Chen T, Lu D, Kan MY, Cui P. Understanding and Classifying Image Tweets. Presented at: Proceedings of the 21st ACM International Conference on Multimedia (MM); October 21-25, 2013; Barcelona, Spain p. 718-784. [doi: 10.1145/2502081.2502203]
28. Chen T, SalahEldeen HM, He X, Kan MY, Lu D. VELDA: Relating an Image Tweet's Text and Images. Presented at: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI); January 25-30, 2015; Austin, Texas, USA p. 30-36.
29. Borth D, Ji R, Chen T, Breuel T, Chang SF. Large-scale Visual Sentiment Ontology and De-tectors Using Adjective Noun Pairs. Presented at: Proceedings of the 21st ACM International Conference on Multimedia (MM); October 21-25, 2013; Barcelona, Spain p. 223-232. [doi: 10.1145/2502081.2502282]
30. You Q, Luo J, Jin H, Yang J. Joint Visual-Textual Sentiment Analysis with Deep Neural Networks. Presented at: Proceedings of the 23rd ACM International Conference on Multimedia (MM); October 26-30, 2015; Brisbane, Australia p. 1071-1074. [doi: 10.1145/2733373.2806284]
31. Can EF, Oktay H, Manmatha R. Predicting Retweet Count using Visual Cues. Presented at: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM); October 27-November 01, 2013; San Francisco, California, USA p. 1481-1484. [doi: 10.1145/2505515.2507824]
32. Cappallo S, Mensink T, Snoek CG. Latent Factors of Visual Popularity Prediction. Presented at: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval (ICMR); June 23-26, 2015; Shanghai, China p. 195-202. [doi: 10.1145/2671188.2749405]
33. Bian J, Yang Y, Zhang H, Chua TS. Multimedia summarization for social events in microblog stream. IEEE Trans. Multimedia 2015 Feb;17(2):216-228. [doi: 10.1109/TMM.2014.2384912]
34. Cai H, Yang Y, Li X, Huang Z. What are popular: exploring twitter features for event detection, tracking and visualization. Presented at: Proceedings of the 23rd ACM International Conference on Multimedia (MM); October 26-30, 2015; Brisbane, Australia p. 89-98. [doi: 10.1145/2733373.2806236]
35. Boididou C, Papadopoulos S, Kompatsiaris Y, Schifferes S, Newman N. Challenges of Computational Verification in Social Multimedia. Presented at: Proceedings of the 23rd International Conference on World Wide Web (WWW); April 7-11, 2014; Seoul, Republic of Korea p. 743-748. [doi: 10.1145/2567948.2579323]
36. Gupta A, Lamba H, Kumaraguru P, Joshi A. Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy. Presented at: Proceedings of 2nd International Workshop on Privacy and Security in Online Social Media (PSOSM); May 13-17, 2013; Rio, Brazil p. 729-736. [doi: 10.1145/2487788.2488033]
37. Yanai K, Kawano Y. Twitter Food Photo Mining and Analysis for One Hundred Kinds of Foods. Presented at: Proceedings of the 15th Pacific-Rim Conference on Advances in Multimedia Information Processing (PCM); December 1-4, 2014; Kuching, Malaysia p. 22-32. [doi: 10.1007/978-3-319-13168-9_3]
38. Zhao N, Hong R, Wang M, Hu X, Chua TS. Searching for Recent Celebrity Images in Microblog Platform. Presented at: Proceedings of the 22nd ACM International Conference on Multimedia (MM); November 3-7, 2014; Orlando, FL, USA p. 841-844. [doi: 10.1145/2647868.2655011]
39. Lin H, Jia J, Guo Q, Xue Y, Li Q, Huang J, et al. User-level Psychological Stress Detection from Social Media Using Deep Neural Network. Presented at: Proceedings of the 22nd ACM International Conference on Multimedia (MM); November 3-7, 2014; Orlando, FL, USA p. 507-516. [doi: 10.1145/2647868.2654945]
40. Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding vaccine refusal: why we need social media now. Am J Prev Med 2016 Apr;50(4):550-552 [FREE Full text] [doi: 10.1016/j.amepre.2015.10.002] [Medline: 26655067]
41. Social Media for Public Health. Vaccination Sentiment and Relevance Tweets URL: http://www.socialmediaforpublichealth.org/resources [accessed 2018-03-09] [WebCite Cache ID 6xnV9oKWr]
42. Twitter Statuses Lookup API. URL: https://developer.twitter.com/en/docs/tweets/post-and-engage/api-reference/get-statuses-lookup [accessed 2018-03-09] [WebCite Cache ID 6xnVtUR3k]
43. Grangier X. Python-Goose - Article Extractor URL: https://github.com/grangier/python-goose [accessed 2018-03-09] [WebCite Cache ID 6r2lrNhya]
44. Facebook Sharing API. URL: https://developers.facebook.com/docs/graph-api/reference/v2.9/url [accessed 2018-03-09] [WebCite Cache ID 6xnWE5BKM]

45. Zauner C. Upper Austria University of Applied Sciences. 2010. Implementation and Benchmarking of Perceptual Image Hash Functions URL: http://www.phash.org/docs/pubs/thesis_zauner.pdf [accessed 2018-03-09] [WebCite Cache ID 6wpvggH0Z]
46. Wang A, Chen T, Kan MY. Re-tweeting from a linguistic perspective. Presented at: Proceedings of the Second Workshop on Language in Social Media (LSM); June 7, 2012; Montreal, Canada p. 46-55.
47. Tesseract Open Source OCR Engine URL: https://github.com/tesseract-ocr/tesseract [accessed 2018-03-09] [WebCite Cache ID 6r1VW6WRh]
48. Face++ URL: https://www.faceplusplus.com/ [accessed 2018-03-09] [WebCite Cache ID 6wpvaLp9i]
49. Zhou E, Cao Z, Yin Q. Naive-deep face recognition: Touching the limit of LFW benchmark or not? URL: https://arxiv.org/abs/1501.04690 [accessed 2018-03-09] [WebCite Cache ID 6xnWiITr0]
50. Sight Hound. Age, gender and emotion benchmarks URL: https://www.sighthound.com/technology/age-gender-emotion/benchmarks [accessed 2018-03-09] [WebCite Cache ID 6uBuVQiib]
51. Clarifai. URL: https://www.clarifai.com [accessed 2018-03-09] [WebCite Cache ID 6r1VsJUTk]
52. Clarifai. Clarifai is at the forefront of the deep learning revolution URL: https://www.clarifai.com/technology [accessed 2018-03-09] [WebCite Cache ID 6uAKZJPD5]
53. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Mar 01;3:993-1022 [FREE Full text]
54. Griffiths TL, Steyvers M. Finding scientific topics. Proc Natl Acad Sci U S A 2004 Apr 06;101 Suppl 1:5228-5235 [FREE Full text] [doi: 10.1073/pnas.0307752101] [Medline: 14872004]
55. Miller M, Banerjee T, Muppalla R, Romine W, Sheth A. What are people tweeting about zika? An exploratory study concerning its symptoms, treatment, transmission, and prevention. JMIR Public Health Surveill 2017 Jun 19;3(2):e38 [FREE Full text] [doi: 10.2196/publichealth.7157] [Medline: 28630032]
56. Mehrotra R, Sanner S, Buntine W, Xie L. Improving lda topic models for microblogs via tweet poolingautomatic labeling. Presented at: Proceedings of the 36th International ACM SIGIR Conference on Research Development in Information Retrieval (SIGIR); July 28-August 01, 2013; Dublin, Ireland p. 889-892. [doi: 10.1145/2484028.2484166]
57. Hong L, Dan O, Davison BD. Predicting Popular Messages in Twitter. Presented at: Proceedings of the 20th International Conference Companion on World Wide Web (WWW); March 28-April 01, 2011; Hyderabad, India p. 57-58. [doi: 10.1145/1963192.1963222]
58. Naveed N, Gottron T, Kunegis J, Alhadi AC. Bad news travels fast: A content-based analysis of interestingness on twitter. Presented at: Proceedings of the 3rd International Web Science Conference (WebSci); June 15-17, 2011; Koblenz, Germany. [doi: 10.1145/2527031.2527052]
59. Petrovic S, Osborne M, Lavrenko V. RT to Win! Predicting Message Propagation in Twitter. Presented at: Proceedings of the 5th International AAAI Conference on Web and Social Media (ICWSM); July 17–21, 2011; Barcelona, Spain.
60. Machajdik J, Hanbury A. Affective image classification using features inspired by psychologyart theory. Presented at: Proceedings of the 18th ACM International Conference on Multimedia (MM); October 25-29, 2010; Firenze, Italy p. 83-92. [doi: 10.1145/1873951.1873965]
61. van de Weijer J, Schmid C, Verbeek J. Learning Color Names from Real-World Images. Presented at: Proceedings of the 20th IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 17-22 2007; Minneapolis, MN, USA p. 1-8. [doi: 10.1109/CVPR.2007.383218]
62. Valdez P, Mehrabian A. Effects of color on emotions. J Exp Psychol Gen 1994 Dec;123(4):394-409. [Medline: 7996122]
63. Zhao X, Zhu F, Qian W, Zhou A. Impact of Multimedia in Sina Weibo: Popularity and Life Span. Presented at: Proceedings of Joint Conference of 6th Chinese Semantic Web Symposium (CSWS) and the First Chinese Web Science Conference (CWSC); November 28-30, 2012; Shenzhen, China p. 55-65. [doi: 10.1007/978-1-4614-6880-6_5]
64. Bakhshi S, Shamma DA, Gilbert E. Faces engage us: Photos with faces attract more likes and comments on Instagram. Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI); April 26-May 01, 2014; Toronto, Canada p. 965-974. [doi: 10.1145/2556288.2557403]
65. Gelley B, John A. Do I Need To Follow You? Examining the Utility of the Pinterest Follow Mechanism. Presented at: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW); March 14-18, 2015; Vancouver, Canada p. 1751-1762. [doi: 10.1145/2675133.2675209]

Abbreviations

API: application programming interface
HPV: human papillomavirus
LDA: Latent Dirichlet Allocation
MMR: measles, mumps, and rubella
OCR: optical character recognition
SVM: Support Vector Machine

Edited by N Collier, M Conway, V Lampos, A Culotta; submitted 15.06.17; peer-reviewed by E Yom-Tov, P Massey, S Doan; comments to author 01.09.17; revised version received 27.10.17; accepted 29.01.18; published 03.04.18.

Please cite as:
Chen T, Dredze M
Vaccine Images on Twitter: Analysis of What Images are Shared
J Med Internet Res 2018;20(4):e130
URL: http://www.jmir.org/2018/4/e130/
doi:10.2196/jmir.8221
PMID:29615386

Original Paper

Harnessing Reddit to Understand the Written-Communication Challenges Experienced by Individuals With Mental Health Disorders: Analysis of Texts From Mental Health Communities

Albert Park[1], PhD; Mike Conway[1], PhD

Department of Biomedical Informatics, School of Medicine, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Albert Park, PhD
Department of Biomedical Informatics
School of Medicine
University of Utah
421 Wakara Way #140
Salt Lake City, UT, 84108-3514
United States
Phone: 1 206 743 7843
Fax: 1 801 581 4297
Email: alpark1216@gmail.com

Abstract

Background: Mental disorders such as depression, bipolar disorder, and schizophrenia are common, incapacitating, and have the potential to be fatal. Despite the prevalence and gravity of mental disorders, our knowledge concerning everyday challenges associated with them is relatively limited. One of the most studied deficits related to everyday challenges is language impairment, yet we do not know how mental disorders can impact common forms of written communication, for example, social media.

Objective: The aims of this study were to investigate written communication challenges manifest in online mental health communities focusing on depression, bipolar disorder, and schizophrenia, as well as the impact of participating in these online mental health communities on written communication. As the control, we selected three online health communities focusing on positive emotion, exercising, and weight management.

Methods: We examined lexical diversity and readability, both important features for measuring the quality of writing. We used four well-established readability metrics that consider word frequencies and syntactic complexity to measure writers' written communication ability. We then measured the lexical diversity by calculating the percentage of unique words in posts. To compare lexical diversity and readability among communities, we first applied pairwise independent sample *t* tests, followed by *P* value adjustments using the prespecified Hommel procedure to adjust for multiple comparison. To measure the changes, we applied linear least squares regression to the readability and lexical diversity scores against the interaction sequence for each member, followed by pairwise independent sample *t* tests and *P* value adjustments. Given the large sample of members, we also report effect sizes and 95% CIs for the pairwise comparisons.

Results: On average, members of depression, bipolar disorder, and schizophrenia communities showed indications of difficulty expressing their ideas compared with three other online health communities. Our results also suggest that participating in these platforms has the potential to improve members' written communication. For example, members of all three mental health communities showed statistically significant improvement in both lexical diversity and readability compared with members of the OHC focusing on positive emotion.

Conclusions: We provide new insights into the written communication challenges faced by individuals suffering from depression, bipolar disorder, and schizophrenia. A comparison with three other online health communities suggests that written communication in mental health communities is significantly more difficult to read, while also consisting of a significantly less diverse lexicon. We contribute practical suggestions for utilizing our findings in Web-based communication settings to enhance members' communicative experience. We consider these findings to be an important step toward understanding and addressing everyday written communication challenges among individuals suffering from mental disorders.

(J Med Internet Res 2018;20(4):e121) doi:10.2196/jmir.8219

KEYWORDS

mental health; depression; depressive disorder, major; depressive disorder; bipolar disorder; bipolar; bipolar and related disorders; schizophrenia; schizotypal personality disorder; schizophrenia spectrum and other psychotic disorders; consumer health information; informatics; information science; social support; psychosocial support system; community networks; self-help groups; communications media

Introduction

Mental disorders are common, incapacitating [1], and account for many years of lost productivity [2]. In addition, serious mental disorders [3] such as depression [4], bipolar disorder [5], and schizophrenia [6] have the potential to be fatal because of the increased risk of suicide. Despite the prevalence and gravity of mental disorders, our knowledge concerning everyday challenges associated with these conditions is relatively limited, especially when compared with many physical conditions.

One of the most studied deficits related to everyday challenges for individuals suffering from depression, bipolar disorder, and schizophrenia is language impairment [7-12]. Researchers of these mental disorders have long suspected language impairment because of deficits in frontal lobe functioning [10,13], which controls both emotion regulation and language processing. Language impairment is typically measured through one's performance in semantic processing tasks (ie, determining semantic relationships between a word, phrase, or category [14-16] or differentiating real words from pseudowords [17,18] based on an individual's semantic network [19]) and verbal fluency tasks (ie, production of words from phonemic or semantic categories [20-22]). Despite the importance of language in everyday life, these studies do not illustrate daily challenges associated with language impairment. Moreover, generalizability remains uncertain because of small sample size [7,8], with inconsistent results regarding language impairment or frontal lobe activities [10,23,24].

Despite its potential for devastating disability, it is unclear how language impairment manifests in common forms of written communication, for example, social media communication. With increasing use of technology comes increasing opportunity to write. For instance, in 2015, 84% of American adults used the internet [25], and one of the most frequent uses of the internet is written communication [26], including communication on social media. Nearly two-thirds of American adults use social media, roughly a tenfold increase from a decade ago [27]. A few social media platforms and online mental health communities within Reddit, for example, have become a popular venue for individuals suffering from mental disorders [28]. Reddit supports throwaway and unidentifiable accounts, which can protect users from social discrimination surrounding mental disorders [29-31] and allow honest discussions that may not be appropriate on other social media sites such as Facebook [32]. Reddit also provides contextual information that is relatively limited in other popular social media platforms (eg, Twitter), because of length limitations.

It is also known that effortful tasks (ie, requiring attention) such as expressing thoughts via writing are more difficult than automatic tasks (ie, not requiring attention) for individuals suffering from depression and bipolar disorder, whereas both types of tasks are equally difficult for schizophrenia patients [33,34]. From previous studies on mental disorders and Reddit [30,35-37], we can infer that individuals suffering from mental disorders also frequently engage in written communication, yet the written communication challenges faced by individuals in online mental health communities remain unknown.

Examining important features of writing provides an opportunity to assess members' written communication skills and any associated linguistic challenges. For example, a study on writing quality used linguistic features such as lexical diversity, syntactic complexity, and word frequency to predict the quality of writing [38]. In different studies, ease of reading (ie, simple and clear writing) [39] and text cohesion with respect to text flow [40] were suggested as some of the most determinant features of writing quality.

We can examine these features to assess online mental health community members' written communication challenges. More specifically, less lexical diversity and poor readability in posts can be a sign of language impairment. Research on language impairment has linked significantly less lexical diversity with specific language impairments [41]. Similarly, poor sentence structure and difficulties with organization and articulating ideas, which can be described as insufficient readability [42], were also associated with language impairment [43].

Readability metrics have been long-studied or used in the field of communication [39,44], education [45], and informatics [46-58], including social media writing [58]. Readability metrics provide quantitative estimates of the ease with which readers can comprehend a written text. Typically, they are given as an estimated US grade level by measuring the linguistic characteristics of a given text [59]. Moreover, readability metrics, although rudimentary, consider two of the three aforementioned features associated with writing quality: word frequencies [39,57,59,60] and syntactic complexity [38,39,45,57,59-61]. From the perspectives of the writers and their writing quality, readability metrics can measure the writers' ability to present ideas simply in a straightforward manner. According to one of the developers of the readability metrics, higher readability scores can indicate needless complexity [44] or writing challenges, such as organization and articulating ideas. Language impairment can hinder writers' ability to simply articulate ideas with ease, while using a less diverse lexicon. Moreover, one benefit of using these readability metrics is that they are computationally simple and relatively straightforward to apply. Thus, we use readability along with lexical diversity (ie, the third writing quality feature) of posts as a proxy for written communication challenges among individuals suffering from depression, bipolar disorder, and schizophrenia.

Though mental health and language impairment have been studied extensively [7-12], less is known about written communication challenges manifested in social media, as well

as the effects of long-term participation in online mental health communities on written communication challenges among individuals suffering from depression, bipolar disorder, and schizophrenia disorder. Understanding written communication challenges among these individuals has implications for treating mental disorders, managing online mental health communities, and conducting future research. Despite the importance in clinical, practical, and public policy implications for mental health, to our knowledge, the investigation of written communication challenges utilizing communication in online mental health communities has not been the focus of previous research on mental health.

We aim to fill this gap in the literature with this study and address two research questions (RQ):

> RQ1: To what extent do written communication challenges manifest in online mental health communities focusing on depression, bipolar disorder, and schizophrenia? As the control, we selected three online health communities (OHCs): one with less emotional challenges and two with less medical or technical terminology.
>
> RQ2: How would acts of participation (ie, posting to interact with other members) in online mental health communities impact members' written communication?

Methods

Community Platform

The data for this study consist of submissions and their associated comments from Reddit's several topically focused subcommunities called subreddits. Submissions are posts that start a conversation, and comments are posts that reply to submissions or other comments. Reddit is a highly popular social media platform with more than 82.5 billion page views, 73 million submissions, and 725 million associated comments from 88,700 active subreddits in 2015 [62]. In addition to Reddit's popularity, Reddit has features suitable for protecting mental health community members' identity (eg, throwaway and unidentifiable accounts). Thus, we examined submissions and comments (*posts* from here on out to maintain clarity) from Reddit to investigate written communication challenges among individuals suffering from potentially stigmatized conditions.

Subreddit Selection

r/depression, r/bipolar, and r/schizophrenia, to our knowledge, are the largest and most active subreddits for their respective mental disorders [63-65]. In May 2017, r/depression has been active for 8 years with 178,921 subscribers [63], r/bipolar has 24,724 subscribers and was formed 8 years ago [64], and r/schizophrenia has 7036 subscribers and has been active for 7 years [65]. Thus, we selected r/depression, r/bipolar, and r/schizophrenia as the main communities of interest for investigating the written communication challenges faced by individuals in online mental health communities.

To understand the significance of written communication among r/depression, r/bipolar, and r/schizophrenia members, we selected r/happy [66], r/loseit [67], and r/bodybuilding [68] for the controls. We first selected r/happy, a subreddit that was created to share positive thoughts and happy stories. The subreddit has been active for 9 years with 116,441 subscribers as of May 2017 [66]. Members of most OHCs experience emotional challenges [69-71] from the distress of living with—or being diagnosed with—a serious condition. However, we looked for an OHC that is not directly related to mental disorders, especially depression, to help ensure that this control group's written communication challenges are not related to mental distress even as a secondary symptom. Thus, we selected the largest and most active, positive, emotion-focused subreddit in Reddit.

We selected a second OHC, r/loseit, to bolster the quality of our findings. r/loseit is a subreddit focusing on weight management and has been a community for 6 years with 425,934 subscribers [67]. We purposely selected a community without a substantial amount of medical or technical terminology because a high level of difficult medical or technical terminology can skew the readability of posts. Although it may be impossible to select OHCs without any medical or technical terminology, one study of r/loseit characterized the most-discussed topics of the community as ordinary health information and management strategies, which can be described without complex medical or technical terminology (eg, food, clothing, physical appearance, workouts, and calorie counting) [72]. Moreover, unlike r/happy, r/loseit contains a substantial amount of emotional support [73], which can indicate that the members are facing emotional challenges similar to many OHCs. Thus, we selected r/loseit, the largest weight management community in Reddit, as a second control group.

We selected a third OHC, r/bodybuilding, in which members are dedicated to passion-centric activities, exercising, and muscular development. The bodybuilding community has 259,743 subscribers and has been active for 9 years [68]. A previous study suggested that members of an online bodybuilding community exchange a considerable level of emotional support (eg, motivational support and competition preparation support) and informational support (eg, training regimes and diets) [74]. The general discussion topics among bodybuilding community members could be relatively similar to the discussion topics among members of r/loseit; however, the two communities could consist of vastly different individuals with respect to health-related goals and habits. Thus, we include r/bodybuilding, the largest and most active muscular development community in Reddit, as the last control group.

Data

First, we used a dataset [75] (publicly available posts from October 2007 to May 2015) that was collected and archived by a Reddit member and has been used in several previous studies [36,76,77]. Second, we extracted posts made in r/depression, r/bipolar, r/schizophrenia, r/happy, r/loseit, and r/bodybuilding. We excluded posts that were marked as *[deleted]* in our analyses. Third, we removed posts with less than five words to help ensure the posts have expressive content and thoughts. Many posts in online communities are short—for example, one-word answering posts (eg, "yes" and "sure") that can be viewed as automatic tasks rather than effortful tasks. These

posts can skew the results; thus, we removed posts with less than five words. Fourth, to restrict our investigation to regular members (ie, exclude throwaway accounts or infrequent members) of the communities, we confined our analysis to members (ie, unique member IDs) who have four or more meaningful posts (ie, posts with five or more words) in the specific subreddit. In a different study [78], a similar threshold was used to determine lurkers who are not yet regularly contributing members. We used a similar threshold to identify regular members. We summarize the OHC dataset in Table 1.

The research reported in this study was exempted from review by the University of Utah's institutional review board (IRB; ethics committee; IRB 00076188) under Exemption 2 as defined in US Federal Regulations 45 CFR 46.101(b).

Research Question 1: Analysis for Communication Challenges in Social Media

To understand how language impairment manifests in written communication, we first measure the readability of posts. Readability of posts assesses writers' ability to simply and clearly present ideas. To assess readability, we used *Flesch-Kincaid grade level* [60], *Simple Measure of Gobbledygook (SMOG) index* [59], *Gunning Fog index* [39], and *Linsear Write formula* [61], all of which are widely used metrics in readability studies [47-56]. Even though readability metrics have been shown to correlate with one another [46], different readability metrics can still generate a range of results. To increase the reliability of our results, we calculated the mean of the four readability metrics, following the procedures of previous studies [47,48]. Additionally, we used *min-max normalization* in our analyses to give equal weight to each readability metric (*readability score* from here on out to maintain clarity); however, we also report the complete readability results by each readability metric and the mean before the normalization. To automatically perform the readability analysis, we used the open-source Python *textstat* package [79].

To calculate the mean of readability scores for each subreddit, we first calculated the mean of readability scores for individual members, then we calculated the mean for each subreddit. Next, we normalized the mean of readability scores for individual members based on minimum and maximum values of the specific communities. This two-step process is to prevent one prolific member skewing the mean of a subreddit. We then measured the lexical diversity by calculating the percent of unique words in posts (ie, the number of unique words divided by the number of total words) with the same two-step process, excluding the normalization process.

To compare readability scores and lexical diversity among different subreddits, we first conducted pairwise independent sample *t* tests, followed by *P* value adjustments using the prespecified Hommel procedure [80] to adjust for multiple comparisons. Given the large sample of members, we also reported effect sizes (*d*) using Cohen *d* [81], as well as 95% CI for the pairwise comparisons, following suggestions of a previous study [82]. The effect sizes were interpreted as *d* (.01)=very small, *d* (.2)=small, *d* (.5)=medium, *d* (.8)=large, *d* (1.2)=very large, and *d* (2.0)=huge [81,83]. We used the open-source R *lsr* package to measure the effect size [84].

To bolster our findings, we manually examined the validity of using readability scores for the purpose of measuring communication challenges. Because high readability scores can also indicate sophisticated language with complex sentence structure, we manually analyzed a randomly selected sample of 120 posts (ie, 20 posts from each subreddit) after controlling for the post lengths and readability scores: 60 posts with high readability scores (ie, top 5% readability scores of a respective subreddit) and 60 posts with low readability scores (ie, bottom 5% readability scores of a respective subreddit). Furthermore, we manually assigned these posts into high and low readability groups to compare readability scores against manual judgments.

Research Question 2: Analysis for Change of Communication Over Time in Social Media

To measure the change of readability and lexical diversity of posts made by each member participating in the six subreddits, we first calculated the readability scores and lexical diversity of individual posts. Then, we organized each post's readability score and lexical diversity according to the posting time per-member basis for the subreddit. Next, we applied linear least squares regression to them against the interaction sequence (ie, determined by the posting time) for each member. We performed linear least squares regression against the interaction sequence rather than time because we are interested in the change caused by each interaction rather than time. We reported the mean of slopes for readability scores and lexical diversity to reflect the overall changes in members in each of the six subreddits. Next, we applied pairwise independent sample *t* tests and the Hommel procedure. We then reported effect sizes and 95% CIs as we did in RQ1. For both analyses, we also reported a comparison among r/happy, r/loseit, and r/bodybuilding to deepen our understanding of the effects of emotional challenges in language impairment.

Table 1. Summary of the dataset.

Subreddit	Dates	Number of posts	Number of members
r/depression	December 2008 to May 2015	526,470	34,685
r/bipolar	January 2010 to May 2015	146,328	5019
r/schizophrenia	October 2012 to May 2015	22,273	896
r/happy	January 2008 to May 2015	70,516	6433
r/loseit	July 2010 to May 2015	1,054,949	46,367
r/bodybuilding	August 2009 to May 2015	724,190	18,927

Results

Research Question 1: Analysis for Communication Challenges in Social Media

We captured the mean and SE for (1) individual readability scores measured by four different metrics, (2) mean readability scores of the four metrics, (3) normalized mean readability scores of the four metrics, (4) lexical diversity, and (5) the total number of words in posts for each of the five communities (Table 2). On average, posts from r/schizophrenia were found to be the most difficult to read (ie, highest normalized readability scores), followed by posts from r/bipolar, r/depression, r/loseit, r/happy, and then r/bodybuilding. Lexical diversity showed a similar trend. On average, posts from r/happy had the most diverse lexicon, followed by posts from r/bodybuilding, r/loseit, r/bipolar, r/schizophrenia, and then r/depression. Figure 1 presents a scatter plot of the mean readability scores and lexical diversity among six different subreddits.

We then conducted pairwise independent sample *t* tests to compare readability scores and lexical diversity of each subreddit to understand the differences between two subreddits. Pairwise comparisons of normalized readability scores among subreddits are shown in Table 3.

Posts from r/bodybuilding, r/happy, and r/loseit were statistically significantly more simply written than posts from r/depression, r/bipolar, and r/schizophrenia in terms of syntactic complexity and word frequency that were measured in readability. The effect sizes were also in between medium to huge when readability scores of r/happy, r/loseit, and r/bodybuilding were compared to readability scores of r/depression, r/bipolar, and r/schizophrenia. Table 3 summarizes these findings.

Pairwise comparisons of lexical diversity showed similar results (Table 4). Posts from r/happy and r/bodybuilding used a significantly more diverse lexicon than the posts from r/depression, r/bipolar, and r/schizophrenia. The effect sizes ranged between very large to huge. Posts from r/loseit also had a significantly more diverse lexicon and had medium to large effect sizes than the posts from the three mental health subreddits. Differences in lexical diversity among posts from the three mental health subreddits had very small to small effect sizes. The lexical diversity differences between posts from r/bipolar and r/depression, as well as between r/schizophrenia and r/depression were statistically significant; however, posts from r/bipolar and r/schizophrenia were not significantly different. Interestingly, a significant difference with large to very large effect size of lexical diversity was found between the posts from r/happy and r/loseit as well. Table 4 summarizes findings on lexical diversity differences.

In our manual analyses, we found that both high and low readability score posts resembled common internet communication and was void of sophisticated writing. However, we encountered several inadequately articulated posts, many in the form of run-on sentence structure. Using inadequate articulation as a guide, we manually assigned 120 posts into high or low readability groups. The manual assessment agreed with the readability score 68% of the time (82 out of 120). The readability score and manual assessment had higher agreement in posts from mental health subreddits compared with the control groups. Mental health subreddits, r/depression, r/bipolar, and r/schizophrenia, had 80%, 80%, and 90% agreement, respectively. Conversely, the control subreddits, r/happy, r/loseit, and r/bodybuilding, had 40%, 70%, and 50% agreement, respectively.

Research Question 2: Analysis for Change of Communication Over Time in Social Media

To understand the effects of participating in online mental health communities with respect to their written communication, we applied linear least squares regression to readability scores and lexical diversity against the interaction sequence.

Members of the three mental health subreddits showed improvement in both readability scores (ie, negative slope for improvement) and lexical diversity (ie, positive slope for improvement). Among the mental health subreddits, r/bipolar showed the most improvement, followed by r/depression and r/schizophrenia for readability scores. For lexical diversity, members improved in order of r/bipolar, r/schizophrenia, and then r/depression. Members of r/bodybuilding had the biggest improvement in readability scores, and members of r/loseit also improved in both readability scores and lexical diversity. Members of r/happy only improved in lexical diversity (Table 5).

Table 2. Communication challenges in members. Variables are reported as the mean (SE) of readability scores, normalized mean of readability scores, lexical diversities, and the total number of words in posts for each community. SMOG: Simple Measure of Gobbledygook.

Subreddit	Flesch-Kincaid grade, mean (SE)	SMOG index, mean (SE)	Gunning Fog index, mean (SE)	Linsear Write formula, mean (SE)	Four metrics, mean (SE)	Four metrics, normalized mean (SE)	Lexical diversity, mean (SE)	Total number of words in posts, mean (SE)
r/happy	4.83 (0.03)	1.61 (0.02)	16.76 (0.03)	5.22 (0.02)	7.11 (0.02)	0.06 (0.0003)	0.93 (0.001)	29.10 (0.25)
r/bodybuilding	5.03 (0.02)	1.92 (0.01)	17.21 (0.02)	5.90 (0.01)	7.51 (0.01)	0.05 (0.0001)	0.92 (0.0003)	34.01 (0.16)
r/loseit	4.83 (0.01)	2.90 (0.01)	16.76 (0.01)	6.06 (0.01)	7.64 (0.01)	0.08 (8.3e-05)	0.88 (0.0003)	52.37 (0.16)
r/depression	5.53 (0.01)	3.74 (0.01)	17.05 (0.01)	6.69 (0.01)	8.25 (0.01)	0.09 (0.0001)	0.84 (0.0004)	76.24 (0.29)
r/bipolar	5.88 (0.03)	3.92 (0.02)	17.80 (0.03)	6.72 (0.03)	8.58 (0.02)	0.13 (0.0004)	0.85 (0.001)	69.76 (0.62)
r/schizophrenia	6.67 (0.08)	4.17 (0.07)	18.55 (0.08)	7.26 (0.09)	9.16 (0.06)	0.16 (0.001)	0.85 (0.002)	72.10 (1.65)

Figure 1. An overview of mean readability scores and lexical diversity among the six subreddits. The gray dotted lines represent the mean of the axes.

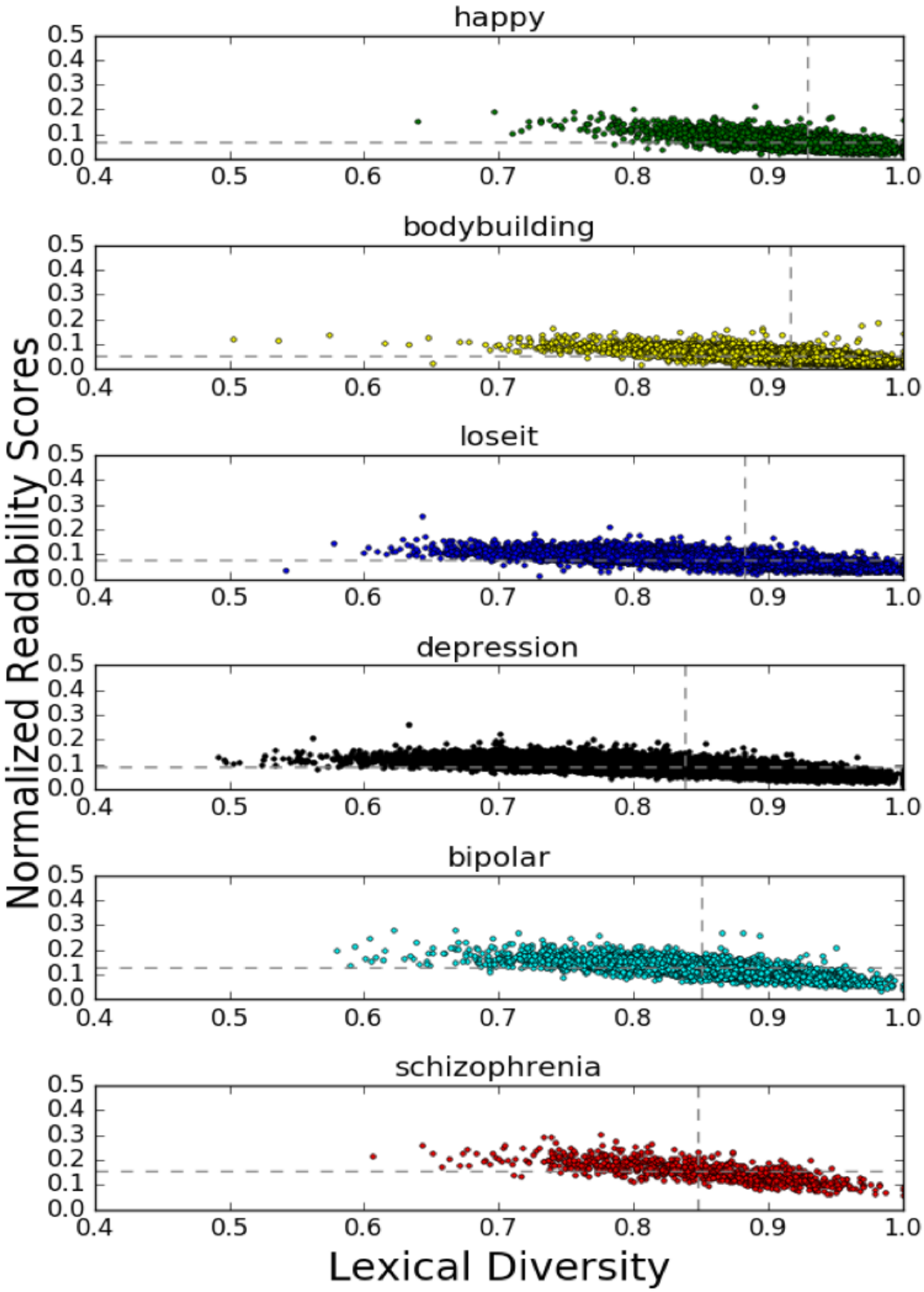


Table 3. Pairwise *t* test of the normalized average scores of four metrics.

Subreddit Comparison (ordered by readability scores)	*t* value	*P* value	Adjusted *P* value (Hommel)	95% CI	Effect size (*d*)
r/schizophrenia					
vs r/bipolar	22.52	<.001	<.001	0.03-0.03	0.97 (large-very large)
vs r/depression	52.70	<.001	<.001	0.07-0.07	3.18 (huge)
vs r/loseit	62.71	<.001	<.001	0.08-0.08	4.43 (huge)
vs r/happy	68.90	<.001	<.001	0.09-0.09	3.46 (huge)
vs r/bodybuilding	81.41	<.001	<.001	0.10-0.11	5.64 (huge)
r/bipolar					
vs r/depression	85.94	<.001	<.001	0.04-0.04	1.70 (very large-huge)
vs r/loseit	116.99	<.001	<.001	0.05-0.05	2.62 (huge)
vs r/happy	116.71	<.001	<.001	0.06-0.06	2.26 (huge)
vs r/bodybuilding	169.66	<.001	<.001	0.07-0.08	3.64 (huge)
r/depression					
vs r/loseit	92.43	<.001	<.001	0.01-0.01	0.67 (medium-large)
vs r/happy	71.60	<.001	<.001	0.02-0.02	1.08 (large-very large)
vs r/bodybuilding	222.10	<.001	<.001	0.04-0.04	1.90 (very large-huge)
r/loseit					
vs r/happy	32.69	<.001	<.001	0.01-0.01	0.55 (medium-large)
vs r/bodybuilding	163.41	<.001	<.001	0.02-0.02	1.39 (very large-huge)
r/happy					
vs r/bodybuilding	43.57	<.001	<.001	0.01-0.01	0.74 (medium-large)

Table 4. Pairwise *t* test of lexical diversity.

Subreddit Comparison (ordered by readability scores)	*t* value	*P* value	Adjusted *P* value (Hommel)	95% CI	Effect size (*d*)
r/happy					
vs r/bodybuilding	20.83	<.001	<.001	0.01 to 0.01	0.30 (small-medium)
vs r/loseit	80.58	<.001	<.001	0.05 to 0.05	0.89 (large-very large)
vs r/bipolar	80.04	<.001	<.001	0.08 to 0.08	1.57 (very large-huge)
vs r/schizophrenia	36.89	<.001	<.001	0.08 to 0.09	1.80 (very large-huge)
vs r/depression	143.16	<.001	<.001	0.09 to 0.09	1.42 (very large-huge)
r/bodybuilding					
vs r/loseit	84.41	<.001	<.001	0.03 to 0.03	0.66 (medium-large)
vs r/bipolar	74.10	<.001	<.001	0.06 to 0.07	1.40 (very large-huge)
vs r/schizophrenia	31.70	<.001	<.001	0.06 to 0.07	1.54 (very large-huge)
vs r/depression	163.08	<.001	<.001	0.08 to 0.08	1.30 (very large-huge)
r/loseit					
vs r/bipolar	36.78	<.001	<.001	0.03 to 0.03	0.59 (medium - large)
vs r/schizophrenia	16.09	<.001	<.001	0.03 to 0.04	0.64 (medium-large)
vs r/depression	100.61	<.001	<.001	0.04 to 0.05	0.74 (medium-large)
r/bipolar					
vs r/schizophrenia	1.21	.23	.23	−0.002 to 0.01	0.05 (very small-small)
vs r/depression	13.63	<.001	<.001	0.01 to 0.01	0.19 (very small-small)
r/schizophrenia					
vs r/depression	4.41	<.001	<.001	0.01 to 0.01	0.14 (very small-small)

Table 5. Writing quality changes in members. Variables are reported as the mean (SE) of slopes for readability scores, normalized mean of slopes for readability scores, slope of lexical diversities, and slope of the total number of words in posts for each community. SMOG: Simple Measure of Gobbledygook.

Subreddit	Flesch-Kincaid grade, mean (SE)	SMOG index, mean (SE)	Gunning Fog index, mean (SE)	Linsear Write formula, mean (SE)	Four metrics, mean (SE)	Four metrics, normalized mean (SE)	Lexical diversity, mean (SE)	Total number of words in posts, mean (SE)
r/happy	−0.01 (0.005)	0.0002 (0.01)	−0.004 (0.003)	−0.01 (0.01)	−0.01 (0.01)	0.13 (0.56)	0.34 (0.23)	−0.002 (0.001)
r/bodybuilding	−0.05 (0.01)	−0.21 (0.10)	−0.06 (0.04)	−0.29 (0.22)	−0.10 (0.03)	−15.22 (5.17)	3.85 (0.34)	−0.01 (0.001)
r/loseit	−0.002 (0.005)	−0.08 (0.04)	−0.02 (0.02)	−0.02 (0.01)	−0.03 (0.01)	−5.37 (1.74)	0.93 (0.88)	−0.005 (0.0006)
r/depression	−0.03 (0.004)	−0.06 (0.004)	−0.01 (0.003)	−0.04 (0.004)	−0.05 (0.005)	−5.59 (0.32)	2.40 (0.10)	−0.01 (0.0003)
r/bipolar	−0.09 (0.01)	−0.14 (0.02)	−0.03 (0.01)	−0.12 (0.02)	−0.14 (0.02)	−9.12 (1.02)	5.04 (0.48)	−0.01(0.001)
r/schizophrenia	−0.02 (0.02)	−0.09 (0.02)	0.005 (0.02)	−0.07 (0.04)	−0.07 (0.03)	−4.69 (1.25)	3.89 (0.89)	−0.01 (0.002)

Table 6. Pairwise *t* test of changes of the normalized average readability scores.

Subreddit Comparison (ordered by readability scores)	*t* value	*P* value	Adjusted *P* value (Hommel)	95% CI	Effect size (*d*)
r/happy					
vs r/schizophrenia	3.52	<.001	.01	2.13 to 7.51	0.11 (very small-small)
vs r/loseit	3.01	.003	.03	1.92 to 9.08	0.02 (very small-small)
vs r/depression	8.87	<.001	<.001	4.46 to 6.99	0.10 (very small-small)
vs r/bipolar	7.98	<.001	<.001	6.98 to 11.53	0.16 (very small-small)
vs r/bodybuilding	2.95	.003	.03	5.17 to 25.54	0.02 (very small-small)
r/schizophrenia					
vs r/loseit	0.32	.75	.90	−3.52 to 4.88	0.002 (very small)
vs r/depression	0.70	.49	.90	−1.64 to 3.44	0.02 (very small-small)
vs r/bipolar	2.75	.01	.05	1.27 to 7.59	0.07 (very small-small)
vs r/bodybuilding	1.98	.05	.24	0.11 to 20.95	0.02 (very small-small)
r/loseit					
vs r/depression	0.12	.90	.90	−3.25 to 3.69	0.0008 (very small)
vs r/bipolar	1.86	.06	.31	−0.20 to 7.70	0.01 (very small-small)
vs r/bodybuilding	1.81	.07	.35	−0.84 to 20.53	0.02 (very small-small)
r/depression					
vs r/bipolar	3.31	.001	.01	1.44 to 5.62	0.06 (very small-small)
vs r/bodybuilding	1.86	.06	.31	−0.52 to 19.77	0.02 (very small-small)
r/bipolar					
vs r/bodybuilding	1.16	.25	.90	−4.22 to 16.42	0.01 (very small)

To understand the significance of the changes in readability scores and lexical diversity, we compared the changes that occurred in the three mental health subreddits against r/happy, r/bodybuilding, and r/loseit via pairwise independent sample *t* tests. The overall comparisons of readability scores among subreddits are shown in Table 6.

Subreddit comparisons indicate that the readability of posts by members of all three mental health subreddits improved significantly more than members of r/happy. Yet, the effect sizes for those comparisons were very small to small. Moreover, only the readability of posts by members of r/bipolar improved significantly more than posts by members of r/depression and r/schizophrenia, with very small to small effects among the pairwise comparison of three mental health subreddits.

Members of r/bipolar also had the most improvement in terms of lexical diversity and significantly more than members of r/depression, r/loseit, and r/happy, albeit the effect sizes were very small to small (Table 7). Furthermore, members of r/schizophrenia and r/depression improved significantly more than members of r/happy; however, no significant difference was found against r/loseit.

Table 7. Pairwise *t* test of lexical diversity changes.

Subreddit Comparison (ordered by readability scores)	*t* value	*P* value	Adjusted *P* value (Hommel)	95% CI	Effect size (*d*)
r/bipolar					
vs r/schizophrenia	1.14	.25	.76	−0.83 to 3.13	0.03 (very small-small)
vs r/bodybuilding	2.01	.04	.22	0.03 to 2.36	0.03 (very small-small)
vs r/depression	5.35	<.001	<.001	1.68 to 3.61	0.12 (very small-small)
vs r/loseit	4.09	<.001	<.001	2.14-6.08	0.02 (very small-small)
vs r/happy	8.77	<.001	<.001	3.65 to 5.75	0.18 (very small-small)
r/schizophrenia					
vs r/bodybuilding	0.04	.96	.96	−1.82 to 1.91	0.001 (very small)
vs r/depression	1.68	.09	.38	−0.26 to 3.24	0.08 (very small-small)
vs r/loseit	2.37	.02	.12	0.51 to 5.41	0.02 (very small-small)
vs r/happy	3.88	<.001	<.001	1.75 to 5.35	0.18 (very small-small)
r/bodybuilding					
vs r/depression	4.05	<.001	<.001	0.75 to 2.15	0.05 (very small-small)
vs r/loseit	3.08	.002	.02	1.06 to 4.77	0.02 (very small-small)
vs r/happy	8.45	<.001	<.001	2.69 to 4.32	0.08 (very small-small)
r/depression					
vs r/loseit	1.66	.10	.39	−0.27 to 3.21	0.01 (very small-small)
vs r/happy	8.16	<.001	<.001	1.56 to 2.55	0.11 (very small-small)
r/loseit					
vs r/happy	0.64	.52	.96	−1.2 to 2.37	0.003 (very small)

Discussion

Principal Findings

We examined the issue of written communication challenges using readability and lexical diversity of posts from publicly accessible online mental health communities on Reddit. We found that on average, members of depression, bipolar disorder, and schizophrenia subreddits wrote posts that are significantly more difficult to read and had significantly less lexical diversity when compared with three other OHCs focusing on positive emotion, exercising, and weight management.

We also found that as members of mental health communities participated more in the community, they wrote posts that were easier to read with more lexical diversity. Interestingly, members of other OHCs also improved, with the exception of readability scores of r/happy members. Only r/bipolar members showed statistically significant improvement in lexical diversity compared with members of the two other OHCs (r/happy and r/loseit), while showing statistically significant improvement compared with r/happy in terms of readability scores. Compared with r/happy members, r/depression and r/schizophrenia members also significantly improved in both examined features.

Another interesting finding is readability scores and lexical diversity of r/loseit, in which members could have depressive symptoms because of the distress of being overweight. The readability scores and lexical diversity of r/loseit were in between r/happy and three mental health subreddits. Still, the posts from r/loseit were statistically significantly easier to read with more lexical diversity (medium to huge effect sizes) compared with the three mental health subreddits. However, posts from r/loseit were statistically significantly harder to read (medium to large effect size), with less lexical diversity (large to very large effect size) compared with r/happy. Members of r/bodybuilding and r/happy wrote more similar to one another than to members of r/loseit in terms of readability scores and lexical diversity.

Despite the possible language impairment faced by members of mental health communities, their real-life communication challenges are unknown. To our knowledge, this is the first study to show mental health community members' written communication challenges occurring in the real world using social media.

Practical Implication for Online Communication and Mental Health

Our analyses suggest that members of online mental health communities could encounter incoherent texts because of the language impairment of other members. Automatically correcting misspellings [85], simplifying language [86], and improving text coherence [87] in posts could enhance the readability of posts and the overall experience of participating in these communities.

Many online communities, including many Reddit's subreddits, utilize moderators to regulate content and support members. A number of automated systems have been suggested to assist

moderators and reduce moderator burden [88]. Similarly, an adaptation of our automatic analysis method could be a basis for detecting individuals whose lexical diversity and readability of posts are worsening in massive scale networks. This could indicate worsening of mental disorder symptoms, and such a feature could alert and allow moderators to provide timely support.

We also showed the potential for improving written communication via more frequent writing in online mental health communities. Designing features of online mental health communities for the purpose of improving written communication can enhance the everyday life of individuals suffering from mental conditions. For example, a place for expressive writing can improve their symptoms [89,90] and possibly help with their written communication challenges.

User Privacy

Research using publicly accessible social media data (such as Reddit) is typically granted exemption from review by IRBs in the US context; however, ethical considerations such as privacy remain critical [91-93]. In this paper, we do not report any user identifiable information to protect user privacy (eg, direct quotations and usernames).

Limitation and Future Directions

Our study has several limitations. A number of confounding factors such as individuals' premorbid-intelligence, -verbal skill and -education level, as well as demographic and geographical characteristics [9] could influence the writing quality other than language impairment. Other possible confounding factors associated with group dynamics and mental health conditions include the communication practices and cultures of specific subreddits, as well as medication and substance use of individuals suffering from mental health conditions. Furthermore, we assumed that high readability scores are reflecting inadequate articulation or organization by writers. Although inadequate articulation and organization can increase readability scores, high readability scores can also be because of sophisticated language and complex sentence structure. However, we did not encounter sophisticated writing in our manual assessment, and it is unlikely that such sophistication and complexity are highly prevalent in everyday communication. Similarly, we do not know how readability scores were influenced by common online communication attributes such as slang, abbreviation, community nomenclature, and misspellings [85], or how lexical diversity was impacted by number of topics and change of topics [94]. However, these online communication attributes are more likely to occur in all subreddits, and thus, affecting the readability scores in a similar manner. Reddit is a widely used platform more frequently used by young males [95,96] in English-speaking nations [96]. Despite more user activities from English speaking nations (85%) [96], it is unclear how participation by English as second language speakers is affecting the results. Additionally, members who choose to participate in r/depression, r/bipolar, and r/schizophrenia are not necessarily representative of their respective populations and are subject to selection bias. Similarly, we do not have any evidence that members of these three mental health subreddits are clinically diagnosed; the severity of their condition is unknown, and overlapping memberships could exist in these subreddits. However, one of the main limitations of previous studies were small sample sizes [7,8], which could be the underlying reason for the inconsistent results [10,23,24]. Thus, given the size of r/depression, r/bipolar, and r/schizophrenia, the prevalence and gravity of mental disorders, the increasing popularity of social media, and the potential challenges associated with daily use of social media make Reddit an interesting platform to study.

Although beyond the scope of this study, further investigation regarding readability metrics may be needed for more accurately determining the grade reading level [46]. We selected readability metrics based on the literature in which the metrics have been validated or used [44,46-56]. However, we noticed a disparity among the metrics. For example, readability scores by *Gunning Fog index* were far greater than the other three metrics. *SMOG index* resulted in readability scores that were less than the other metrics. Despite the apparent differences, the scores were correlated with one another as a previous study suggested [46], and we used the mean of normalized scores of four readability metrics to strengthen the reliability of our findings. Due to the consistent statistical results, we believe that these four metrics can measure the general difficulty of readability. We also acknowledge that our large sample size could have inflated the statistical significance levels. Thus, we reported 95% CIs and used effect sizes when interpreting the results. Another interesting future direction would be to investigate why members are improving and longitudinal changes in written communication with respect to prolonged participation in online mental health communities. In this study, we only examined the overall impact of participation in online mental health communities; however, understanding how members are improving their written communication skills could potentially inform the design of related patient education programs.

Conclusions

We provide new insights into the written communication challenges faced by individuals suffering from depression, bipolar disorder, and schizophrenia. A comparison of mental health communities to three other OHCs suggests that writings in mental health communities were significantly more difficult to read, while consisting of a significantly less diverse lexicon. Our findings also suggest that participating in these subreddits has the potential to improve members' written communication over time. We contribute practical suggestions for utilizing our findings in online communication settings to enhance members' communicative experience. We consider these findings to be an important step toward understanding written communication challenges among individuals suffering from mental disorders.

Acknowledgments

We restricted our analysis to publicly available discussion content. The study was exempted from review by the University of Utah's Institutional Review Board (Ethics Committee) [IRB 00076188]. AP's contribution to this research was supported by the National Library of Medicine of the National Institutes of Health under training grant T15 LM007124. MC's contribution to this research was supported by the National Library of Medicine of the National Institutes of Health under award numbers R00LM011393 & K99LM011393. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Kessler RC, Aguilar-Gaxiola S, Alonso J, Chatterji S, Lee S, Ormel J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. Epidemiol Psichiatr Soc 2009;18(1):23-33 [FREE Full text] [Medline: 19378696]
2. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 09;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
3. Hirokawa S, Kawakami N, Matsumoto T, Inagaki A, Eguchi N, Tsuchiya M, et al. Mental disorders and suicide in Japan: a nation-wide psychological autopsy case-control study. J Affect Disord 2012 Oct;140(2):168-175. [doi: 10.1016/j.jad.2012.02.001] [Medline: 22391515]
4. Takahashi Y. Depression and suicide. Japan Med Assoc J 2001;44(8):359-363 [FREE Full text]
5. Oquendo MA, Waternaux C, Brodsky B, Parsons B, Haas GL, Malone KM, et al. Suicidal behavior in bipolar mood disorder: clinical characteristics of attempters and nonattempters. J Affect Disord 2000 Aug;59(2):107-117. [Medline: 10837879]
6. Melle I, Johannesen JO, Friis S, Haahr U, Joa I, Larsen TK, et al. Early detection of the first episode of schizophrenia and suicidal behavior. Am J Psychiatry 2006 May;163(5):800-804. [doi: 10.1176/ajp.2006.163.5.800] [Medline: 16648319]
7. Moritz S, Birkner C, Kloss M, Jahn H, Hand I, Haasen C, et al. Executive functioning in obsessive-compulsive disorder, unipolar depression, and schizophrenia. Arch Clin Neuropsychol 2002 Jul;17(5):477-483. [Medline: 14592001]
8. Fossati P, Amar G, Raoux N, Ergis AM, Allilaire JF. Executive functioning and verbal memory in young patients with unipolar depression and schizophrenia. Psychiatry Res 1999 Dec 27;89(3):171-187. [Medline: 10708264]
9. Suslow T. Estimating verbal intelligence in unipolar depression: comparison of word definition and word recognition. Nord J Psychiatry 2009;63(2):120-123. [doi: 10.1080/08039480802316010] [Medline: 18985516]
10. Klumpp H, Deldin P. Review of brain functioning in depression for semantic processing and verbal fluency. Int J Psychophysiol 2010 Feb;75(2):77-85. [doi: 10.1016/j.ijpsycho.2009.10.003] [Medline: 19815035]
11. Fossati P, Guillaume LB, Ergis AM, Allilaire JF. Qualitative analysis of verbal fluency in depression. Psychiatry Res 2003 Jan 25;117(1):17-24. [Medline: 12581817]
12. Fossati P, Harvey PO, Le Bastard G, Ergis AM, Jouvent R, Allilaire JF. Verbal memory performance of patients with a first depressive episode and patients with unipolar and bipolar recurrent depression. J Psychiatr Res 2004;38(2):137-144. [Medline: 14757327]
13. Franke P, Maier W, Hardt J, Frieboes R, Lichtermann D, Hain C. Assessment of frontal lobe functioning in schizophrenia and unipolar major depression. Psychopathology 1993;26(2):76-84. [doi: 10.1159/000284803] [Medline: 8321896]
14. Bedny M, Hulbert JC, Thompson-Schill SL. Understanding words in context: the role of Broca's area in word comprehension. Brain Res 2007 May 18;1146:101-114. [doi: 10.1016/j.brainres.2006.10.012] [Medline: 17123486]
15. Seghier ML, Lazeyras F, Pegna AJ, Annoni JM, Zimine I, Mayer E, et al. Variability of fMRI activation during a phonological and semantic language task in healthy subjects. Hum Brain Mapp 2004 Nov;23(3):140-155. [doi: 10.1002/hbm.20053] [Medline: 15449358]
16. D'Arcy RC, Connolly JF, Service E, Hawco CS, Houlihan ME. Separating phonological and semantic processing in auditory sentence processing: a high-resolution event-related brain potential study. Hum Brain Mapp 2004 May;22(1):40-51. [doi: 10.1002/hbm.20008] [Medline: 15083525]
17. Binder JR, McKiernan KA, Parsons ME, Westbury CF, Possing ET, Kaufman JN, et al. Neural correlates of lexical access during visual word recognition. J Cogn Neurosci 2003 Apr 01;15(3):372-393. [doi: 10.1162/089892903321593108] [Medline: 12729490]
18. Xiao Z, Zhang JX, Wang X, Wu R, Hu X, Weng X, et al. Differential activity in left inferior frontal gyrus for pseudowords and real words: an event-related fMRI study on auditory lexical decision. Hum Brain Mapp 2005 Jun;25(2):212-221. [doi: 10.1002/hbm.20105] [Medline: 15846769]
19. Collins AM, Loftus EF. A spreading-activation theory of semantic processing. Psychol Rev 1975 Oct;82(6):407-428 [FREE Full text] [doi: 10.1037/0033-295X.82.6.407]

20. Borkowski JG, Benton AL, Spreen O. Word fluency and brain damage. Neuropsychologia 1967 May;5(2):135-140 [FREE Full text] [doi: 10.1016/0028-3932(67)90015-2]
21. Frith CD, Friston KJ, Liddle PF, Frackowiak RS. A PET study of word finding. Neuropsychologia 1991 Dec;29(12):1137-1148 [FREE Full text] [doi: 10.1016/0028-3932(91)90029-8]
22. Basso A, Burgio F, Prandoni P. Semantic category and initial letter word fluency in left-brain-damaged patients. Eur J Neurol 1997 Nov;4(6):544-550 [FREE Full text] [doi: 10.1111/j.1468-1331.1997.tb00404.x]
23. Frith CD, Friston KJ, Herold S, Silbersweig D, Fletcher P, Cahill C, et al. Regional brain activity in chronic schizophrenic patients during the performance of a verbal fluency task. Br J Psychiatry 1995 Sep;167(3):343-349. [Medline: 7496643]
24. Heinrichs RW, Zakzanis KK. Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. Neuropsychology 1998 Jul;12(3):426-445. [Medline: 9673998]
25. Perrin A, Duggan M. Pew Internet. 2015. Americans' Internet access Internet: 2000-2015 URL: http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015 [WebCite Cache ID 6qDiVhx1k]
26. Fox S, Rainie L. Pew Internet.: Pew Internet & American Life Project; 2014. The Web at 25 in the U.S. The overall verdict: The internet has been a plus for society and an especially good thing for individual users URL: http://www.pewinternet.org/2014/02/27/the-web-at-25-in-the-u-s [WebCite Cache ID 6Y5L8D2oE]
27. Perrin A. Pew Internet.: Pew Internet & American Life Project; 2015. Social Media Usage : 2005-2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015 [WebCite Cache ID 6lwiynC0A]
28. Conway M, O'Connor D. Social media, big data, and mental Health: current advances and ethical implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]
29. Corrigan P. How stigma interferes with mental health care. Am Psychol 2004 Oct;59(7):614-625. [doi: 10.1037/0003-066X.59.7.614] [Medline: 15491256]
30. De CM, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. 2014 Presented at: Proceedings of ICWSM, AAAI; June 1, 2014; Ann Arbor, Michigan, USA p. 71-80.
31. Barney LJ, Griffiths KM, Banfield MA. Explicit and implicit information needs of people with depression: a qualitative investigation of problems reported on an online depression support forum. BMC Psychiatry 2011 May 18;11:88 [FREE Full text] [doi: 10.1186/1471-244X-11-88] [Medline: 21592335]
32. Newman M, Lauterbach D, Munson S, Resnick P, Morris M. It's not that I don't have problems, I'm just not putting them on Facebook. : ACM Press; 2011 Presented at: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work; March 19-23, 2011; Hangzhou, China p. 341-350 URL: http://dl.acm.org/citation.cfm?id=1958824.1958876 [doi: 10.1145/1958824.1958876]
33. Hartlage S, Alloy LB, Vázquez C, Dykman B. Automatic and effortful processing in depression. Psychol Bull 1993 Mar;113(2):247-278. [Medline: 8451334]
34. Roy-Byrne PP, Weingartner H, Bierer LM, Thompson K, Post RM. Effortful and automatic cognitive processes in depression. Arch Gen Psychiatry 1986 Mar;43(3):265-267. [Medline: 3954546]
35. 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]
36. Gkotsis G, Oellrich A, Hubbard TJ, Dobson RJ, Liakata M, Velupillai S. The language of mental health problems in social media. : Association for Computational Linguistics Presented at: Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 16, 2016; San Diego, CA p. 63-73 URL: https://www.researchgate.net/profile/George_Gkotsis/publication/303365632_The_language_of_mental_health_problems_in_social_media/links/573e2d0408ae9f741b3009ff.pdf
37. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach. Comput Human Behav 2018 Jan;78:98-112. [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
38. McNamara DS, Crossley SA, McCarthy PM. Linguistic features of writing quality. Writ Commun 2009 Nov 24;27(1):57-86 [FREE Full text] [doi: 10.1177/0741088309351547]
39. Gunning R. The Technique of Clear Writing. New York, NY, US: McGraw-Hill International Book Co; 1952.
40. Halliday MA, Hasan H. Cohesion in English. London, UK: Longman Group Ltd; 1976.
41. Watkins RV, Kelly DJ, Harbers HM, Hollis W. Measuring children's lexical diversity: differentiating typical and impaired language learners. J Speech Hear Res 1995 Dec;38(6):1349-1355. [Medline: 8747826]
42. Todirascu A, François T, Gala N, Fairon C, Ligozat AL, Bernhard D. Coherence and cohesion for the assessment of text readability. Presented at: International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2013); October 15-16, 2013; Marseille, France p. 11-19.
43. Dockrell JE, Lindsay G, Connelly V. The impact of specific language impairment on adolescents' written text. Except Child 2009 Jul;75(4):427-446 [FREE Full text] [doi: 10.1177/001440290907500403]
44. Gunning R. The fog index after twenty years. J Bus Commun 2016 Sep 16;6(2):3-13 [FREE Full text] [doi: 10.1177/002194366900600202]

45. Bormuth JR. : US Department of Health, Education, and Welfate; Office of Education Bureau of Research; 1969. Development of Readability Analyses URL: https://files.eric.ed.gov/fulltext/ED029166.pdf [accessed 2018-03-17] [WebCite Cache ID 6xyidJ2yM]
46. Meade CD, Smith CF. Readability formulas: cautions and criteria. Patient Educ Couns 1991 Apr;17(2):153-158 [FREE Full text] [doi: 10.1016/0738-3991(91)90017-Y]
47. Park A, Zhu SH, Conway M. The readability of electronic cigarette health information and advice: a quantitative analysis of web-based information. JMIR Public Health Surveill 2017 Jan 06;3(1):e1 [FREE Full text] [doi: 10.2196/publichealth.6687] [Medline: 28062390]
48. Eltorai AE, Naqvi SS, Ghanian S, Eberson CP, Weiss AC, Born CT, et al. Readability of invasive procedure consent forms. Clin Transl Sci 2015 Dec;8(6):830-833 [FREE Full text] [doi: 10.1111/cts.12364] [Medline: 26678039]
49. Meade CD, Byrd JC. Patient literacy and the readability of smoking education literature. Am J Public Health 1989 Feb;79(2):204-206. [Medline: 2913844]
50. Malouff J, Gabrilowitz D, Schutte N. Readability of health warnings on alcohol and tobacco products. Am J Public Health 1992 Mar;82(3):464. [Medline: 1536374]
51. 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]
52. D'Alessandro DM, Kingsley P, Johnson-West J. The readability of pediatric patient education materials on the World Wide Web. Arch Pediatr Adolesc Med 2001 Jul;155(7):807-812. [Medline: 11434848]
53. Yin HS, Gupta RS, Tomopoulos S, Wolf MS, Mendelsohn AL, Antler L, et al. Readability, suitability, and characteristics of asthma action plans: examination of factors that may impair understanding. Pediatrics 2013 Jan;131(1):e116-e126. [doi: 10.1542/peds.2012-0612] [Medline: 23209106]
54. Risoldi CZ, Gregory P, Wilson A. Readability of consumer health information on the internet: a comparison of U.S. government-funded and commercially funded websites. J Health Commun 2012;17(9):1003-1010. [doi: 10.1080/10810730.2011.650823] [Medline: 22512714]
55. Agarwal N, Hansberry DR, Sabourin V, Tomei KL, Prestigiacomo CJ. A comparative analysis of the quality of patient education materials from medical specialties. JAMA Intern Med 2013 Jul 08;173(13):1257-1259. [doi: 10.1001/jamainternmed.2013.6060] [Medline: 23689468]
56. Terblanche M, Burgess L. Examining the readability of patient-informed consent forms. Open AccessJ Clin Trials 2010 Oct:157 [FREE Full text] [doi: 10.2147/OAJCT.S13608]
57. Pitler E, Nenkova A. Revisiting readability: A unified framework for predicting text quality. Presented at: EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing; October 25-27, 2008; Honolulu, Hawaii p. 186-195 URL: http://dl.acm.org/citation.cfm?id=1613715.1613742
58. Agichtein E, Castillo C, Donato D, Gionis A, Mishne G. Finding high-quality content in social media. : ACM Press Presented at: Proceedings of the international conference on Web search and web data mining - WSDM '08; February 11-12, 2008; Palo Alto, California, USA p. 183 URL: http://portal.acm.org/citation.cfm?doid=1341531.1341557
59. McLaughlin GH. SMOG grading: a new readability formula. J Read 1969;12(8):639-646 [FREE Full text]
60. Kincaid JP, Fishburne RP, Chissom BS, Rogers RL. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Millington, TN: Naval Technical Training Command; Feb 1975.
61. Wikipedia. Linsear Write Readability Formula Internet URL: https://en.wikipedia.org/wiki/Linsear_Write [accessed 2017-05-14] [WebCite Cache ID 6qSbnXtuc]
62. Reddit. 2015. Reddit in 2015 URL: http://www.redditblog.com/2015/12/reddit-in-2015.html [accessed 2016-01-12] [WebCite Cache ID 6eTHN0TFD]
63. Reddit. 2017. r/depression in May 2017 URL: https://www.reddit.com/r/depression/search?q=depression [accessed 2017-05-11] [WebCite Cache ID 6qOM1RXcE]
64. Reddit. 2017. r/bipolar in May 2017 URL: https://www.reddit.com/search?q=bipolar [accessed 2017-05-11] [WebCite Cache ID 6qOMetQpg]
65. Reddit. 2017. r/schizophrenia in May 2017 URL: https://www.reddit.com/search?q=schizophrenia [accessed 2017-05-11] [WebCite Cache ID 6qOO4tqYc]
66. Reddit. 2017. r/happy in May 2017 URL: https://www.reddit.com/search?q=happy [accessed 2017-05-11] [WebCite Cache ID 6qOPnUJ8f]
67. Reddit. 2017. r/loseit in May 2017 URL: https://www.reddit.com/search?q=loseit [accessed 2017-05-11] [WebCite Cache ID 6qOQ3l4QA]
68. Reddit. 2017. r/bodybuilding in Oct 2017 URL: https://www.reddit.com/search?q=bodybuilding [accessed 2017-10-14] [WebCite Cache ID 6uCFDwTJ5]
69. Park A, Hartzler AL, Huh J, McDonald DW, Pratt W. Homophily of vocabulary usage: beneficial effects of vocabulary similarity on online health communities participation. AMIA Annu Symp Proc 2015;2015:1024-1033 [FREE Full text] [Medline: 26958240]

70. Hwang KO, Ottenbacher AJ, Green AP, Cannon-Diehl MR, Richardson O, Bernstam EV, et al. Social support in an Internet weight loss community. Int J Med Inform 2010 Jan;79(1):5-13 [FREE Full text] [doi: 10.1016/j.ijmedinf.2009.10.003] [Medline: 19945338]
71. Wang Y, Kraut R, Levine J. To stay or leave? : ACM Press Presented at: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12; February 11-15, 2012; Seattle, Washington, USA p. 833-842 URL: http://dl.acm.org/citation.cfm?doid=2145204.2145329 [doi: 10.1145/2145204.2145329]
72. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Couto Silva AP, Meira Jr 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]
73. Cunha TO, Weber I, Haddadi H, Pappa G. The Effect of Social Feedback in a Reddit Weight Loss Community. New York, NY, USA: ACM Press Presented at: DH '16 Proceedings of the 6th International Conference on Digital Health Conference; April 11-13, 2016; Montréal, Québec, Canada p. 99-103 URL: http://dl.acm.org/citation.cfm?doid=2896338.2897732
74. Ploderer B, Howard S, Thomas P. Being Online, Living Offline: The Influence of Social Ties over the Appropriation of Social Network Sites. : ACM Press Presented at: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work; November 08-12, 2008; San Diego, CA, USA p. 333-342 URL: http://portal.acm.org/citation.cfm?id=1460618
75. Reddit_Member. Reddit. I have every publicly available Reddit comment for research URL: https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/; [WebCite Cache ID 6kgAuNxDE]
76. Park A, Conway M. Towards tracking opium related discussions in social media. Online J Public Health Inform 2017 May 02;9(1):e73 [FREE Full text] [doi: 10.5210/ojphi.v9i1.7652]
77. Park A, Conway M. Tracking Health Related Discussions on Reddit for Public Health Applications. Presented at: Annu Symp proceedings AMIA Symp; November 3-8, 2017; Washington D.C URL: https://knowledge.amia.org/65881-amia-1.3897810/t003-1.3901461/f003-1.3901462/2730999-1.3901643/2730999-1.3901644?timeStamp=1520105203311
78. Nonnecke B, Preece J. Lurker demographics. New York, NY, USA: ACM Press Presented at: Proceedings of the SIGCHI conference on Human Factors in Computing Systems; April 01-06, 2000; The Hague, The Netherlands p. 73-80 URL: http://dl.acm.org/citation.cfm?id=332409 [doi: 10.1145/332040.332409]
79. Bansal S, Aggarwal C. Python Software Foundation. Textstat URL: https://pypi.python.org/pypi/textstat [accessed 2016-09-23] [WebCite Cache ID 6kjFlWIwD]
80. Wright SP. Adjusted P-values for simultaneous inference. Biometrics 1992 Dec;48(4):1005 [FREE Full text] [doi: 10.2307/2532694]
81. Cohen J. Statistical Power Analysis for the Behavioral Sciences. (2nd ed). Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
82. Lin M, Lucas HC, Shmueli G. Research commentary—too big to fail: large samples and the p-value problem. Inf Syst Res 2013 Dec;24(4):906-917 [FREE Full text] [doi: 10.1287/isre.2013.0480]
83. Sawilowsky SS. New e ect size rules of thumb. J Mod Appl Stat Methods 2009;8(2):597-599.
84. Navarro D. : Citeseer; 2014. Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners Version 0.5 URL: https://cran.r-project.org/web/packages/lsr/lsr.pdf [WebCite Cache ID 6qjwho9sl]
85. Park A, Hartzler AL, Huh J, McDonald DW, Pratt W. Automatically detecting failures in natural language processing tools for online community text. J Med Internet Res 2015 Aug 31;17(8):e212 [FREE Full text] [doi: 10.2196/jmir.4612] [Medline: 26323337]
86. Carroll JB, Minnen G, Pearce D, Canning Y. Simplifying text for language-impaired readers. Presented at: Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics; June 08-12, 1999; Bergen, Norway p. 269-270 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.5993&rep=rep1&type=pdf
87. Smith CA, Hetzel S, Dalrymple P, Keselman A. Beyond readability: investigating coherence of clinical text for consumers. J Med Internet Res 2011 Dec 02;13(4):e104 [FREE Full text] [doi: 10.2196/jmir.1842] [Medline: 22138127]
88. Calvo RA, Hussain MS, Milne D, Nordbo K, Hickie I, Danckwerts P. Augmenting online mental health support services. In: Integrating Technology in Positive Psychology Practice. Hershey, PA: IGI Global; 2016:82-103.
89. Krpan KM, Kross E, Berman MG, Deldin PJ, Askren MK, Jonides J. An everyday activity as a treatment for depression: the benefits of expressive writing for people diagnosed with major depressive disorder. J Affect Disord 2013 Sep 25;150(3):1148-1151 [FREE Full text] [doi: 10.1016/j.jad.2013.05.065] [Medline: 23790815]
90. Koopman C, Ismailji T, Holmes D, Classen CC, Palesh O, Wales T. The effects of expressive writing on pain, depression and posttraumatic stress disorder symptoms in survivors of intimate partner violence. J Health Psychol 2005 Mar;10(2):211-221. [doi: 10.1177/1359105305049769] [Medline: 15723891]
91. Conway M. Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. J Med Internet Res 2014 Dec 22;16(12):e290 [FREE Full text] [doi: 10.2196/jmir.3617] [Medline: 25533619]
92. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]

93. 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]
94. Park A, Hartzler AL, Huh J, Hsieh G, McDonald DW, Pratt W. "How Did We Get Here?": topic drift in online health discussions. J Med Internet Res 2016 Nov 02;18(11):e284 [FREE Full text] [doi: 10.2196/jmir.6297] [Medline: 27806924]
95. Duggan M, Smith A. Pewinternet. 2013. 6% of online adults are Reddit users URL: http://www.pewinternet.org/2013/07/03/6-of-online-adults-are-reddit-users/ [accessed 2016-02-12] [WebCite Cache ID 6fEEqzeJp]
96. Bogers T, Wernersen R. How 'Social' are Social News Sites? Exploring the Motivations for Using Reddit.com. : iSchools; 2014 Presented at: iConference 2014 Proceedings; March 4-7, 2014; Berlin, Germany p. 329-344 URL: https://www.ideals.illinois.edu/handle/2142/47295

Abbreviations

IRB: institutional review board
OHC: online health community
RQ: research question
SMOG: Simple Measure of Gobbledygook

Edited by N Collier, M Conway, V Lampos, A Culotta; submitted 16.06.17; peer-reviewed by S Champlin, G Lim, A Elmessiry; comments to author 01.09.17; revised version received 09.12.17; accepted 03.01.18; published 10.04.18.

Please cite as:
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;20(4):e121
URL: http://www.jmir.org/2018/4/e121/
doi:10.2196/jmir.8219
PMID:29636316

Original Paper

Utilization of a Clinical Trial Management System for the Whole Clinical Trial Process as an Integrated Database: System Development

Yu Rang Park[1,2,3], PhD; Young Jo Yoon[4], MSc; HaYeong Koo[1], BS; Soyoung Yoo[3,5], PhD; Chang-Min Choi[1,6,7], MD, PhD; Sung-Ho Beck[1,8], MD, PhD; Tae Won Kim[1,7], MD, PhD

[1]Clinical Research Center, Asan Institute of Life Sciences, Asan Medical Center, Seoul, Republic Of Korea

[2]Department of Biomedical Informatics, Asan Medical Center, Seoul, Republic Of Korea

[3]Health Innovation Bigdata Center, Seoul, Republic Of Korea

[4]Office of Clinical Research Information, Asan Institute of Life Sciences, Asan Medical Center, Seoul, Republic Of Korea

[5]Human Research Protection Center, Asan Institute of Life Sciences, Asan Medical Center, Seoul, Republic Of Korea

[6]Department of Pulmonology and Critical Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic Of Korea

[7]Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic Of Korea

[8]Clinical Trial Center, Asan Institute of Life Sciences, Asan Medical Center, Seoul, Republic Of Korea

Corresponding Author:
Tae Won Kim, MD, PhD
Clinical Research Center
Asan Institute of Life Sciences
Asan Medical Center
88 Olympic-ro 43-gil, Songpa-gu
Seoul, 05505
Republic Of Korea
Phone: 1 230103910
Fax: 1 230106961
Email: twkimmd@amc.seoul.kr

Abstract

Background: Clinical trials pose potential risks in both communications and management due to the various stakeholders involved when performing clinical trials. The academic medical center has a responsibility and obligation to conduct and manage clinical trials while maintaining a sufficiently high level of quality, therefore it is necessary to build an information technology system to support standardized clinical trial processes and comply with relevant regulations.

Objective: The objective of the study was to address the challenges identified while performing clinical trials at an academic medical center, Asan Medical Center (AMC) in Korea, by developing and utilizing a clinical trial management system (CTMS) that complies with standardized processes from multiple departments or units, controlled vocabularies, security, and privacy regulations.

Methods: This study describes the methods, considerations, and recommendations for the development and utilization of the CTMS as a consolidated research database in an academic medical center. A task force was formed to define and standardize the clinical trial performance process at the site level. On the basis of the agreed standardized process, the CTMS was designed and developed as an all-in-one system complying with privacy and security regulations.

Results: In this study, the processes and standard mapped vocabularies of a clinical trial were established at the academic medical center. On the basis of these processes and vocabularies, a CTMS was built which interfaces with the existing trial systems such as the electronic institutional review board health information system, enterprise resource planning, and the barcode system. To protect patient data, the CTMS implements data governance and access rules, and excludes 21 personal health identifiers according to the Health Insurance Portability and Accountability Act (HIPAA) privacy rule and Korean privacy laws. Since December 2014, the CTMS has been successfully implemented and used by 881 internal and external users for managing 11,645 studies and 146,943 subjects.

Conclusions: The CTMS was introduced in the Asan Medical Center to manage the large amounts of data involved with clinical trial operations. Inter- and intraunit control of data and resources can be easily conducted through the CTMS system. To our knowledge, this is the first CTMS developed in-house at an academic medical center side which can enhance the efficiency of clinical trial management in compliance with privacy and security laws.

(J Med Internet Res 2018;20(4):e103) doi:10.2196/jmir.9312

KEYWORDS

clinical trial; information systems; academic medical center; information technology; privacy

Introduction

Background

Clinical trials can pose potential risks and hurdles in communication between parties due to the many stakeholders involved, namely pharmaceutical companies, clinical research organizations (CROs), health authorities, ethical committees or institutional review boards (IRBs), courier vendors, and academic medical centers [1-4]. Given the wide scope and high volume of participants in clinical trials conducted nowadays, efficiency is a critical issue at site level during the trial [1,4]. These challenges arise due to the various authorities involved which gives rise to conflicting administrative processes, dysfunctional communications with the IRB, limitation of real-time data access for both investigators and authorities to consistently keep patients on track, limited personnel and/or infrastructural resources, noncompliance due to flaws in reporting, and omission of major events for awards and stipends. Furthermore, poor monetary management problems such as billing and checking problems can occur. These challenges can lead to subsequent controversy in future compliance [4-9].

To address the challenges outlined above, pharmaceutical companies and CROs have developed supportive tools such as clinical trial management systems (CTMSs) [2]. CTMSs enable the centralization of the clinical trial process, thereby reducing the number of procedural errors and enhancing communications among multiple stakeholders by providing timely metrics updates on a real-time basis [10]. Academic medical centers, however, have reported that proper utilization of commercial CTMSs poses certain operational hurdles [6,9-11].

Objectives

As the academic medical center must manage the patient's sensitive data compared with the pharmaceutical company, the security of the system should be considered in development. Moreover, academic medical centers have to make extra efforts to establish a system that is able to link pre-existing systems such as the health information system (HIS; subject management), electronic institutional review board (e-IRB; study approval), and enterprise resource planning (ERP) program (contract and budget of clinical trial) [12]. Therefore, the academic medical center can gain greater benefits from utilizing a centralized CTMS as its integrated research database, because copious volumes of medical data are often collected and saved in separately-operated systems. The overall objective of this study was to describe the methods, considerations, and recommendations for the development and utilization of CTMS as a consolidated research database in an academic medical center. We focused on the standardization of clinical trial processes and terminologies, enabling system interface among the legacy systems, and compliance with regulations of security and privacy of clinical data to better manage the whole process of clinical trials performed at academic medical centers.

Methods

Overview

This CTMS was launched in a tertiary hospital, Asan Medical Center (AMC). AMC is one of the largest academic medical centers in Asia, housing more than 2700 inpatient beds, and having around 10,000 patients visiting the outpatient clinics per day on average. Moreover, around 1100 clinical studies are initiated and conducted in the center every year, and it is one of the centers with the highest numbers of clinical trials in Korea. AMC has obtained and retained full accreditation from the Association for the Accreditation of Human Research Protection Program since 2013, and the AMC IRB has received accreditation from the Forum for Ethical Review Committees in the Asian and Western Pacific Region since 2006.

Since 1997, the AMC clinical trial center (CTC) has been designated as a specialized institution for conducting clinical trials. AMC CTC was designated as the Global Center of Excellence by the Ministry of Health and Welfare in 2012.

Task Force Activities for System Development

The first step in the system development was to identify the tasks and longitudinal challenges that should be handled by the system, thereby establishing detailed system requirements. To identify and verify the system requirements from various stakeholders for developing a site-level CTMS, AMC organized a task force for 14 months from November 2013 to December 2014. The task force consisted of 17 stakeholders in clinical trial and research (including principal investigators, clinical research coordinator, clinical research associate, pharmacist, contract specialist, budget specialist information technology (IT) experts, relevant IRBs, and information protection subcommittee). All authors of this paper participated in this task force. Following 6 topics are discussed in this task force: (1) define clinical trial processes at the site- and unit-specific level, (2) define data governance and access rules for each member and member's unit, (3) define and match controlled vocabulary and annotation rules to represent clinical trial data, (4) define system requirements for each department, (5) define an interfacing strategy between CTMS and legacy systems (e-IRB, ERP, and HIS), and (6) define data protection strategies based on international and domestic law. For this purpose, 56 stakeholders were interviewed 25 times, and the resulting data

were used by the CTMS task force for developing site-level CTMS. On the basis of these results from CTMS task force, we define a strategy, a process, and guidelines for building and managing site-level CTMS.

System Development

The basic architecture for building CTMS was defined as follows, which involved all members of the aforementioned CTMS task force. The first step to building CTMS was defining the trial processes at the site level. The next step was to define and extract vocabulary from actual clinical trial data. The vocabulary set was matched with the standard terminology, namely the Clinical Data Interchange Standards Consortium (CDISC)-controlled vocabulary version P29 [13]. We then identified the tasks of all departments involved in the clinical trial and collected the system requirements through the analysis of necessary functions and data for each task (needed for computerization and collaboration, characteristics of generated data, access to data, etc). Table 1 shows the categories and features for management of overall clinical trials at the site level.

Finally, we designed the system architecture for developing CTMS at the site level, which includes modeling the database scheme, specific user interface (UI), and user experience (UX) design. In all the above system development processes, we referenced official guidelines from the US Food and Drug Administration (FDA) and the Australian Therapeutic Goods Administration (TGA). The US FDA has outlined a guidance document addressing computerized systems used in clinical investigations, otherwise known as 21 Part 11 [14] and the Australian TGA has developed a set of notes on Good Clinical Practice (GCP) [15]. They are used to protect research subjects, manage consent, and comply with reporting obligations to regulatory agencies such as IRB, US FDA, and Korean Ministry of Food and Drug Safety. Each function of the CTMS is intended to ensure that key stakeholders in clinical trials effectively comply with GCP and FDA 21 Part 11, the regulation for managing electronic clinical trial data. For example, we implemented the single sign-on function for authority check, leaving a log of all events occurring in the CTMS and managing the training log of the participants. To protect personal information, the Health Insurance Portability and Accountability Act (HIPAA) and Korean Privacy Act have been followed when designing and building the system.

A total of 12 months was spent on system development. The first 3 months were spent for collecting unit and user requirements, 2 months were spent on designing systems such as the database scheme, UI, UX, and terminology, 5 months were spent on developing, and the remaining 2 months were spent on system deployments. The evaluation of this system was carried out by monitoring the change of clinical test management flow in the schedule management for subjects according to CTMS introduction. AMC CTMS was built in the IBM Advanced Interactive eXecutive 6.1 server using JAVA 6, Oracle 11g, Apache 2.2.27, and Weblogic 10.3.6.0. CTMS can be accessed via all major internet browsers such as Chrome, Safari, Internet Explorer, Edge, and Firefox. Communication between server and system was secured with a Hypertext Transfer Protocol Secure in combination with a Secure Sockets Layer or Transport Layer Security protocol.

Table 1. The categories and features for management of overall clinical trials at the site level. The clinical trial management system (CTMS) features are divided into 3 types: inputting data, receiving an interface through a different system, and automatically calculating.

Category	Features	Description
Study management	Study list[a]	The study listed in the CTMS[b] is an IRB[c] approved study. IRB and CTMS are shared daily with the system interface. Additional study may be registered by the user
	Status history[d]	Ability to show a record of all events (additions, modifications, deletions) taking place in the study
	Milestone management[d]	Ability to manage timeline of study (eg, Regulatory Complete)
	Contract management[d]	Ability to manage clinical trial contract information (contract timeline, stakeholders, negotiations, etc)
	Budget management[d]	Management functions for budgeting and execution of research funds
	Document management[d]	Ability to manage all documents generated during clinical trials (separated by department)
Site management	Site management[d]	Ability to manage information about another site when carrying out multisite research by ARO[e]
	Communication and contact information management[d]	Ability to manage communication with a site or other organization by ARO
	Investigational product[d]	Ability to manage information about interventional product, vendor, sponsor by ARO
Subject management	Subject management[f]	Ability to manage subject information automatically interface with HIS[g]
	Subject scheduling calendar[f]	Ability to manage the subject's schedule is automatically displayed in calendar form through the patient management function
	Quality control management[d]	Ability to manage the quality control of subject includes informed consent, adverse event, protocol deviation, and inclusion or exclusion criteria managing for each subject
	Healthy volunteer announcement list[d]	Ability to manage the registration of healthy volunteer recruitment announcement and healthy volunteer pool management
Clinical monitoring	IRB regulatory list[d]	IRB approval and related data management ability for each site in case of multisite research
	MFDS[h] regulatory list[d]	MFDS approval and related data management ability
	Site visit list[d]	Ability to manage information about site visit status and results
	Protocol deviation list[d]	Ability to manage information related to site-specific protocol deviation
	SAE[d,i] list	Ability to manage information related to site-specific SAE
External request management	Feasibility request management[d]	Ability to manage feasibility request from sponsors
	PRIMS[d,j]	Provides clinical trial advisor request management from external pharmaceutical companies
	Clinical trial management of medical devices[d]	Provides clinical trial management functions related to medical devices coming from external organizations
Resource management	Medical equipment management[a]	Ability to manage medical device calibration information
	Drug management[d]	Ability to manage the clinical drug import and export
	Biomaterial management[a]	Ability to manage the biomaterial obtained during clinical trials
	Monitoring room management[d]	Ability to manage spaces for monitoring by external clinical research associate (CRA)
	Investigator profile[d]	Ability to manage investigator profile of our organization
Education management	Education management[d]	Ability to manage training for clinical trial worker in inside and outside organization

Category	Features	Description
	SOP[d,k] Education	Supports SOP management, training, and automatic notification by the unit
Unit management	Unit member management[d]	Support for unit member management, CTMS page access right and position management
	Member education management[d]	Demonstrate the data generated by the education manager for the education management of unit members
	To-do list management[e]	Provides management functions for tasks to be performed by each user. The task is automatically generated by CTMS such as request confirmation to obtain clinical patient consent form

[a]The interface.
[b]CTMS: Clinical Trial Management System.
[c]IRB: Institutional Review Board.
[d]Inputted directly by the user.
[e]ARO: academic research organization.
[f]Automatic calculation.
[g]HIS: Health Information System.
[h]MFDS: Ministry of Food and Drug Safety.
[i]SAE: Serious Adverse Event.
[j]PRIMS: Preclinical and eaRly ClInical Support program.
[k]SOP: Standard Operating Procedure.

Results

We successfully developed and introduced the CTMS to AMC to manage the large amount of data involved in clinical trial operations. This system was designed for inter- and intraunit control of data and resources to improve efficiency in clinical trial management.

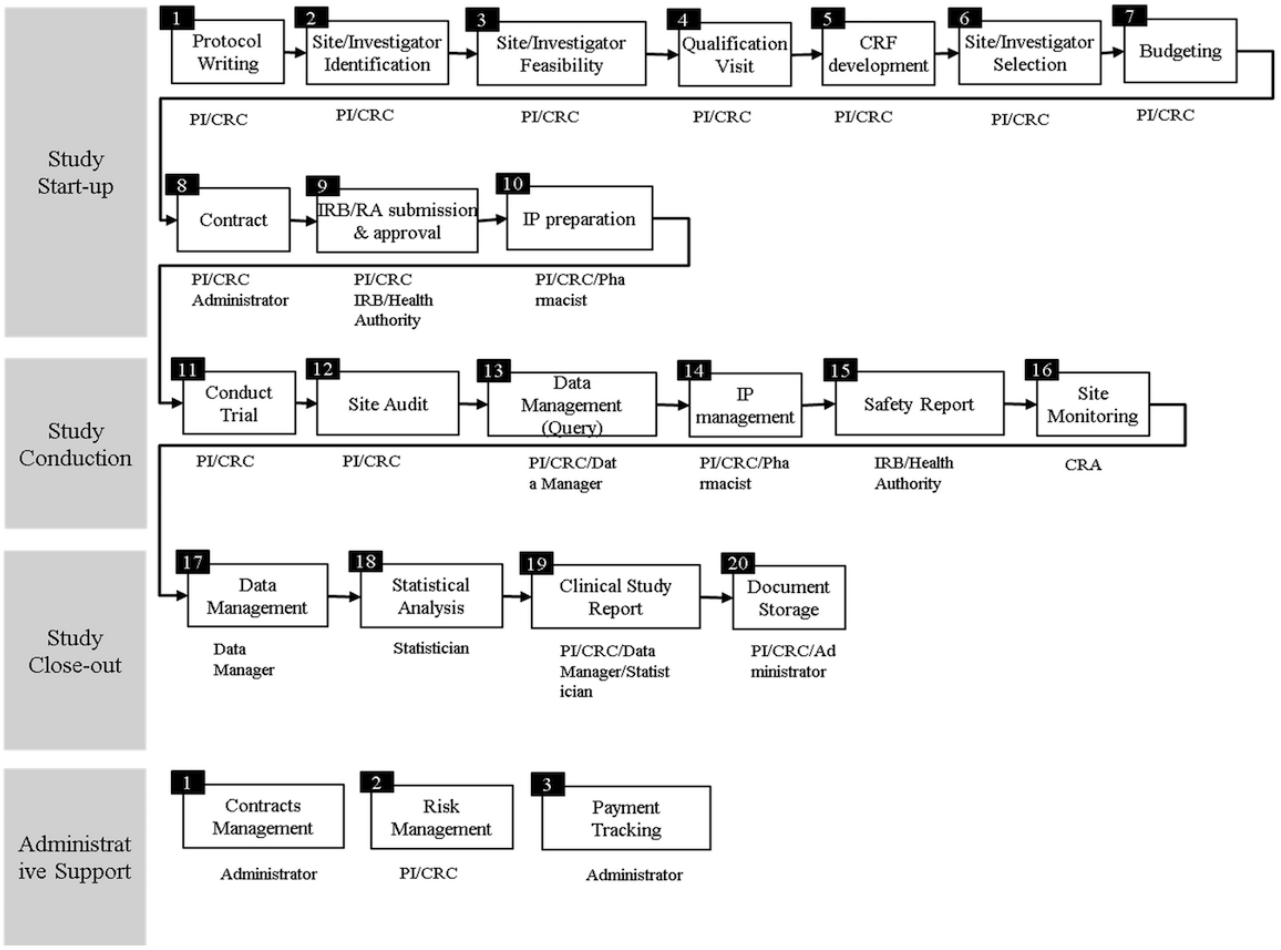
Site Level Clinical Trial Process

The general processes of a clinical trial at the site level were established and grouped into 4 important milestones (namely study start-up, study conduction, study close out, and administrative support) and then further categorized into 23 internal processes (Figure 1). We defined 4 groups of milestones to include all start-up activities from protocol writing to intellectual property preparation, clinical trial conduction step including site monitoring, trial close-out step to cover document storage process, and, finally, administrative support that covers contract management and payment tracking support for the overall clinical trial process.

Controlled Vocabulary for Clinical Trial

Ensuring the uniformity of terms is crucial when creating interdepartmental collaboration systems. In this study, a set of related terms for each process were defined after examining each department's clinical trial processes to ensure the consistency. We extracted 1354 terms at the raw level by dividing them into 23 individual clinical trial processes (defined in the Results section), and into 7 clinical trial tasks, namely study management, resource management, healthy volunteer, report management, education management, user and organization management, and administrative management. A team consisting of 1 medical doctor, 2 clinical research coordinators (CRCs), 2 medical records technicians, and 1 IT professional mapped the raw level terminology used in the local terms used at each site and produced 643 mapped terms. The defined 643 terms were created in 2 languages—Korean and English—and the terms were used throughout the construction of the CTMS system. We also performed a mapping with the CDISC Controlled Terminology, which is a representative term for clinical trials, to verify the representativeness of the term. As a result, almost two-thirds of the terms (421/643, 65.4%) were verified to have been mapped correctly.

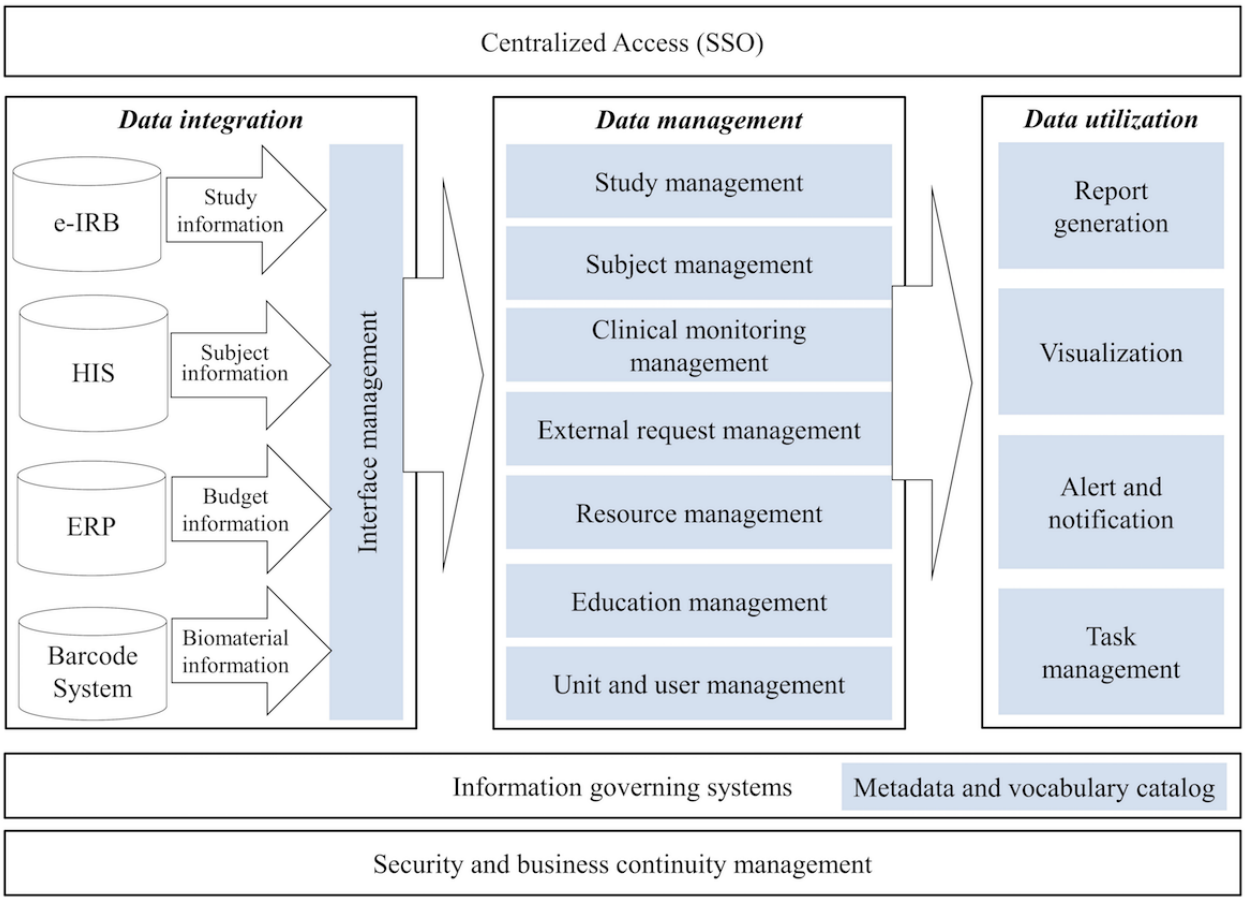
Figure 1. A designed clinical trial process for the Clinical Trial Management System. CRA: clinical research associate, CRC: clinical research coordinator, CRF: case report form, IP: intellectual property, IRB: institutional review board, PI: principal investigator.



System Architecture

The AMC CTMS is an all-in-one system containing all functions needed for operating site-level clinical trials and is a Web-based single sign-on system. As shown in Figure 2, the AMC CTMS is mainly composed of 3-layered domains, namely data integration, management, and utilization. First, the data integration layer of the system includes functions that interface data between the CTMS and legacy systems such as e-IRB, HIS, ERP, and barcode system in the hospital. Data for each existing system are unified and transmitted to the CTMS according to the system-specific interfacing cycle. Depending on the characteristics of the legacy system, the interface methods with the CTMS are different: (1) large scale systems such as the HIS and ERP systems were made through a separate interface server, (2) e-IRB as a Web-based system is linked to a Web service via the Application Programming Interface, and (3) the barcode system as a standalone system is linked through a database to database link. The data management layer of the system includes 7 management applications for basic study management and for managing operational data for clinical trials at the site level. Finally, the data utilization layer of the system is composed of the following 4 applications—report generation, visualization, alert and notification, and task management. The 4 applications in this domain are based on operating data in the CTMS to support clinical trial operations, such as notifications of subject visits to the hospital.

Figure 2. System architecture for Clinical Trial Management System. e-IRB: electronic institutional review board, ERP: enterprise resource planning, HIS: health information system, SSO: Single Sign-On.



Data Governance and Access Rules

As AMC CTMS was designed to be used by various people both inside and outside the clinical trial site, the CTMS task force created rigorous data governance and access rules to protect both the system and the data. In terms of data access authority, the basic unit of the CTMS is the study, and the access right of the study is defined according to the IRB approval. Page access rights are classified according to departmental business characteristics. Every page is clearly distinguished between the subjects of management and use. For example, the Biomaterial Management page is the subject of management of the clinical pharmacokinetics laboratory, but the subject of use is all researchers and CTC members. External user registration is the responsibility of the system administrator, and only for companies previously registered with IRB. Furthermore, access rights are also granted in a limited manner.

Privacy and Security

To protect sensitive patient information, 21 personal health identifiers (PHIs) were excluded from the CTMS to comply with international and domestic law. Table 2 shows the 21 PHIs adopted by AMC for developing a de-identified Clinical Data Warehouse. As a result, a CTMS user can identify the subject based on the subject's number assigned for each clinical trial and their date of birth. Systematically, the patient's registration number in the hospital is encrypted and stored in the CTMS database, which is linked to the HIS data for automatic data transferring. A more detailed description of the 21 PHIs determined by AMC has been described previously [16,17].

Table 2. The 21 Personal Health identifiers adopted by the Asan Medical Center from Table 1 in Shin et al reference number [17] (adapted with permission).

Number	Identifier	Remarks	Reference
1	Name	Excludes physician's name, includes information regarding friends and relatives.	HIPAA[a] safe harbor; HIPAA LDS[b]
2	Address	Smaller than the submunicipal level divisions (Dong, -Eup, and -Myeon).	HIPAA safe harbor; HIPAA LDS
3	Phone number	Includes mobile phone and fax numbers.	HIPAA safe harbor; HIPAA LDS
4	Email address		HIPAA safe harbor; HIPAA LDS
5	Resident registration number		Korean Personal Information Protection Act
6	Foreigner registration number		Korean Personal Information Protection Act
7	Passport number		Korean Personal Information Protection Act
8	Health insurance policy number		HIPAA safe harbor; HIPAA LDS
9	Bank account number		HIPAA safe harbor; HIPAA LDS
10	Credit card number		HIPAA safe harbor
11	Certificate or license number	Driver's license	Korean Personal Information Protection Act; HIPAA safe harbor; HIPAA LDS
12	Vehicle license plate number		HIPAA safe harbor; HIPAA LDS
13	Patient identifier	Medical record numbers	HIPAA safe harbor
14	Hospital membership ID	Hospital homepage, referral system	Korean Act on Promotion of Information and Communication Network Utilization and Information Protection
15	Hospital employee number		HIPAA safe harbor
16	IP address		HIPAA safe harbor; HIPAA LDS
17	URL		HIPAA safe harbor; HIPAA LDS
18	Biometric identifier	Fingerprints, retina, vein, voice prints, and personally identifiable genetic information	HIPAA safe harbor; HIPAA LDS
19	Full-face photographic images and any comparable images		HIPAA safe harbor; HIPAA LDS
20	Birth date (allowing year and month)	For example, July 1960 can be used, but July 4, 1960, should be used as July 1960	HIPAA safe harbor
21	Other unique identifying numbers	Pathology numbers	HIPAA safe harbor

[a]HIPAA: Health Insurance Portability and Accountability Act.

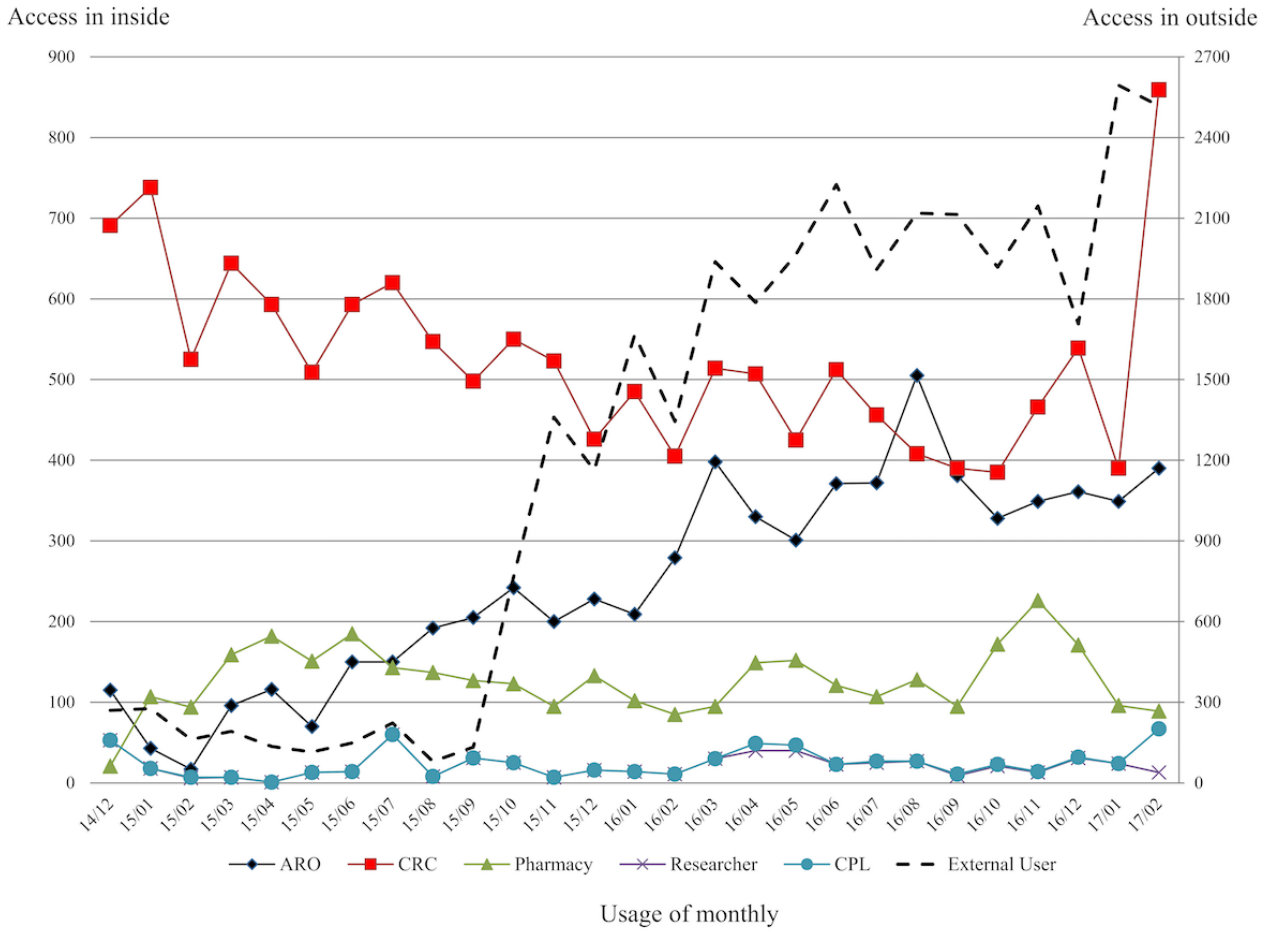
[b]LDS: Limited Data Set.

The CTMS system goes through a pre-inspection in accordance with the privacy principle guidelines. The privacy principle guidelines are broken down into 64 items, and this is performed on a nonregular basis by the Ministry of Government Administration and Home Affairs of South Korea. A simulation hacking test was conducted based on 3 steps, namely information gathering, vulnerability categorization, and penetration phase, by configuring the major vulnerabilities of the system from the 2013 Open Web Application Security Project's 10 weak points, and the National Intelligence Service's 8 weak points.

Evaluation

The evaluation of the CTMS was performed by examining scenarios related to schedule management for subjects in clinical trials. Before the introduction of the AMC CTMS, clinical trial stakeholders proceeded with their own research in isolation. For example, to manage subject schedules, the CRC used a personalized tool such as Google Calendar, and the record was not reused. In the CTMS, each research study is assigned a specific study design (Multimedia Appendix 1). If the user is registered as a participating researcher upon approval of the IRB study, the assignment can also be accessed in the CTMS.

Figure 3. Monthly access trend of Clinical Trial Management System by user's occupation. ARO: academic research organization, CRC: clinical research coordinator, CPL: clinical pharmacokinetics laboratory.



When a subject is recruited, the patient's schedule is shown in a calendar format in the study design data, which is essentially controlled by the CRC, although the study participant has the right to retrieve these data (Multimedia Appendix 2). If there is a schedule for export of clinical trial drug, these data will be forwarded to the pharmacy in advance and a record of the amount of exported drug is stored in the CTMS (Multimedia Appendix 3). As a result, within the CTMS, all stakeholders can use and view the same data with varying privileges in terms of patient schedule management.

Deployments

Since December 2014, the CTMS has been successfully implemented at AMC. To utilize the data before the introduction of the system, data loading work was carried out 2 months before the system opened (approximately 8000 studies and 100,000 patients' information with related data). As a result, by March 31 2017, a total of 11,645 studies have been managed by 881 users in the AMC CTMS. Moreover, a total of 146,943 subjects have been enrolled in the AMC CTMS, including those who were managed by the pre-existing HIS. There are 1316 external users, including CRA from CROs and trainees. Figure 3 shows the monthly access trend of the CTMS according to user's department. We observed distinct differences in the CTMS access rates according to the occupation of the user in the hospital. For example, the ARO showed a steadily increasing pattern (R^2=.81), whereas the CRC decreased, but the access rate suddenly increased due to the implementation of additional functions for CRC. The utilization rate for external users showed a dramatically increasing overall pattern (R^2=.85).

Discussion

Principal Findings

Clinical trial data management and data quality control pose challenging tasks in many organizations participating in clinical trials [1,3], especially in academic medical centers where the environment is optimized for patient care and not clinical trials [4,18]. In pharmaceutical companies and CROs, CTMS are generally developed to match the characteristics of the tasks performed by the institution. Many academic medical centers are not only concerned with conducting clinical trials, but also have an obligation to collect and retain data related for the whole clinical trial period while maintaining a sufficiently high level of data quality [1,4,12]. In recent years, the obligation to monitor the ARO function has been added for better management of investigator-initiated trials.

Although the introduction of commercial or open-source systems can be considered at academic medical centers, many existing studies have noted limitations in using these systems at this type of medical center [6,9-11,19,20]. The biggest problem with

commercial or open source CTMSs is that it is difficult to customize a fixed workflow [10,19,20]. These problems, therefore, make it difficult for such a CTMS to manage a variety of clinical trials, which is often necessary at academic medical centers, or clinical trials involving many stakeholders simultaneously. To address these burdens of covering a wide scope of work, we designed and constructed a site-level CTMS according to the following four important aspects. First, we established an agreement among the stakeholders involved in clinical trials. For this, we formed a CTMS task force with various stakeholders participating in clinical trials for 14 months, a period during which we covered all phases of CTMS design, implementation, and deployment. During the 14 months of the task force operation, interviews on various departments were conducted to organize and integrate the requirements for designing CTMS.

Second, a standardized clinical trial process was established and the unified terminology at our academic medical center was listed. As mentioned in the Methods section, different types of clinical trials were performed at the site with different study phases and therapeutic areas. There were also some cases in which only a part of the clinical trial was conducted according to the contract. To arrange these heterogeneous tasks in clinical trials, institutional standardization of clinical trial processes and terminology was defined.

Third, we built strong interfaces between the CTMS and legacy clinical trial support systems such as HIS, e-IRB, ERP, and biomaterial management system. On the system side, the most important aspect of site-level CTMS is its seamless operation with the legacy systems. We conducted a thorough review with the hospital's IT development team to determine which data would be linked at any given time and whether there were any potential privacy or security issues with the associated data.

Lastly, the CTMS was designed in accordance with domestic and international security and privacy protection regulations for compliance perspectives. As an operating system that deals with sensitive patient information, the system was designed based on various laws and regulations in Korea and abroad, and it has been verified as secure by the hospital's IT development team.

Strengths and Limitations

The primary advantage of utilizing a CTMS at an academic medical center is that it makes it easier to track clinical trial progression and capture the whole process of the clinical trial with minimal human efforts using a standardized IT system. As clinical trials are performed by multiple stakeholders, it is crucial to have all stakeholders consistently updated of any progress and to be on the same page with every decision made. The conventional way of reporting clinical trial progress to stakeholders often resulted in miscommunications as stakeholders delivered messages and updates via their preferred way of communication. This could range from phone calls to emails, or from fax to hard copies of letters. This mode of communication lends itself to the omission of important stakeholders from the loop of communication, and various stakeholders could often miss data if important information is not communicated in a timely manner. To address these issues, many pharmaceutical companies have actively tried to build standardized communication tools by recruiting support from their IT development teams. In this regard, CTMSs have proven to be an effective tool for managing clinical trials. Principal investigators who often simultaneously conduct multiple clinical trials received reports on the progress of all ongoing clinical trials on a real-time basis and were able to effectively manage their trials that involved dozens of responding staffs, such as sub-investigators, CRCs, CRAs, pharmacists, and administrative staffs, as they make use of the CTMS as a single consolidated database. As principal investigators could follow the study progress on a real-time basis, it became easier to promptly address any issues identified during the course of clinical trials.

Moreover, the CTMS significantly assisted for the leadership in academic medical centers because the CTMS tracks the performance of multi-stakeholders and extracts data that reflects utilization of resources, holding issues, bottleneck steps, and so on. Instead of receiving dispersed reports from each individual function, it is now possible for the leadership of academic medical centers to track the performance of each study at any time, to resolve any issues, and to make informed management decisions.

One of the limitations of this study is that the feasibility of the CTMS development is only confirmed by a single academic medical center. Although the volume of clinical trials is large and the CTMS covers different types of clinical trials that reflect different standard operating procedures of various pharmaceutical companies, its data were extracted only from a single-frame platform. Thus, data from academic medical centers of varying institution size and heterogeneous IT infrastructures need to be investigated in a future study. Recently, the FDA has created a Clinical Trial Transformation Initiative to improve the quality and efficiency of clinical trials [21]. This initiative is underway for safety reporting and data monitoring committee project to improve clinical trial safety. Further research is needed to improve the CTMS to flexibly reflect these regulatory policy changes.

Conclusions

As the academic medical center has a responsibility and obligation to conduct and manage clinical trials while maintaining a sufficiently high level of quality, it is necessary to build an IT system for supporting standardized clinical trial processes and to comply with relevant regulations. In this paper, we propose the methods, considerations, and recommendations for development and utilization of a CTMS as a consolidated research database in an academic medical center. We have outlined the benefits of adopting a CTMS based on specific scenarios, but further studies on efficiency and accuracy based on data are needed.

Acknowledgments

This study was supported by a grant from the Korean Health Technology R&D Project, Ministry of Health & Welfare (no. HI14C1061), the Institute for Information & communications Technology Promotion (IITP) project, Korea government (MSIT; No. 2017M3A9B6061832) and the Korean Health Technology R&D Project, Ministry of Health & Welfare (no. HI17C1862). We thank Dr Joon Seo Lim from the Scientific Publications Team at Asan Medical Center for his editorial assistance in preparing this manuscript.

Conflicts of Interest

None declared

Multimedia Appendix 1

Screenshot of study design on the Asan Medical Center (AMC) Clinical Trial Management System (CTMS). Sensitive items in terms of study information and privacy protection were masked. The original webpage shows only the Korean menu names. They were translated into English for international users.

[JPG File, 435KB - jmir_v20i4e103_app1.jpg]

Multimedia Appendix 2

Screenshot of calendar of subject scheduling in the Asan Medical Center (AMC) Clinical Trial Management System (CTMS). Sensitive items in terms of study information and privacy protection were masked. The original webpage shows only the Korean menu names. They were translated into English for international users.

[JPG File, 508KB - jmir_v20i4e103_app2.jpg]

Multimedia Appendix 3

Screenshot of drug management in Asan Medical Center (AMC) Clinical Trial Management System (CTMS). Sensitive items in terms of study information and privacy protection were masked. The original webpage shows only the Korean menu names. They were translated into English for international users.

[JPG File, 422KB - jmir_v20i4e103_app3.jpg]

References

1. Lu Z, Su J. Clinical data management: Current status, challenges, and future directions from industry perspectives. Open Access J Clin Trials 2010;2:93-105.
2. Lu Z. Information technology in pharmacovigilance: Benefits, challenges, and future directions from industry perspectives. Drug Healthc Patient Saf 2009;1:35-45 [FREE Full text] [Medline: 21701609]
3. Dilts DM, Sandler AB. Invisible barriers to clinical trials: the impact of structural, infrastructural, and procedural barriers to opening oncology clinical trials. J Clin Oncol 2006 Oct 01;24(28):4545-4552. [doi: 10.1200/JCO.2005.05.0104] [Medline: 17008693]
4. Rubin E, Lazar D. AAHCDC. 2009. Clinical Trials Offices: What's New in Research Administration URL: http://www.aahcdc.org/Publications-Resources/Series/Issue-Briefs/View/ArticleId/15955/Clinical-Trials-Offices-What-s-New-In-Research-Administration [accessed 2018-02-22] [WebCite Cache ID 6xOy8goDS]
5. Pihlstrom BL, Barnett ML. Design, operation, and interpretation of clinical trials. J Dent Res 2010 Aug;89(8):759-772. [doi: 10.1177/0022034510374737] [Medline: 20581353]
6. Fegan GW, Lang TA. Could an open-source clinical trial data-management system be what we have all been looking for? PLoS Med 2008 Mar 04;5(3):e6 [FREE Full text] [doi: 10.1371/journal.pmed.0050006] [Medline: 18318594]
7. Delawi D, Dhert WJA, Oner FC. Conducting a European multi-center trial: first experiences with the new EU Clinical Trials Directive from an academic perspective. Eur Spine J 2008 Aug;17(8):1113-1115 [FREE Full text] [doi: 10.1007/s00586-008-0699-5] [Medline: 18551324]
8. Grimes DA, Hubacher D, Nanda K, Schulz KF, Moher D, Altman DG. The Good Clinical Practice guideline: a bronze standard for clinical research. Lancet 2005;366(9480):172-174. [doi: 10.1016/S0140-6736(05)66875-4] [Medline: 16005342]
9. Raptis DA, Mettler T, Fischer MA, Patak M, Lesurtel M, Eshmuminov D, et al. Managing multicentre clinical trials with open source. Inform Health Soc Care 2014 Mar;39(2):67-80. [doi: 10.3109/17538157.2013.812647] [Medline: 24517456]
10. Leroux H, McBride S, Gibson S. On selecting a clinical trial management system for large scale, multi-centre, multi-modal clinical research study. Stud Health Technol Inform 2011;168:89-95. [Medline: 21893916]

11. Zhao W, Pauls K. Architecture design of a generic centralized adjudication module integrated in a web-based clinical trial management system. Clin Trials 2016 Apr;13(2):223-233 [FREE Full text] [doi: 10.1177/1740774515611889] [Medline: 26464429]
12. DFHCC. OnCore Information URL: http://www.dfhcc.harvard.edu/research/clinical-research-support/clinical-trials-research-informatics-office/services-support/oncore-information/ [accessed 2018-02-22] [WebCite Cache ID 6xOyhfkfO]
13. Clinical Data Interchange Standards Consortium. 2017 Dec 22. Controlled Terminology URL: https://www.cdisc.org/standards/semantics/terminology [accessed 2018-02-22] [WebCite Cache ID 6xOz0hVqW]
14. Electronic Code of Federal Regulations. 2018 Feb 23. URL: https://www.ecfr.gov/cgi-bin/text-idx?SID=3ee286332416f26a91d9e6d786a604ab&mc=true&tpl=/ecfrbrowse/Title21/21tab_02.tpl [accessed 2018-06-07]
15. Therapeutic Goods Administration. 2018 Feb 05. ICH Guideline for Good Clinical Practice URL: https://www.tga.gov.au/publication/note-guidance-good-clinical-practice [accessed 2018-02-21] [WebCite Cache ID 6xOzd9Scp]
16. Shin S, Lyu Y, Shin Y, Choi HJ, Park J, Kim W, et al. Lessons Learned from Development of De-identification System for Biomedical Research in a Korean Tertiary Hospital. Healthc Inform Res 2013 Jun;19(2):102-109 [FREE Full text] [doi: 10.4258/hir.2013.19.2.102] [Medline: 23882415]
17. Shin S, Park YR, Shin Y, Choi HJ, Park J, Lyu Y, et al. A De-identification method for bilingual clinical texts of various note types. J Korean Med Sci 2015 Jan;30(1):7-15 [FREE Full text] [doi: 10.3346/jkms.2015.30.1.7] [Medline: 25552878]
18. Croghan IT, Viker SD, Limper AH, Evans TK, Cornell AR, Ebbert JO, et al. Developing a clinical trial unit to advance research in an academic institution. Contemp Clin Trials 2015 Nov;45(Pt B):270-276 [FREE Full text] [doi: 10.1016/j.cct.2015.10.001] [Medline: 26454064]
19. Shah J, Rajgor D, Pradhan S, McCready M, Zaveri A, Pietrobon R. Electronic data capture for registries and clinical trials in orthopaedic surgery: open source versus commercial systems. Clin Orthop Relat Res 2010 Oct;468(10):2664-2671 [FREE Full text] [doi: 10.1007/s11999-010-1469-3] [Medline: 20635174]
20. Choi B, Drozdetski S, Hackett M, Lu C, Rottenberg C, Yu L, et al. Usability comparison of three clinical trial management systems. AMIA Annu Symp Proc 2005:921 [FREE Full text] [Medline: 16779208]
21. Clinical Trials Transformation Initiative. 2017. URL: https://www.ctti-clinicaltrials.org/ [accessed 2018-02-22] [WebCite Cache ID 6xQNBX5W3]

Abbreviations

ARO: academic research organization
AMC: Asan Medical Center
CDISC: Clinical Data Interchange Standards Consortium
CRA: clinical research associate
CRC: clinical research coordinator
CRO: clinical research organization
CTC: clinical trial center
CTMS: clinical trial management system
e-IRB: electronic institutional review board
ERP: enterprise resource planning
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
HIS: health information system
LDS: Limited Data Set
TGA: Therapeutic Goods Administration
UI: user interface
UX: user experience

Edited by G Eysenbach; submitted 31.10.17; peer-reviewed by S Tao, SY Shin, C Chen; comments to author 06.12.17; revised version received 19.12.17; accepted 22.01.18; published 24.04.18.

Please cite as:
Park YR, Yoon YJ, Koo H, Yoo S, Choi CM, Beck SH, Kim TW
Utilization of a Clinical Trial Management System for the Whole Clinical Trial Process as an Integrated Database: System Development
J Med Internet Res 2018;20(4):e103
URL: http://www.jmir.org/2018/4/e103/
doi:10.2196/jmir.9312
PMID:29691212

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org