Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 21 (2019), Issue 2 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Editorial

Reviews

Viewpoints

Original Papers

Corrigenda and Addendas

Editorial

Digital Trespass: Ethical and Terms-of-Use Violations by Researchers Accessing Data From an Online Patient Community

Emil Chiauzzi[1], PhD; Paul Wicks[1], PhD

PatientsLikeMe, Inc, Cambridge, MA, United States

Corresponding Author:
Emil Chiauzzi, PhD
PatientsLikeMe, Inc
160 Second Street
Cambridge, MA, 02142
United States
Phone: 1 617 499 4003
Email: echiauzzi@outlook.com

Abstract

With the expansion and popularity of research on websites such as Facebook and Twitter, there has been increasing concern about investigator conduct and social media ethics. The availability of large data sets has attracted researchers who are not traditionally associated with health data and its associated ethical considerations, such as computer and data scientists. Reliance on oversight by ethics review boards is inadequate and, due to the public availability of social media data, there is often confusion between public and private spaces. In addition, social media participants and researchers may pay little attention to traditional terms of use. In this paper, we review four cases involving ethical and terms-of-use violations by researchers seeking to conduct social media studies in an online patient research network. These violations involved unauthorized scraping of social media data, entry of false information, misrepresentation of researcher identities of participants on forums, lack of ethical approval and informed consent, use of member quotations, and presentation of findings at conferences and in journals without verifying accurate potential biases and limitations of the data. The correction of these ethical lapses often involves much effort in detecting and responding to violators, addressing these lapses with members of an online community, and correcting inaccuracies in the literature (including retraction of publications and conference presentations). Despite these corrective actions, we do not regard these episodes solely as violations. Instead, they represent broader ethical issues that may arise from potential sources of confusion, misinformation, inadequacies in applying traditional informed consent procedures to social media research, and differences in ethics training and scientific methodology across research disciplines. Social media research stakeholders need to assure participants that their studies will not compromise anonymity or lead to harmful outcomes while preserving the societal value of their health-related studies. Based on our experience and published recommendations by social media researchers, we offer potential directions for future prevention-oriented measures that can be applied by data producers, computer/data scientists, institutional review boards, research ethics committees, and publishers.

(J Med Internet Res 2019;21(2):e11985) doi:10.2196/11985

KEYWORDS

ethical issues; social media; data sharing; privacy; informed consent; data protection; data anonymization

Introduction

According to the Pew Research Center [1], the majority of Americans use social media websites such as Facebook (68%) and YouTube (75%), with roughly a quarter to one-third using other sites such as Snapchat, Instagram, LinkedIn, and Twitter. The sheer volume of data arising has proved to be an inviting target for both social good and ethically questionable practices alike. Social media data have driven important public health research, including monitoring disease outbreaks [2], predicting health risk behaviors [3], accessing hard-to-reach populations [4], health promotion [5], user health-communication patterns [6], and mutual medical data sharing between patients [7]. Some researchers have adopted a more participatory approach by engaging high-risk groups such as drug users to detect trends and encourage harm reduction [8]. The analysis of these data has ushered in a variety of innovative analytic techniques such as natural language processing, network analysis, deep learning, and geolocation to provide further insight into these large datasets [9].

With such a rapidly evolving landscape, this area has been no stranger to ethical controversy [10,11]. Ethical questions have

arisen in highly publicized cases such as the Facebook social contagion study [12,13], the release of an OKCupid dataset of 70,000 users [14], and most recently, the use of 50 million user profiles on Facebook by Cambridge Analytica during the 2016 US presidential campaign [15]. In each of these cases, large quantities of user profile data compromised user privacy or manipulated users through targeted messaging.

Academic reviews suggest that there is "widespread neglect" of ethical considerations by social media researchers [16], such as inadequate informed consent, lack of researcher boundaries, reposting of personally identifiable content, and deliberate misrepresentation or deception [16] [17,18,19]. A recent study found that online searches of verbatim Twitter quotes found in journal articles can be tracked back to individual users 84% of the time [17], despite users' lack of awareness of this sharing, resistance to being studied, and desire to consent to these practices [18,19]. Some researchers misrepresent themselves or engage in deception to engage with social media participants [20]. Many researchers assume that social media data are in the public domain, obviating the need for consent altogether [21].

There may be several reasons for these challenges. First, researchers conducting studies in the United States may believe that approval by institutional review boards (IRBs) is sufficient for addressing ethical considerations. A recent review of 156 academic studies mining social media found that ethical considerations were limited only to minimum requirements for IRB approval, rather than broader ethical considerations (eg, privacy, public and private spaces, and original contexts for providing data) [16]. Only 13 of 156 (8%) studies mentioned ethical considerations beyond IRB approval. Researchers relying on the US Department of Health and Human Services (USDHHS) Common Rule guideline to bypass informed consent when research does not involve an intervention or uses "existing data sets" inadequately address these ethical concerns [22], which may be further amplified in stigmatized conditions such as mental illness [23]. Second, IRB members may lack consensus among themselves on the IRB review process in social computing research, the need for informed consent, their own regulatory obligations, and criteria for evaluating social media projects on a case-by-case basis [24]. Third, considering social media users as just another class of traditional "human subjects" misses the mark. Critics also suggest that traditional definitions of terms such as "human subject," "informed consent," privacy, ownership of data, terms of use, and private and public settings are too narrow for online contexts [21]. Finally, current regulations such as the USDHHS Common Rule emphasize risk mitigation at the initial stages of research—study design and data collection—rather than at later stages that involve access to and dissemination of data [25]. Informed consent collected at a single point in time may not account for the "drift" that occurs in a participant's willingness to share data [26].

Outside of the United States, there are a wide variety of national research ethics governing bodies and over 1000 laws, regulations, and standards that provide oversight for human subjects research in 130 countries [27]. The rigor of ethical review varies widely across countries. In Europe, ethics review is generally stringent and managed through national bioethics agencies, health ministries, food and drug safety organizations, national research committees, etc [27]. Ethical review processes in countries such as China are less well developed, with a lack of standardization in operating procedures, professional ethics training, protection of vulnerable groups, and privacy safeguards [28]. In both of these scenarios, issues of privacy, data trustworthiness, and consent have yet to be resolved, even with the advent of the European Union General Data Protection Regulation (GDPR) [29]. Research ethics committees (RECs) often lack the expertise to evaluate technical standards, methodologies, data ownership, and group-level ethical harms in big data studies [30]. Taken together, these issues suggest that international ethical review frameworks continue to be highly challenged by the current dynamic social media research environment.

Second, accessing and de-identifying social media data is not difficult. Data transgressions can be enabled by the ready availability of user data combined with the dissemination of "scraping" technologies that allow easy extraction [31]. Data scraping and de-anonymizing can be accomplished by individuals with no more than basic programming and statistics skills [32]. Unfortunately, privacy has been considered a "binary value"—either public or private [33]—rather than a continuum [34]. While some researchers assume that information shared in public spaces is inherently available for public consumption and may be used without consent, it is important to respect the nature of the data, collection context, and user expectations [33]. Identifiability should not be regarded as a binary value (either "public" or "private"), but as a continuum based on the nature and extent of the data [33]. Attempts at de-identification are a necessary but insufficient to ensure safe use of data [34], with some researchers warning that true de-identification is a "false promise" [35]. Re-identification has been accomplished with relatively limited data available such as Netflix subscriber movie ratings [32] or simple demographics [36].

Third, the perception that big data are somehow "objective" and can be analyzed independent of context is an illusion [37,38,39]. Social media users post information for reasons differing widely from what researchers may imagine. For example, within the PatientsLikeMe platform [40], patients adopt a broader definition of "treatments" than clinicians and researchers. For patients, treatments may include "pets" and "handicapped parking stickers" just as much as medications, medical procedures, and therapies. Faulty data assumptions and researcher biases may cascade into poorly built algorithms that lead to ultimate inaccurate (and possible harmful) conclusions, termed by O'Neil [41] as "weapons of math destruction." It is important not to dissociate the data from the *people* behind them [33]. Even when aggregate data are used and no individual identification has been made, researchers need to be sensitive to the potential psychological and behavioral consequences of findings (particularly with stigmatized or vulnerable groups) as well as the scale and generalizability of conclusions [23,42]. There is a risk of type I error when findings are overgeneralized [43], thus requiring more mixed methods and longitudinal data gathering [42].

Fourth, health research has traditionally been conducted by researchers trained in human subject ethics and overseen by

established ethics panels. However, the recent growth of "big data" sets in health has attracted computer science researchers who may be less well versed or monitored with regard to key ethical issues [10]. Wright [44] warns that many computer scientists are skirting the ethical traditions of medical and social science professionals, who abide by guidelines such as the Belmont Report [45] and the USDHHS Common Rule [46]. Buchanan et al [47] suggest that computer science researchers "may not fully understand or believe that their projects align with the same ethical concerns that pertain to human subjects, such as the minimization of risk or harm to individuals, confidentiality, privacy, or just recruitment methods."

Several questions arise in this context. How do these ethical violations occur? How are these violations discovered and remedied by data producers? Most importantly, what corrective actions can and should be taken to prevent violations that compromise the privacy of social media users? In order to address these questions, we share four cases involving ethical and terms-of-use violations that highlight the four challenges described above. These violations involved the use, interpretation/misinterpretation, and dissemination of patient self-reported data and forum posts available at PatientsLikeMe [40]. In this manuscript, our goal is to utilize these cases as a springboard to protect patient privacy while finding ways of meeting investigators' legitimate public health research objectives.

Case Studies: Real-World Experiences From an Online Health Community

The following four cases provide examples of ethical and methodological issues that arise when researchers gather social media data without observing the website's terms of use. These cases have been selected as representatives of the breadth of issues encountered over 12 years (typically, at a rate of one or two per year) at PatientsLikeMe, an online patient community devoted to research (Textbox 1).

Each of the cases illustrates a different set of ethical problems. We have applied the health-related research ethics guidelines created by the Council for International Organizations of Medical Sciences (CIOMS) in conjunction with the World Health Organization [48] as a primary framework for these cases.

For the purposes of this paper, we will distinguish the ethical violations from terms-of-use violations, which represent a lack of adherence to website-specific policies or approval to conduct research-related activities. The lack of attention to terms of use by prospective users accessing various websites and apps has been well documented [49] and should be distinguished from the ethical violations noted earlier. Terms-of-use violations may include participation in ways that do not conform to the purpose of the forum, posting false content, unauthorized scraping of data, or a lack of authorization to conduct research by the data producer. There is certainly potential for these concepts to overlap, particularly on websites that involve the sharing of personal health information. Table 1 describes the types of violations as well as CIOMS guidelines that apply to the cases in this manuscript.

Because we aimed to remain transparent, we emailed the prepublication manuscript to the researchers represented in Cases 2-4 below (Case 1 has already been publicized in the national press). After providing 1 month for responses but receiving none, we moved forward with the final manuscript. We have not named specific researchers or papers in Cases 2-4 in order to preserve their anonymity.

Case 1: Large-Scale Data Scraping by Commercial Market Researchers

Background

In a well-publicized 2009 incident reported in the Wall Street Journal [50], staff at the company Nielsen Media sought to understand how patients with mental health conditions talked about the company. The company created an unauthorized account on PatientsLikeMe and used automated "scraper" software to begin copying open-text discussion data from the message board forums. In total, they harvested about 5% of the mood disorder forum's qualitative discussion content for an undisclosed commercial client. Our team detected the scraping software, suspended the account (and three others linked to it) shortly after it was initiated, and emailed the company to ask them to stop.

Relevant Terms of Use and Ethical Guidelines

Because this was considered "market research," no IRB was involved. For market researchers, the level of ethical oversight is not the same as that for academic researchers in most studies. However, professional bodies such as the Market Research Association state that members should "Protect the rights of respondents, including the right to refuse to participate in part or all of the research process," among other guidelines [54]. Market researchers may need to develop their own standards related to health-data gathered online or endorse existing guidelines. For example, the Association of Internet Researchers recommends that researchers obtain consent from either participants individually or community owners [21]. Harvesting sensitive data from people with mental health issues also warrants consideration of vulnerable populations; without proper procedures in place to ensure data were handled correctly, there is a risk of re-identification. Scraping only the visible data (as opposed to accessing a full dataset) risks drawing spurious or biased conclusions.

Response

We emailed the company with a cease-and-desist letter. PatientsLikeMe sent a private message to its entire membership describing the incident and wrote a blog post about it. As a result, about 200 members decided to close their accounts. Six months later, reporters at the Wall Street Journal investigated the story as part of a series looking at scraping activity on the Web, and the incident was reported on the newspaper's front page [50].

Textbox 1. Description of PatientsLikeMe.

PatientsLikeMe is an online community of over 600,000 people living with about 2900 medical conditions including amyotrophic lateral sclerosis, mood disorders, HIV, and rare diseases [51]. As part of the membership, individuals who are interested in joining the site are asked to review our user agreement [52] and privacy policy [53]. The user agreement describes acceptable lawful use, inappropriate posting practices, and restriction of content use within the site. The privacy policy provides clear and transparent communication about data as well as rights to see, correct, and delete data; get notified if data are stolen; and request that data processing stop.

The PatientsLikeMe privacy policy is written in plain language and allows patients to review, correct, or delete their data. Patients may self-report their conditions, treatments, symptoms, and patient-reported outcome measures (reporting as much or as little as they like) and are able to look at aggregated reports to help decide how they might better manage their condition. Most individual "profiles" are only viewable to other members, who must "log in" to the site after registering with an email address, while some aggregate data reports are viewable from the "logged out" part of the website. Although patients are comfortable anonymously sharing their data with vetted researchers [29], there are many ways in which an uninformed external researcher could misinterpret the way data are collected or be unaware of known biases that are familiar to our internal researchers. In addition, anyone entering "fake data" can potentially trigger negative consequences for data quality and, potentially, even patient safety.

PatientsLikeMe has adopted this model because patients lack access to information that can affect their treatment decisions. Sharing "real world" data allows patients, providers, and researchers to collaborate in evaluating current treatment effectiveness, gaps in treatment, and potential new and better treatments. This collaboration can speed the pace of research and improve health care delivery. To facilitate this mission, PatientsLikeMe is funded through investment, as well as commercial and academic research partnerships, rather than advertising or member fees. Because of the serious nature of health data, PatientsLikeMe has been committed to applying these data responsibly toward patient-centered goals and implementing a "data for good" philosophy. Responsible big data research seeks soundness and accuracy of data while maximizing good and minimizing harm [33].

Table 1. Case violations and Council for International Organizations of Medical Sciences guidelines.

Violation type	Case 1 - Commercial scraping	Case 2 - De-anonymization of forum user	Case 3 - Fake profile data	Case 4 - Multiple scraper bots	Relevant CIOMS[a] guideline number
PLM[b] terms-of-use violations					
Not a patient, caregiver, health care professional, or visitor with legitimate reasons to participate[c]	✓	✓	✓	✓	7, 22
Posting false content[d]			✓		4, 11
Use of any robot, spider, scraper, or other automated means to access the site or content[e]	✓	✓		✓	7, 12, 22
Lack of research authorization by PLM[f]	✓	✓	✓	✓	7, 8, 9, 10, 22, 25
Ethical violations					
De-identifying patient data in any way		✓			4, 11, 14, 15, 22
Inadequate/no informed consent	✓	✓		✓	9, 10, 12, 22
False identification or misrepresentation			✓	✓	4, 22
Verbatim use of user posts		✓			4, 11, 12, 14, 15, 22

[a]CIOMS: Council for International Organizations of Medical Sciences.

[b]PLM: PatientsLikeMe.

[c]PLM user agreement: “To become a member and access the area on this Site reserved for members (the ‘Member Area’), PatientsLikeMe requires that you are either a (a) diagnosed patient of the particular community you are joining or a parent or legal guardian acting for such a patient who is under 18 years of age or incapacitated, (b) caregiver for a patient eligible to join such community, (c) health care professional (e.g. doctor, nurse, health researcher, etc.), (d) guest with legitimate, non-commercial reasons to participate in the community and who agrees to respect the privacy and preserve the dignity of all community participants or (e) guest as authorized by a PatientsLikeMe member or employee.”

[d]PLM user agreement: “Members shall not post or upload any information or other content on the Site that (a) is false, inaccurate or misleading; (b) is obscene or indecent; (c) infringes any copyright, patent, trademark, trade secret or other proprietary rights or rights of publicity or privacy of any party; or (d) is defamatory, libelous, threatening, abusive, hateful, or contains pornography.”

[e]PLM user agreement: “You may not use any robot, spider, scraper, or other automated means to access the Site or content or services provided on the Site for any purposes.”

[f]PLM user agreement: “Please note that under our terms of service, you are not permitted to capture or utilize data from within the site nor to solicit members through our forums or private message to take part in your study.”

Resolution

In the Wall Street Journal article, a company representative stated, **“**It was a bad legacy practice that we don't do anymore...It’s something that we decided is not acceptable, and we stopped.” Corrective efforts included upgrading our automated scraper-detection software, clarifying how commercial researchers could contact PatientsLikeMe for authorization, determining which actions are permissible and not permissible on the site, and sustaining communication with our members about the implications for their data and further participation on the site.

Case 2: De-anonymization of Individual Forum Members

Background

Around 2014, computer science researchers at a European university developed an algorithm that could be used to de-identify highly sensitive medical data, which individuals might choose to share on social networks in order to reduce their risk of personal identification. The system involved automated methods for determining the “identifying information content” of a given piece of data (ie, “I’m a woman living with a mental health condition for the past two years” vs “my name is Susan and I was diagnosed with bipolar disorder in Boston on June 2, 2016”). In order to illustrate their approach, they provided in their manuscript a verbatim text quote from a member discussing how they came to be diagnosed with HIV. The authors published their study, whereupon a Google Scholar Alert notified us that the research had taken place.

Relevant Terms of Use and Ethical Guidelines

No formal ethics review was conducted, which may have contributed to the oversight. In terms of accessibility, while this story was “shared online,” it was on a private profile accessible only to other patients logged into the site. Searching for the

verbatim text within the logged-in area of PatientsLikeMe quickly identified the member concerned. Although de-identification is never foolproof (and indeed, this was the point of the study itself), if the patient had decided to change his/her mind and delete the data or close their PatientsLikeMe account, the quote and the patient's association with it could have persisted permanently within the scientific literature. CIOMS guideline 22 on the use of data obtained from the online environment states, "When researchers use the online environment and digital tools to obtain data for health-related research they should use privacy-protective measures to protect individuals from the possibility that their personal information is directly revealed or otherwise inferred when datasets are published, shared, combined or linked [55]." Additional considerations should have been given, as HIV is a highly stigmatized condition.

Response

During other similar incidents in the past, reaching out solely to the authors or their institutions often failed to yield a response. As a result, we emailed the authors and the journal editor with our concerns to ensure this issue would be dealt with appropriately.

Resolution

As no specific patient data were mentioned in the papers, no data were scraped from the site. The focus was a theoretical algorithm, and all parties quickly realized their error. A partial retraction was agreed upon to replace the verbatim quote with a synthetic quote. PatientsLikeMe notified the member concerned. Although CIOMS guideline 22 speaks to research in the online environment, the guidance is general instead of recommendations for best practices for every platform. More specific advice for preventing risk to patients can be found from NatCen's Social Research guidance, which recommends "(Test) the traceability of a tweet or post and (take) responsible steps to inform the user and protect their identity, if desired. Best practices include paraphrasing instead of verbatim quotes and not using an individual's handle/user name."

Case 3: Researcher Misrepresentation and Fake Profile Data

Background

Researchers at a European university secured a grant to investigate the extent to which users of social networks thoroughly read and consider the "terms of use" of social networks like PatientsLikeMe. To test this in controlled conditions, 20 students were asked to register accounts on PatientsLikeMe and complete fake data from a prespecified set of instructions. Focus groups held with the students later revealed that most of them had not read the terms of use. The authors published their study, whereupon a Google Scholar Alert notified us that the research had taken place 10 months before. Both grant funding and REC approval were sought and granted for this study, despite the lack of a "letter of support" from PatientsLikeMe as a potential collaborator.

Relevant Terms of Use and Ethical Guidelines

Deceptive practices such as researchers misleading participants about their identity are never acceptable, and we were surprised that an REC had approved such activity. In our case, researchers prompted students to enter fake data into a system requiring log-in, which is used by patients, regulators, and health care professionals to guide practice and conduct medical research. CIOMS guideline 1 states, "Although scientific and social value are the fundamental justification for undertaking research, researchers, sponsors, research ethics committees and health authorities have a moral obligation to ensure that all research is carried out in ways that uphold human rights, and respect, protect, and are fair to study participants and the communities in which the research is conducted. Scientific and social value cannot legitimate subjecting study participants or host communities to mistreatment, or injustice."

Response

We emailed the authors, REC, and funders with our concerns. The researchers stated that they did not think they needed permission for a "publicly available forum" and emphasized that the focus of their research was not medical but informational, focusing on the "terms of use" rather than the data of the PatientsLikeMe members themselves. A number of discussions and arguments had to be put forward to explain to the researchers *why* this behavior was wrong in the first place; one analogy we used was that while students *could* pretend to be sick patients in a hospital waiting room in order to conduct research on the clarity of signage within the institution, this would quickly be understood as unethical.

The researchers thought their activities were "outside the logged-in" parts of the site (which they were not) and that students had never re-accessed their accounts after the initial study (which they had). The REC agreed that entering false data was suboptimal behavior, admitted to confusion around some of the complex technical issues surrounding online research, and agreed this was an area they would learn more about in future. The funding body claimed that as the institution had its own REC, they had no further responsibility to check that the permissions were in place.

Resolution

As no specific patient data were mentioned in the papers, no data were scraped from the site, and the focus was indeed the understanding of the "terms of use." We agreed with the authors that a partial retraction in rescinding the name of PatientsLikeMe from their papers would be enough, along with assurances that this would not happen again. We also agreed that our terms and conditions could benefit from clarification. Between this experience and the recent enactment of the European GDPR, work is underway currently to clarify patients' rights in terms of privacy and access to their data and to make explicitly clear that just because patients share their data within the community, it does not grant researchers the right to use the data. When making a determination of whether a community is public or private, the researcher should consider the availability of information to the general public, member perceptions of

privacy, sensitivity of content, record permanence, and the intended audience of the study [20].

Case 4: Repeated Scraping Through Multiple Accounts

Background

Computer science researchers at an Asian university sought to build a neural network capable of determining whether side effects that members were attributing to a treatment they were taking might, in fact, be symptoms of their condition; for example, "trouble sleeping" might be caused by their *depression* rather than *a drug* they were taking. In order to gather test data, they created an account on PatientsLikeMe and began "scraping" data from patient profiles with automated software. When our security systems were tripped by the software activity, their account was closed. Over the following 2 weeks, multiple, seemingly related, accounts were created, many from "disposable" email accounts, in order to continue scraping, which were closed as soon as we identified them. With data from over 5000 users, they prepared a manuscript for a computer science conference to be presented a year later, comparing the reported experience of patients to a third-party data source and describing their algorithm. The authors published a preconference proceeding, whereupon a Google Scholar Alert notified us that the research had taken place 10 months before.

Relevant Terms of Use and Ethical Guidelines

Multiple CIOMS guidelines appeared to be breached, including respect for rights (guideline 1, no permission or consent was requested), balancing individual risks and benefits to participants (guideline 4, no steps were taken to minimize harm to patients), community engagement (guideline 7, the data were gathered covertly), consent (guidelines 9 and 10, no consent was requested or exempted), use of health data (guideline 12, patients' response to treatment was scraped and analyzed), vulnerable persons (guideline 15, the focus included members with severe mental health issues), online environment (guideline 22, the researchers did not inform the community), and ethics committee review (guideline 23, this work did not undergo formal ethics review). The researchers did not appreciate that using a logged-in account was crossing a boundary nor that active shut down of their accounts by our security team was a "no entry" signal. In our discussion, the researchers appeared to feel that because the emphasis of their research was neural networks, they were "far" from medical data. More traditional medical researchers would have had to undergo quite considerable ethical oversight, consent, and data privacy policies to access similar data from a hospital or insurer. Building systems that used such algorithms to judge the soundness of a patient report risked diminishing the fidelity of patients' lived experiences; many, if not most, patient experiences with disease and treatments cannot be found in medical texts, and few medical researchers would assume that divergence meant that the patients were automatically "wrong."

Response

We emailed the authors, conference chairs, and chair of their department with our concerns and requested full retraction of the paper, identification of all scraper accounts, and deletion of all data. The researchers stated that they had only accessed "public" parts of the site, denied having used multiple scraper accounts, said that the data had been held securely, and requested they be allowed to anonymize the data source. In mitigation, they claimed that the paper had received positive peer reviews from the community. Initially, the conference chairs were against retraction based on their judgement that no "material harm" had been done to PatientsLikeMe, that scraping the data was technically easy for a researcher to perform, and that it was unclear whether any laws had been violated. However, further careful investigation by our security team revealed that over 50 "bot" accounts were created from the same rather narrow geographical region during a time period consistent with the conduct of the methods detailed in the paper. On further discussion, the authors admitted that "maybe" an intern had done this. However, scientific record keeping was lacking, as no systematic records had been kept to verify this.

Resolution

The authors apologized and deleted all locally held data. The conference chairs accordingly decided that the authors had not been truthful, and therefore, the study was retracted from the conference proceedings. PatientsLikeMe notified the members concerned. Because the authors were not forthcoming about their activities, our security team had to exhaust significant resources in determining which accounts were bots and which users' data had been accessed, and in refuting the authors' claim. In addition, significant management resources were consumed communicating with the authors and other parties, and communication resources were used in messaging the affected users.

Discussion

Overview

These case studies highlight the broad challenges that arise when researchers gather social media data without prior authorization. The current literature on social media ethics emphasizes issues such as "terms of use," "informed consent," and "data privacy," but the practical implications of these infractions creates ripple effects on patients and the staff responsible for protecting their data. Researchers may gather data to satisfy their scientific goals but should balance these with potential adverse effects on patients, the company affected by the terms of use violation, and the validity of their research enterprise. The lack of informed consent and respect for privacy deprives potential participants of choice regarding the use of their data. Once personal health data have been accessed in an unauthorized or unethical manner, the wide availability of powerful search tools create additional threats to patient privacy.

We believe there are many ways in which the analysis of social media data can contribute to the public good as well as inform individuals about ways to improve and maintain their health. However, the lack of equitable data access, underlying biases in data interpretation, and inadequate transparency between those who provide and those who analyze data risks squanders the many potential advantages of algorithmic decision making [56]. Throughout these cases, we believe that researchers based their treatment of study participants' data on several false assumptions that violated a number of ethical guidelines.

Faulty Assumption 1: "The Internet" Is Not Subject to Ethical Review

Throughout our experiences, we perceived the sense that data (and the "social media users" contributing them) are less worthy of respect or protection when users participate online as opposed to when the same "patients" receive care in a brick-and-mortar health institution like a hospital. To add to the matter, members of ethics review boards may not consider social media studies to be human subject research under current legal definitions and may not believe that data scraping requires informed consent [57]. In our view, social media and "big data" research is not ethically exceptional and should be treated in the same manner as traditional forms of research [57]. Of the cases reported here, only Case 2 obtained ethical approval, and even then, the behaviors exhibited fell short of what we could consider ethical. Terminology may cloud matters, as existing guidelines may confine themselves only to "biomedical" or "medical" fields, which may lead some researchers to exclude their projects from ethical oversight on the basis that their focus or their branch of study is computer science, business, or design. However, CIOMS [48] uses the broader term "health-related research" to encourage greater inclusiveness rather than focusing on researchers' occupation or training. Online contexts should be compared to offline analogues to highlight potential considerations that may affect informed consent; if it was not acceptable to do something in a hospital waiting room, doing it on the internet does not absolve researchers of responsibility. We believe that interpreting the USDHHS Common Rule for "existing data set" as "free access to any health data set on the Internet" is a faulty assumption.

Faulty Assumption 2: Social Media Spaces Are "Public"

In our discussions with individuals involved in the cases reported here, we encountered a lack of cultural sensitivity to the "perceived privacy" of individuals choosing to share information within a "closed network" as opposed to an open forum. It is probably best to take a conservative approach and consider that any content requiring an email for access may not be considered public by a site's users.

Where trespasses were acknowledged, they were claimed to be justified by good intentions. For example, while few would argue in favor of the potentially good intentions of gathering and analyzing social media posts in Case 1 to try and understand mental health problems, such good intentions do not act as blanket absolution from ethical considerations such as consent, privacy, de-identification, or minimization of harms. In the real world, reading and analyzing the diaries or written correspondence of patients with mental health problems would not be deemed acceptable even if they were left unsecured.

Faulty Assumption 3: Data Can Be Analyzed Independent of Context

Although large datasets may appear alluring by their sheer scale, in practice, they can introduce larger errors of interpretation by inspiring false confidence in the conclusions drawn. In Case 4, the researchers were unaware that there was a host of additional contextual data recorded about how patients had multiple comorbidities and understood the purpose of their medications or that they may have been using some treatments for off-label purposes rather than their standard indications [59]. The absence on their team of trained health professionals also obscured important context about the relationship between a condition's symptoms and the common side effects of treatments used for the condition. Without understanding the sampling of a data set, the limits of meaningful questions and interpretations may not be observed [37]. Scientifically, data scraping without context may result in potentially inaccurate algorithms that may get reported and reused in application, leading to potentially harmful consequences [41]. Our discussions with researchers revealed a general lack of care and rigor that would be of scientific concern even without the ethical considerations. We explained the importance of understanding the context and structure of the data that were scraped in order to produce meaningful scientific results and requested a retraction of questionable findings and interpretations to avoid contaminating the literature.

Faulty Assumption 4: Computer Science Research Does Not Need to Abide by Health Research Guidelines When It Is Only Accessing "Data"

While computer science researchers were responsible for only Case 4 reported here, computer science practitioners are responsible for the bulk of our other unreported cases, confirming Wright's [38] assertions that the field needs to adjust its practices before further incidents undermine their social license to practice. Computer scientists are "largely focused on the care and feeding of electronic devices" and may have different conceptions of what constitutes a "human subject"—a living person or data that are representative of a living person [42]. Involving computer scientists on ethics review boards may be an effective way of encouraging ownership of ethics issues from the inside out as well as assuring more technology expertise in medical and other studies. This would also encourage more complete paper trails when untangling ethics transgressions.

Appropriately Resolving Terms-of-Use Violations

We have shared our experiences, in part, to guide other practitioners in the field. Unfortunately, the effects on data reporting may be difficult to detect and may not be caught until publications and conference papers appear. The resolution of the scientific inaccuracies and communications, as well as deletion of scraped data, often required difficult conversations over extended periods. We recommend that data producers develop their own standard operating procedures and hold practice scenarios when responding to violations.

For instance, because substantial time and effort are devoted to research planning, execution, and publication, a recently published or in-process journal article represents a considerable "sunk cost." As a result, researchers, funders, conference organizers, and journal editors may apply pressure to data producers to "allow" publications to proceed with corrections rather than retract findings. Over the course of the cases experienced by our team, nearly every supervisor, institution, conference chair, or publisher challenged in the case of a violation first asked (politely) for clemency, forgiveness, "retrospective consent," or even "post-hoc ethical approval."

Rather than adopt a punitive philosophy, we respectfully reminded these researchers of our responsibilities to patients who are our members and from whom we have earned social license to use and maintain their data responsibly. However, having policies and prepared communications in place early on would reduce the burden on staff members who may find such interactions challenging.

Limitations

Our report contains several limitations. First, the authors are employees of a for-profit company and therefore have a conflict of interest in "protecting" network data. We hope to encourage similar experiences by others in the academic or nonprofit sphere to share their experiences. Second, the cases reported here are relatively brief and due to our desire to preserve anonymity where possible, there is little additional detail for interested readers. Third, as a complex and emerging area, our conclusions are necessarily editorial rather than evidence based. For example, future work could survey social network users whose data have been shared without their consent. Finally, the individuals described herein may not feel they have an adequate "right to reply"; we would welcome divergent views on the topics we have outlined here.

Future Directions: Prevention Rather Than Cure

Based on the need to maximize benefits while limiting potential harm in social media research, we believe that there are several potential strategies that can be pursued. First, rationales such as social media data use for "public benefit" and "public interest" need to be carefully defined [34]. Investigator transparency is as critical in social media research as in traditional forms of research; researchers should disclose their presence, not misrepresent themselves, and be truthful about the risks and benefits of their studies [48,58]. Researchers should exercise extreme caution in adapting or combining data sets for potentially invasive purposes. Common sense strategies such as avoiding the reuse of verbatim quotes should be adopted. It is therefore critical for researchers to understand the nature of the data source that they are accessing. As part of ethics protocol submission, investigators should certify that they are complying with the terms of use of the targeted research websites or justify to ethics review boards why their methodologies fail to comply [57]. In addition, ethics review boards should include members with strong knowledge of online research and computer science methodologies, so that applications for ethical approval can be vetted more carefully [47].

When data have been collected without authorization, there should be standard operating procedures developed and followed with regard to how data obtained without authorization should be managed, deleted, and verified. University information technology departments could take a lead in this regard. Further attention to ethical issues in computer and data science training and conduct may help prevent the violations discussed in this paper while recognizing the value of important research questions. Data producers (such as PatientsLikeMe) and data scientists can enhance each other's work if an appropriate dialogue can take place. Data producers can adopt a proactive stance by finding ways to curate and expand access to views of their data (such as through application programming interfaces), so that important scientific research can be encouraged while minimizing ethical and terms-of-use violations. In order to meet the needs of computer science and other researchers, PatientsLikeMe has started investigating ways to provide tools for researchers to interrogate data sets in order to yield insights with less risk to member privacy.

Such strategies would only be the beginning of addressing social media privacy challenges, but we welcome further enhancement of and feedback on these ideas. A group of data scientists recently reported on a crowdsourced "Hippocratic Oath for Data Science" [60] that calls upon their peers to "Ensure that all data practitioners take responsibility for exercising ethical imagination in their work, including considering the implication of what came before and what may come after, and actively working to increase benefit and prevent harm to others."

Acknowledgments

The authors would like to thank the following reviewers of the manuscript: James Heywood, Benjamin Heywood, Steve Hammond, Greg Ploussios, and John Torous. We would also like to thank the following members of the PatientsLikeMe Ethics & Compliance Advisory Board for their input: Sally Okun (Board Chair), Hans van Delden, Letitia Browne-James, and Gary Rafaloff.

Authors' Contributions

Both authors contributed to the conceptualization, writing, and review of this manuscript.

Conflicts of Interest

Both authors are employed by and own stock options in PatientsLikeMe, Inc. Paul Wicks also is a academic section editor for JMIR Publications.

References

1. Smith A, Anderson M. Pew Research Center - Internet & Technology. Social Media Use in 2018 URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2019-02-14]
2. Herrera JL, Srinivasan R, Brownstein JS, Galvani AP, Meyers LA. Disease Surveillance on Complex Social Networks. PLoS Comput Biol 2016 Dec;12(7):e1004928 [FREE Full text] [doi: 10.1371/journal.pcbi.1004928] [Medline: 27415615]

3. Young SD. Social Media as a New Vital Sign: Commentary. J Med Internet Res 2018 Apr 30;20(4):e161 [FREE Full text] [doi: 10.2196/jmir.8563] [Medline: 29712631]
4. 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]
5. Hudnut-Beumler J, Po'e E, Barkin S. The Use of Social Media for Health Promotion in Hispanic Populations: A Scoping Systematic Review. JMIR Public Health Surveill 2016 Jul 11;2(2):e32 [FREE Full text] [doi: 10.2196/publichealth.5579] [Medline: 27400979]
6. Hswen Y, Naslund JA, Chandrashekar P, Siegel R, Brownstein JS, Hawkins JB. Exploring online communication about cigarette smoking among Twitter users who self-identify as having schizophrenia. Psychiatry Res 2017 Dec;257:479-484 [FREE Full text] [doi: 10.1016/j.psychres.2017.08.002] [Medline: 28841509]
7. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
8. Barratt M, Lenton S. Beyond recruitment? Participatory online research with people who use drugs. International Journal of Internet Research Ethics 2010;3:69-86 [FREE Full text]
9. Yeung D. Social Media as a Catalyst for Policy Action and Social Change for Health and Well-Being: Viewpoint. J Med Internet Res 2018 Mar 19;20(3):e94 [FREE Full text] [doi: 10.2196/jmir.8508] [Medline: 29555624]
10. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [Medline: 11701577]
11. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. Am Psychol 2004;59(2):105-117. [doi: 10.1037/0003-066X.59.2.105] [Medline: 14992637]
12. Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Dec 17;111(24):8788-8790 [FREE Full text] [doi: 10.1073/pnas.1320040111] [Medline: 24889601]
13. Verma IM. Editorial expression of concern: Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jul 22;111(29):10779 [FREE Full text] [doi: 10.1073/pnas.1412469111] [Medline: 24994898]
14. Zimmer M. Wired Internet. 2016. OkCupid Study Reveals the Perils of Big-Data Science URL: https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/ [accessed 2019-01-15] [WebCite Cache ID 75RkTCmXJ]
15. Rosenberg M, Confessore N, Cadwalladr C. New York Times. 2018 Mar 17. How Trump Consultants Exploited the Facebook Data of Millions URL: https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html [accessed 2019-02-15] [WebCite Cache ID 75Rkg10nu]
16. Taylor J, Pagliari C. Mining social media data: How are research sponsors and researchers addressing the ethical challenges? Research Ethics 2017 Oct 26;14(2):1-39. [doi: 10.1177/1747016117738559]
17. Ayers J, Caputi T, Nebeker C, Dredze M. Don't quote me: reverse identification of research participants in social media studies. npj Digital Med 2018 Aug 2;1(1):30 [FREE Full text] [doi: 10.1038/s41746-018-0036-2]
18. Fiesler C, Proferes N. "Participant" Perceptions of Twitter Research Ethics. Social Media + Society 2018 Mar 10;4(1):205630511876336. [doi: 10.1177/2056305118763366]
19. Williams ML, Burnap P, Sloan L. Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation. Sociology 2017 Dec;51(6):1149-1168 [FREE Full text] [doi: 10.1177/0038038517708140] [Medline: 29276313]
20. Roberts L. Ethical Issues in Conducting Qualitative Research in Online Communities. Qualitative Research in Psychology 2015 Jan 29;12(3):314-325. [doi: 10.1080/14780887.2015.1008909]
21. Markham A, Buchanan E. AOIR. Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0) URL: http://aoir.org/reports/ethics2.pdf [accessed 2019-02-15]
22. Metcalf J, Keller E, Boyd D. Perspectives on Big Data, Ethics, and Society. 2016. Council on Big Data, Ethics, and Society URL: https://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/ [accessed 2019-02-15]
23. 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]
24. Vitak J, Proferes N, Shilton K, Ashktorab Z. Ethics Regulation in Social Computing Research: Examining the Role of Institutional Review Boards. J Empir Res Hum Res Ethics 2017 Dec;12(5):372-382. [doi: 10.1177/1556264617725200] [Medline: 28831844]
25. Vayena E, Gasser U, Wood A, O'Brien D, Altman M, See T, et al. Washington and Lee Law Rev. 2016. Elements of a new ethical framework for big data research URL: https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=1040&context=wlulr-online [accessed 2019-02-15] [WebCite Cache ID 75QcBlZhE]
26. Hutton L, Henderson T. I didn't sign up for this!: Informed consent in social network research. 2015 Presented at: Proc Ninth Int AAAI Conf Web Soc Media Internet; 2015; University of Oxford, Oxford, UK p. 178-187 URL: https://research-repository.st-andrews.ac.uk/handle/10023/6691

27. Office for Human Research Protections, HHS.gov. 2018. International Compilation of Human Research Standards Internet URL: https://www.hhs.gov/ohrp/international/compilation-human-research-standards/index.html [accessed 2019-02-15] [WebCite Cache ID 75RklFkWl]
28. TRUST - Equitable Research Partnerships. The Chinese Ethical Review System and its Compliance Mechanisms URL: http://trust-project.eu/wp-content/uploads/2016/03/Chinese-Ethics-Review-System.pdf [accessed 2019-01-15] [WebCite Cache ID 75Rkq4cze]
29. Hand DJ. Aspects of Data Ethics in a Changing World: Where Are We Now? Big Data 2018 Sep 01;6(3):176-190 [FREE Full text] [doi: 10.1089/big.2018.0083] [Medline: 30283727]
30. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. PLoS One 2018;13(10):e0204937 [FREE Full text] [doi: 10.1371/journal.pone.0204937] [Medline: 30308031]
31. Glez-Peña D, Lourenço A, López-Fernández H, Reboiro-Jato M, Fdez-Riverola F. Web scraping technologies in an API world. Brief Bioinform 2014 Sep;15(5):788-797. [doi: 10.1093/bib/bbt026] [Medline: 23632294]
32. Narayanan A, Felten E. RandomWalker. 2014. No silver bullet: De-identification still doesn?t work URL: http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf [accessed 2019-02-15] [WebCite Cache ID 75Rkz6t0z]
33. Zook M, Barocas S, Boyd D, Crawford K, Keller E, Gangadharan SP, et al. Ten simple rules for responsible big data research. PLoS Comput Biol 2017 Dec;13(3):e1005399 [FREE Full text] [doi: 10.1371/journal.pcbi.1005399] [Medline: 28358831]
34. Bishop L. UK Data Service. 2017. Big data and data sharing:Ethical issues URL: https://www.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethical-issues.pdf [accessed 2019-02-15] [WebCite Cache ID 75Rl0EgLk]
35. Ohm P. UCLA Law Review. 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization URL: https://www.uclalawreview.org/broken-promises-of-privacy-responding-to-the-surprising-failure-of-anonymization-2/ [accessed 2019-02-15]
36. Sweeney L. Data Privacy Lab. 2000. Simple demographics often identify people uniquely URL: http://dataprivacylab.org/projects/identifiability/paper1.pdf [accessed 2019-02-15] [WebCite Cache ID 75RlAkgEt]
37. Boyd D, Crawford K. Critical Questions for Big Data. Information, Communication & Society 2012 Jun;15(5):662-679. [doi: 10.1080/1369118X.2012.678878]
38. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. Sci Eng Ethics 2016 Apr;22(2):303-341. [doi: 10.1007/s11948-015-9652-2] [Medline: 26002496]
39. Nissenbaum H. A contextual approach to privacy online. Daedalus 2011;140(4):48 [FREE Full text]
40. PatientsLikeMe. URL: https://www.patientslikeme.com/ [accessed 2019-02-08]
41. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality And Threatens Democracy. USA: Crown; 2019.
42. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. J Med Internet Res 2017 Oct 31;19(10):e353 [FREE Full text] [doi: 10.2196/jmir.6426] [Medline: 29089287]
43. Daniulaityte R, Carlson R, Falck R, Cameron D, Perera S, Chen L, et al. "I just wanted to tell you that loperamide WILL WORK": a web-based study of extra-medical use of loperamide. Drug Alcohol Depend 2013 Jun 01;130(1-3):241-244 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.11.003] [Medline: 23201175]
44. Wright D. Research ethics and computer science. In: Proceedings of the 24th Annual Conference on Design of Communication. 2006 Oct 18 Presented at: Proceedings of the 24th Annual Conference on Design of Communication; 2006; Myrtle Beach, SC, USA. [doi: 10.1145/1166324.1166369]
45. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Office for Human Research Protections, HHS.gov. 2014 Dec. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html [accessed 2019-02-15]
46. Government Publishing Office. 2017. The Common Rule URL: https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/pdf/2017-01058.pdf [accessed 2019-02-15]
47. Buchanan E, Aycock J, Dexter S, Dittrich D, Hvizdak E. Computer Science Security Research and Human Subjects: Emerging Considerations for Research Ethics Boards. Journal of Empirical Research on Human Research Ethics 2011 Jun;6(2):71-83. [doi: 10.1525/jer.2011.6.2.71]
48. Council for International Organizations of Medical Sciences. 2016. International Ethical Guidelines for Health-Related Research Involving Humans URL: http://www.sciencedirect.com/science/article/B6VC6-45F5X02-9C/2/e44bc37a6e392634b1cf436105978f01 [accessed 2019-01-15] [WebCite Cache ID 75RlC6opS]
49. Obar J, Oeldorf-Hirsch A. The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. 2018 Presented at: TPRC 44 44th Res Conf Commun Inf Internet Policy; 2018; George Mason University, Fairfax, VA URL: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3208371_code962598.pdf?abstractid=2757465&mirid=1
50. Angwin J, Stecklow S. The Wall Street Journal. 2010. 'Scrapers' Dig Deep for Data on Web URL: https://www.wsj.com/articles/SB10001424052748703358504575544381288117888 [accessed 2019-02-15] [WebCite Cache ID 75RlH8dpN]

51. Brownstein CA, Brownstein JS, Williams DS, Wicks P, Heywood JA. The power of social networking in medicine. Nat Biotechnol 2009 Oct;27(10):888-890. [doi: 10.1038/nbt1009-888] [Medline: 19816437]
52. PatientsLikeMe. Terms and Conditions of Use URL: https://www.patientslikeme.com/about/user_agreement [accessed 2019-02-07] [WebCite Cache ID 7610OXnGz]
53. PatientsLikeMe. Welcome to the privacy policy URL: https://www.patientslikeme.com/about/privacy [accessed 2019-02-07] [WebCite Cache ID 7610TrtAA]
54. Insights Association. 2013. MRA Code of Marketing Research Standards URL: https://www.insightsassociation.org/sites/default/files/misc_files/mra_code.pdf [accessed 2019-02-15]
55. CIOMS. 2009. International Ethical Guidelines for Epidemiological Studies Internet URL: https://cioms.ch/wp-content/uploads/2017/01/International_Ethical_Guidelines_LR.pdf [accessed 2019-02-15] [WebCite Cache ID 75RlKEMi0]
56. Olhede S, Wolfe P. The growing ubiquity of algorithms in society: implications, impacts and innovations. Philos Trans A Math Phys Eng Sci 2018 Sep 13;376(2128):1-16 [FREE Full text] [doi: 10.1098/rsta.2017.0364] [Medline: 30082309]
57. Metcalf J, Crawford K. Big Data and Society. 2016. Where are Human Subjects in Big Data Research? The Emerging Ethics Divide URL: https://ssrn.com/abstract=2779647 [accessed 2019-01-15] [WebCite Cache ID 75RlRKRE9]
58. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using Social Media as a Research Recruitment Tool: Ethical Issues and Recommendations. Am J Bioeth 2017 Mar;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
59. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res 2011 Jan;13(1):e6 [FREE Full text] [doi: 10.2196/jmir.1643] [Medline: 21252034]
60. Community Principles on Ethical Data Practices. URL: https://datapractices.org/community-principles-on-ethical-data-sharing/ [accessed 2019-02-07] [WebCite Cache ID 7611cDeny]

Abbreviations

CIOMS: Council for International Organizations of Medical Sciences
GDPR: General Data Protection Regulation
IRB: institutional review board
REC: research ethics committee
USDHHS: United States Department of Health and Human Services

Edited by G Eysenbach; submitted 20.08.18; peer-reviewed by E Buchanan, A Jobin, C Fiesler, R Daniulaityte; comments to author 27.09.18; revised version received 16.11.18; accepted 03.02.19; published 21.02.19.

Please cite as:
Chiauzzi E, Wicks P
Digital Trespass: Ethical and Terms-of-Use Violations by Researchers Accessing Data From an Online Patient Community
J Med Internet Res 2019;21(2):e11985
URL: http://www.jmir.org/2019/2/e11985/
doi:10.2196/11985
PMID:30789346

Review

The Messages Presented in Electronic Cigarette–Related Social Media Promotions and Discussion: Scoping Review

Kahlia McCausland[1], BSc; Bruce Maycock[1], PhD; Tama Leaver[2], PhD; Jonine Jancey[1], PhD

[1]Collaboration for Evidence, Research and Impact in Public Health, School of Public Health, Curtin University, Bentley, Australia

[2]School of Media, Creative Arts and Social Inquiry, Curtin University, Bentley, Australia

Corresponding Author:
Kahlia McCausland, BSc
Collaboration for Evidence, Research and Impact in Public Health
School of Public Health
Curtin University
Kent Street
Bentley, 6102
Australia
Phone: 61 92667382 ext 7382
Email: kahlia.mccausland@curtin.edu.au

Abstract

Background: There has been a rapid rise in the popularity of electronic cigarettes (e-cigarettes) over the last decade, with growth predicted to continue. The uptake of these devices has escalated despite inconclusive evidence of their efficacy as a smoking cessation device and unknown long-term health consequences. As smoking rates continue to drop or plateau in many well-developed countries, transnational tobacco companies have transitioned into the vaping industry and are now using social media to promote their products. Evidence indicates e-cigarettes are being marketed on social media as a harm reduction alternative, with retailers and manufacturers utilizing marketing techniques historically used by the tobacco industry.

Objective: This study aimed to identify and describe the messages presented in e-cigarette–related social media (Twitter, YouTube, Instagram, and Pinterest) promotions and discussions and identify future directions for research, surveillance, and regulation.

Methods: Data sources included MEDLINE, Scopus, ProQuest, Informit, the *Journal of Medical Internet Research*, and Google Scholar. Included studies were published in English between 2007 and 2017, analyzed content captured from e-cigarette–related social media promotions or discussions, and reported results for e-cigarettes separately from other forms of tobacco and nicotine delivery. Database search ceased in October 2017. Initial searches identified 536 studies. Two reviewers screened studies by title and abstract. One reviewer examined 71 full-text articles to determine eligibility and identified 25 studies for inclusion. This process was undertaken with the assistance of the Web-based screening and data extraction tool—Covidence. The review was registered with the Joanna Briggs Institute (JBI) Systematic Reviews database and followed the methodology for JBI Scoping Reviews.

Results: Several key messages are being used to promote e-cigarettes including as a safer alternative to cigarettes, efficacy as a smoking cessation aid, and for use where smoking is prohibited. Other major marketing efforts aimed at capturing a larger market involve promotion of innovative flavoring and highlighting the public performance of vaping. Discussion and promotion of these devices appear to be predominantly occurring among the general public and those with vested interests such as retailers and manufacturers. There is a noticeable silence from the public health and government sector in these discussions on social media.

Conclusions: The social media landscape is dominated by pro-vaping messages disseminated by the vaping industry and vaping proponents. The uncertainty surrounding e-cigarette regulation expressed within the public health field appears not to be reflected in ongoing social media dialogues and highlights the need for public health professionals to interact with the public to actively influence social media conversations and create a more balanced discussion. With the vaping industry changing so rapidly, real-time monitoring and surveillance of how these devices are discussed, promoted, and used on social media is necessary in conjunction with evidence published in academic journals.

(J Med Internet Res 2019;21(2):e11953) doi:10.2196/11953

KEYWORDS

electronic nicotine delivery systems; social media; public health; review

Introduction

There has been a dramatic rise in the popularity of electronic cigarettes (e-cigarettes) since the first commercialized product was developed in China in 2003 [1,2]. It is now estimated that there are 35 million e-cigarette users globally (including heat not burn tobacco products) [3], with this rapid growth predicted to continue. According to BIS Research [4], the global e-cigarette industry will experience an annual growth of more than 22% until 2025, reaching a total market value of US $50 billion dollars at this time.

Since the advent of first generation e-cigarettes, which closely resemble traditional cigarettes in appearance and size, they have been the center of much debate. It has been suggested that these devices may be a less harmful alternative to smoking [5], provide health benefits to smokers who switch completely to them [6], lessen cigarette cravings [7], and facilitate smoking cessation [8]. However, promotion of e-cigarettes may also encourage nonsmokers, particularly young people, to initiate use [9,10], facilitate experimentation with traditional tobacco products [11], and undermine tobacco control efforts [12]. Recent studies also suggest that e-cigarette use is associated with negative health consequences [13,14] and may not facilitate adult smokers to quit at rates higher than smokers who do not use these products [15]. These contrasting arguments are evident in Web-based marketing by e-cigarette retailers and manufacturers [16], along with social media discussions about e-cigarettes [17]. Furthermore, the lack of agreement among countries on the population-level impact of these devices and how they should be regulated [2] (eg, UK Royal College of Physicians identifies e-cigarettes as a public health strategy, whereas the World Health Organization and the US Surgeon General see them as presenting potentially new health problems [18]) may cause confusion among consumers and the public in general. This, therefore, highlights the importance of examining social media as it offers opportunities to attract new users, promote continued use, and build brand loyalty.

Traditionally dominated by small start-up companies, the e-cigarette market has experienced rapid growth and transition, and more recently, large manufacturers and transnational tobacco companies have come to dominate the market. Major tobacco companies have entered the vaping industry by either acquiring e-cigarette companies and brands or developing their own products. Major tobacco companies now involved in the vaping industry include British American Tobacco, Imperial Tobacco, the Altria Group, Reynolds American, Philip Morris International, and Japan Tobacco International [19]. These companies have benefited from large advertising and marketing budgets, which enable promotion across the World Wide Web [20].

A significant portion of e-cigarette business is conducted on the internet [21], with most existing e-cigarette companies operating websites or other Web-based selling systems [22]. Sources suggest that e-cigarette manufacturers are careful to distance their products from tobacco [23] by using techniques such as aesthetic appeal, including attractiveness, coolness, colors, and innovative packaging and flavor variations. In addition, websites and social media accounts have been found to exhibit price promotions, and competitions and discount coupons; there is also evidence of celebrity endorsements and sports sponsorship [24].

An accurate understanding of the types of e-cigarette messages social media users are exposed to, and who is disseminating this information can assist in the development of appropriate surveillance to inform future policy and regulations. A scoping review was, therefore, undertaken to identify and describe the messages presented in e-cigarette–related social media (Twitter, YouTube, Instagram, and Pinterest) promotions and discussions.

Methods

Scoping Review Overview

The review was registered prospectively with the Joanna Briggs Institute (JBI) Systematic Reviews database (May 5, 2017) and proposed methods specified in advance in a protocol [25]. The scoping review adhered to the methods manual developed by the JBI [26].

Objectives

This scoping review aimed to identify and describe the messages presented in e-cigarette–related social media (Twitter, YouTube, Instagram, and Pinterest) promotions and discussions and identify future directions for research, surveillance, and regulation.

Inclusion Criteria

Included studies had to examine and analyze e-cigarette–related social media promotions and discussions. Studies needed to clearly identify the social media platform under investigation. Studies reporting multiple social media platforms were excluded unless results for each platform were reported separately. This was so the results for each social media platform could be extracted and reported, making it possible to clearly identify similarities and differences between the platforms. Studies identifying other tobacco products (eg, tobacco cigarette, snus, chewing tobacco, or hookah) were excluded unless e-cigarettes were also examined and reported separately. In addition, studies that did not distinguish between e-cigarettes and other forms of tobacco and nicotine delivery were excluded. Studies examining traditional media (eg, television and newspaper) were excluded unless social media platforms were also examined and reported separately. Studies were limited to the following countries: the United Kingdom, the Unites States of America, New Zealand, Australia, and Canada. These countries were selected as they are all developed countries and e-cigarette use is well established [27]. The review considered only peer-reviewed primary research studies published in English in the last 10 years (2007-2017); this period correlates with the approximate time that e-cigarettes were first introduced to the Unites States and Europe [28].

Table 1. Summary of excluded studies subject to full-text review with reason (N=48).

Reason for exclusion	Studies (n)
Excluded at full-text review	
Wrong study design (ie, does not examine a social media platform or code for account type, theme, or sentiment)	12
Does not report electronic cigarettes (e-cigarettes) in the results	7
Results for different social media platforms not reported separately	2
Publication type	4
Country of study	1
Excluded at data extraction	
Wrong study design	14
Results for e-cigarettes not reported separately	2
Results for different social media platforms not reported separately	1
A specific population is examined (ie, people with mental illness)	2
Country of study	3

Search Strategy and Study Selection

Overall, 5 databases were searched (MEDLINE, Scopus, ProQuest, Informit, and Google Scholar) using the following terms:

("electronic cigarette" OR e-cigarette OR "electronic nicotine delivery system" OR "personal vapo?ri?er" OR "electronic nicotine delivery device" OR "vape pen" OR "smokeless tobacco" OR "electric cigarette" OR "electric nicotine delivery system" OR "electric nicotine delivery device" OR e-hookah OR e-juice OR e-liquid OR vaping) AND ("social media" OR internet OR online OR YouTube OR Facebook OR Instagram OR Twitter OR "online media" OR "digital media" OR "social networking") AND ("content analysis" OR "content evaluation" OR message OR meaning OR coding OR "media analysis" OR "textual analysis").

In addition, the search strategy was entered as a nested Boolean search into Google Scholar, with the first 200 results examined for eligibility and subject to the screening process outlined below. Preliminary searches located relevant studies published in the *Journal of Medical Internet Research*, a hand-search of this journal was, therefore, also undertaken.

Retrieved references from each database were imported into EndNote X7 (Clarivate Analytics) [29] reference management software, with duplicate references removed before being imported into Covidence [30]. Covidence is a Web-based software platform that streamlines the production of systematic reviews by supporting the key steps in the review process [30]. Studies were assessed for inclusion, examined initially by title and abstract. Full-text articles were retrieved for those studies that appeared to meet the inclusion criteria or if further examination was required to determine eligibility. Moreover, 2 reviewers (KM and JJ) independently screened all titles and abstracts to determine their eligibility. The primary reviewer (KM) then undertook full-text screening. These processes were assisted by the Web-based screening and data extraction tool—Covidence [30]. Finally, the reference list of all articles subject to full-text review was screened to determine possible inclusion of additional studies. Identified studies were assessed for suitability based on full-text review undertaken by the primary reviewer. A summary of excluded studies subject to full-text review and the reason for exclusion is provided in Table 1.

Extraction of Results

The relevant content from each study was extracted using a data extraction pro forma, which included title, author, publication year, country of study, aim/purpose of study, social media platform, sample size, study design/methods, results, and key findings that relate to the review question. Included studies were required to have developed coding categories for content including one or more of the following: account type, themes, and sentiment. Account type characterizes the publisher of the social media post; theme reflects the domain of the actual content conveyed, such as the categories of health, smoking cessation, and regulation; and sentiment reflects the stance expressed in a social media post toward e-cigarettes, related products or its users, whether positive, neutral, or negative. To ensure data extraction consistency, 2 reviewers (KM and JJ), independent of one another, extracted data from the same 5 studies using the data extraction pro forma. The reviewers then met to determine whether the extraction approach was consistent. The primary reviewer (KM) then went on to extract data from the remaining studies unaccompanied.

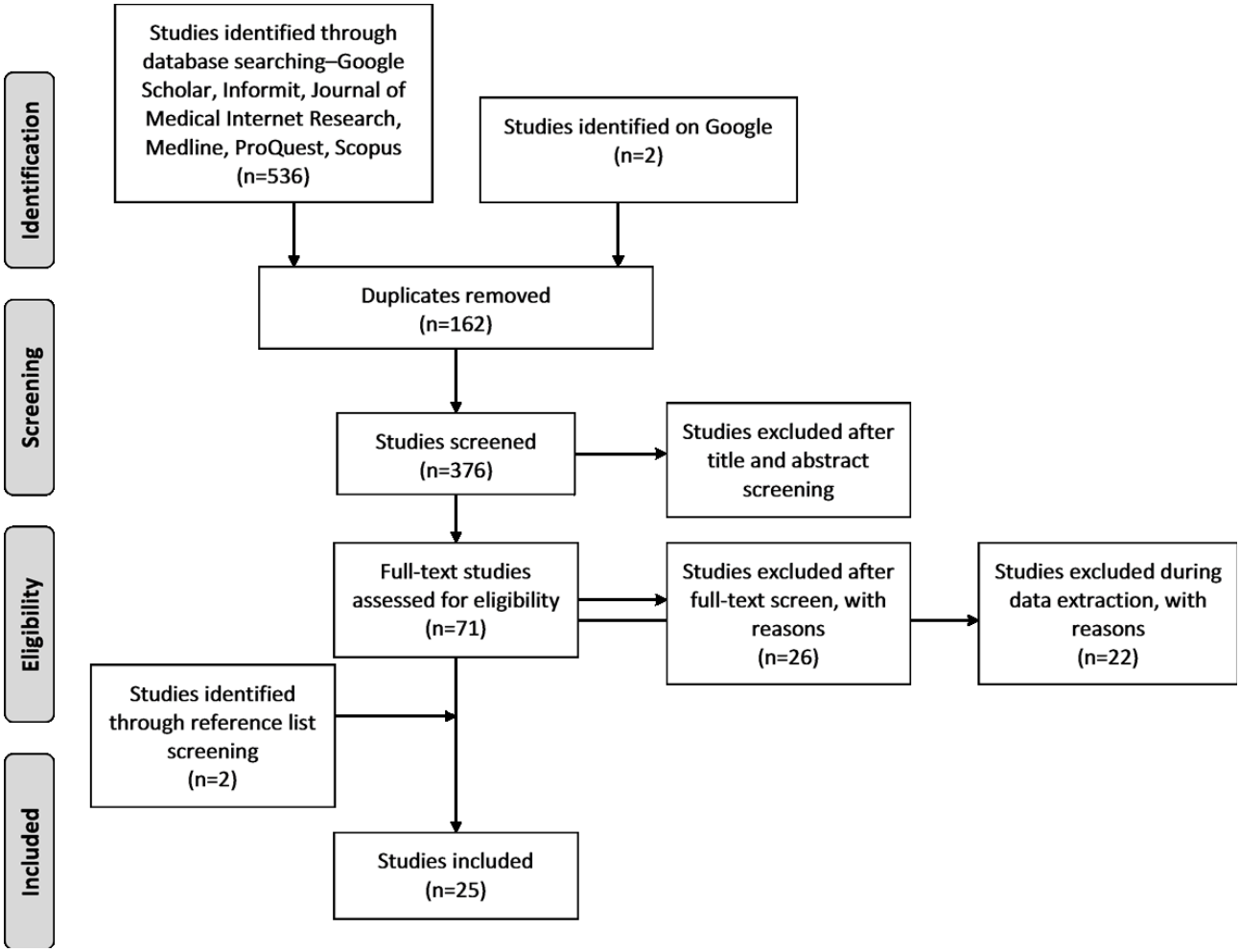
Results

Description of Included and Excluded Studies

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram detailed in Figure 1 presents the number of studies at each stage of the review process.

A total of 25 studies were identified for inclusion in this review. A total of 18 studies analyzed Twitter data [16,17,31-46]; 4 examined YouTube including videos [47-49] and data associated with videos, such as video tags, titles, or descriptions [50]; and 3 studies investigated images on Instagram and Pinterest [51-53].

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Twenty-four studies were conducted in the United States [16,17,31-41,43-53] and one in Canada [42] (Table 2).

Sample Size and Data Coding

The sample size of included studies varied widely, even within social media platforms (Twitter, YouTube, Pinterest, and Instagram), with the platform under investigation influencing the coding method used (Table 2). Methods used included hand coding [31,32,36-39,41,42,46-53] or machine learning [44], or a combination of the 2 methods [16,17,33-35,40,43,45]. Hand coding involved one or more human coders categorizing data, whereas machine learning used algorithms to give computers the ability to code data, although hand coding was usually used for an initial subset of data to help refine the algorithm to improve its accuracy [16,33-35,43,45]. Studies that analyzed text (ie, tweets from Twitter and YouTube video descriptions) predominantly employed hand coding for smaller samples (eg, <3000) [31,32,36-39,42,46], and a combination of hand coding and machine learning [16,33-35,40,43,45] or only machine learning [44] for larger samples, the largest being 1,669,123 tweets. Three studies did, however, hand code samples of over 10,000 [17,41,50]. All studies that analyzed images (ie, Pinterest, Instagram, and YouTube) did so by hand coding samples of between 46 and 2208 posts and videos (Table 2).

Account Type

A total of 12 studies [16,31,33,37,39,41-44,47-48,53] used coding to identify the source (account type) of the social media data collected in their sample, most commonly informed by data found in account profiles (eg, bio, location, and profile photo) and preceding social media posts (Table 3). The most common account types coded for were personal [16,31,33,39,41,42,44,47,48,53] and commercial [33,37,39,41,42,44,48,53], of which these account types represented up to 82.68% (104,283/126,127) [44] and 89.73% (66,102/73,672) [33] of some samples, respectively. Overall, 3 studies coded for government, foundation or not for profit organizations, [39,41] and public health and health care [42] accounts. All were studies of Twitter discussions that unanimously reported that tweets from these account types represented less than 3% of their sample size (1.0%, 5/500 for [39]; 0.08%, 8/10,128 for [41], and 1.3%, 4/300 and 3.3%, 10/300 for [42]). Overall, 3 studies coded for fake accounts, of which 2 reported these accounts represented similar percentages of their sample (6.90%, 699/10,128 for [41] and 9.7%, 29/300 for [42]), whereas the third found an overabundance (80.7%, n not provided) [16].

Table 2. Description of included studies, sample size, and coding method.

Authors, year, country	Social media platform	Sample size	Coding method
Burke-Garcia et al, 2017, United States [39]	Twitter	1000 tweets	Hand coding
Lazard et al, 2017, United States [40]	Twitter	4629 tweets	Machine learning and hand coding
Allem et al, 2017, United States [31]	Twitter	2192 tweets	Hand coding
Ayers et al, 2017, United States [17]	Twitter	11,600 tweets	Hand coding
Dai et al, 2017, United States [45]	Twitter	757,167 tweets	Hand coding and machine learning
Clark et al, 2016, United States [16]	Twitter	850,000 tweets	Hand coding, machine learning, and hedonometrics
van der Tempel et al, 2016, Canada [42]	Twitter	600 tweets	Hand coding
Han et al, 2016, United States [35]	Twitter	1,021,561 tweets	Hand coding and machine learning
Jo et al, 2016, United States [36]	Twitter	2847 tweets	Hand coding
Kavuluru et al, 2016, United States [43]	Twitter	224,000 tweets	Hand coding and machine learning
Sowles et al, 2016, United States [37]	Twitter	1156 tweets	Hand coding
Unger et al, 2016, United States [38]	Twitter	1519 tweets	Hand coding
Lazard et al, 2016, United States [44]	Twitter	126,127 tweets	Machine learning
Cole-Lewis et al, 2015, United States [41]	Twitter	10,128 tweets	Hand coding
Kim et al, 2015, United States [34]	Twitter	1,669,123 tweets	Hand coding and machine learning
Harris et al, 2014, United States [32]	Twitter	683 tweets	Hand coding
Huang et al, 2014, United States [33]	Twitter	73,672 tweets	Handing coding and machine learning
Prochaska et al, 2012, United States [46]	Twitter	153 accounts	Hand coding
Sears et al, 2017, United States [47]	YouTube	46 videos	Hand coding
Basch et al, 2016, United States [48]	YouTube	99 videos	Hand coding
Merianos et al, 2016, United States [49]	YouTube	55 videos	Hand coding
Huang et al, 2016, United States [50]	YouTube	28,089 videos tags, titles, or descriptions	Hand coding
Lee et al, 2017, United States [51]	Instagram and Pinterest	1800 images	Hand coding
Chu et al, 2016, United States [52]	Instagram	2208 posts	Hand coding
Laestadius et al, 2016, United States [53]	Instagram	85 posts	Hand coding

Themes

All 25 included studies coded for themes (Table 4). Health, safety, and harms was the most coded for theme in this review [17,31-33,35,38,39,41,43,46-51,53]; however, various descriptions for health, safety, and harms were used (eg, health, harm reduction, and harms encompassing both the health benefits and consequences of e-cigarette use). Additional themes frequently cited were smoking cessation [16,17,33,35,37, 39,41-43,46,47,49,50,53]; product types and characteristics [16,17,32-37,39,41,43,49-51]; advertisement, promotion, and marketing [16,31,38,39,41,42,44,45,48,51,52]; regulation, policy, and government [31,32,35,39-42,49,50]; price promotions, discounts, coupons, giveaways, and competitions [16,33,34,36,37,44,50]; and smoke-free, use indoors or where cigarettes are banned [17,35,40,43,47,49].

Sentiment

Of the 25 studies, a total of 12 coded for sentiment [16,31,32,38-43,45,47,49] (Table 5). Overall, 3 studies made the distinction when coding for message attitude [38,42,45] rather than emotional tone or affective content.

Table 3. Coded category—account type.

Account type	Studies, n (%)	References
Personal (general public, individuals, organic, and user-generated)	10 (40)	[16,31,33,39,41,42,44,47,48,53]
Commercial (marketing, tobacco or electronic cigarette [e-cigarette] company or retailer)	8 (32)	[33,37,39,41,42,44,48,53]
Press, media, or news (verifiable press or other prominent media sources of information, such as blogs)	3 (12)	[41,42,48]
Fake (hacked, bots, and automated)	3 (12)	[16,41,42]
Professional (television studio or network, production company, or organization)	2 (8)	[47,48]
Government, foundation, or not for profit organization	2 (8)	[39,41]
Proponents (sales or marketing agencies and individuals who advocate or specifically identify themselves as vapers)	2 (8)	[43,44]
Celebrity or public figure	2 (8)	[41,42]
Unknown or other	2 (8)	[31,37]
Public health, health care	1 (4)	[42]
Vaping-related handle (vaping-related term in handle name or Twitter bio)	1 (4)	[37]
Personal accounts with industry ties	1 (4)	[42]
E-cigarette community movement	1 (4)	[41]
General entity (company, store, or advocacy group)	1 (4)	[31]

Table 4. Coded category—themes.

Themes	Studies, n (%)	References
Health, safety, and harms	16 (64)	[17,31-33,35,38,39,41,43,46-51,53]
Health	10 (40)	[31,33,38,46-51,53]
Safety	5 (20)	[17,32,33,48,50]
Harms	2 (8)	[48,49]
Harm reduction	2 (8)	[35,43]
Health and safety	1 (4)	[41]
Health and health consequence	1 (4)	[39]
Smoking cessation	14 (56)	[16,17,33,35,37,39,41-43,46,47,49,50,53]
Product types and characteristics	14 (56)	[16,17,32-37,39,41,43,49-51]
Advertisement, promotion, marketing	11 (44)	[16,31,38,39,41,42,44,45,48,51,52]
Regulation, policy, government	9 (36)	[31,32,35,39-42,49,50]
Price promotions, discounts, coupons, giveaways, competitions	7 (28)	[16,33,34,36,37,44,50]
Smoke-free, use indoors or where cigarettes are banned	6 (24)	[17,35,40,43,47,49]
More economical than smoking	5 (20)	[17,42,47,49,53]
Social status, acceptance	4 (16)	[17,38,47,51]
Cleaner than tobacco, environment friendly, no/minimal odor	4 (16)	[17,47,49,53]
First or second person experience, use, opinion, or purchases	4 (16)	[39,42,52,53]
Recreation, customization, tricks	3 (12)	[47,51,53]
Other/unknown	3 (12)	[31,38,39]
Product image	2 (8)	[37,52]
Craving	2 (8)	[41,42]
Illicit substance use in e-cigarettes	2 (8)	[41,51]
Personal opinion	2 (8)	[42,45]
News articles and updates	2 (8)	[42,44]
Tastes good	2 (8)	[42,49]
Getting others started, encouraging use, offering advice	2 (8)	[40,42]
Second-hand smoke	2 (8)	[47,49]
Cessation devices or gateway products for youth to establish nicotine addictions	2 (8)	[44,49]
Text	1 (4)	[52]
Lies/propaganda	1 (4)	[32]
Science (studies)	1 (4)	[32]
Issue salience	1 (4)	[32]
Underage e-cigarette use	1 (4)	[41]
E-cigarette use in association with other addictive substances (eg, alcohol, caffeine)	1 (4)	[41]
Parental e-cigarette use	1 (4)	[41]
Places of use	1 (4)	[34]
Neutral information	1 (4)	[42]
Humor	1 (4)	[42]
Just starting e-cigarettes	1 (4)	[42]
Advocating e-cigarettes	1 (4)	[42]
Attempt to engage other Twitter users	1 (4)	[42]
Using or comparing with other substances/nicotine replacement therapies	1 (4)	[42]

Themes	Studies, n (%)	References
Presence of identity or community	1 (4)	[53]
Technology (modern products, information about science behind the products)	1 (4)	[47]
Celebrity, model	1 (4)	[51]
Meme	1 (4)	[51]
Anti-smoking	1 (4)	[51]
Utilization patterns	1 (4)	[39]
Consumer endorsement	1 (4)	[39]
Money (taxes, small businesses)	1 (4)	[31]
Addiction to e-cigarettes	1 (4)	[49]
Reactions to e-cigarette policies and questions about e-cigarette health risk claims	1 (4)	[44]
Similar to real cigarettes	1 (4)	[49]

Table 5. Coded category—sentiment.

Sentiment	Studies, n (%)	References
Emotional tone or affective content		
Positive or negative	5 (20)	[16,41-43,47]
Positive or negative valence	2 (8)	[39,40]
Pro or anti	2 (8)	[31,49]
Pro- or anti-policy	1 (4)	[32]
Neutral	7 (28)	[31,39,41,42,45,47,49]
Unable to tell	1 (4)	[32]
Message attitude		
Pro or con	1 (4)	[42]
Pro or anti	1 (4)	[38]
Supportive or against	1 (4)	[45]
Neutral or do not know	3 (12)	[38,42,45]

Discussion

Principal Findings

Data Coding

The coding methods used were hand coding, machine learning, or a combination of the two. Compared with hand coding, machine learning can rapidly code large amounts of data; however, hand coding undertaken by humans may more accurately discriminate the complexities and subtleties of language [54]. Although hand coding can be subject to individual bias, the development of codes grounded in literature and achieving acceptable levels of inter-rater reliability can assist to reduce this [55]. Studies that require the determination of subtle differences in language or context may, therefore, be better placed to employ hand coding for a smaller sample of data, whereas studies that rely less on context could employ machine learning to code larger samples [55]. The increased complexity of interpreting visual social media (eg, YouTube, Instagram, and Pinterest) meant all studies of these platforms employed hand coding [47-49,51-53].

Account Type

Personal

Studies included in this review reported dissemination of diverse e-cigarette messaging by predominantly commercial social media accounts [33,53]; however, other studies discovered conversations occurring among personal accounts dominating the social media landscape [31,41,42,44,47,48]. Personal accounts were found to be discussing, endorsing, and promoting various themes, most commonly marketing [41,48,53], smoking cessation [33,42,44], recreation and technology [47,53], and first-person experience and opinion [41,42]. This is particularly important as individuals may be less critical of material posted by other consumers compared with retailers [56,57] and may be more easily persuaded by other individuals they know, given their relative closeness and potentially increased perception of source credibility [58,59].

Commercial

Several key marketing strategies were found to be used by commercial social media accounts. These included the use of popular hashtags that enabled marketing messages to *piggy back*

on trending topics and increase dissemination reach [42], use of fake user accounts to disseminate spam and favorable views [33,42], and the offer of price promotions and product giveaways [33,44,53]. Social media networking and marketing efforts undertaken by the vaping industry may have contributed to the rapid rise in popularity of e-cigarettes, the extent of which has been demonstrated by the findings in this review. It has also been hypothesized by some researchers that the lack of regulatory standards on social media may be playing an ever-increasing role in the diffusion of tobacco products and prosmoking messages [60].

Government, Foundation or Not for Profit Organizations, and Public Health and Health Care

Of the studies that coded for government, foundation, or not for profit accounts [39,41,42], limited public health–related messaging was identified, and activity from these account types represented less than 3% of samples. These findings indicate more action from public health and government to communicate the potential harms and benefits of alternate nicotine delivery products via social media is required to balance the information shared on these platforms.

Fake

Most tweets produced by accounts classified as fake were found to promote e-cigarettes as effective smoking cessation aids, either by emulating first-person anecdotes or linking to news articles or other Web-based media [41,42], with some accounts potentially using computer programs to generate and post content automatically [33,34].

The general tweet structure from an automated bot is provided here [16]:

> *@USER [I,We] [tried, pursued] to [give up, quit] smoking. Discovered BRAND electronic cigarettes and quit in [#] weeks. [Marvelous,Amazing,Terrific]! URL*
>
> *@USER It's now really easy to [quit,give up] smoking (cigarettes).—these BRAND electronic cigarettes are lots of [fun,pleasure]! URL*
>
> *@USER electronic cigarettes can assist cigarette smokers to quit, it's well worth the cost URL*

This type of spamming suggests that there are remunerations to be gained by steering potential online consumers to certain retail websites [34].

Themes

Health, Safety, and Harms

All studies that coded for health, safety, and harms reported that e-cigarettes are being referred to as *healthier* and *safer* than traditional tobacco products on social media [17,31-33,35,38,39,41,43,46-51,53]. Provided that scientific evidence about the safety of e-cigarettes is largely inconclusive, marketing claims that use words such as *safer* to describe their products could contribute to confusion about their overall safety, especially among youth. Promoting a product by claiming that it is healthier than tobacco smoking, the leading cause of preventable death, is therefore controversial and may only have merit when targeting smokers who are contemplating quitting or reducing use [61].

There is indication that an individual's perception of a substance's potential harms and benefits and their behavior of use is influenced by the availability of information discussing the health effects of that substance [62]. A recent analysis reports that 34.20% (8433/24,658) of American youth sampled believe that e-cigarettes are less harmful than cigarettes, and 45.00% (11,096/24,658) are not sure [63].

Example *safety* coded tweets are displayed in the following excerpt [32]:

> *RT @ChiPublicHealth: Electronic cigs contain a dangerous, addictive drug & should be regulated like other nicotine products #ecigtruths*
>
> *@ChiPublicHealth it's not about being safe, it's about being SAFER than the alternative #EcigsSaveLives it's about HARM REDUCTION #Casaa*

Smoking Cessation

Over half (56%, 14/25) of studies included in this review found evidence of e-cigarettes being promoted as a smoking cessation tool [16,17,33,35,37,39,41-43,46,47,49,50,53], although their efficacy as such is yet to be determined [8]. However, some research indicates much smaller proportions of e-cigarette advertisements are now endorsing these devices as quit aids [37,42], and cited reasons for use by vapers have significantly shifted away from smoking cessation (43.00%, 1247/2900 in 2012 vs 29.00%, 841/2900 in 2015) toward use to increase social image (21.00%, 609/2900 in 2012 vs 37.00%, 1073/2900 in 2015) [17]. Of concern is that these results suggest that e-cigarette uptake is not solely driven by a desire among smokers to quit smoking [64].

Product Types or Characteristics

Overall, 14 studies coded for e-cigarette product characteristics such as brands, flavors, and nicotine content, and of these, the majority (86%, 12/14) [16,17,32,35-37,39,41,43,49-51] coded for the mention or depiction of electronic cigarette juice (e-juice) flavors. Marketing social media posts and videos were most commonly found to promote the vast array of e-juice flavors available on the market [16,35,37,43,49], a strategy historically used to entice new tobacco consumers [65], especially youth [66]. As a result of mounting evidence that flavored tobacco products facilitate youth smoking [67], these products (aside from menthol) were effectively banned in 2009 [68]. However, no such ban currently exists for e-cigarettes with thousands of flavors available for purchase [22]. Some research suggests that flavors appeal to adult smokers and may aid smoking cessation [69,70]; nevertheless, increasing evidence demonstrating that flavors also attract youth to the e-cigarette market is mounting [71-73], which is a cause for concern as nicotine addiction has been found to cause problems with adolescent brain development [74]. Studies have found flavor profiles (eg, tobacco and menthol) that are more appealing to some adults may have minimal appeal to young people [69,75]. It has, therefore, been proposed as a harm reduction strategy that these flavors be promoted to adults to assist tobacco substitution,

whereas restricting those flavors that appeal most to youth (eg, fruits and deserts) [76].

Advertisement, Promotion, and Marketing

A concern of e-cigarette promotion on social media is the high level of cross-platform interaction (ie, using apps to post content across several social media platforms) [33], and given the sizeable youth presence on these platforms [77] provides an avenue to invite nonsmokers, youth in particular, to experiment with and instigate use. However, just because youth are utilizing social media does not inevitably mean they are subjected to e-cigarette marketing, as they would need to *follow* these accounts, be exposed through their social networks (ie, followers or those they are following), or browse via direct searches [34]. Recent studies have, however, found that e-cigarette users learn about vaping and these devices from the internet and social media [78,79]; therefore, monitoring how e-cigarettes are promoted on these platforms is incredibly important.

Regulation, Policy, and Government

Messages against government regulation were found to be most prominent [31], for example:

> *Wow, CA DPH thinks it acceptable to deceive the ppl it is supposed to serve: #stillblowingsmoke? no #notblowingsmoke Don't let the FDA go without making your voice heard…#vapecommunity #vape #ecig #notblowingsmoke #ecigssavelive*

Many antiregulation posts expressed concern over the motivations for wanting e-cigarettes regulated, suggesting policy makers were only concerned about these devices because tobacco revenue would decrease if people started using them and that policy represents the teaming of government and industry such that the Food and Drug Administration (FDA) deeming rule would work only to enhance the power of Big Tobacco [31,32,40,42,44]. The uncertainty surrounding e-cigarette regulation expressed within the public health field appears not to be reflected in ongoing social media dialogues [41] and highlights the need for public health professionals to interact with the public to actively influence social media conversations and create a more balanced discussion [40,44].

Price Promotions

This review provides evidence of the existence of e-cigarette marketing on social media, of which a substantial portion includes price promotions, discounts, coupons, free trials, giveaways, and competitions [16,33,34,36,37,44,50]. These types of incentives can persuade potential consumers to make a purchase and assist vendors to create a loyal customer base [80], which has already been demonstrated for tobacco [81,82]. It is well documented that smoking behaviors react to changes in cigarette prices [83], and in response, tobacco control efforts have sought to eradicate the use of these incentives [84]. Similarly, studies have reported that e-cigarette sales are very responsive to price variation, and implementing policy to limit price promotions, free-trials, and giveaways could lead to significant behavior change and uptake [85]. People who use e-cigarettes regularly cite smoking cessation as their motivation for vaping initiation; for this group of people, price promotions that enable affordability of these products longer term could be viewed as appropriate [37], although evidence supporting the use of these devices as a smoking cessation aid is still out for debate [8].

Smoke-Free and Use Indoors or Where Cigarettes Are Banned

A major drawback of cigarettes is the smoke they emit, which is known to contain thousands of chemicals dangerous to human health [86], and for this reason, cigarettes are now subject to smoking bans and smoke-free policies all over the world [87]. Studies included in this review found that e-cigarette proponents frequently highlight the smoke-free aspect of vaping and that these devices can be used where tobacco is currently restricted [17,35,40,43,47,49]. Marketing that accentuates that e-cigarettes can be used *anywhere* may undermine enforcement of smoke-free policies and tobacco control efforts [12] and expose nonusers to toxins [13]. A survey of a representative sample of American adults found that increased frequency of exposure to e-cigarette advertising was associated with lower support for policies that restrict use in public places [88]. These results suggest the need for more publicly available information regarding the chemical composition and possible health consequences of inhaling second-hand vapor [38].

Recreation

Less commonly coded for, however an important aspect of vaping to recognize is recreation, which was coded for among image-based social media (ie, Instagram, Pinterest, and YouTube) studies [47,51,53]. These studies commonly reported depictions of customization and modification of e-cigarette devices for both functional and aesthetic purposes and of vapers exhaling large plumes of vapor (known as cloud chasing) and performing vape tricks. Depiction of these vaping practices could contribute to the normalization of vaping, as images and videos represent these acts as fun and more commonplace and socially accepted than is in reality [52,89], with many posts accompanied by hashtags signifying community and social identity [53]. For example [31]:

> *What's your favourite #vaping trick? #VapeTricks #Vapelife #VapeOn #NotBlowingSmoke*

Many hashtags emerge from users themselves through an organic user-led process [90], with research suggesting substance –focused hashtags can serve as an "addiction bond" [91].

Social media posts and videos mentioning different product characteristics (eg, flavors, mods, and illicit substances) and displaying consumers' ability to choose and modify aspects of their vaping experience indicates that customization and recreation largely contributes to e-cigarette discourse on social media and may have contributed to their rapid increase in popularity [47,50-52].

Sentiment

Studies which coded for sentiment and did not specifically state they were coding for message attitude most commonly reported that their sample was positively skewed toward e-cigarettes, their users, and antiregulation [31,32,40-43,47,49], whereas studies that coded for message attitude reported predominantly

neutral attitude [38,42,45]. Social media posts from accounts with vested interests (eg, commercial or automated) and the general public were found to present positive messages related to e-cigarettes [16,41-43,49], whereas news- and health-related accounts provided messages that were least positive or neutral [41,42,49].

Examples of positive, negative, and neutral tweets are provided here [39]:

> *Medical professionals surveyed. Overwhelmingly prefer #vaping to smoking. #vape #vaplyfe #the http://t.co/tcKsX6Dc0S http://t.co/tiJBNZjBZa*
>
> *RT @StopVaping RETWEET this if your not VAPING today because you want to live.*
>
> *Vaping in the United States has eclipsed cigarette smoking in some age groups. #Vaping #eCigarettes #Rosemont http://t.co/wzgVT0p2C1*

The proliferation of social media platforms and *Big Data* analytics provides the opportunity to explore and monitor people's perceptions of e-cigarettes in real time and what fuels opinion over time [41,45]. The studies included in this review could be used to establish a sentiment baseline for public health professionals to develop campaigns and interventions [41] and act as supplementary data to traditional surveys [45].

In agreement with Lienemann et al [55], when coding for sentiment, clarity and comparability across studies could be enhanced by distinguishing between attitude and emotion. For example, social media data can be provaping; however, it can have a negative emotional tone.

Recommendations for Research

Given the volume of personal accounts found to be discussing, endorsing, and promoting various aspects of vaping, further research to determine who the *loudest* social media accounts are in the sense that their material is being seen and shared most frequently and how this material is influencing other social media users' perceptions and use of e-cigarettes is, therefore, warranted [41]. The perceived safety of these products may also be a contributing factor in the increasing trend of vaping among adult never smokers and former smokers [92]. Research is, therefore, required to determine the implications of claims promoting e-cigarettes as a superior product on audience perception and use [47].

The use and depiction of vaping for recreation raises questions about the promotion of these devices as a hobby or socializing opportunity [64]. As such, it may be valuable to investigate the degree to which the vaping industry is targeting nonsmoking youth who may have an interest in vaping for enjoyment or as a hobby rather than a smoking cessation tool [37,93].

Furthermore, the US FDA has recognized the impact of e-cigarettes recently, ratifying a rule (August 8, 2016) that extended their regulatory authority to all tobacco products, including e-cigarettes. These regulations restrict youth access by prohibiting the sale of e-cigarettes to those aged under 18 years, embargos the use of free samples for promotion, and states e-cigarette products must now require a health warning [94]. These restrictions highlight the need for continued research and monitoring of social media commercialization of these products and for this issue to be placed on public health and policy agendas.

Limitations

The review did not assess the quality of the evidence presented in each study, rather provided an overview regardless of quality as per the methodology outlined in the Manual for Scoping Reviews by JBI [26]. The search strategy included several popular terms used to describe e-cigarettes; however, keywords including emerging and variations of slang terms may have been overlooked and therefore, resulted in an incomplete retrieval of identified research. Furthermore, it is possible that additional studies relevant to the research question may have been identified if alternative databases were searched.

The reviewed material reflects a general bias toward certain social media platforms such as Twitter as its data are mostly public and easily accessible to researchers, whereas Facebook and other platforms are not [95]. This is not an indication that Facebook and other platforms are not spaces where e-cigarettes are discussed, but only that these activities are not visible to researchers.

Conclusions

The social media landscape is being dominated by provaping messages disseminated by the vaping industry and vaping proponents, whereas the uncertainty surrounding e-cigarette regulation expressed within the public health field appears not to be reflected in ongoing social media dialogues. Latest generation e-cigarettes are resembling less and less their first generation *cig-a-like* counterparts and are being promoted not only as a smoking cessation device and safer alternative to smoking but also as a recreational activity whereby you can create your own unique vaping experience with the use of flavors, device modification, and vape tricks. With the industry changing so rapidly, real-time monitoring and surveillance of how these devices are discussed, promoted, and used on social media is necessary in conjunction with evidence published in academic journals. The need for real-time monitoring and surveillance also highlights the need to close the chasm between research and practice [96]. Some government agencies have recognized and are attempting to bridge this gap by introducing research translation initiatives, annual conferences, education programs, and more varied communications [97,98] as they attempt to move evidence through the publication pipeline faster and more efficiently. However, Departments of Health may well have to start thinking about investing in real-time monitoring and surveillance to interact with the public to actively influence social media conversations and create a more balanced discussion with regard to e-cigarettes.

Acknowledgments

This work was supported by an Australian Government Research Training Program Scholarship. The Scholarship is provided by the Commonwealth of Australia to support general living costs for students undertaking Research Doctorate studies. The funder is not involved in any aspect of the project and has had no input in the interpretation or publication of the study results.

Authors' Contributions

KM, JJ, and BM conceptualized the research. KM drafted the manuscript, and JJ and BM aided in developing the research question and study methods. All authors contributed meaningfully to editing and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Grana R, Benowitz N, Glantz S. Center for Tobacco Control Research and Education. 2013. Background paper on e-cigarettes (electronic nicotine delivery systems) URL: https://cloudfront.escholarship.org/dist/prd/content/qt13p2b72n/qt13p2b72n.pdf?t=n53zve [accessed 2018-11-26] [WebCite Cache ID 71hgCoQCm]
2. Kennedy RD, Awopegba A, De León E, Cohen JE. Global approaches to regulating electronic cigarettes. Tob Control 2017;26(4):440-445 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053179] [Medline: 27903958]
3. Tobacco in Australia. Melbourne, Australia: Cancer Council Victoria; 2017. The e-cigarette market URL: http://www.tobaccoinaustralia.org.au/chapter-18-harm-reduction/18b-1-the-ecigarettemarket [accessed 2018-08-16] [WebCite Cache ID 71hgXbjyG]
4. BIS Research. Electronic cigarette & e vapor (vaporizer) market research reports. Fremont, California: BIS Research; 2016. URL: https://bisresearch.com/industry-report/electronic-cigarette-market-size-forecast.html [accessed 2018-08-16] [WebCite Cache ID 71hgwi5Tu]
5. Public Health England. E-cigarettes: An evidence update. A report commissioned by Public Health England. London, United Kingdom: Public Health England; 2015. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733022/Ecigarettes_an_evidence_update_A_report_commissioned_by_Public_Health_England_FINAL.pdf [accessed 2018-08-08] [WebCite Cache ID 71hh939UY]
6. Farsalinos KE, Polosa R. Safety evaluation and risk assessment of electronic cigarettes as tobacco cigarette substitutes: a systematic review. Ther Adv Drug Saf 2014;5(2):67-86 [FREE Full text] [doi: 10.1177/2042098614524430] [Medline: 25083263]
7. Bullen C, McRobbie H, Thornley S, Glover M, Lin R, Laugesen M. Effect of an electronic nicotine delivery device (e cigarette) on desire to smoke and withdrawal, user preferences and nicotine delivery: randomised cross-over trial. Tob Control 2010;19:98-103. [Medline: 2037858]
8. Hartmann-Boyce J, Begh R, Aveyard P. Electronic cigarettes for smoking cessation. Br Med J 2018;360:j5543. [Medline: 29343486]
9. Roditis M, Delucchi K, Cash D, Halpern-Felsher B. Adolescents' perceptions of health risks, social risks, and benefits differ across tobacco products. J Adolesc Health 2016;58(5):558-566 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.01.012] [Medline: 27107909]
10. Choi K, Forster JL. Beliefs and experimentation with electronic cigarettes: a prospective analysis among young adults. Am J Prev Med 2014;46(2):175-178 [FREE Full text] [doi: 10.1016/j.amepre.2013.10.007] [Medline: 24439352]
11. Best C, Haseen F, Currie D, Ozakinci G, MacKintosh AM, Stead M, et al. Relationship between trying an electronic cigarette and subsequent cigarette experimentation in Scottish adolescents: a cohort study. Tob Control 2017;27:373-378 [FREE Full text] [doi: 10.1136/tobaccocontrol-2017-053691] [Medline: 28735273]
12. World Health Organization. Moscow, Russia: WHO; 2014. Electronic nicotine delivery systems: Report by WHO URL: http://www.who.int/iris/handle/10665/147110 [accessed 2018-08-08] [WebCite Cache ID 71hhnjYHB]
13. Grana R, Benowitz N, Glantz SA. E-cigarettes: a scientific review. Circulation 2014;129(19):1972-1986 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.007667] [Medline: 24821826]
14. Alzahrani T, Pena I, Temesgen N, Glantz SA. Association between electronic cigarette use and myocardial infarction. Am J Prev Med 2018;55(4):455-461. [doi: 10.1016/j.amepre.2018.05.004] [Medline: 30166079]
15. Weaver SR, Huang J, Pechacek TF, Heath JW, Ashley DL, Eriksen MP. Are electronic nicotine delivery systems helping cigarette smokers quit? Evidence from a prospective cohort study of U.S. adult smokers, 2015-2016. PLoS One 2018;13(7):e0198047 [FREE Full text] [doi: 10.1371/journal.pone.0198047] [Medline: 29985948]
16. Clark EM, Jones CA, Williams JR, Kurti AN, Norotsky MC, Danforth CM, et al. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. PLoS One 2016;11(7):e0157304 [FREE Full text] [doi: 10.1371/journal.pone.0157304] [Medline: 27410031]

17. Ayers JW, Leas EC, Allem J, Benton A, Dredze M, Althouse BM, et al. Why do people use electronic nicotine delivery systems (electronic cigarettes)? A content analysis of Twitter, 2012-2015. PLoS One 2017;12(3):e0170702 [FREE Full text] [doi: 10.1371/journal.pone.0170702] [Medline: 28248987]
18. Jancey J, Maycock B, McCausland K, Howat P. E-Cigarettes: implications for health promotion in the Asian Pacific region. Asia Pac J Public Health 2018;30(4):321-327. [doi: 10.1177/1010539518762855] [Medline: 29521153]
19. Tobacco Tactics. Tobacco Tactics. 2018. E-cigarettes URL: http://www.tobaccotactics.org/index.php/E-cigarettes [accessed 2018-08-16] [WebCite Cache ID 71hi5J27h]
20. Chapman M. Campaign Live. 2014. Wolves in sheep's clothing? Big Tobacco and its takeover of the e-cigarette market URL: https://www.campaignlive.co.uk/article/wolves-sheeps-clothing-big-tobacco-its-takeover-e-cigarette-market/1281851 [accessed 2018-08-08] [WebCite Cache ID 71hmcfeo2]
21. IFF Research. United Kingdom: HM Revenue and Customs; 2016. Understanding the online e-cigarette market URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/565928/Understanding_Online_Ecigarette_market_HMRC-research.pdf [accessed 2018-08-08] [WebCite Cache ID 71hiuUvoT]
22. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014;23(Suppl 3):iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]
23. de Andrade M, Hastings G, Angus K, Dixon D, Purves R. Cancer Research UK. 2013. The marketing of electronic cigarettes in the UK URL: https://www.cancerresearchuk.org/sites/default/files/cruk_marketing_of_electronic_cigs_nov_2013.pdf [accessed 2018-11-26] [WebCite Cache ID 71hjRexM2]
24. Cancer Research UK. 2014. The marketing of electronic cigarettes in the UK URL: https://www.cancerresearchuk.org/sites/default/files/cruk_marketing_of_electronic_cigs_nov_2013.pdf [accessed 2018-08-16] [WebCite Cache ID 71hjYOo64]
25. McCausland K, Maycock B, Jancey J. The messages presented in online electronic cigarette promotions and discussions: a scoping review protocol. BMJ Open 2017;7(11):e018633 [FREE Full text] [doi: 10.1136/bmjopen-2017-018633] [Medline: 29122804]
26. The Joanna Briggs Institute. The Joanna Briggs Institute. 2015. Methodology for JBI scoping reviews URL: http://joannabriggs.org/assets/docs/sumari/Reviewers-Manual_Methodology-for-JBI-Scoping-Reviews_2015_v2.pdf [accessed 2018-11-26] [WebCite Cache ID 71hjiNpxO]
27. International Tobacco Control Policy Evaluation Project. International Tobacco Control Policy Evaluation Project. Ontario, Canada: Department of Psychology, University of Waterloo; 2016. Countries URL: http://www.itcproject.org/countries [accessed 2018-11-26] [WebCite Cache ID 71hkQyLbZ]
28. Consumer Advocates for Smoke Free Alternatives Association. 2017. Historical timeline of electronic cigarettes URL: http://casaa.org/historical-timeline-of-electronic-cigarettes/ [accessed 2018-11-26] [WebCite Cache ID 71hkY5rRH]
29. Clarivate Analytics. 2016. URL: https://endnote.com/ [accessed 2018-11-28] [WebCite Cache ID 74GSYYIuU]
30. Covidence. 2018. URL: https://www.covidence.org/home [accessed 2018-11-26] [WebCite Cache ID 71hkmjsyC]
31. Allem J, Escobedo P, Chu K, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. Tob Control 2017;26(2):226-229. [doi: 10.1136/tobaccocontrol-2015-052757] [Medline: 26956467]
32. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. J Med Internet Res 2014;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
33. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014;23(Suppl 3):iii26-iii30 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]
34. Kim AE, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, et al. Using Twitter data to gain insights into e-cigarette marketing and locations of use: an infoveillance study. J Med Internet Res 2015;17(10):e251 [FREE Full text] [doi: 10.2196/jmir.4466] [Medline: 26545927]
35. Han S, Kavuluru R. Exploratory analysis of marketing and non-marketing e-cigarette themes on Twitter. In: Social Informatics. Cham: Springer; 2016:307-322.
36. Jo CL, Kornfield R, Kim Y, Emery S, Ribisl KM. Price-related promotions for tobacco products on Twitter. Tob Control 2016;25(4):476-479. [doi: 10.1136/tobaccocontrol-2015-052260] [Medline: 26130449]
37. Sowles SJ, Krauss MJ, Connolly S, Cavazos-Rehg PA. A content analysis of vaping advertisements on Twitter, November 2014. Prev Chronic Dis 2016;13:E139 [FREE Full text] [doi: 10.5888/pcd13.160274] [Medline: 27685432]
38. Unger JB, Escobedo P, Allem J, Soto DW, Chu K, Cruz T. Perceptions of secondhand e-cigarette aerosol among Twitter users. Tob Regul Sci 2016;2(2):146-152 [FREE Full text] [doi: 10.18001/TRS.2.2.5] [Medline: 28090560]
39. Burke-Garcia A, Stanton C. A tale of two tools: reliability and feasibility of social media measurement tools examining e-cigarette twitter mentions. IMU 2017;8:8-12. [doi: 10.1016/j.imu.2017.04.001]
40. Lazard AJ, Wilcox GB, Tuttle HM, Glowacki EM, Pikowski J. Public reactions to e-cigarette regulations on Twitter: a text mining analysis. Tob Control 2017;26(e2):e112-e116. [doi: 10.1136/tobaccocontrol-2016-053295] [Medline: 28341768]
41. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. J Med Internet Res 2015;17(10):e243 [FREE Full text] [doi: 10.2196/jmir.4969] [Medline: 26508089]

42. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. Int J Public Health 2016;61(2):249-256. [doi: 10.1007/s00038-016-0791-2] [Medline: 26841895]
43. Kavuluru R, Sabbir AK. Toward automated e-cigarette surveillance: spotting e-cigarette proponents on Twitter. J Biomed Inform 2016;61:19-26 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.006] [Medline: 26975599]
44. Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, Bernhardt JM. E-cigarette social media messages: a text mining analysis of marketing and consumer conversations on Twitter. JMIR Public Health Surveill 2016;2(2):e171 [FREE Full text] [doi: 10.2196/publichealth.6551] [Medline: 27956376]
45. Dai H, Hao J. Mining social media data for opinion polarities about electronic cigarettes. Tob Control 2017;26(2):175-180. [doi: 10.1136/tobaccocontrol-2015-052818] [Medline: 26980151]
46. Prochaska JJ, Pechmann C, Kim R, Leonhardt JM. Twitter=quitter? An analysis of Twitter quit smoking social networks. Tob Control 2012;21(4):447-449 [FREE Full text] [doi: 10.1136/tc.2010.042507] [Medline: 21730101]
47. Sears CG, Walker KL, Hart JL, Lee AS, Siu A, Smith C. Clean, cheap, convenient: promotion of Electronic cigarettes on YouTube. Tob Prev Cessat 2017;3:1-7 [FREE Full text] [doi: 10.18332/tpc/69393] [Medline: 28725876]
48. Basch CH, Mongiovi J, Hillyer GC, MacDonald Z, Basch CE. YouTube videos related to e-cigarette safety and related health risks: implications for preventing and emerging epidemic. Public Health 2016;132:57-59. [doi: 10.1016/j.puhe.2015.12.003] [Medline: 26826891]
49. Merianos AL, Gittens OE, Mahabee-Gittens EM. Depiction of health effects of electronic cigarettes on YouTube. J Subst Use 2016;21(6):614-619 [FREE Full text] [doi: 10.3109/14659891.2015.1118565] [Medline: 28217030]
50. Huang J, Kornfield R, Emery SL. 100 million views of electronic cigarette YouTube videos and counting: quantification, content evaluation, and engagement levels of videos. J Med Internet Res 2016;18(3):e67 [FREE Full text] [doi: 10.2196/jmir.4265] [Medline: 26993213]
51. Lee AS, Hart JL, Sears CG, Walker KL, Siu A, Smith C. A picture is worth a thousand words: electronic cigarette content on Instagram and Pinterest. Tob Prev Cessat 2017;3:119 [FREE Full text] [doi: 10.18332/tpc/74709] [Medline: 28815224]
52. Chu K, Allem J, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2017;26:575-578. [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
53. Laestadius LI, Wahl MM, Cho YI. #Vapelife: an exploratory study of electronic cigarette use and promotion on Instagram. Subst Use Misuse 2016;51(12):1669-1673. [doi: 10.1080/10826084.2016.1188958] [Medline: 27484191]
54. Nelson L, Burk D, Knudsen M, McCall L. Sociol Methods Res. 2018. The future of coding: a comparison of hand-coding and three types of computer-assisted text analysis methods URL: https://pdfs.semanticscholar.org/c56d/e1efda6358f535d07ee07051b80cff9b9c74.pdf?_ga=2.134716721.1990280664.1543398844-1967781214.1525692408 [accessed 2018-11-28] [WebCite Cache ID 74GT3Y3pP]
55. Lienemann BA, Unger JB, Cruz TB, Chu K. Methods for coding tobacco-related Twitter data: a systematic review. J Med Internet Res 2017;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]
56. Steyn P, Ewing M, van Heerden G, Pitt L, Windisch L. From whence it came. Int J Advert 2015;30(1):133-160. [doi: 10.2501/IJA-30-1-133-160]
57. Romito LM, Hurwich RA, Eckert GJ. A snapshot of the depiction of electronic cigarettes in YouTube videos. Am J Health Behav 2015;39(6):823-831. [doi: 10.5993/AJHB.39.6.10] [Medline: 26450550]
58. Westerman D, Spence P, Van Der Heide B. Social media as information source: recency of updates and credibility of information. J Comput-Mediat Comm 2013;19(2):171-183. [doi: 10.1111/jcc4.12041]
59. Dubois E, Gaffney D. The multiple facets of influence: identifying political influentials and opinion leaders on Twitter. Am Behav Sci 2014;58(10):1260-1277. [doi: 10.1177/0002764214527088]
60. Elkin L, Thomson G, Wilson N. Connecting world youth with tobacco brands: YouTube and the internet policy vacuum on Web 2.0. Tob Control 2010;19(5):361-366. [doi: 10.1136/tc.2010.035949] [Medline: 20739706]
61. Schraufnagel DE. Electronic cigarettes: vulnerability of youth. Pediatr Allergy Immunol Pulmonol 2015;28(1):2-6 [FREE Full text] [doi: 10.1089/ped.2015.0490] [Medline: 25830075]
62. Berg CJ, Stratton E, Schauer GL, Lewis M, Wang Y, Windle M, et al. Perceived harm, addictiveness, and social acceptability of tobacco products and marijuana among young adults: marijuana, hookah, and electronic cigarettes win. Subst Use Misuse 2015;50(1):79-89 [FREE Full text] [doi: 10.3109/10826084.2014.958857] [Medline: 25268294]
63. Amrock SM, Zakhar J, Zhou S, Weitzman M. Perception of e-cigarette harm and its correlation with use among U.S. adolescents. Nicotine Tob Res 2015;17(3):330-336 [FREE Full text] [doi: 10.1093/ntr/ntu156] [Medline: 25125321]
64. Soule EK, Lopez AA, Guy MC, Cobb CO. Reasons for using flavored liquids among electronic cigarette users: a concept mapping study. Drug Alcohol Depend 2016;166:168-176 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.07.007] [Medline: 27460860]
65. Physicians for a Smoke-Free Canada. Ottawa: FASC; 2009. Flavoured tobacco: How the industry baits its prey URL: http://www.smoke-free.ca/pdf_1/2009/Flavoured-Jun2.pdf [accessed 2018-11-26] [WebCite Cache ID 71hktzwtn]
66. National Cancer Institute. 2008. The role of the media in promoting and reducing tobacco use URL: https://cancercontrol.cancer.gov/brp/tcrb/monographs/19/m19_complete_accessible.pdf [WebCite Cache ID 71hl0g5So]

67. Lewis MJ, Wackowski O. Dealing with an innovative industry: a look at flavored cigarettes promoted by mainstream brands. Am J Public Health 2006;96(2):244-251. [doi: 10.2105/AJPH.2004.061200] [Medline: 16380563]
68. US Department of Health and Human Services. US Food & Drug Administration. 2018. Menthol and other flavors in tobacco products URL: https://www.fda.gov/TobaccoProducts/Labeling/ProductsIngredientsComponents/ucm2019416.htm [accessed 2018-11-26] [WebCite Cache ID 71hl9Axvk]
69. Farsalinos KE, Romagna G, Tsiapras D, Kyrzopoulos S, Spyrou A, Voudris V. Impact of flavour variability on electronic cigarette use experience: an internet survey. Int J Environ Res Public Health 2013;10(12):7272-7282 [FREE Full text] [doi: 10.3390/ijerph10127272] [Medline: 24351746]
70. Dawkins L, Turner J, Roberts A, Soar K. 'Vaping' profiles and preferences: an online survey of electronic cigarette users. Addiction 2013;108(6):1115-1125. [doi: 10.1111/add.12150] [Medline: 23551515]
71. Kong G, Morean ME, Cavallo DA, Camenga DR, Krishnan-Sarin S. Reasons for electronic cigarette experimentation and discontinuation among adolescents and young adults. Nicotine Tob Res 2015;17(7):847-854 [FREE Full text] [doi: 10.1093/ntr/ntu257] [Medline: 25481917]
72. Kim H, Lim J, Buehler SS, Brinkman MC, Johnson NM, Wilson L, et al. Role of sweet and other flavours in liking and disliking of electronic cigarettes. Tob Control 2016;25(Suppl 2):ii55-ii61 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053221] [Medline: 27708124]
73. Ambrose BK, Day HR, Rostron B, Conway KP, Borek N, Hyland A, et al. Flavored tobacco product use among us youth aged 12-17 years, 2013-2014. J Am Med Assoc 2015;314(17):1871-1873. [doi: 10.1001/jama.2015.13802] [Medline: 26502219]
74. Dwyer JB, McQuown SC, Leslie FM. The dynamic effects of nicotine on the developing brain. Pharmacol Ther 2009;122(2):125-139 [FREE Full text] [doi: 10.1016/j.pharmthera.2009.02.003] [Medline: 19268688]
75. Morean ME, Butler ER, Bold KW, Kong G, Camenga DR, Cavallo DA, et al. Preferring more e-cigarette flavors is associated with e-cigarette use frequency among adolescents but not adults. PLoS One 2018;13(1):e0189015 [FREE Full text] [doi: 10.1371/journal.pone.0189015] [Medline: 29300749]
76. Pepper JK, Ribisl KM, Brewer NT. Adolescents' interest in trying flavoured e-cigarettes. Tob Control 2016;25(Suppl 2):ii62-ii66 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053174] [Medline: 27633762]
77. Reach of leading social media and networking sites used by teenagers and young adults in the United States as of February. New York, United States of America: Statista, Inc; 2018. URL: https://www.statista.com/statistics/199242/social-media-and-networking-sites-used-by-us-teenagers/ [accessed 2018-11-26] [WebCite Cache ID 71hlCn40c]
78. Emery SL, Vera L, Huang J, Szczypka G. Wanna know about vaping? Patterns of message exposure, seeking and sharing information about e-cigarettes across media platforms. Tob Control 2014;23(Suppl 3):iii17-iii25 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051648] [Medline: 24935893]
79. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. Am J Prev Med 2011;40(4):448-453. [doi: 10.1016/j.amepre.2010.12.007] [Medline: 21406279]
80. Cheney MK, Gowin M, Wann TF. Vapor store owner beliefs and messages to customers. Nicotine Tob Res 2016;18(5):694-699. [doi: 10.1093/ntr/ntv129] [Medline: 26113622]
81. Cornelius ME, Driezen P, Fong GT, Chaloupka FJ, Hyland A, Bansal-Travers M, et al. Trends in the use of premium and discount cigarette brands: findings from the ITC US Surveys (2002-2011). Tob Control 2014;23(Suppl 1):i48-i53 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051045] [Medline: 24092600]
82. Cummings KM, Hyland A, Lewit E, Shopland D. Use of discount cigarettes by smokers in 20 communities in the United States, 1988-1993. Tob Control 1997;6(Suppl 2):S25-S30 [FREE Full text] [Medline: 9583649]
83. Chaloupka FJ, Cummings KM, Morley CP, Horan JK. Tax, price and cigarette smoking: evidence from the tobacco documents and implications for tobacco company marketing strategies. Tob Control 2002;11(Suppl 1):I62-I72 [FREE Full text] [Medline: 11893816]
84. World Health Organization. Geneva, Switzerland: WHO; 2003. WHO Framework Convention on Tobacco Control URL: http://apps.who.int/iris/bitstream/handle/10665/42811/9241591013.pdf?sequence=1 [accessed 2018-11-26] [WebCite Cache ID 71hlVcYTw]
85. Huang J, Tauras J, Chaloupka FJ. The impact of price and tobacco control policies on the demand for electronic nicotine delivery systems. Tob Control 2014;23(Suppl 3):iii41-iii47 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051515] [Medline: 24935898]
86. Campbell M, Ford C, Winstanley M. Tobacco in Australia. Melbourne, Australia: Cancer Council Victoria; 2017. The health effects of secondhand smoke URL: http://www.tobaccoinaustralia.org.au/chapter-4-secondhand/4-2-what-is-in-secondhand-smoke/ [accessed 2018-11-26] [WebCite Cache ID 71hlhT1MI]
87. World Health Organization. Geneva, Switzerland: WHO; 2007. Guidelines on the protection from exposure to tobacco smoke URL: http://www.who.int/fctc/cop/art%208%20guidelines_english.pdf [accessed 2018-11-26] [WebCite Cache ID 71hmASTEb]
88. Tan AS, Bigman CA, Sanders-Jackson A. Sociodemographic correlates of self-reported exposure to e-cigarette communications and its association with public support for smoke-free and vape-free policies: results from a national survey

of US adults. Tob Control 2015;24(6):574-581 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051685] [Medline: 25015372]
89. Lerman K, Yan X, Wu X. The "majority illusion" in social networks. PLoS One 2016;11(2):e0147617 [FREE Full text] [doi: 10.1371/journal.pone.0147617] [Medline: 26886112]
90. Highfield T, Leaver T. A methodology for mapping Instagram hashtags. First Monday 2015;20(1):1-11. [doi: 10.5210/fm.v20i1.5563]
91. Zappavigna M. Enacting identity in microblogging through ambient affiliation. Discourse Commun 2013;8(2):209-228. [doi: 10.1177/1750481313510816]
92. McMillen RC, Gottlieb MA, Shaefer RM, Winickoff JP, Klein JD. Trends in electronic cigarette use among U.S. adults: use is increasing in both smokers and nonsmokers. Nicotine Tob Res 2014;17(10):1195-1202. [doi: 10.1093/ntr/ntu213] [Medline: 25381306]
93. Measham F, O'Brien K, Turnbull G. "Skittles & Red Bull is my favourite flavour": e-cigarettes, smoking, vaping and the changing landscape of nicotine consumption amongst British teenagers – implications for the normalisation debate. Drug-Educ Prev Polic 2016;23(3):224-237. [doi: 10.1080/09687637.2016.1178708]
94. US Department of Health and Human Services. US Food & Drug Administration. 2017. The facts on the FDA's new tobacco rule URL: https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm506676.htm [accessed 2018-11-26] [WebCite Cache ID 71hlS8rv8]
95. Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C. Peter Lang. 2014. Twitter and society URL: https://tinyurl.com/y74uq2uu [accessed 2018-11-26] [WebCite Cache ID 71hmK81WG]
96. Green LW. Closing the chasm between research and practice: evidence of and for change. Health Promot J Austr 2014;25(1):25-29. [doi: 10.1071/HE13101] [Medline: 24666557]
97. National Health and Medical Research Council. 2000. How to put the evidence into practice: Implementation and dissemination strategies URL: https://nhmrc.gov.au/about-us/publications/how-put-evidence-practice-implementation-and-dissemination-strategies#block-views-block-file-attachments-content-block-1 [accessed 2018-11-26] [WebCite Cache ID 71hmOvfIu]
98. Public Health England. 2015. Doing, supporting and using public, health research. The Public Health England Strategy for Research, Translation and Innovation URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/453993/Research_translation_and_innovation_strategy_final.pdf [accessed 2018-11-26] [WebCite Cache ID 71hmTkx9T]

Abbreviations

e-cigarette: electronic cigarette
e-juice: electronic cigarette juice
FDA: Food and Drug Administration
JBI: Joanna Briggs Institute

Edited by G Eysenbach; submitted 16.08.18; peer-reviewed by E Clark, J Colditz, JP Allem; comments to author 13.10.18; revised version received 24.10.18; accepted 25.10.18; published 05.02.19.

Please cite as:
McCausland K, Maycock B, Leaver T, Jancey J
The Messages Presented in Electronic Cigarette–Related Social Media Promotions and Discussion: Scoping Review
J Med Internet Res 2019;21(2):e11953
URL: https://www.jmir.org/2019/2/e11953/
doi:10.2196/11953
PMID:30720440

Review

Digital Interventions to Reduce Sedentary Behaviors of Office Workers: Scoping Review

Yitong Huang[1,2], MSc; Steve Benford[2], PhD; Holly Blake[3,4], PhD

[1]Horizon Centre for Doctoral Training, University of Nottingham, Nottingham, United Kingdom

[2]Mixed Reality Laboratory, School of Computer Science, University of Nottingham, Nottingham, United Kingdom

[3]School of Health Sciences, University of Nottingham, Nottingham, United Kingdom

[4]National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham, United Kingdom

Corresponding Author:
Yitong Huang, MSc
Horizon Centre for Doctoral Training
University of Nottingham
B06, Computer Science
Wollaton Road
Nottingham, NG8 1BB
United Kingdom
Phone: 44 115 823255
Email: yitong.huang@nottingham.ac.uk

Abstract

Background: There is a clear public health need to reduce office workers' sedentary behaviors (SBs), especially in the workplace. Digital technologies are increasingly being deployed in the workplace to measure and modify office workers' SBs. However, knowledge of the range and nature of research on this topic is limited; it also remains unclear to what extent digital interventions have exploited the technological possibilities.

Objective: This study aimed to investigate the technological landscape of digital interventions for SB reduction in office workers and to map the research activity in this field.

Methods: Terms related to SB, office worker, and digital technology were applied in various combinations to search Cochrane Library, Joanna Briggs Institute Database of Systematic Reviews, MEDLINE, PsycARTICLES, PsycINFO, Scopus, Association for Computing Machinery Digital Library, Engineering index Compendex, and Google Scholar for the years 2000 to 2017. Data regarding the study and intervention details were extracted. Interventions and studies were categorized into development, feasibility and/piloting, evaluation, or implementation phase based on the UK Medical Research Council (MRC) framework for developing and evaluating complex interventions. A novel framework was developed to classify technological features and annotate technological configurations. A mix of quantitative and qualitative approaches was used to summarize data.

Results: We identified 68 articles describing 45 digital interventions designed to intervene with office workers' SB. A total of 6 common technological features had been applied to interventions with various combinations. Configurations such as "information delivery and mediated organizational and social support" and "digital log and automated tailored feedback" were well established in evaluation and implementation studies; in contrast, the integration of passive data collection, connected devices, and ATF or scheduled prompts was mostly present in development and piloting research.

Conclusions: This review is the first to map and describe the use of digital technologies in research on SB reduction in office workers. Interdisciplinary collaborations can help to maximize the potential of technologies. As novel modes of delivery that capitalize on embedded computing and electronics, wireless technologies have been developed and piloted in engineering, computing, and design fields, efforts can be directed to move them to the next phase of evaluation with more rigorous study designs. Quality of research may be improved by fostering conversations between different research communities and encouraging researchers to plan, conduct, and report their research under the MRC framework. This review will be particularly informative to those deciding on areas where further research or development is needed and to those looking to locate the relevant expertise, resources, and design inputs when designing their own systems or interventions.

(J Med Internet Res 2019;21(2):e11079) doi:10.2196/11079

KEYWORDS

telemedicine; sedentary behavior; workplace; technology; internet; microcomputers

Introduction

Background

Sedentary behaviors (SBs) are activities that require very low energy expenditure of less than 1.5 metabolic equivalents and typically involve lying down and sitting [1]. Excessive SB is recognized as an exposure to a risk factor different from a lack of moderate to vigorous physical activity (MVPA), as an individual who engages in 150 min of exercise every week can still spend the majority of the remaining waking hours in SB. Reducing SB may require approaches very different from those required to increase physical activity (PA), as sedentary time can accumulate unintentionally in a broad range of contexts such as during leisure time, transportation, and in the workplace. Although a recent meta-analysis [2]indicates that 60 to 75 min of MVPA per day seems to offset the increased risks of mortality associated with sitting for more than 8 hours per day, this amount of MVPA is notably beyond the recommended levels of MVPA in most public health guidelines [3,4]. More importantly, mounting evidence suggests reducing SB, especially prolonged episodes of SB, has its own benefits on metabolic and musculoskeletal health, and potentially on other health conditions [5-8].

A number of studies [9-11] have found that office workers' within-work time is characterized by more prolonged SB with fewer breaks than nonwork time; sedentary work contributes significantly to overall sedentary exposure of office workers and, thus, the health risks. A recent statement by an international panel of experts highlighted the need for interventions that target the reduction of prolonged SB in this setting and population, for both better health and productivity outcomes [12]. In this paper, we focus on a potential solution: SB interventions delivered with digital technologies.

According to the Oxford Dictionary, digital technologies refer to technologies involving or relating to the use of computer technology, which includes tools, systems, devices, and resources that generate, store, or process data in the form of digital signals. The past decades have seen an exponential growth of computing power at affordable prices. This has resulted in an increasing variety of digital gadgets (eg, personal computer, tablets, smartphones, wearables, and Internet of Things) that a person is exposed to and interacts with on a day-to-day basis. This presents health intervention designers and researchers with a wider range of device choices that offer different form factors and features. Indeed, digital health has demonstrated great promise in a range of clinical settings and populations in terms of behavioral measurement and intervention delivery (eg, pediatric care [13] and mental health [14]).

However, when it comes to digital SB interventions, the behavioral target of "being less sedentary" and the use of digital media seem to present us with a paradox here. First, the increase in sedentary occupations and sedentariness at work in itself is closely related to the evolution of digital technology, which enables more work to be completed at desks without manual labor or even light PA. Second, a recent study [15] has found that information and communication technologies (ICTs) have supported new break activities in the office (eg, checking social media during mini-breaks or watching videos over lunch breaks) while evoking negative feelings at the same time. The researchers used the term "screen guilt" to describe office workers' need to disconnect from screen-based ICTs during breaks for both physical and psychological well-being.

This has led us to rethink what (or even whether) digital features should be incorporated when designing SB interventions. The intersection of digital health and SB has attracted a lot of research interest and accumulated a large body of interdisciplinary research in recent years. As a first step in our own research on exploiting novel digital technologies for the delivery of workplace SB interventions, we wanted to review the literature on this topic in a systematic manner, to map the current technological landscape and research activities conducted in different disciplines, and to determine research gaps in terms of utilizing and innovating technologies for workplace SB interventions.

Previous Reviews

To date, 7 systematic reviews on SB interventions targeting adults have been published [16-22]. This section overviews which aspects of the topic have been addressed in those reviews and which areas require more secondary research.

All the reviews were inclusive of all SB reduction interventions regardless of the presence of digital elements. Chau and colleagues [16] reviewed workplace studies published up to April 2009 and identified only 6 eligible studies that included sitting as an outcome measure. Only 2 types of digital media were covered (emails [23-25] and pedometer [23]). Measurement of SB was self-reported in all 6 studies, none of which found significant intervention effect on sitting reduction. The result was inconclusive with respect to the most appropriate intervention approach or delivery mode because of disparate study designs and delivery modes across studies. With a similar inclusion criterion as Chau and colleagues' [16], a more recent review [17] by Shrestha and colleagues identified 20 eligible workplace studies published up to June 2015. The analysis was focused on comparing the effects of different intervention components with absence of these components or alternative components. Only a small part of the analysis was pertinent to digital interventions. First, it compared the effect of computer prompts plus information counseling on sitting reduction with information counseling only, based on data from 3 studies [26-28]. Second, it compared the effect of different contents in e-newsletters on sitting reduction, based on 1 study [29]. The findings from both analyses were nonsignificant or inconclusive, given the low quality of evidence. Commissaris and colleagues [18] specifically reviewed workplace SB interventions aimed to influence workers' SB while doing productive work. As a small part of their analyses, they compared 6 interventions including self-monitoring of SB and/or PA (using devices such as pedometers) with 4 interventions not including

self-monitoring and suggested that self-monitoring seemed to be ineffective in improving SB/PA at work. Another review of workplace SB interventions by Chu and colleagues [19] included 26 studies published up to December 2015 and classified them based on intervention strategies into 3 categories: (1) environmental strategies, (2) educational/ behavioral strategies (involving educational program and point-of-choice motivational signs), and (3) combined strategies. They concluded from subgroup analyses that interventions combining multiple components resulted in the greatest sitting reduction, followed by environmental strategies. However, the review did not distinguish digital and nondigital delivery of intervention strategies within each category. Similar to Chu and colleagues' review [19], Gardener and colleagues' review [30] was also focused on intervention strategies, but with a broader scope (ie, including nonworkplace studies) and a more fine-grained coding scheme based on the underlying intervention functions [31] and behavior change techniques (BCTs) [32]. They found that education, persuasion, environmental restructuring, and training were the most promising intervention functions and that self-monitoring, problem solving, and changing the social or physical environment were particularly promising BCTs for reducing SB. Martin and colleagues' review [21] was also inclusive of nonworkplace interventions. It was suggested that interventions targeting SB only and lifestyle change might be more promising than those targeting PA only or a combination of PA and SB, which was similar to the conclusion reached in Prince and colleagues' review [22].

Although shedding light on intervention strategies and components effective for reducing workplace SB, those reviews fell short in 2 aspects.

First, they did not differentiate diverse ways an intervention strategy/component could be digitally implemented and delivered. For instance, for the same strategy of point-of-choice prompts, the actual quantities of prompts received and noticed by participants may differ significantly depending on whether the break reminder was delivered on workstation screens, by smartphone notifications, or via tactile feedback from wearable devices. Apart from specific technological features, how different features were applied in combination and in support of each other is also worthy of attention. For instance, just-in-time adaptive intervention (JITAI), an approach that employs context-aware sensing and computing to detect the behavioral context and tailor the intervention in real time, can address the dynamically changing needs of individuals much better than a traditional intervention delivering static content with a fixed schedule [33]. Knowledge of such nuances in technological design is important as they may lead to considerable difference in the quality and quantity of interventions delivered to participants, making outcomes incomparable across studies.

Second, none of the abovementioned reviews included the engineering and computer science literature, despite the rapid prototyping and piloting of novel technologies within these fields that may become or inform the next generation of digital interventions. An exploratory search of this body of literature has found an abundance of user-centered design research [34] on technologies targeting SB reduction in office workers. Those studies, although employing very different study designs from clinical trials, have gathered valuable data about design-related outcomes (technological feasibility, usability, and acceptability) usually by involving stakeholders from the outset of intervention development. The findings do not only inform technology design but also give an indication of the potential user uptake, attitude, and adherence to different intervention technologies should they be moved to later stages of development and evaluation. As yet, awareness of the size and location of this body of evidence is lacking.

This Review

In summary, although previous reviews have touched on the technological design in SB interventions, there is a need for a review that is dedicated to this topic and that encompasses a wider range of literature. Specifically, the following questions can be explored:

1. How have digital technologies been used in interventions to reduce office workers' SB at work?
2. What research has been done on them and what development phases have they reached?
3. Where does the research gap lie as to utilizing and innovating digital technologies for SB interventions targeting office workers?

In view of the above, we selected the approach of scoping review, which is a particularly useful tool to synthesize findings established with different study designs and to address broader topics than those addressed by systematic reviews (eg, effectiveness) [35].

The review will be reported with the following structure. Considering the complexity of this topic, we will first review existing classifications and frameworks proposed from several disciplines to describe digital technologies for behavior change. Second, we present the search and review method. In the Results section, we first provide a quantitative summary of studies and interventions identified in this review. Then, we narratively overview the range of research conducted on interventions with different technological designs and summarize the findings pertinent to the technological features. Finally, we discuss findings and suggest avenues for future research.

This review is not aimed to estimate the efficacy of interventions with or without certain digital components, which should be addressed by further systematic reviews once the technological landscape is laid out. Neither is this review focused on comparing the capabilities and limitations of various brands of technological devices, which have been featured in other studies [36,37]. Instead, the main objective here is to scope research across different fields through review of the technological features present in interventions and mapping different research activities (eg, design-led research, feasibility studies, and experimental studies) onto different stages in the process of intervention development and evaluation. Another objective is to synthesize the design-related findings (eg, satisfaction, usability, acceptability, feasibility, and engagement) of digital interventions, which were overlooked in previous reviews.

Existing Frameworks and Classifications for Digital Health Technologies

The technological aspect of digital health has been discussed under several umbrella terms such as persuasive technology (PT)/system [38,39], behavioral intervention technology (BIT) [40], and mode of delivery (MoD) for behavior change interventions (BCIs) [41]. Here, we review frameworks that categorize digital health technologies based on physical manifestations and functions (both high-level functional roles and specific system features).

On the Basis of Physical Manifestations

PT, a technology intentionally designed to change a person's attitude or behavior, has been categorized into desktop-based, artifact-based, and environment-based systems, based on form factors [42]. Desktop-based systems are those only accessible through traditional personal computers and include Web pages and emails designed for desktop viewing and computer software. Artifact-based systems are usually portable and may include smartphones, wearable devices, and physically embodied agents, such as robot toys. Environment-based systems refer to computing systems built into the physical space or fixed to facilities to capture behaviors of users of the space or facility and to deliver point-of-choice persuasions, such as a system built into a public restroom to detect and encourage handwashing-with-soap behaviors of all toilet users [43].

On the Basis of Roles and Functions

The functional triad of PT [39] describes 3 general roles a computer can play in its interaction with the user, namely, a tool that increases user abilities, a medium that delivers content to create experience, and a social actor that evokes social responses especially with animate characteristics.

More recently, detailed system functionalities have been identified that explicitly or implicitly support those roles. For instance, the persuasive system design (PSD) model [38] suggested design principles under the following 4 categories: (1) primary task support, which includes reducing complex behaviors into simpler ones, tunneling experience, tailoring and personalization, self-monitoring, simulation, and rehearsal; (2) dialogue support, including positive reinforcement, reminders, suggestions, similarity, liking, and social role; (3) credibility, including expertise, authority, and trustworthiness; and (4) social support, by mediating social interactions and social influences. Some of these principles correspond to functional roles in the functional triad. For example, the principle of "reduction (ie, reducing complex behavior into simple tasks helps users perform the target behavior) and "self-monitoring (ie, providing means for users to track their performance or status) both enable the system to play the role of a tool. The principle of "simulation (ie, enable users to immediately observe the link between cause and effect) and "social facilitation (ie, providing means for discerning other users who are performing the same behavior) support the role of a medium; the principle of "social role (ie, adopt a virtual social role) can be directly mapped onto the role of a social actor in the functional triad. It should be noted that although PSD has the merit of supporting requirement engineering, it does not follow a clear hierarchical structure, and the design principles are a combination of behavior change strategies (eg, self-monitoring), functional elements (eg, simulation), and nonfunctional characteristics (eg, similarity and credibility).

Webb and colleagues [41] developed a novel scheme to code modes of delivering internet-based health BCIs into 3 broad categories: (1) automated functions, including the use of an enriched information environment, automated tailored feedback on progress, and automated follow-up reminders and tips; (2) communicative functions, including mediating communication with advisors and peers; and (3) use of supplementary modes. Similar concepts were termed as BIT elements by Mohr and colleagues [40], referring to actual technical instantiations in the intervention that the user interacts with. In addition to those functional components included in Webb's coding scheme, Mohr and colleagues [40] listed BIT elements appearing in more recent apps such as passive data collection (PDC; ie, data collected with smartphone sensors or external devices or through application programming interfaces [APIs] from other available sources) and logs (ie, data entry field facilitating self-monitoring).

All the abovementioned frameworks will be considered with adaptations wherever necessary in our analysis of the technological aspects of interventions to be reviewed.

Methods

Search and Selection

An interdisciplinary literature search was conducted of the following databases: Ovid MEDLINE, Cochrane library, JBI database of systematic reviews, Association for Computing Machinery digital library and Engineering index Compendex. Table 1 lists the databases searched in each field.

Synonyms and subject headings relating to the following terms were applied in various combinations: *office worker, sedentary behavior, technology, workplace* (see Multimedia Appendix 1 for example search strategy). Reference lists of existing reviews [16-22] on workplace SB reduction and PA promotion were hand searched to identify additional eligible studies.

Title, abstracts, and full text of retrieved articles were reviewed for eligibility by applying the following criteria: (1) having office workers in the study sample; (2) targeting SB during work or had proxy measures of workplace SB (objective and/or self-report daily sitting of office workers); (3) involving digital technologies such as mobile and computer apps, digital multimedia contents, wearable activity trackers, and other devices with sensing and computing capabilities in the production, delivery, and/or customization of intervention contents; (4) published in peer-reviewed scientific journals/conference proceedings between 2000 and 2017; and (5) published in the English language.

Table 1. Databases searched in each field.

Fields	Databases
Medical and health sciences	Ovid MEDLINE, Cochrane library, and JBI[a] database of systematic reviews
Computing and engineering	Association for Computing Machinery digital library and Engineering index Compendex
Interdisciplinary	Scopus

[a]JBI: Joanna Briggs Institute.

Observational studies without administering or developing any intervention were excluded, though design research with an explicit intent to inform the development of digital SB interventions was included. Studies were also excluded if digital technologies were only used for purposes other than intervention delivery, such as using digital tools for pre-and poststudy assessments without feeding the data into the intervention content in any way.

Data Extraction

Full articles of eligible studies were reviewed to extract the following information where possible: publication data (authors, years, countries where the study was conducted, or where the first author was based if the study country was not specified), primary target behavior (SB vs PA vs others), intervention details, study details (eg, study type, participants, data collection methods, and duration), intervention development and research phase, technological features and configurations, and outcomes. Emphasis was placed on 2 types of outcomes pertinent to the design and use of technology: design-related outcomes informative for future iterations of intervention, which typically included satisfaction, usability, technical and process feasibility (eg, reach, dose, and fidelity of delivery), acceptability, engagement, and interactions with the technology, and user-related outcomes such as change in SB, PA, work performance, and perceived enablers for changes.

On the basis of the UK Medical Research Council (MRC) framework for developing and evaluating complex interventions, we categorized the whole article or sections of the articles into respective research phases: development, feasibility and piloting, evaluation, and implementation; we also categorized the intervention based on the phase reported in the latest publication about the intervention (Table 2).

We adapted existing classification frameworks to derive our own coding scheme to annotate the technological aspect of each intervention (Table 3). The framework was primarily based on the BIT model [40], which complemented with elements from other coding schemes/frameworks introduced previously, to cover a broader range of technologies and to reflect the specialty of the workplace setting (eg, the addition of “mediated organizational support and social influences). Each code in the classification system can be viewed as a distinct technological feature (eg, a data log) implemented to deliver one or more intervention component (eg, self-monitoring of behaviors). A series of codes joined by “and” were used to annotate a technological configuration where several features were integrated to deliver 1 or more intervention component. For instance, an intervention that offered tailored feedback on progress based on users’ self-reported daily step counts was annotated with “DL and ATF. Notably, “Scheduled prompts (SP) delivered according to real-time user status passively captured by sensing technologies (“PDC and SP) are inherently different from SPs that interrupt users at fixed times throughout the day regardless of the users’ actual sitting time; hence, an additional code of “JITAI” was used to annotate “PDC and SP configurations to highlight the fact that the JITAI approach was present.

Table 2. Definitions of the development and research phases.

Phase	Definition and examples
Development phase	Studies could be one of the following: (1) reporting the design and development process of the intervention, following approaches such as Intervention Mapping, participatory design and user-centered design, (2) laboratory studies investigating design-related outcomes (feasibility, usability, and user experience) before the intervention has reached a deployable state of development, and (3) short in-the-wild deployment studies evaluating specific intervention components within a functional prototype before investing in further development.
Feasibility and piloting phase	Studies focused on investigating design-related outcomes of an intervention after it has reached a relatively complete stage of development, where user-related outcomes (behavior change, health and well-being, and productivity) were often measured as secondary outcomes with smaller sample sizes and less rigorous study designs.
Evaluation phase	Studies using a larger sample size and more rigorous study designs to assess important user-related outcomes and establish the efficacy of interventions.
Implementation phase	The intervention has already gone through the evaluation phase and has been used in practice for some time (eg, ≥2 years). As many implementation efforts are not reported, it was expected that this phase would have low representation.

Table 3. Links between our codes and categories from existing frameworks.

Codes with descriptions	BIT[a] elements [40]	Roles in the functional triad [44]	MoD[b] for internet-based interventions [41]
Digital logs (DL): technology provides a convenient way for the user to enter data, which can be a mobile phone diary for self-monitoring of behaviors or a Web-based questionnaire assessing current behavior and psychological determinants of behaviors.	Log	Tool	N/A[c]
Passive data collection (PDC): use wearable, smartphone-based and environment-based objective monitors to obtain time-stamped SB[e] records automatically.	Passive data collection	Tool	N/A
Connected devices (CD): one or more external sensing device is connected either wirelessly or with a cable to a central computing device.	N/A	Tool	N/A
Scheduled prompts (SP): break reminders delivered either at fixed intervals or with some schedule adaptive to the real-time user status.	Notification push	Tool, medium and/or social actor	Automated functions: automated follow-up messages (reminders)
Information delivery (ID): one or more forms of digital media with varying richness (text, links, testimonials, videos, or games) is used to present information that is usually static over time (eg, health facts, scripted motivational messages, and practical suggestions).	Information delivery	Medium and/or social actor	Automated functions: use of enriched information environment (eg, links, testimonials, videos, games); Use of supplementary modes (eg, emails and other digital media)
Automated tailored feedback (ATF): feedback on individual behaviors and progress, such as personalized goal setting and recommendations, that usually require some calculations of data input from DL or PDC.	Reports and visualization	Medium and/or social actor	Automated functions: automated tailored feedback based on individual progress
Mediated organizational support and social influences (MOSSI): emails conveying managers' approval, online forums facilitating communication and/or competition among program participants, and other digital means of linking the participant to other individuals for the purpose of social influences or organizational support. (Email access to the support from a consultant or coach should be coded under ID instead)	Messaging	Medium	Communicative functions: access to peer-to-peer support

[a]BIT: behavioral intervention technology.

[b]MOD: mode of delivery.

[c]N/A: not applicable.

[d]BCT: behavior change technique.

[e]SB: sedentary behavior.

Data Synthesis

Results on study characteristics (ie, publication data, study design, MRC development and research phase, and participants) and intervention characteristics (ie, target behavior, theoretical underpinning, technological design, and MRC development and research phase) were quantitatively summarized and presented using descriptive statistics.

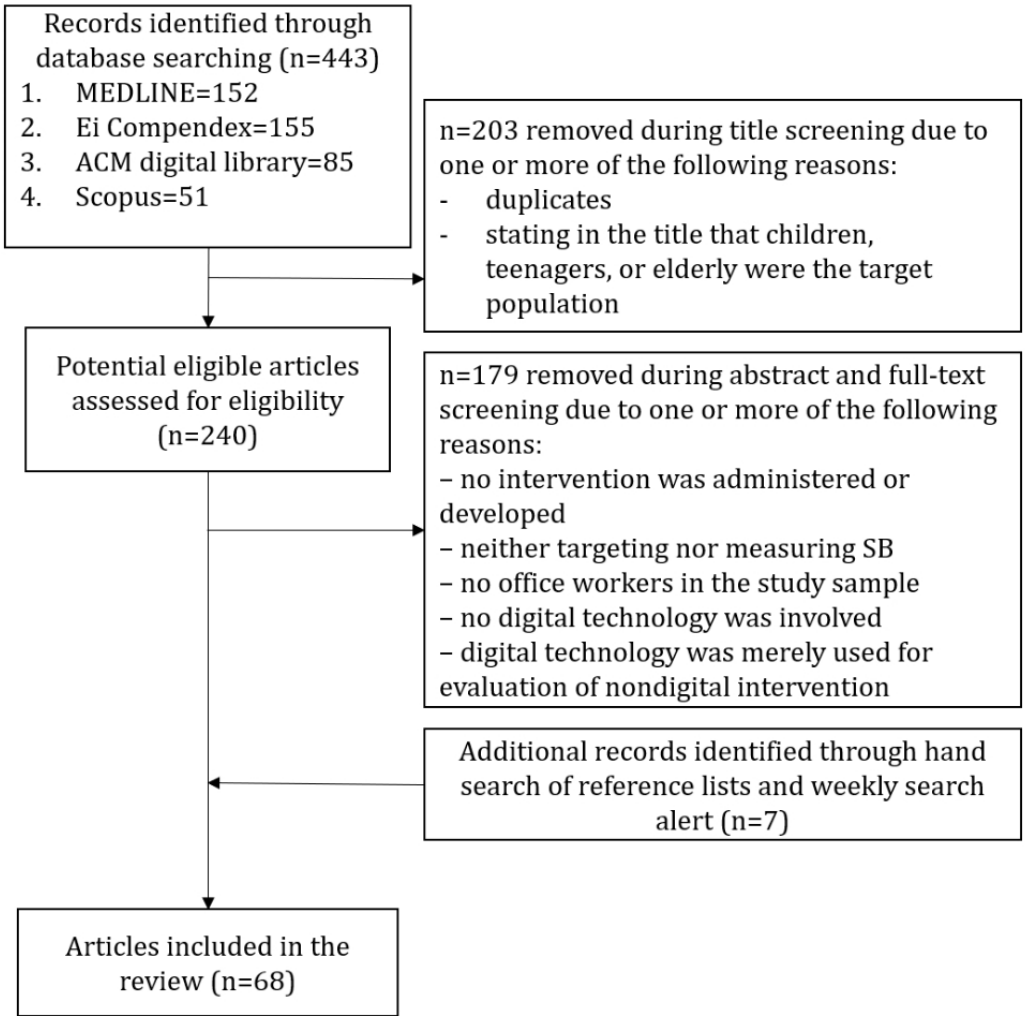
Due to the heterogeneity of study design (eg, interviews, laboratory testing, and randomized controlled trials [RCTs]) and outcomes (eg, design inspirations, usability, engagement, and effectiveness), meta-analysis of specific outcomes across studies was not suitable. Instead, a primarily qualitative approach was used to summarize the research under each category of technological configuration, with a focus on design-related findings and implications, which were most relevant to the research questions of our interest.

Results

Overview

A total of 68 articles were included in this review (Figure 1), corresponding to 45 unique interventions. Each article was counted as a separate study, even if it was focused on a different aspect of the same research project reported in another article.

Figure 1. Search and screening results.



Study Characteristics

Publication Data

As shown in Figure 2, there is an overall upward trend in the number of articles published on this topic over the past two decades or so, with 2014 being the most fruitful year. Overall, 66 published articles represented research that was conducted in 16 countries, in addition to 2 articles that reported international studies conducted in 64 countries [45] and 3 countries (the United Kingdom, Australia, and Spain) [23], respectively. The most represented countries were Australia (n=19 articles), the United States (n=17), the Netherlands (n=8), and the United Kingdom (n=4). Another 7 European countries (eg, Austria, Spain, Portugal, Belgium, Germany, Switzerland, and Finland) were represented in a total of 20 articles.

In terms of publication avenues, the included articles were published in 40 different scientific journals and proceedings. Divided by disciplines, 42 articles were published in the field of medical and health sciences, 13 in engineering and computing (including ergonomics and human factors), and 13 in interdisciplinary journals or conferences (eg, *PloS One*), out of which 6 were in the interdisciplinary field of digital health (eg, *Journal of Medical Internet Research*).

Study Design

For experimental studies, 25 articles reported RCTs (including cluster RCTs), 4 reported randomized crossover studies, 4 reported before-and-after studies with control or comparison group(s), and 10 reported before-and-after studies without control or comparison group(s). In addition to those traditional experimental designs, 9 articles reported descriptive quantitative process data (eg, fidelity of delivery, reach, usage pattern of the technology, and compliance to break prompts), 11 articles reported qualitative data reflecting participants/stakeholders' perspectives (eg, pre-and poststudy interviews), and 19 articles reported the design and development of the technology.

Note that the above categories were not mutually exclusive as 1 article could include both quantitative and qualitative results and report both the design process and an evaluation study.

Development and Research Phase

All 68 articles featured complex interventions according to the MRC definition. Table 4 shows the number of articles categorized into each intervention development phase based on the MRC framework. Except for 2 articles that reported both the development and piloting phase [46,47], each article was assigned 1 category.

Figure 2. Number of articles by year of publication and country of study.

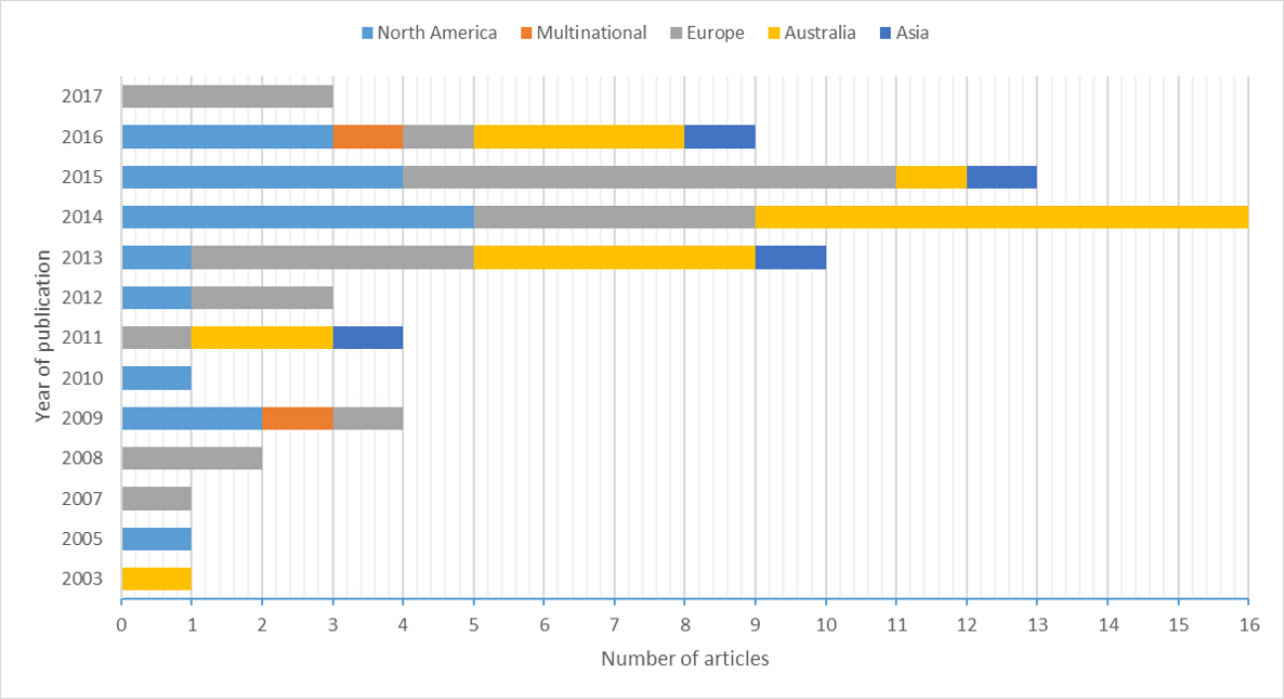


Table 4. Distribution of articles by development phases.

Phase of intervention development	Articles (n)
Development	19
Feasibility and piloting	34
Evaluation	10
Implementation	7

Participants

All studies included participants employed in office-based jobs. Indeed, most studies recruited participants from office-based workplaces covering different sectors and worksite sizes, although the majority of studies were conducted in universities and public-sector worksites. Only a few design and development studies recruited via local newspaper, social media, and from participant pools, resulting in a mixture of office workers and unemployed participants (eg, the study by Rabbi and colleagues [46]—13 students and 4 office workers; the studies by Bond and colleagues and Thomas and colleagues [48,49]—12 retired/employed and 18 office workers; the study by Mukhtar and Belaid [50]—2 graduate students and 2 faculty members; and the study by He Q and Agu [51]—6 students and 2 colleagues).

Overall, 63 studies recruited participants regardless of body mass index, whereas 5 studies targeted overweight and obese adults [48,49,52-54]; all studies but 1 [55] included both female and male participants. Except for 1 design and development study where sample size was not reported, sample sizes ranged from 1 [56] to 91 [57] among development studies, 3 [55] to 412 [58] among piloting studies, 153 [59] to 631 [60] among evaluation studies, and 291 [61] to 69,291 [45] among implementation studies.

Intervention Characteristics

Target Behavior

Of all 45 interventions, 18 interventions (27 articles) focused primarily on SB reduction, 14 (22 articles) targeted a combination of SB reduction and other behaviors (eg, PA promotion, diet management, posture correction, prompting social interactions with colleagues, and general lifestyle change), and 13 (19 articles) targeted other behaviors (eg, posture correction and PA promotion) without an SB reduction element in the intervention design but reported SB change as a secondary behavioral outcome.

Theoretical Underpinning

Overall, 19 interventions were underpinned by at least one theory, which included the theory of planned behavior (n=5), social cognitive theory (n=4), social ecological model (n=4), the stages of change or transtheoretical model (n=4), and theories of habits (n=3). The development of 3 interventions followed frameworks (eg, Intervention Mapping) that supported theory-based intervention design [57,62,63].

Technological Design and Development Phase

Multimedia Appendix 2 provides details about the technological features and configurations implemented in each intervention, the methods used to study those interventions, and study

outcomes. Table 5 presents summative results on different technological features/configurations in relation to the development and research phase based on MRC framework.

Summary of Design-Related Findings

Information Delivery and Mediated Organizational Support and Social Influences

The use of digital media for "information delivery was prevalent among reviewed interventions and was sometimes integrated with the feature of "mediated organizational support and social influences ("ID and MOSSI). A long-standing use case of this was motivational messages sent from the managers' email addresses to convey organizational support and endorsement for the program [59,64-66]. In other cases, "ID and MOSSI were implemented in the form of online discussion forums or social networking sites to encourage individuals to share experiences with peers and to foster social support or team competition [45,60,61,67,68].

Two-thirds of the "ID and MOSSI" interventions had moved beyond development and piloting phases, with 6 interventions [23,59,60,69-71] having reached the evaluation phase and 2 [45,61] having reached the implementation phase. There was consistent evidence for positive user-related outcomes (eg, reduction in SB and increase in PA and work productivity) across studies [23,59,60,67,70,71], except for the study by van Berkel and colleagues [69], which delivered a lifestyle intervention with a small component focused on SB reduction and yielded nonsignificant intervention effects on SB or other lifestyle behaviors.

The only published development work on "ID and MOSSI" configuration was novel in applying ambient and affective interfaces to persuasion. A system called "PerFrame was created to play footages of the users' close friend performing expressions showing either approval or disapproval, depending on whether the users' behavior was healthy or not [72].

Digital Log and Automated Tailored Feedback

Integration of "digital log and "automated tailored feedback was another common configuration ("DL and ATF), as such systems took user inputs and generated feedback accordingly. These ranged from textual advice tailored to psychological constructs assessed with a simple Web-based questionnaire [24,73,74] to sophisticated visualization and simulations tools providing feedback on outcomes of self-reported behaviors such as daily step counts [45,60,67] and PA [75,76].

Although only 8 interventions were identified in this category, half of them [24,60,67,74] had reached the evaluation phase and one [45] the implementation phase. All reported SB reduction in the intervention group over time, though only 2 [60,67] reported significant between-group (intervention vs control) difference in SB reduction.

Table 5. Summative results on technological design and development phase.

Technological design	Total, n (%)	Development, n (%)	Feasibility and piloting, n (%)	Evaluation, n (%)	Implementation, n (%)
Overall	45 (100)	13 (29)	21 (47)	8 (18)	3 (7)
ID[a]	36 (100)	9 (25)	17 (47)	8 (22)	2 (6)
DL[b]	14 (100)	1 (7)	5 (36)	5 (36)	3 (21)
PDC[c]	39 (100)	12 (31)	18 (46)	6 (15)	3 (8)
CD[d]	12 (100)	6 (50)	5 (42)	1 (8)	—[e]
SP[f]	28 (100)	13 (46)	14 (50)	1 (4)	—
ATF[g]	29 (100)	9 (31)	12 (41)	6 (21)	2 (7)
MOSSI[h] and ID	12 (100)	1 (8)	3 (25)	6 (50)	2 (17)
PDC and ATF	26 (100)	9 (35)	11 (42)	4 (15)	2 (8)
PDC and SP (JITAI[i])	19 (100)	13 (68)	5 (26)	1 (5)	—
Using on-board sensors	8 (100)	6 (75)	2 (59)	—	—
Using connected sensing devices ("CD, PDC, and SP")	11 (100)	7 (64)	3 (27)	1 (9)	—

[a]ID: information delivery.

[b]DL: digital log.

[c]PDC: passive data collection.

[d]CD: connected device.

[e]no intervention found in the category

[f]SP: scheduled prompts.

[g]ATF: automated tailored feedback.

[h]MOSSI: mediated organizational support and social influences.

[i]JITAI: just-in-time adaptive intervention.

Several studies have examined design-related outcomes such as user engagement and experience of the "DL and ATF" platform. For instance, in the study by Compernolle and colleagues [74], it was reported that 86% of the participants in the intervention condition requested computer-tailored feedback and advice and that the majority rated the advice positively; in contrast, in the study by Marshall and colleagues [24], only half of the participants visited the website for tailored feedback and even fewer used the website for a second time. Although both platforms delivered stage-based advice tailored to participants' self-reported PA and psychological determinants of PA, it could be the provision of pedometers in the study by Compernolle and colleagues [74] that made a difference.

Despite a lack of evidence showing "DL and ATF" as the efficacious component causing SB reduction, it was reported as a key mechanism of behavior change in several qualitative studies. Participants in the study by Bort-Roig and colleagues [77] highlighted the motivational value of being able to view logged data through visual graphics on a website and gain feedback; the study by Cooley and colleagues [78] interviewed 15 participants, who suggested that the mere act of logging nonpurposeful physical activities during breaks changed their perceptions of what constituted exercise—they also thought the automated feedback on progress helped them set up goals.

Passive Data Collection and Automated Tailored Feedback

Replacing "digital log with "passive data collection to provide input for "automated tailored feedback is a more technologically advanced configuration ("PDC and ATF"), as it capitalizes on automated sensing technologies and activity detection algorithms. Smartphones and pedometers were the 2 most frequently used devices for this configuration.

A number of smartphone apps incorporated data from on-board accelerometers or utilized Android APIs for real-time activity classification. Feedback was usually offered in the form of a dashboard with a break timer, daily accumulative active and inactive minutes, and/or a lifelog of activity episodes in chronological order [47,48,50,51]. Practical issues with this technological approach were identified, such as "phone battery drained quickly because of the accelerometer use and "users did not always carry the phone with them [47,51,79].

Pedometers were often used to provide instant and simplistic feedback on PA (eg, [74]). They were also used as a support tool (1) alongside DL to enhance the accuracy of self-report PA and (2) alongside MOSSI to provide the metric for team-based competition [28,45,52,54,66,71,80,81]. Participants generally considered the technological monitoring tool very helpful [54,77] and an evidence for organizational investment in staff health [82].

Notably, only 6 [23,45,59,67,74,81] out of the 25 "PDC and ATF" had reached the evaluation and implementation phases, 5 of which were pedometer-based interventions. Most interventions that used smartphone for both "PDC and ATF" were in the development and piloting phase.

Development research conducted in this space was innovative and informative in several aspects. First, machine learning was applied to classify activities and generate suggestions based on the users' past behavioral patterns, which were found to yield stronger intention to follow than generic suggestions [46]. Second, the likeability of different forms of feedback was explored: "at-a-glance and real-time display of summative data was perceived as useful and motivating by users [48,51]; potential features demanded by users were visual feedback on the health outcomes of SB, accurate and reliable data sources, and the control over the collection and sharing of their data feedback with colleagues [83].

Passive Data Collection and Scheduled Prompts (Just-in-Time Adaptive Interventions)

Passively collected data were utilized in 19 interventions to determine when to trigger prompts. Those were coded as "PDC and SP in addition to "JITAI in a bracket to be differentiated from the 9 SP interventions that prompted users at fixed times throughout the day [52,55,76,84-87]. Smartphone was the top-choice device used in this category, followed by desktop computers. A few studies used other connected devices (CD), which will be discussed in the "CD, PDC, and SP" configuration category.

Overall, 18 out of 19 "PDC and SP" interventions were in the development and piloting phase. This body of research produced outcomes particularly relevant to this review.

First, the studies were fruitful in identifying the optimum modality, frequency, and manner for interrupting users in the middle of sedentary work. Van Dantzig and colleagues [47] suggested the textual content of the persuasive messages was unimportant and a timely tactile notification on the smartphone might be just sufficient. Thomas and Bond [49] conducted a randomized crossover study with audible break prompts delivered from a smartphone app for 1 week in each of the 3 conditions: (1) a 3-min break prompt after 30 continuous sedentary minutes, (2) a 6-min break prompt after 60 sedentary minutes, and (3) a 12-min break prompt after 120 sedentary minutes. It was discovered that the 3- and 6-min conditions resulted in the greatest number and sum duration of walking breaks, the best and fastest compliance with prompts; from the users' perspective, the 6-min condition was the most preferred one [48]. Mukhtar and Belaid [50] found that reminders delivered with variable intervals adaptive to the duration of the last inactive episode were preferred by users to reminders delivered with fixed intervals. In terms of manner, some interventions adopted a so-called "passive prompt approach, in which the screen was locked unless the user complied with the suggestions, whereas others followed an "active prompt approach by allowing the user to snooze or dismiss the prompt and carry on work. Although higher odds of compliance were recorded in the passive prompts condition than in the active prompts condition in 1 study [75], user annoyance with the passive prompt approach was also reported [78].

Second, the research was innovative in applying "quick-and-dirty design methods to piloting novel intervention approaches and studying potential usability issues without large investment in development. For instance, in the abovementioned PerFrame study, a so-called "Wizard of Oz paradigm was applied to control the system output. That is, instead of

implementing complex Computer Vision algorithms, the researcher observed the users' sitting posture via a camera and remotely controlled which video footages to play [72]. In another example, researchers drew on a range of design research techniques such as diary, scenario, and technology probe to elicit user feedback on the design idea of an emotionally expressive robot, which would otherwise take a long period of development before getting users' input [88].

Connected Devices, Passive Data Collection, and Scheduled Prompts

Within the "PDC and SP configuration category, 11 intervention delivery systems employed an even more technologically advanced feature, by drawing on data from externally *CD*.

Only 1 "CD, PDC, and SP" intervention had moved to the evaluation phase [59]. The study compared an intervention including a wearable activity tracker that made the smartphone prompts responsive to real-time user status with an intervention without the external device. Although there were no significant between-condition differences in prolonged sitting reduction, a 70.5% uptake of the waist-worn activity tracker was encouraging.

The development and piloting research in this space extended our knowledge of devices and media that can be possibly used for delivering SB interventions.

Several peripheral sensing devices with various form factors were incorporated in interventions reviewed, including cushions on chairs to monitor sitting time [53,56], wearables to capture activities and postures [59,89,90], and sensors attached to workstations to infer sedentary time from workstation use time [47,91].

A number of data transfer technologies were used to establish connectivity between devices. Bluetooth technology was commonly used for wireless communications between portable devices, for instance, between an Android/iOS device and a nearby peripheral sensing device [59,92]. Some early studies used mobile networks to send text messages from a server to a mobile phone as a way of prompting users [47,89]. Universal Serial Bus (USB) and other cable-based connections were often utilized in systems for which portability was not crucial. For instance, the studies by Van Dantzig and colleagues, Slootmaker and colleagues, Ferreira and colleagues, and Carr and colleagues [47,89,91,93] used USB-type protocols for sending environment-based sensor data to the users' workstations, where the prompts were scheduled and delivered. USB protocol was also used in early prototypes of connected systems [56,94] to actuate novel user interfaces (eg, mechanically controlled sculpture and ambient light) from an Arduino, which is an open-source platform for creating interactive electronic objects.

Pros and cons of different technologies were explored. Wadhwa and colleagues [79] examined the technological feasibility and social acceptability of mobile versus environment-based sensing. The authors proposed a triggered-sensing approach to replace some mobile sensing with infrastructure sensing to extend battery life of mobile sensors; in addition, they analyzed users' response latencies to different prompts and found a slight user preference for mobile-based notifications to workstation-based ones. Haller and colleagues [95] connected a posture sensing chair to 3 different types of media for delivering prompts (onscreen graphic feedback, tactile feedback from the chair itself, and physical feedback delivered by a plastic plant that became droopy to represent bad posture of the user); the result was in favor of the physical feedback, as it required the shortest time to return to the main task after the prompted activity and was rated by users as least disturbing. Along the same line of reasoning, several design studies assessed the technological feasibility, ease of understanding, usability, and likeability of ambient displays, such as programmable sculptures that changed shape [56,91], or ambient lights that altered color [94,96] to reflect users' sedentary time and remind the user to take breaks. Nonetheless, although all the researchers suggested the need for longer-term experiments to establish the viability of their design approaches, no published follow-up studies were found.

Discussion

Principal Findings

This review sought to inform its readership about the research activity and technological landscape in the field of digital SB interventions for office workers and to identify research gaps and collaborative opportunities that could be further exploited.

This paper, first of all, serves as a roadmap that indicates the range and location of the literature on this topic. A total of 68 articles describing 45 interventions were identified. Although only a few studies were capable of providing definitive evidence (25 RCTs, of which only 9 were qualified as "evaluation phase studies), this is to be expected in an expanding field of interest with a lot of efforts to bring in novel technological features and configurations. In terms of geographic distribution, we observed that the development and piloting work conducted in this field was located across the globe, whereas evaluation/implementation research tended to be concentrated in specific countries and was usually associated with large national research initiatives (eg, Australia: "Stand up Australia and "Global Corporate Challenge ; the Netherlands: "Vitality in Practice ; and Spain: "Walk@WorkSpain). Some of those projects were also fruitful in generating publications, partly because they followed a phased approach to conducting and reporting the development, piloting, and evaluation of complex interventions as recommended by the MRC guidance ("Stand up Australia [63,65,70,97,98] and "Vitality in Practice" (VIP) project [57,99]). In terms of disciplines where research on this topic can be located, we demonstrated the added value of searching for articles outside medical and health sciences databases. Finally, we found many SB reduction elements embedded in interventions targeting other behaviors such as posture correction or PA promotion. Indeed, only 18 interventions in this review solely targeted SB reduction.

Second, this review provides an overview of the current technological landscape in this field, with a novel coding scheme constructed specially for this purpose. As shown in Table 5, configurations like "ID and MOSSI and "DL and ATF have mostly been researched in the evaluation and implementation phase. Less investment in development or piloting was observed, probably because those configurations typically used

technologies merely as media to exchange information that was traditionally delivered with print media or face-to-face communications, and hence, less complex computational model or infrastructure design was needed. In contrast, research on interventions that delivered SP or ATF based on PDC ("PDC and SP [JITAI]," "PDC and ATF"), in particular with sensors from CDs ("CD, PDC, and SP"), mostly remained in the development and piloting phase.

Notably, although validated PDC devices, such as the ActivPAL (PAL Technologies Ltd, Glasgow, United Kingdom) and ActiGraph (LLC, Pensacola, FL, USA), were widely used for outcome measurement [27,28,52,59,65,70,80,84,87,97,98,100], they were seldom integrated with other technological features as part of the intervention delivery system in the studies reviewed. This might be because early models of the ActivPAL and ActiGraph devices were not equipped with any output module (eg, a screen) to let wearers, or even researchers, receive feedback on SB during the monitoring period; neither were the stored data accessible to third-party apps or devices in real time for implementation of JITAI. This may, in turn, demotivate deployment of those devices beyond the assessment period (usually 1 week or 5 workdays), which could otherwise collect data throughout the whole study period and generate valuable insights into the process of change, as demonstrated in several studies [47,49,97]. This situation should soon be improved with the latest ActiGraph GT9X Link (LLC, Pensacola, FL, USA) and SitFIT (PAL Technologies Ltd, Glasgow, United Kingdom) devices that come with screens for instant feedback on behaviors and Bluetooth modules for communication with external devices.

That was why we coded CDs separately and considered it a very important trend that could potentially catalyze a paradigm shift in the use of data in behavior change. Not to mention easier integration of multiple data sources to make interventions more relevant to the context, CD greatly expands the range of interfaces and media that can be used to deliver SPs to users. We identified exploratory work on developing and piloting ambient displays to deliver break reminders subtly [72,94,96]. The technological advancements in the field of Tangible, Embedded, and Embodied Interactions presents new promise for this line of research, as mechanically controlled objects have been created [56,95] or designed [88,91] as a creative and pleasant way to persuade users into taking breaks and caring for their own health.

Implications

In addition, 2 notable blank spots can be identified in Table 5, suggesting areas where evidence is lacking and more investigations are warranted.

One is the dearth of research on interventions utilizing CDs, especially in evaluation and implementation phases. Research opportunities exist in exploiting wireless connectivity to make interventions more relevant to individual users and contexts. Manufacturers of well-validated PDC devices are starting to provide Software Development Kits (SDKs), such as the new ActiGraph Link SDK, which allows third-party apps or devices to stream the PDC devices' raw data in real time or near real time. This is very encouraging; however, no studies have been published featuring interventions using such SDKs to exploit the value of CD. To achieve this, collaborations between health scientists, computer scientists, and engineers from both academia and the industry need to be fostered.

Another notable blank spot in Table 5 is the lack of research on SPs beyond the piloting/feasibility phase. Considering the numerous innovative break-prompting installations that have been developed and piloted in engineering and computer science, efforts could be directed toward moving them to the next phase of evaluation with a more rigorous study design. This line of research is promising for 2 reasons. First, research suggests in-the-moment guidance that prompts smaller yet more frequent changes in existing behavior has a potential for greater impact than suggestions only tailored to overall behaviors periodically (eg, daily energy burnt) [46]. However, there is a lack of knowledge about the opportune manner of prompting office workers in the moment of sedentary work. Second, as the cost of embedded electronics is dropping, it becomes increasingly possible to scale up interventions delivered with novel technological devices, such as those systems reminding users subtly by changing ambient light or shape of physical artifacts [56,94,96].

Finally, upon reflection of using the MRC framework and conducting this review as an interdisciplinary team, we have realized differences in the understanding of "development and a lack of connection between different communities. There are encouraging examples where researchers followed through and published more than 1 stage of developing, piloting, and evaluating an intervention [57,62,63]. However, when it comes to the design and development of technologies for delivering interventions, it appears that health and behavioral scientists without technical backgrounds are less involved or interested. Meanwhile, although technological innovations are taking place in the fields of engineering and design, there seems to be a lack of mechanisms in place to feed design-related findings into other fields or move the novel technologies downstream to the evaluation phase.

It requires more thinking as to how to better connect and empower 2 communities—the community with expertise in intervention content development and evaluation and the community with capacities to design, develop, and study technologies with users. The answer to the question is beyond the scope of this review. Nevertheless, as a starting point, researchers from all disciplines can familiarize themselves with the MRC guidance and position their research in the big picture of developing and evaluating complex interventions. Health and behavioral scientists can also get more involved in user-centered research and have more inputs to early-stage technology innovations.

Limitations

The aim of this review was to scope the research activities and describe the technology design in SB interventions targeting office workers; as such, we did not intend to compare or synthesize the behavior change outcomes across interventions with meta-analysis. In addition, our review used a single code for PDC and focused on its integration with other technological features. The measurement and self-monitoring properties of

different devices used in those studies could have been coded with a more fine-grained coding scheme. However, we deemed this unnecessary, because a scoping review specifically on devices for self-monitoring SB and PA [36] was published during our data extraction phase and the authors of that review had coded the devices in terms of wear locations, outcomes measured, the type of feedback available, and various other measurement and self-monitoring properties

Conclusions

This review demonstrates the prevalent and diverse use of digital technologies in SB interventions targeting office workers. The use of technology to deliver information, to mediate organizational support and social influences, and to provide feedback based on self-reported data is well established in this field. More research is needed to exploit wireless connectivity between devices to make interventions more adaptive to the users' current state and context. Novel media interfaces for delivering subtle prompts are being innovated and are worth more attention. Opportunities exist to improve the utility of future research by encouraging interdisciplinary conversations and collaborations, potentially under the MRC framework for the development and evaluation of complex interventions.

Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council (grant number EP/L015463/1), Horizon Centre for Doctoral Training at the University of Nottingham, and Unilever UK Ltd.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies.

[PDF File (Adobe PDF File), 66KB - jmir_v21i2e11079_app1.pdf]

Multimedia Appendix 2

Summary of technological design, study method and key results.

[PDF File (Adobe PDF File), 57KB - jmir_v21i2e11079_app2.pdf]

References

1. Pate RR, O'Neill JR, Lobelo F. The evolving definition of "sedentary". Exerc Sport Sci Rev 2008 Oct;36(4):173-178. [doi: 10.1097/JES.0b013e3181877d1a] [Medline: 18815485]
2. Ekelund U, Steene-Johannessen J, Brown W, Fagerland M, Owen N, Powell K, et al. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. The Lancet 2016 Sep;388(10051):1302-1310 [FREE Full text] [doi: 10.1016/S0140-6736(16)30370-1] [Medline: 25246403]
3. World Health Organization. Recommended population levels of physical activity for health. In: Global recommendations on physical activity for health. Geneva: WHO Press; 2010:23-28.
4. National Health Services UK. Physical activity guidelines for adults URL: https://www.nhs.uk/live-well/exercise// [accessed 2018-12-06] [WebCite Cache ID 74T2AP2Go]
5. Tremblay MS, Colley RC, Saunders TJ, Healy GN, Owen N. Physiological and health implications of a sedentary lifestyle. Appl Physiol Nutr Metab 2010 Dec;35(6):725-740. [doi: 10.1139/H10-079] [Medline: 21164543]
6. Henson J, Yates T, Biddle SJ, Edwardson CL, Khunti K, Wilmot EG, et al. Associations of objectively measured sedentary behaviour and physical activity with markers of cardiometabolic health. Diabetologia 2013 May;56(5):1012-1020. [doi: 10.1007/s00125-013-2845-9] [Medline: 23456209]
7. Bankoski A, Harris TB, McClain JJ, Brychta RJ, Caserotti P, Chen KY, et al. Sedentary activity associated with metabolic syndrome independent of physical activity. Diabetes Care 2011 Feb;34(2):497-503 [FREE Full text] [doi: 10.2337/dc10-0987] [Medline: 21270206]
8. Lynch BM. Sedentary behavior and cancer: a systematic review of the literature and proposed biological mechanisms. Cancer Epidemiol Biomarkers Prev 2010 Nov;19(11):2691-2709 [FREE Full text] [doi: 10.1158/1055-9965.EPI-10-0815] [Medline: 20833969]
9. Parry S, Straker L. The contribution of office work to sedentary behaviour associated risk. BMC Public Health 2013 Apr 04;13:296 [FREE Full text] [doi: 10.1186/1471-2458-13-296] [Medline: 23557495]
10. Waters CN, Ling EP, Chu AH, Ng SH, Chia A, Lim YW, et al. Assessing and understanding sedentary behaviour in office-based working adults: a mixed-method approach. BMC Public Health 2016 Apr 27;16:360 [FREE Full text] [doi: 10.1186/s12889-016-3023-z] [Medline: 27117178]

11. De Cocker K, Duncan MJ, Short C, van Uffelen JG, Vandelanotte C. Understanding occupational sitting: prevalence, correlates and moderating effects in Australian employees. Prev Med 2014 Oct;67:288-294. [doi: 10.1016/j.ypmed.2014.07.031] [Medline: 25117522]
12. Buckley JP, Hedge A, Yates T, Copeland RJ, Loosemore M, Hamer M, et al. The sedentary office: an expert statement on the growing case for change towards better health and productivity. Br J Sports Med 2015 Nov;49(21):1357-1362. [doi: 10.1136/bjsports-2015-094618] [Medline: 26034192]
13. 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 Jun 18;12(2):e22 [FREE Full text] [doi: 10.2196/jmir.1390] [Medline: 20562092]
14. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
15. Skatova A, Bedwell B, Shipp V, Huang Y, Young A, Rodden T, et al. The role of ICT in office work breaks. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York, USA: ACM New York; 2016 Presented at: CHI '16; 2016; San Jose, California, USA p. 3049-3060 URL: https://dl.acm.org/citation.cfm?doid=2858036.2858443 [doi: 10.1145/2858036.2858443]
16. Chau JY, der Ploeg HP, van Uffelen JG, Wong J, Riphagen I, Healy GN, et al. Are workplace interventions to reduce sitting effective? A systematic review. Prev Med 2010 Nov;51(5):352-356. [doi: 10.1016/j.ypmed.2010.08.012] [Medline: 20801153]
17. Shrestha N, Kukkonen-Harjula KT, Verbeek JH, Ijaz S, Hermans V, Bhaumik S. Workplace interventions for reducing sitting at work. Cochrane Database Syst Rev 2016 Mar 17;3:CD010912. [doi: 10.1002/14651858.CD010912.pub3] [Medline: 26984326]
18. Commissaris DA, Huysmans MA, Mathiassen SE, Srinivasan D, Koppes LL, Hendriksen IJ. Interventions to reduce sedentary behavior and increase physical activity during productive work: a systematic review. Scand J Work Environ Health 2016 May 01;42(3):181-191. [doi: 10.5271/sjweh.3544] [Medline: 26683116]
19. Chu AH, Ng SH, Tan CS, Win AM, Koh D, Müller-Riemenschneider F. A systematic review and meta-analysis of workplace intervention strategies to reduce sedentary time in white-collar workers. Obes Rev 2016 May;17(5):467-481. [doi: 10.1111/obr.12388] [Medline: 26990220]
20. Gardner B, Phillips LA, Judah G. Habitual instigation and habitual execution: definition, measurement, and effects on behaviour frequency. Br J Health Psychol 2016 Sep;21(3):613-630. [doi: 10.1111/bjhp.12189] [Medline: 26991427]
21. Martin A, Fitzsimons C, Jepson R, Saunders DH, van der Ploeg HP, Teixeira PJ, EuroFIT Consortium. Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis. Br J Sports Med 2015 Aug;49(16):1056-1063. [doi: 10.1136/bjsports-2014-094524] [Medline: 25907181]
22. Prince SA, Saunders TJ, Gresty K, Reid RD. A comparison of the effectiveness of physical activity and sedentary behaviour interventions in reducing sedentary time in adults: a systematic review and meta-analysis of controlled trials. Obes Rev 2014 Nov;15(11):905-919 [FREE Full text] [doi: 10.1111/obr.12215] [Medline: 25112481]
23. Gilson ND, Puig-Ribera A, McKenna J, Brown WJ, Burton NW, Cooke CB. Do walking strategies to increase physical activity reduce reported sitting in workplaces: a randomized control trial. Int J Behav Nutr Phys Act 2009 Jul 20;6:43 [FREE Full text] [doi: 10.1186/1479-5868-6-43] [Medline: 19619295]
24. Marshall AL, Leslie ER, Bauman AE, Marcus BH, Owen N. Print versus website physical activity programs: a randomized trial. Am J Prev Med 2003 Aug;25(2):88-94. [Medline: 12880874]
25. Plotnikoff RC, McCargar LJ, Wilson PM, Loucaides CA. Efficacy of an e-mail intervention for the promotion of physical activity and nutrition behavior in the workplace context. Am J Health Promot 2005;19(6):422-429. [doi: 10.4278/0890-1171-19.6.422] [Medline: 16022206]
26. Pedersen SJ, Cooley PD, Mainsbridge C. An e-health intervention designed to increase workday energy expenditure by reducing prolonged occupational sitting habits. Work 2014;49(2):289-295. [doi: 10.3233/WOR-131644] [Medline: 23787256]
27. Evans RE, Fawole HO, Sheriff SA, Dall PM, Grant PM, Ryan CG. Point-of-choice prompts to reduce sitting time at work: a randomized trial. Am J Prev Med 2012 Sep;43(3):293-297. [doi: 10.1016/j.amepre.2012.05.010] [Medline: 22898122]
28. Swartz AM, Rote AE, Welch WA, Maeda H, Hart TL, Cho YI, et al. Prompts to disrupt sitting time and increase physical activity at work, 2011-2012. Prev Chronic Dis 2014 May 01;11:E73 [FREE Full text] [doi: 10.5888/pcd11.130318] [Medline: 24784909]
29. Gordon A. A theory-based pilot study to decrease sitting time in the workplace. Arizona, US: Arizona State University; 2013. URL: https://repository.asu.edu/attachments/114413/content/Gordon_asu_0010N_13144.pdf [accessed 2018-12-06] [WebCite Cache ID 74T3a6BnO]
30. 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. [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
31. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

32. 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]
33. Nahum-Shani I, Smith NS, Tewari A, Witkiewitz K, Collins LM, Spring B, et al. Just-in-time adaptive interventions (JITAIs): An organizing framework for ongoing health behavior support. Pennysylvania State, USA: The Methodology Center; 2014. URL: https://methodology.psu.edu/media/techreports/14-126.pdf [accessed 2018-12-06] [WebCite Cache ID 74T42uKfr]
34. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001;55:587-634. [doi: 10.1006/ijhc.2001.0503]
35. The Joanna Briggs Institute. 2015. Methodology for JBI scoping reviews URL: http://joannabriggs.org/assets/docs/sumari/Reviewers-Manual_Methodology-for-JBI-Scoping-Reviews_2015_v2.pdf [accessed 2018-12-06] [WebCite Cache ID 74T4BdK8O]
36. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 04;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
37. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
38. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1):485-500.
39. Fogg BJ. Persuasive technologies. Commun ACM 1999 May;42(5):26-29. [doi: 10.1145/301353.301396]
40. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
41. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
42. King P, Tester J. The landscape of persuasive technologies. Commun ACM 1999 May;42(5):31-38. [doi: 10.1145/301353.301398]
43. Judah G, Aunger R, Schmidt WP, Michie S, Granger S, Curtis V. Experimental pretesting of hand-washing interventions in a natural setting. Am J Public Health 2009 Oct;99(Suppl 2):S405-S411. [doi: 10.2105/AJPH.2009.164160] [Medline: 19797755]
44. Fogg BJ. Persuasive computers: perspectives and research directions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1998 Presented at: CHI '98; April 18-23, 1998; Los Angeles, California, USA p. 225-232. [doi: 10.1145/274644.274677]
45. Ganesan AN, Louise J, Horsfall M, Bilsborough SA, Hendriks J, McGavigan AD, et al. International mobile-health intervention on physical activity, sitting, and weight: the Stepathlon Cardiovascular Health Study. J Am Coll Cardiol 2016 May 31;67(21):2453-2463 [FREE Full text] [doi: 10.1016/j.jacc.2016.03.472] [Medline: 27050185]
46. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
47. Van Dantzig S, Geleijnse G, van Halteren A. Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquit Comput 2012 Jul 12;17(6):1237-1246. [doi: 10.1007/s00779-012-0588-0]
48. Bond DS, Thomas JG, Raynor HA, Moon J, Sieling J, Trautvetter J, et al. B-MOBILE--a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. PLoS One 2014 May 25;9(6):e100821 [FREE Full text] [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]
49. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: implications for JITAI optimization. Health Psychol 2015 Dec;34S:1261-1267 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]
50. Mukhtar H, Belaid D. Using adaptive feedback for promoting awareness about physical activeness in adults. 2013 Presented at: 10th International Conference on Ubiquitous Intelligence & Computing and 2013 IEEE 10th International Conference on Autonomic & Trusted Computing; December 18-21, 2013; Vietri sul Mere, Italy p. 638-643. [doi: 10.1109/UIC-ATC.2013.99]
51. He Q, Agu E. On11: an activity recommendation application to mitigate sedentary lifestyle. In: Proceedings of the 2014 workshop on physical analytics. 2014 Presented at: WPA '14; June 16, 2014; Bretton Woods, New Hampshire, USA. [doi: 10.1145/2611264.2611268]
52. Júdice PB, Hamilton MT, Sardinha LB, Silva AM. Randomized controlled pilot of an intervention to reduce and break-up overweight/obese adults' overall sitting-time. Trials 2015 Nov 02;16:490 [FREE Full text] [doi: 10.1186/s13063-015-1015-4] [Medline: 26525049]

53. Gilson ND, Ng N, Pavey TG, Ryde GC, Straker L, Brown WJ. Project Energise: using participatory approaches and real time computer prompts to reduce occupational sitting and increase work time physical activity in office workers. J Sci Med Sport 2016 Nov;19(11):926-930. [doi: 10.1016/j.jsams.2016.01.009] [Medline: 26922132]
54. Carr LJ, Karvinen K, Peavler M, Smith R, Cangelosi K. Multicomponent intervention to reduce daily sedentary time: a randomised controlled trial. BMJ Open 2013 Oct 18;3(10):e003261 [FREE Full text] [doi: 10.1136/bmjopen-2013-003261] [Medline: 24141969]
55. Green N, Sigurdsson S, Wilder DA. Decreasing bouts of prolonged sitting among office workers. J Appl Behav Anal 2016 Sep;49(3):717-722. [doi: 10.1002/jaba.309] [Medline: 27150262]
56. Jafarinaimi N, Forlizzi J, Hurst A, Zimmerman J. Breakaway: an ambient display designed to change human behavior. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems. 2005 Presented at: CHI EA '05; April 02-07, 2005; Portland, OR, USA p. 1945-1948. [doi: 10.1145/1056808.1057063]
57. Coffeng JK, Hendriksen IJ, Duijts SF, Proper KI, van Mechelen W, Boot CR. The development of the Be Active & Relax "Vitality in Practice" (VIP) project and design of an RCT to reduce the need for recovery in office employees. BMC Public Health 2012 Aug 02;12:592 [FREE Full text] [doi: 10.1186/1471-2458-12-592] [Medline: 22852835]
58. Coffeng JK, Boot CR, Duijts SF, Twisk JW, van Mechelen W, Hendriksen IJ. Effectiveness of a worksite social & physical environment intervention on need for recovery, physical activity and relaxation; results of a randomized controlled trial. PLoS One 2014 Dec 26;9(12):e114860 [FREE Full text] [doi: 10.1371/journal.pone.0114860] [Medline: 25542039]
59. Brakenridge CL, Fjeldsoe BS, Young DC, Winkler EA, Dunstan DW, Straker LM, et al. Evaluating the effectiveness of organisational-level strategies with or without an activity tracker to reduce office workers' sitting time: a cluster-randomised trial. Int J Behav Nutr Phys Act 2016 Nov 04;13(1):115 [FREE Full text] [doi: 10.1186/s12966-016-0441-3] [Medline: 27814738]
60. Sternfeld B, Block C, Quesenberry Jr CP, Block TJ, Husson G, Norris JC, et al. Improving diet and physical activity with ALIVE: a worksite randomized trial. Am J Prev Med 2009 Jun;36(6):475-483. [doi: 10.1016/j.amepre.2009.01.036] [Medline: 19460655]
61. Aittasalo M, Livson M, Lusa S, Romo A, Vähä-Ypyä H, Tokola K, et al. Moving to business - changes in physical activity and sedentary behavior after multilevel intervention in small and medium-size workplaces. BMC Public Health 2017 Apr 17;17(1):319 [FREE Full text] [doi: 10.1186/s12889-017-4229-4] [Medline: 28415993]
62. van Berkel J, Proper KI, Boot CR, Bongers PM, van der Beek AJ. Mindful "Vitality in Practice": an intervention to improve the work engagement and energy balance among workers; the development and design of the randomised controlled trial. BMC Public Health 2011 Sep 27;11:736 [FREE Full text] [doi: 10.1186/1471-2458-11-736] [Medline: 21951433]
63. Neuhaus M, Healy GN, Fjeldsoe BS, Lawler S, Owen N, Dunstan DW, et al. Iterative development of Stand Up Australia: a multi-component intervention to reduce workplace sitting. Int J Behav Nutr Phys Act 2014 Feb 21;11:21 [FREE Full text] [doi: 10.1186/1479-5868-11-21] [Medline: 24559162]
64. Neuhaus M, Eakin EG, Straker L, Owen N, Dunstan DW, Reid N, et al. Reducing occupational sedentary time: a systematic review and meta-analysis of evidence on activity-permissive workstations. Obes Rev 2014 Oct;15(10):822-838. [doi: 10.1111/obr.12201] [Medline: 25040784]
65. Healy GN, Eakin EG, Lamontagne AD, Owen N, Winkler E, Wiesner G, et al. Reducing sitting time in office workers: short-term efficacy of a multicomponent intervention. Prev Med 2013 Jul;57(1):43-48. [doi: 10.1016/j.ypmed.2013.04.004] [Medline: 23597658]
66. Gilson N, McKenna J, Cooke C, Brown W. Walking towards health in a university community: a feasibility study. Prev Med 2007 Feb;44(2):167-169. [doi: 10.1016/j.ypmed.2006.09.012] [Medline: 17156836]
67. Puig-Ribera A, Bort-Roig J, González-Suárez AM, Martínez-Lemos I, Giné-Garriga M, Fortuño J, et al. Patterns of impact resulting from a 'sit less, move more' web-based program in sedentary office employees. PLoS One 2015 Apr 01;10(4):e0122474 [FREE Full text] [doi: 10.1371/journal.pone.0122474] [Medline: 25830782]
68. Carroll EA, Czerwinski M, Roseway A, Kapoor A, Johns P, Rowan K, et al. Food and mood: just-in-time support for emotional eating. 2013 Presented at: Humaine Association Conference on Affective Computing and Intelligent Interaction; September 2-5, 2013; Geneva, Switzerland. [doi: 10.1109/ACII.2013.48]
69. van Berkel J, Boot CR, Proper KI, Bongers PM, van der Beek AJ. Effectiveness of a worksite mindfulness-based multi-component intervention on lifestyle behaviors. Int J Behav Nutr Phys Act 2014 Jan 27;11:9 [FREE Full text] [doi: 10.1186/1479-5868-11-9] [Medline: 24467802]
70. Healy GN, Eakin EG, Owen N, Lamontagne AD, Moodie M, Winkler EA, et al. A cluster randomized controlled trial to reduce office workers' sitting time: effect on activity outcomes. Med Sci Sports Exerc 2016 Sep;48(9):1787-1797. [doi: 10.1249/MSS.0000000000000972] [Medline: 27526175]
71. Puig-Ribera A, Bort-Roig J, Giné-Garriga M, González-Suárez AM, Martínez-Lemos I, Fortuño J, et al. Impact of a workplace 'sit less, move more' program on efficiency-related outcomes of office employees. BMC Public Health 2017 May 16;17(1):455 [FREE Full text] [doi: 10.1186/s12889-017-4367-8] [Medline: 28511642]
72. Obermair C, Reitberger W, Meschtscherjakov A, Lankes M, Tscheligi M. perFrames: persuasive picture frames for proper posture. In: Persuasive Technology. Berlin, Heidelberg: Springer; 2008.

73. De Cocker K, De Bourdeaudhuij I, Cardon G, Vandelanotte C. Theory-driven, web-based, computer-tailored advice to reduce and interrupt sitting at work: development, feasibility and acceptability testing among employees. BMC Public Health 2015 Sep 24;15:959 [FREE Full text] [doi: 10.1186/s12889-015-2288-y] [Medline: 26404915]
74. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a web-based, computer-tailored, pedometer-based physical activity intervention for adults: a cluster randomized controlled trial. J Med Internet Res 2015 Feb 09;17(2):e38 [FREE Full text] [doi: 10.2196/jmir.3402] [Medline: 25665498]
75. Cooley D, Pedersen S. A pilot study of increasing nonpurposeful movement breaks at work as a means of reducing prolonged sitting. J Environ Public Health 2013;2013:128376 [FREE Full text] [doi: 10.1155/2013/128376] [Medline: 23690798]
76. Mainsbridge CP, Cooley PD, Fraser SP, Pedersen SJ. The effect of an e-health intervention designed to reduce prolonged occupational sitting on mean arterial pressure. J Occup Environ Med 2014 Nov;56(11):1189-1194 [FREE Full text] [doi: 10.1097/JOM.0000000000000243] [Medline: 25376414]
77. Bort-Roig J, Martin M, Puig-Ribera A, González-Suárez Á, Martínez-Lemos I, Martori JC, et al. Uptake and factors that influence the use of 'sit less, move more' occupational intervention strategies in Spanish office employees. Int J Behav Nutr Phys Act 2014 Dec 10;11:152 [FREE Full text] [doi: 10.1186/s12966-014-0152-6] [Medline: 25490857]
78. Cooley D, Pedersen S, Mainsbridge C. Assessment of the impact of a workplace intervention to reduce prolonged occupational sitting time. Qual Health Res 2014 Jan;24(1):90-101. [doi: 10.1177/1049732313513503] [Medline: 24231074]
79. Wadhwa R, Chugh A, Kumar A, Singh M, Yadav K, Eswaran S, et al. SenseX: design and deployment of a pervasive wellness monitoring platform for workplaces. In: Service-Oriented Computing. Berlin, Heidelberg: Springer; Nov 16, 2015:427-443.
80. Parry S, Straker L, Gilson ND, Smith AJ. Participatory workplace interventions can reduce sedentary time for office workers--a randomised controlled trial. PLoS One 2013 Nov 12;8(11):e78957 [FREE Full text] [doi: 10.1371/journal.pone.0078957] [Medline: 24265734]
81. Macniven R, Engelen L, Kacen MJ, Bauman A. Does a corporate worksite physical activity program reach those who are inactive? Findings from an evaluation of the Global Corporate Challenge. Health Promot J Austr 2015 Aug;26(2):142-145. [doi: 10.1071/HE14033] [Medline: 25917174]
82. Gilson N, McKenna J, Cooke C. Experiences of route and task-based walking in a university community: qualitative perspectives in a randomized control trial. J Phys Act Health 2008;5(Suppl 1):S176-S182. [Medline: 18364522]
83. Mohadis HM, Ali NM. Designing persuasive application to encourage physical activity at workplace among older workers. New Jersey, USA: Institute of Electrical and Electronics Engineers Inc; 2016 Aug 15 Presented at: 2016 Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP); 21-23 July, 2016; Konya, Turkey p. 126-130 URL: http://aac.asm.org/cgi/pmidlookup?view=long&pmid=25246403 [doi: 10.1109/DICTAP.2016.7544013]
84. Donath L, Faude O, Schefer Y, Roth R, Zahner L. Repetitive daily point of choice prompts and occupational sit-stand transfers, concentration and neuromuscular performance in office workers: an RCT. Int J Environ Res Public Health 2015 Apr 20;12(4):4340-4353 [FREE Full text] [doi: 10.3390/ijerph120404340] [Medline: 25903058]
85. Taylor WC, Paxton RJ, Shegog R, Coan SP, Dubin A, Page TF, et al. Impact of booster breaks and computer prompts on physical activity and sedentary behavior among desk-based workers: a cluster-randomized controlled trial. Prev Chronic Dis 2016 Nov 17;13:E155 [FREE Full text] [doi: 10.5888/pcd13.160231] [Medline: 27854422]
86. Mackenzie K, Goyder E, Eves F. Acceptability and feasibility of a low-cost, theory-based and co-produced intervention to reduce workplace sitting time in desk-based university employees. BMC Public Health 2015 Dec 24;15:1294 [FREE Full text] [doi: 10.1186/s12889-015-2635-z] [Medline: 26703900]
87. Kerr J, Takemoto M, Bolling K, Atkin A, Carlson J, Rosenberg D, et al. Two-arm randomized pilot intervention trial to decrease sitting time and increase sit-to-stand transitions in working and non-working older adults. PLoS One 2016 Jan 06;11(1):e0145427 [FREE Full text] [doi: 10.1371/journal.pone.0145427] [Medline: 26735919]
88. Reeder S, Kelly L, Kechavarzi B, Sabanovic S. Breakbot: a social motivator for the workplace. In: Proceedings of the Conference on Designing Interactive Systems. 2010 Presented at: 8th ACM Conference on Designing Interactive Systems; August 16-20, 2010; Aarhus, Denmark. [doi: 10.1145/1858171.1858184]
89. Slootmaker SM, Chinapaw MJ, Schuit AJ, Seidell JC, Van Mechelen W. Feasibility and effectiveness of online physical activity advice based on a personal activity monitor: randomized controlled trial. J Med Internet Res 2009 Jul 29;11(3):e27 [FREE Full text] [doi: 10.2196/jmir.1139] [Medline: 19674956]
90. El-Sayed B, Farra N, Moacdieh N, Hajj Z. A novel mobile wireless sensing system for realtime monitoring of posture and spine stress. : IEEE; 2011 Feb Presented at: 2011 1st Middle East Conference on Biomedical Engineering; 21-24 February, 2011; Sharjah, United Arab Emirates p. 428-431. [doi: 10.1109/MECBME.2011.5752156]
91. Ferreira MJ, Caraban AK, Karapanos E. Breakout: predicting and breaking sedentary behaviour at work. In: CHI '14 Extended Abstracts on Human Factors in Computing Systems.: ACM; 2014 Presented at: CHI EA '14; April 26-May 01, 2014; Toronto, ON, Canada. [doi: 10.1145/2559206.2581330]
92. Carr LJ, Leonhard C, Tucker S, Fethke N, Benzo R, Gerr F. Total worker health intervention increases activity of sedentary workers. Am J Prev Med 2016 Jan;50(1):9-17. [doi: 10.1016/j.amepre.2015.06.022] [Medline: 26260492]

93. Carr LJ, Walaska KA, Marcus BH. Feasibility of a portable pedal exercise machine for reducing sedentary time in the workplace. Br J Sports Med 2012 May;46(6):430-435. [doi: 10.1136/bjsm.2010.079574] [Medline: 21324889]
94. Mateevitsi V, Reda K, Leigh J, Johnson A. The health bar: a persuasive ambient display to improve the office worker's well being. In: Proceedings of the 5th Augmented Human International Conference. 2014 Presented at: AH '14; March 07-08, 2014; Kobe, Japan. [doi: 10.1145/2582051.2582072]
95. Haller M, Richter C, Brandl P, Gross S, Schossleitner G, Schrempf A, et al. Finding the right way for interrupting people improving their sitting posture. In: Human-Computer Interaction–INTERACT 2011. Berlin, Heidelberg: Springer; 2011:1-17.
96. Fortmann J, Stratmann T, Boll S, Poppinga B, Heuten W. Make me move at work! An ambient light display to increase physical activity. : IEEE; 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops; May 5-8, 2013; Venice, Italy p. 274-277. [doi: 10.4108/icst.pervasivehealth.2013.252089]
97. Stephens SK, Winkler EA, Trost SG, Dunstan DW, Eakin EG, Chastin SF, et al. Intervening to reduce workplace sitting time: how and when do changes to sitting time occur? Br J Sports Med 2014 Jul;48(13):1037-1042. [doi: 10.1136/bjsports-2014-093524] [Medline: 24815544]
98. Neuhaus M, Healy GN, Dunstan DW, Owen N, Eakin EG. Workplace sitting and height-adjustable workstations: a randomized controlled trial. Am J Prev Med 2014 Jan;46(1):30-40. [doi: 10.1016/j.amepre.2013.09.009] [Medline: 24355669]
99. Coffeng JK, Hendriksen IJ, van Mechelen W, Boot CR. Process evaluation of a worksite social and physical environmental intervention. J Occup Environ Med 2013 Dec;55(12):1409-1420. [doi: 10.1097/JOM.0b013e3182a50053] [Medline: 24270291]
100. De Cocker K, De Bourdeaudhuij I, Cardon G, Vandelanotte C. The effectiveness of a web-based computer-tailored intervention on workplace sitting: a randomized controlled trial. J Med Internet Res 2016 May 31;18(5):e96 [FREE Full text] [doi: 10.2196/jmir.5266] [Medline: 27245789]

Abbreviations

API: application programming interface
ATF: automated tailored feedback
BCI: behavior change intervention
BCT: behavior change technique
BIT: behavioral intervention technology
CD: connected device
DL: digital log
ICT: information and communication technology
ID: information delivery
JBI: Joanna Briggs Institute
JITAI: just-in-time adaptive intervention
MoD: mode of delivery
MOSSI: mediated organizational support and social influences
MRC: Medical Research Council
MVPA: moderate to vigorous physical activity
PA: physical activity
PDC: passive data collection
PSD: persuasive system design
PT: persuasive technology
RCT: randomized controlled trial
SB: sedentary behavior
SDK: Software Development Kits
SP: scheduled prompts
USB: Universal Serial Bus

Edited by G Eysenbach; submitted 29.05.18; peer-reviewed by J Murray, A Loveday, J Miller; comments to author 19.07.18; revised version received 30.08.18; accepted 12.09.18; published 07.02.19.

Please cite as:
Huang Y, Benford S, Blake H
Digital Interventions to Reduce Sedentary Behaviors of Office Workers: Scoping Review
J Med Internet Res 2019;21(2):e11079
URL: http://www.jmir.org/2019/2/e11079/
doi:10.2196/11079
PMID:30730294

Viewpoint

A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint

Sascha Miller[1], MSc; Ben Ainsworth[2], PhD; Lucy Yardley[1,3], PhD; Alex Milton[3], PhD; Mark Weal[4], PhD; Peter Smith[5], PhD; Leanne Morrison[1,6], PhD

[1]Center for Clinical and Community Applications of Health Psychology, Department of Psychology, University of Southampton, Southampton, United Kingdom

[2]Department of Psychology, University of Bath, Bath, United Kingdom

[3]School of Psychological Science, University of Bristol, Bristol, United Kingdom

[4]Web and Internet Science Group, School of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom

[5]Department of Social Statistics and Demography, School of Economic, Social and Political Sciences, University of Southampton, Southampton, United Kingdom

[6]Primary Care and Population Sciences, School of Medicine, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Sascha Miller, MSc
Center for Clinical and Community Applications of Health Psychology
Department of Psychology
University of Southampton
Highfield Campus
Southampton, SO17 1BJ
United Kingdom
Phone: 44 07972583059
Email: S.J.Miller@soton.ac.uk

Abstract

Trials of digital interventions can yield extensive, in-depth usage data, yet usage analyses tend to focus on broad descriptive summaries of how an intervention has been used by the whole sample. This paper proposes a novel framework to guide systematic, fine-grained usage analyses that better enables understanding of how an intervention works, when, and for whom. The framework comprises three stages to assist in the following: (1) familiarization with the intervention and its relationship to the captured data, (2) identification of meaningful measures of usage and specifying research questions to guide systematic analyses of usage data, and (3) preparation of datasheets and consideration of available analytical methods with which to examine the data. The framework can be applied to inform data capture during the development of a digital intervention and/or in the analysis of data after the completion of an evaluation trial. We will demonstrate how the framework shaped preparation and aided efficient data capture for a digital intervention to lower transmission of cold and flu viruses in the home, as well as how it informed a systematic, in-depth analysis of usage data collected from a separate digital intervention designed to promote self-management of colds and flu. The Analyzing and Measuring Usage and Engagement Data (AMUsED) framework guides systematic and efficient in-depth usage analyses that will support standardized reporting with transparent and replicable findings. These detailed findings may also enable examination of what constitutes effective engagement with particular interventions.

(J Med Internet Res 2019;21(2):e10966) doi:10.2196/10966

KEYWORDS

behavioral research; internet; health; patient engagement; data analysis

Introduction

Digital interventions are intended to support positive change in a range of health-related outcomes, including psychological, behavioral, educational, social, and environmental [1-3]. They may be delivered using any digital device (eg, phone and computer), making them cost-effective for providers [4,5]. Trialing a digital intervention can yield complex, large-scale datasets containing detailed usage data. If analyzed appropriately, this data is able to provide invaluable detail on how users interact with the intervention and inform our understanding of engagement. Measuring digital intervention

engagement has been described as a multidimensional concept, including the extent to which an intervention is used (eg, amount, frequency, and duration) and the subjective experience of the user as characterized by attention, affect, and interest [6]. As a key element of engagement, in-depth and consistently applied usage analyses are capable of providing invaluable insight into the field of engagement with digital interventions.

Usage analyses frequently examine the extent to which an intervention is used by the whole sample, utilizing variables such as the number of times users logged in, total time spent on an intervention, or number of pages viewed [7]. These broad-level analyses do not always take advantage of the detailed and comprehensive data available and they frequently assume that greater amounts of usage are indicative of higher levels of interaction that lead to increased changes in target behavior [8]. Harnessing the full range of data can instead enable more informative usage variables to be computed or combined, which may answer specific research questions about patterns of usage (ie, who the intervention was used by and how it was used) [8,9]. Recent interest in *effective engagement* considers these individual patterns of usage and the minimum level of engagement necessary for changes in target behavior to occur, including variation across individuals [8,10-12]. Effective engagement is defined as sufficient engagement with the intervention to achieve intended outcomes [8]. For example, a digital intervention designed to lower the transmission of cold and flu provided four sessions of content, requiring the intervention to be accessed on four separate occasions. However, analysis of usage data, reported behavior, and incidences of illnesses revealed that using the first session alone facilitated the required change in behavior to increase positive outcomes [11]. Alternatively, effective engagement may be context dependent, whereby viewing specific content components, in a certain order, or at an appropriate time, is the minimum threshold necessary for change [8,11,12].

During the planning of a digital intervention, processes such as logic models and guiding principles may be used to structure the theoretical underpinning and associated content for the intervention [13-15]. These techniques help identify behavioral determinants (eg, beliefs associated with the target behavior), which may be important in influencing the target behavior; for example, low confidence to manage symptoms or perceived barriers to performing a specific behavior. In order to influence favorably these behavioral determinants, content containing behavior change techniques (BCTs), such as goal setting, feedback on behavior, or habit formation, are incorporated into the intervention [16]. However, by performing analyses which focus solely on broad usage patterns across the whole sample, the opportunity is missed to understand how specific intervention content (eg, BCTs) is used by subgroups with particular characteristics (ie, behavioral determinants) and the extent to which there is a relationship to the target behavior [11,16]. Devising a plan of analysis to answer these questions using the fine-grained data often available from digital interventions enables us to examine the constructs of the logic model and further our understanding of the mechanisms of action underlying successful behavior change [17,18].

Analyzing usage metrics to better understand engagement has been proposed for some time, with the “law of attrition” being one of the first theories to draw attention to the benefits of examining usage data in this way [19]. The importance of the type of content viewed as well as the amount has also been acknowledged [20]. More recently, researchers have advocated using complex log data from digital interventions to further our understanding of engagement [11,12] and to examine relationships between usage, participant characteristics, and health outcomes [21]. However, the importance of providing consistently reported findings that will enable comparison of usage across different digital interventions has also been highlighted [6,8]. Existing guidelines encourage precise and standardized reporting for general analyses of digital interventions [22,23]. The challenge of undertaking efficient and systematic analysis of large datasets without the guidance of a framework is already acknowledged: Sieverink et al detailed the importance of using research questions to guide analysis of log data [24] and Taki et al demonstrated how categorizing different usage metrics can inform our understanding of engagement [25]. However, systematic reviews suggest that these types of analysis of usage data are not yet routinely undertaken [7,26]. This may be due to the absence of a framework that contains comprehensive checklists combining both the systematic breakdown of usage data and the formulation of research questions to structure usage analyses of digital interventions. In addition, without prior identification of necessary data capture processes, the final usage data collected may be unable to answer the research questions posed.

This paper proposes a novel framework to structure the process of analyzing usage associated with a digital intervention by doing the following: (1) drawing together potential measures of usage and identifying which are meaningful to the intervention, (2) generating specific research questions to act as testable hypotheses, and (3) supporting data preparation and selection of methods for analysis. Specifically, the framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) can encourage the collection and/or extraction of data that will explain who used which parts of the intervention at what time and whether that was associated with positive outcomes. The framework focuses on usage as a key component of engagement, but does not aim to encompass all aspects of engagement. Nonetheless, the examination and analyses of usage data, using the framework, can move toward the identification of what constitutes effective engagement. In addition, the framework offers an approach to digital intervention data analysis that can be applied both before and after data collection.

When used during intervention development, the AMUsED framework aids development teams to compile an a priori analysis plan for use after data collection. This allows the opportunity to evaluate whether all necessary data will be collected and whether this is in a suitable format for analysis at a later date. This is particularly pertinent for interventions that are developed with external partners who may be unaware of the theoretically based elements of the intervention and their implications for analyses. When applied after data collection, the framework is especially useful for general orientation when

a researcher is unfamiliar with the intervention or when no advance plan of analysis is available. Using the framework helps focus exploratory usage analyses on addressing the theory underpinning an intervention and the plausible mechanisms of action on target outcomes, aiding more scientifically rigorous analyses. Should an analysis plan be available, the framework facilitates a review to ensure that the plan is still appropriate and aids revision where necessary.

Development of the Framework

The AMUsED framework was initially developed as a means to systematically and rigorously analyze post hoc usage data collected during digital intervention trials. The first author (SM) was tasked with analyzing usage data from *Internet Dr*, a successfully trialed digital intervention (see the case study involving Internet Dr below). This task was challenging because of the author's unfamiliarity with the intervention, the depth and complexity of data collected, and the absence of an existing framework to provide step-by-step guidance on approaching a usage analysis. Stages 1 and 2 of the framework were developed alongside the process of understanding and beginning analyses of usage data collected from Internet Dr. An early version of the framework was presented to a multidisciplinary digital intervention development team with experience across health psychology, primary care, and statistics. The framework was then refined based on the team's input and experiences of applying the framework to their own usage analyses. The value of having a systematic process through which to consider data collection during the development phase of a digital intervention was subsequently noted. The framework was then expanded and applied to the amendment of a second intervention, *Germ Defence* (see the case study involving Germ Defence below). Following this, the framework was presented to the wider scientific community at a national conference in the United Kingdom. Here, the value of using the structure provided by the framework to support collaboration between social scientists and software development companies and identification of necessary data collection processes was recognized. The framework was then shaped further to provide equal weight to both a priori and post hoc analysis needs.

Description of the Analyzing and Measuring Usage and Engagement Data (AMUsED) Framework

Overview

The framework is presented in three stages: (1) familiarization with available datasets, (2) selecting meaningful measures of usage and generating research questions, and (3) preparation for analysis. Each stage is available in checklist format, with generic questions acting as prompts for the researcher to consider in the context of their own specific intervention (see Multimedia Appendices 1-3). It is anticipated that use of the three stages will be iterative depending upon whether the framework is being applied in advance of or after data collection (see Figure 1). For example, when considering appropriate analytical software (Stage 3) during the development phase of an intervention, it may be necessary to reformat how data are recorded to ensure compatibility. Alternatively, analyses of collected data may reveal unexpected patterns of usage, such as repeated visits to a component of content, from which new exploratory research questions can then be generated (Stage 2).

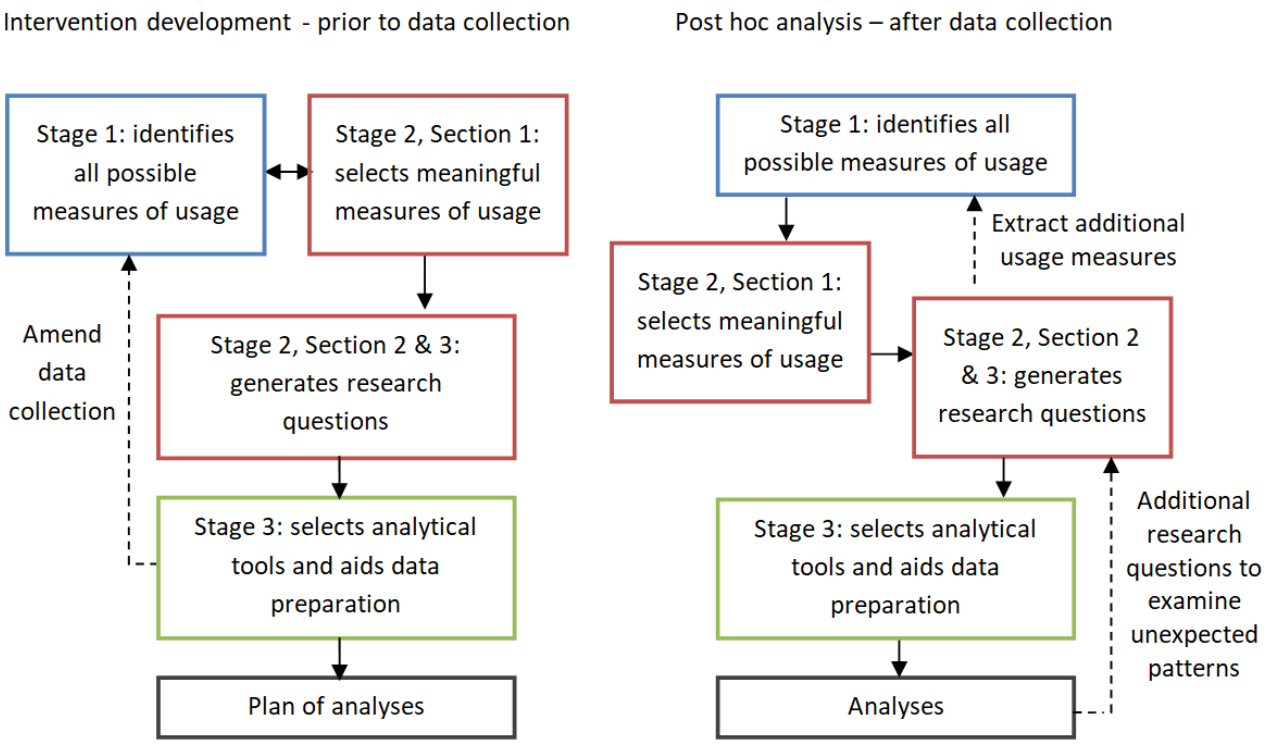
The framework focuses specifically on examining the relationships and associations between measures of usage and user characteristics, theoretical variables, behavior, and/or health-related outcomes. However, it is anticipated that analyses of usage would be considered in the context of a broader process evaluation that may examine how variables other than usage are associated with intervention outcomes [27].

Stage 1: Familiarization With the Data—Identifying Variables

Evaluation of a digital intervention can produce large datasets containing information collected in a variety of formats. It may be necessary to collate relevant data across the datasets and compute new variables before usage analyses can be conducted. To simplify this process, Stage 1 proposes a checklist (see Multimedia Appendix 1) comprising a set of generic questions that will support a comprehensive understanding of the structure, processes, and content of the intervention in relation to data capture, contents of the datasheets, and factors related to trial implementation (eg, participant recruitment, Stage 1, Item 3). When used during the development phase of an intervention, the framework provides the opportunity to record and measure usage data that align with the proposed analysis plan. Ensuring efficient data capture at the outset can remove the need for extensive data cleaning and manipulation. When used for post hoc analysis only, Stage 1 can support the identification of appropriate usage variables and inform subsequent data cleaning and manipulation in preparation for analysis.

The usage data has been grouped into three categories. *Intervention characteristics* describes architecture, content, and expected workflow through the intervention, including intended usage [15] (eg, anticipated number of log-ins, number of available content components, and number of pages within a tunneled section). *Accrued data* covers all data collected during the running of the intervention, such as logs of interactions or log data (eg, date and time of use, pages viewed, and time spent on them) and user-entered data (eg, self-report). *Contextual data* encompasses previous findings related to the intervention development and trial (eg, factors affecting usage) and relevant external factors (eg, national health promotion campaigns).

Figure 1. Stages of the Analyzing and Measuring Usage and Engagement Data (AMUsED) framework. Dotted lines indicate optional paths to revisit if necessary.



Accrued data is usually collected automatically and recorded in datasheets, making the variables easier to extract and analyze. While some variables for intervention characteristics and contextual data may also be captured in this way (eg, number of log-ins), it is anticipated that additional measures may be either conceptual or external, meaning that they would not automatically be recorded as analyzable measures. For example, a digital intervention designed to increase physical activity may include pages addressing barriers to the target behavior, such as not having enough time to exercise or not having access to equipment. Pages containing advice to overcome these barriers may be distributed throughout the intervention, but are based on the same underlying theoretical concept (ie, an intervention characteristic). Therefore, a new variable needs to be created capturing usage of these pages and exposure to the theoretical underpinning. By identifying these measures of usage in advance, it is possible for additional data capture processes to be created, aiding efficient analyses after data collection. For example, depending on the delivery platform, code may be added to an intervention so that users who view these pages are recorded within a unique variable column in the log data. Contextual data is also less likely to be automatically collected and recorded within the study. If a large-scale outbreak of a respiratory infection occurs during the trial of an intervention aimed at reducing transmission of such infections, one might want to assess the potential impact on usage data. As with intervention characteristics, it may be possible to add further measures in order to capture personal experience of the illness or impact in a broader context.

It should be noted that the range of data available for collection may differ depending on the software used to develop and/or deliver the digital intervention. Using the framework during development to determine in advance which data are crucial may facilitate software development or else alternative workarounds. For example, where software is unable to collate total time spent on selected intervention pages, it will be necessary to ensure time spent by page is readily available to collate this after data collection. Interventions will also vary greatly in architecture and structure, depending on design, software, and delivery platform used (eg, website, app, or text based). For example, a *session* (Stage 1, Item 1) may refer to a single log-in, a component of content available across multiple log-ins, or the amount of times a specific activity is accessed. The framework provides a structure broad enough to be applied to different interventions. However, it is not anticipated that all criteria in the checklists will apply to every digital intervention. Where concepts and examples provided do not directly translate, researchers are encouraged to define them as relevant for their intervention and adapt the framework as needed.

Stage 2: Selecting Measures of Usage and Generating Research Questions for Engagement

Overview

Establishing testable hypotheses is the precursor to carrying out systematic analyses. The aim of Stage 2 is to support the generation of specific research questions to drive testing of hypotheses. Stage 2 is divided into three sections to reflect the

increasing complexity of comprehensive usage analyses: Section 1 helps define specific measures of usage (ie, descriptive statistics), while Sections 2 and 3 generate research questions (ie, bivariate and multivariate analyses).

Stage 2, Section 1: Descriptions of Usage Variables

The first section of Stage 2 provides a nonexhaustive list of potential usage measures. Example questions on the checklist (see Multimedia Appendix 2) demonstrate how measures of usage may be constructed.

The abundance of data and potential usage variables can encourage unsystematic data dredging. Identifying and reviewing the range of usage measures available enables researchers to make informed and/or theoretically driven decisions about what will be the most meaningful variables to include in any subsequent analysis plan. The process of familiarization with intervention content and architecture (Stage 1) may highlight considerations when selecting usage variables for analysis. For example, it may become apparent that certain sections of the digital intervention were considered to be of greater importance during the planning process, such as components that are theoretically informed (eg, pages containing BCTs or advice and goal-setting sections). Therefore, analyzing the usage of these pages, specifically, would be more meaningful than analyzing the total number of pages viewed. Alternatively, the intervention logic model may indicate that two theoretically based components are considered to have equal importance, yet they may have differing amounts of content within them, meaning that users would spend more time on one than the other. On that basis, analyzing the time spent on theory-based components may result in misleading conclusions about the impact of usage on health outcomes. In this scenario, a categorical usage metric may be more meaningful (eg, having completed or revisited the component). Analyzing a single usage measure is unlikely to provide a comprehensive understanding of engagement for all users across an intervention. However, combining multiple usage measures in a systematic way will provide a more detailed understanding of how users engaged with the intervention and what patterns of usage are associated with intended outcomes.

During intervention development, this process is undertaken prior to data collection and is therefore based on prospective data identified in Stage 1—bracketed numbers provided in Stage 2, Section 1, indicate their counterpart sections in Stage 1. Given the considerable crossover, we anticipate that these sections will be completed iteratively (see Figure 1). The purpose of Stage 1 is to identify all potential measures of usage available within the data; Stage 2 then narrows down that selection by considering which measures will provide the most informative understanding of usage for a specific intervention. The selection is informed by fundamental elements of the intervention highlighted in the planning process [13,14]. For example, in a digital intervention targeting weight loss, important measures of usage might be identified as entering weekly self-reports of weight, repeat use of recipe component, and time spent watching exercise videos. When carrying out post hoc analyses, descriptive statistics for measures of usage identified in Stage 2 may provide greater insight into which measures will be more informative (see Figure 1).

Stage 2, Sections 2 and 3: Relationships Between Usage, Participant Characteristics, Target Behaviors, and Behavioral Determinants

The remaining two sections of Stage 2 (see Multimedia Appendix 2) will guide the generation of specific research questions to assess how usage might be related to participant characteristics, behavioral determinants, and target behavior. While this stage can be used to generate limitless questions to drive exploration of the data, the framework is instead intended to be used to help select the most important questions that will answer theory-driven hypotheses. Usage variables are considered in relation to participant characteristics (Stage 2, Item 2), target behavior and behavioral determinants (Stage 2, Item 3), and behavior change across the intervention (Stage 2, Item 3). By answering these questions, it is anticipated that patterns of usage that reflect effective engagement with a specific intervention can be described. The moderating effect of demographic, psychosocial, and health factors (Stage 2, Item 2) on the relationship between usage and outcomes are also considered in Section 3 of Stage 2. When defining these variables, it is intended that the framework be adapted to individual interventions; for example, it is possible that a measure of usage (eg, uploading ongoing health monitoring statistics) may also be the intended primary outcome [13].

Stage 3: Preparation for Analysis

The Stage 3 checklist (see Multimedia Appendix 3) supports the process of selecting appropriate types of analyses and analytical software, as well as the data preparation necessary to translate the research questions developed in Stage 2 into a plan of analysis. Generic questions guide the researcher to consider broad issues, such as available resources (Stage 3, Item 1) (eg, timeframe, additional researcher support, and analysis plan for efficacy), more specific issues of selecting appropriate type of analysis and analytical software (Stage 3, Item 2), and data management (Stage 3, Item 3) (eg, amalgamation, manipulation, and cleaning).

Our experience suggests that traditional statistical methods are not always suitable for analyzing the types of research questions generated by the framework. For example, while research is usually powered to analyze efficacy, it is frequently underpowered for the type of subgroup analyses needed for in-depth usage analyses. In addition, whereas analyses of amounts of usage (eg, total time spent or number of log-ins) often lend themselves to traditional methods, examining patterns of usage (eg, movement through pages) requires alternative methods to identify and inform subsequent statistical analysis. Therefore, techniques such as visualization and process mining may be more informative as they can reveal patterns of usage within the data, such as workflow through an intervention, clustering by participant groups, and temporal details [12,28-31]. In applying Stage 3 of the framework, the required data format for any analytical software should be considered.

Application of the Analyzing and Measuring Usage and Engagement Data (AMUsED) Framework: Two Case Studies

Overview

The following section provides researchers with practical examples of how the framework checklists can be applied in advance of or after data collection. The key findings and applications from utilizing the framework are highlighted below. The framework is necessarily comprehensive and completion of the checklists creates a lot of data and information. Indeed, this is the very process by which it supports the generation of systematic and rigorous usage analyses. The completed checklists for both studies have been amalgamated to enable comparison and are available as supplementary data (see Multimedia Appendices 4-6) so researchers using the framework in practice have detailed examples of its use.

Applying the Framework During Development: Germ Defence

Overview

PRImary care trial of a website-based Infection control intervention to Modify Influenza-like illness and respiratory infection Transmission (PRIMIT) was a large randomized controlled trial (RCT) that showed a digital intervention to be effective at lowering the transmission of colds, influenza, and stomach upsets within the home through increased handwashing [32]. The framework was used to inform and structure the process of updating and amending the intervention to make it ready for dissemination as an open-access resource for use by the general public. As part of that process, the intervention was renamed *Germ Defence*. The research team involved with the dissemination was already familiar with the intervention, having worked on the design and evaluation of the PRIMIT study—for full details of the PRIMIT intervention and evaluation trial, please see Little et al [32].

Stage 1: Familiarization With the Data

Applying Stage 1 of the framework supported us in undertaking a detailed review of the original version of Germ Defence (Stage 1, Item 3.2; see Multimedia Appendix 4), along with data collected from the prior RCT (Stage 1, Item 2). This informed crucial updates to the collection of usage data and the generation of research questions, which we describe in the following sections.

Disseminating Germ Defence to the general public required us to strike a balance between obtaining informed consent to collect a minimal amount of data to support evaluation, while still enabling easy access to key aspects of the intervention by users who may be less willing to engage with standard research procedures (Stage 1, Item 3.1). Completing the Stage 1 checklist also allowed us to identify the following: (1) how the intervention and consent procedures should be streamlined (Stage 1, Items 3.1 and 3.2) and (2) what pertinent self-report data should be collected to enrich analyses of the automatically collected usage data and enable comparison with the prior RCT data (Stage 1, Item 1.2).

Stage 2, Section 1: Descriptions of Usage Variables

We reviewed the range of possible usage variables and identified which ones would provide the most informative picture of how Germ Defence was accessed and used during dissemination (see Multimedia Appendix 5). For example, the first component of the intervention contains compulsory tunneled pages, including a section for selecting handwashing goals. Examining dropout across this component and online consent pages, along with repeat use of the goal-setting section, will enable us to understand if and where users disengaged with the intervention. We then compared our list with the data collected from the prior RCT. This identified crucial amendments to the data capture process for Germ Defence that would otherwise have been missed. Specifically, data recorded on use of the goal-setting component was overwritten when revisited, losing both user-entered data and our ability to view movement backward and forward through these pages (Stage 1, Item 1.2). Identifying this issue in advance meant we were able to adapt the back-end processes to ensure the required data were captured.

Stage 2, Sections 2 and 3: Relationships Between Usage, Participant Characteristics, Target Behaviors, and Behavioral Determinants

Completion of Stages 1 and 2 of the framework in parallel helped us to narrow down our selection of usage-related questions to focus on behavioral determinants that were identified to be most strongly correlated with the target behavior in the prior RCT (see Multimedia Appendix 5). Since efficacy of Germ Defence has already been established from the prior RCT, the primary focus of the dissemination phase is to examine patterns of usage “in the wild” and their relationship to baseline user characteristics. The following research questions are a selection of those generated (Stage 2, Item 2):

1. Which pages see the highest amount of dropout, including consent and baseline measures?
2. How do users move through the goal-setting pages and what goals do they select?
3. Are baseline measures for handwashing, level of belief that handwashing will lower infection transmission, and/or belief in the ability to increase handwashing associated with usage?
4. Do users’ perceptions about the risk of infection to themselves or a household member relate to usage?
5. Do the means through which users hear about the website relate to usage?

Self-report data on behavioral and psychological variables will be collected using an optional survey (Stage 2, Item 3) in order to minimize potential dropout. This could be subject to selection bias with significant differences in the characteristics of users choosing to complete or not complete the survey. Any analysis examining the association between usage and behavioral outcome or change in behavioral determinants will be undertaken with caution. However, accessing or completing the survey may be operationalized as a measure of usage (Stage 2, Item 1), providing the opportunity to analyze relationships

between intervention and survey usage (eg, Is viewing more intervention pages associated with completing the survey?) (Stage 2, Item 2). A comparison of baseline characteristics will also enable a check of whether those who complete the follow-up survey are different to those who do not.

Stage 3: Preparation for Analysis

The analytical tools available are SPSS for Windows version 24 (IBM Corp) and LifeGuide Visualisation Tool (University of Southampton) (Stage 3, Item 2; see Multimedia Appendix 6) [28,33]. It is anticipated that there will be insufficient power for definitive hypothesis testing. Patterns of usage (eg, repeat use and dropout across tunneled pages) will be best explored, initially, using visual tools. As Germ Defence has been built using LifeGuide software, the data produced will be compatible with the visualization tool (Stage 3, Item 3). Usage data collected from the intervention will need to be amalgamated and linked with self-report data from the optional survey. Thus, it was necessary to ensure that all users were allocated a unique nonidentifiable numeric ID upon first access so that all data can be linked (Stage 3, Item 3).

Using the research questions developed in Stage 2 and considerations highlighted in Stage 3, a full plan of analysis for Germ Defence was developed to inform efficient and systematic analysis after data collection. Applying the framework helped us prioritize research questions most relevant for the focus of the research (eg, how interventions are accessed and used "in the wild") that would not be undermined by the constraints of using optional self-report measures.

Applying the Framework for Post Hoc Analysis: Internet Dr

Overview

The framework was used to develop an analysis plan for usage data collected during an RCT of *Internet Dr*, a digital intervention to support the self-care of respiratory tract infections (RTIs) and to reduce unnecessary general practitioner (GP) visits. The RCT showed that users with access to Internet Dr were less likely to contact their GP about an RTI than those without access [32]. The usage analyses for Internet Dr will be conducted by researchers who were not involved in the original design, development, and evaluation of the intervention. The framework enabled the researchers to understand the intervention and associated data collection and to construct systematic research questions to investigate usage—for full details of the Internet Dr intervention and evaluation trial, please see Little et al [34].

Stage 1: Familiarization With Data

Internet Dr is structured around three components of theoretically based content. *Doctor's Questions* and *Common Questions* aim to support users who are unsure if their symptoms are serious and whether they are in need of medical treatment (Stage 1, Items 1.2 and 3.2) [35]. *Treatment Options* is intended to increase self-efficacy for users who wish to manage symptoms they are finding distressing (Stage 1, Items 1.2 and 3.2) [36]. Applying Stage 1 aided understanding of how these three components relate to the psychological theories underpinning the intervention and, thus, the proposed determinants of the target behavior (ie, illness perception, health locus of control, willingness to tolerate symptoms, and treatment preferences; see Multimedia Appendix 4).

All content was available whenever users accessed the intervention across a 24-week period in the winter (Stage 1, Item 1.1). However, users were encouraged to log in specifically during periods of illness to help manage their symptoms. Completing the checklist emphasized the importance of recognizing these two distinct purposes for accessing the intervention: (1) to view content while ill and (2) to view content when well, perhaps out of curiosity. These differences in motivation to access the intervention when well or unwell may also be reflected in differences in patterns of usage.

Stage 2, Section 1: Descriptions of Usage Variables

Given the theoretically based content of the three components within the intervention, usage of each was identified as relevant to understanding underlying mechanisms of action (eg, number of users, number of pages viewed, time spent, and number of revisits) (see Multimedia Appendix 5). For the *Doctor's Questions* component, compulsory tunneled pages are completed, leading to illness management advice on the last page (Stage 1, Item 1.1). Therefore, users of this component would, in theory, not benefit unless they had reached the final page, so completion and dropout were identified as important measures of usage for this component. It was also intended by design that users would view *Doctor's Questions* first (Stage 1, Item 1.2). Thus, analyzing the order in which users visited the different components was important to understand whether the intervention was used as intended, as well as how intended versus nonintended order of use was related to users' perceptions of their RTI, their perceived ability to self-manage, and whether they contacted their GP. As differences in users' motivations for accessing the intervention may lead to differences in usage patterns (Stage 1), measures of usage identified in this component (eg, number of pages viewed and time spent) should be described for three situations: usage when ill, usage when well, and across all usage.

Stage 2, Sections 2 and 3: Relationships Between Usage, Participant Characteristics, Target Behaviors, and Behavioral Determinants

Considerations from the previous sections helped form pertinent research questions for the remaining two sections of Stage 2 (see Multimedia Appendix 5). For example, relationships between viewing specific content and theoretical constructs and behavioral determinants will be examined. As motivations for use have been identified as potentially influential, associations between reasons for accessing the intervention and patterns of usage and/or personal characteristics will be explored (Stage 2, Item 2). In addition, we will examine whether users followed the intended navigational paths and whether this was related to visiting their GP (Stage 2, Item 3). Below are some example research questions from the usage analysis plan:

1. Does usage of *Doctor's Questions* differ when intervention access is made when ill compared to when well (eg, starting

the component, number of pages viewed, time spent, and viewing the advice page)?
2. Are baseline personal characteristics associated with intervention use when ill or not ill?
3. Is viewing content during illness associated with lower GP visits?

Finally, we aim to identify whether viewing a specific piece or amount of content at a certain time (eg, when ill) led to a user being less likely to contact their GP.

Stage 3: Preparation for Analysis

SPSS for Windows version 24 (IBM Corp) and LifeGuide Visualisation Tool (University of Southampton) [21,26] will be used for data analysis (Stage 3, Item 2; see Multimedia Appendix 6). There will be sufficient power to analyze average usage of the intervention (eg, by whole sample) and associations with behavioral determinants and target behavior. Subgroup analyses are unlikely to be sufficiently powered (eg, comparing usage and outcomes of users accessing content when ill versus when prompted by completion of interim study measures). Some of the identified patterns of usage include movement through the intervention (Stage 3, Item 1) and will therefore be best explored visually (eg, the order in which the three content components were accessed). The datasheets are compatible with LifeGuide Visualisation Tool (University of Southampton). However, data is spread across several datasheets, requiring extraction, transformation, and amalgamation prior to analysis (Stage 3, Item 3).

Through completion of the three stages of the framework (see Multimedia Appendices 4-6), it was possible to break down a complex digital intervention and develop a comprehensive usage analysis plan, which will help identify what type of usage was successful in supporting self-management, for whom was it most beneficial, and at what time it was most influential.

Comparing Case Studies

Both interventions target behavior associated with RTIs. However, Germ Defence focuses on infection prevention and may be accessed at any time, whereas Internet Dr supports self-management of symptoms while infected with an RTI. Despite the differences in their architecture, content, and function, the framework was suitably generic to be applied to both interventions. Although the same Stage 1 checklist was applied to both interventions, it enabled two completely different processes: for Germ Defence, Stage 1 helped shape structural changes to the intervention and data capture processes; for Internet Dr, Stage 1 enabled understanding of a previously unfamiliar and complex intervention and the accompanying datasheets (see Multimedia Appendix 4). Despite these different requirements, the checklist was comprehensive enough to fulfill both needs and lead to greater insights, such as realizing that a key component of Germ Defence (ie, the goal-setting component) was not capturing data as required, and understanding that Internet Dr was designed to be used during illness, but could be accessed at any time. Completing this first stage was the most complex and time consuming of the three stages for both interventions. However, through the thorough understanding of the intervention gained from Stage 1, the subsequent stages were easier to complete as the information was readily available to fit the generic questions. For example, having identified the theoretical underpinning of the three components of Internet Dr, their related measures of behavioral determinants, and expected relationship to GP contact, generating research questions to examine how usage related to changes in behavior and behavioral determinants was both simple and quick.

Through completing the checklist for Stage 2, Section 1, it is apparent that operationalizing usage in terms of amount was valid for both interventions (eg, number of log-ins, number of pages viewed, and time spent on pages; see Multimedia Appendix 5). This suggests that describing the extent to which an intervention has been used is a necessary first step for examining usage and that the number of pages viewed and time spent on them may inform our understanding of different styles of engagement. For example, spending more time on or revisiting pages may be indicative of higher levels of interaction compared to viewing pages briefly. However, as previously discussed, focusing on broad-based, summative descriptions of usage alone may not be sufficient to understand how the intervention supported change in target behaviors and outcomes. For example, distinguishing and comparing usage of the three theoretically based content components of Internet Dr will aid understanding of the potential mechanisms of action within the intervention.

The differences in research questions generated from Stage 2 highlight the differences in structure between the two interventions. Germ Defence is a stand-alone intervention requiring access only once, with an optional follow-up survey. Once completed, the data generated from using the intervention will provide a snapshot of behavior at that time. On that basis, research questions focus on user characteristics and behavioral determinates and target behavior at baseline (see Multimedia Appendix 5). In contrast, in addition to self-report and log data over a 6-month period, Internet Dr users' GP notes providing information for the year before and after the trial commenced were also collected. This depth and length of duration of data collection enables different research questions, including consideration of behavior prior to the trial and for some months after.

Stage 2 of the framework highlighted the relevance of examining patterns of usage (eg, movement through the intervention) and subgroups usage analyses. As both interventions have insufficient power to analyze subgroup usage, and patterns of usage lend themselves to visual exploration (see Multimedia Appendix 6), this supports the use of contemporary visualization tools in addition to traditional statistical methods.

Both interventions include research questions examining relationships between participant characteristics and usage and whether any of those characteristics moderate the relationship between usage and target behavior. Through comparison of these similar analyses across multiple interventions, it will become possible to build up a pattern of how personal characteristics may influence digital intervention usage, leading to generic learning points to inform future intervention design. This may also be the case for usage analyses of interventions

with similar aims (eg, self-management of illness) or similar theoretical underpinning, behavioral determinants, or BCTs. Once a body of research is assembled, it would be possible to use the framework in advance to structure data capture and analysis so that it is comparable with prior interventions and published research.

Discussion

Overview

The AMUsED framework aims to support detailed and systematic analysis of digital intervention usage. The framework comprises three stages of checklists for researchers to do the following: (1) understand the intervention's design, theoretical underpinning, and data collection processes; (2) define meaningful variables to assess usage and generate both broad and fine-grained research questions to examine relationships between usage, participant characteristics, and target behavior and behavioral determinants; and (3) prepare datasheets and consider appropriate software for analysis.

The framework has been applied to two digital interventions: Germ Defence promotes RTI prevention and Internet Dr supports self-management of RTI symptoms. Using the framework while preparing Germ Defence for public dissemination identified necessary amendments to data capture processes. For Internet Dr, the framework helped guide a research team who were previously unfamiliar with the intervention design and data to devise a comprehensive usage analysis plan. The case studies demonstrate the flexibility of the framework to be applied to different interventions and the advantages of using the framework, both before and after data collection.

Implications

The AMUsED framework checklists provide researchers with easily applied templates for carrying out detailed usage analyses of digital interventions. The framework supports the level of rigor in reporting digital intervention content and findings called for by current guidelines from the UK Medical Research Council, the Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth (CONSORT-EHEALTH), and the Template for Intervention Description and Replication (TIDieR) [22,23,27,37]. The checklists extend upon and draw together previous work on categorizing digital intervention usage data and selecting research questions [19-21,24,25] by providing a systematic and comprehensive process for researchers to follow. The process can be incorporated into existing digital intervention development methods, such as the person-based approach [14] and the behavioral intervention technology model [15], enabling pretesting of data capture processes to support theory-based hypothesis testing.

The framework encourages usage analyses that will broaden our understanding of mechanisms of action underlying a specific digital intervention, explaining the relationships between user characteristics, patterns of usage, and behavior change. Through this process, it may be possible to identify effective engagement, finding the level of usage necessary for a specific intervention in order to change the target behavior [11]. This will lead to digital interventions being developed to be more concise, efficient, and targeted, making them less arduous for the user and supporting higher rates of uptake and engagement.

Limitations

The framework has been developed and tested using Web-based interventions built using the same software [33] that captures extensive log data and has the ability for researchers to write additional code in order to capture tailor-made usage measures. Digital interventions may be delivered across a wide variety of platforms (eg, text messaging, apps, and websites) and developed using different software. This leads to substantial variation in design, the manner in which they are written or coded, and the availability and format of data collected. The framework is flexible enough to be applied across diverse interventions and sufficiently detailed to generate specific testable hypotheses for most digital interventions. However, we welcome other researchers to use the framework and build upon it based on their experience.

The AMUsED framework focuses on the analysis of measures of usage as one facet of engagement. Where objective measures of physiological reactions (eg, cardiac activity and eye tracking) or subjective self-report measures of engagement are available, it is hoped that future research may examine these alongside usage data and develop the framework further to incorporate them, thereby increasing our ability to explain not just the role of usage, but engagement more broadly [6]. It is also our hope that the framework will be applied as part of a mixed-methods approach, triangulating usage analyses with insights and experiences collected qualitatively [6,8,10].

The framework has been applied to ensure adequate data collection when used during the development phase of digital interventions. However, the framework also has the potential to be used to inform study design to answer empirical questions on effective engagement; for example, multiphase optimization strategy (MOST) and sequential multiple assignment randomized trial (SMART) [38]. Although this paper does not address this application, it provides an avenue for future research for the wider application of the framework.

Conclusions

The AMUsED framework offers a systematic process for carrying out in-depth usage analyses. The aim of the framework is to capture and formalize the techniques used by experienced researchers to support researchers who are new to conducting usage analyses, or new to a particular intervention, in deciding how to assess usage data that will be or has been collected. Using the framework will benefit researchers by lowering the possibility of overlooking key questions and making the reporting of usage analyses more efficient, leading to a quicker turnaround for publishing. The checklists provide the means to increase transparency and make findings easier to replicate, while discouraging unsystematic data dredging. The process will also encourage greater detail and consistency in the reporting of usage and engagement, making it easier to apply the findings to a wider context [7] and enabling comparison across different interventions and evaluation studies. The

framework helps to operationalize and measure usage in ways that will better inform our understanding of engagement with a digital intervention, encompassing broad measures of usage by the whole sample, through to specific theory-based usage variables and usage by subgroups based on personal characteristics. It guides insight into which components of an intervention worked and how they interacted with users' personal characteristics. Finally, by using the framework it may be possible to identify the extent of usage required to support changes in behavior and health-related outcomes and, thus, an understanding of what constitutes an effective level of engagement for specific interventions.

Acknowledgments

This work was funded by the Economic and Social Research Council (award number 1692136; Understanding digital intervention engagement: Making sense of large-scale usage data). The *Internet Dr* and *Germ Defence* interventions were developed using LifeGuide software, which was partly funded by the National Institute for Health Research Southampton Biomedical Research Centre. The authors would like to thank the LifeGuide Team and Digital Intervention Group at the University of Southampton, Southampton, United Kingdom.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Stage 1 checklist for the Analyzing and Measuring Usage and Engagement Data (AMUsED) framework.

[PDF File (Adobe PDF File), 968KB - jmir_v21i2e10966_app1.pdf]

Multimedia Appendix 2

Stage 2 checklist for the Analyzing and Measuring Usage and Engagement Data (AMUsED) framework.

[PDF File (Adobe PDF File), 779KB - jmir_v21i2e10966_app2.pdf]

Multimedia Appendix 3

Stage 3 checklist for the Analyzing and Measuring Usage and Engagement Data (AMUsED) framework.

[PDF File (Adobe PDF File), 584KB - jmir_v21i2e10966_app3.pdf]

Multimedia Appendix 4

Stage 1 checklist with case studies.

[PDF File (Adobe PDF File), 1MB - jmir_v21i2e10966_app4.pdf]

Multimedia Appendix 5

Stage 2 checklist with case studies.

[PDF File (Adobe PDF File), 1MB - jmir_v21i2e10966_app5.pdf]

Multimedia Appendix 6

Stage 3 checklist with case studies.

[PDF File (Adobe PDF File), 640KB - jmir_v21i2e10966_app6.pdf]

References

1. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for drug and alcohol abuse and smoking addiction: Is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):178-186. [doi: 10.1016/j.cpr.2010.10.002] [Medline: 21095051]
2. Webb TL, Joseph J, Yardley L, Michie S. Using the Internet to promote health behavior change: A systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

3. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
4. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of eHealth interventions in somatic diseases: A systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
5. Solomon D, Proudfoot J, Clarke J, Christensen H. e-CBT (myCompass), antidepressant medication, and face-to-face psychological treatment for depression in Australia: A cost-effectiveness comparison. J Med Internet Res 2015 Nov 11;17(11):e255 [FREE Full text] [doi: 10.2196/jmir.4207] [Medline: 26561555]
6. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Dec;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
7. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of user engagement with mobile- and Web-delivered self-care interventions for adults with T2DM: A review of the literature. Curr Diab Rep 2016 Dec;16(7):66 [FREE Full text] [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
8. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Dec;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
9. Sheeran P, Klein WMP, Rothman AJ. Health behavior change: Moving from observation to intervention. Annu Rev Psychol 2017 Jan 03;68:573-600. [doi: 10.1146/annurev-psych-010416-044007] [Medline: 27618942]
10. 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 Dec 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
11. Ainsworth B, Steele M, Stuart B, Joseph J, Miller S, Morrison L, et al. Using an analysis of behavior change to inform effective digital intervention design: How did the PRIMIT website change hand hygiene behavior across 8993 users? Ann Behav Med 2017 Jun;51(3):423-431 [FREE Full text] [doi: 10.1007/s12160-016-9866-9] [Medline: 27909944]
12. Morrison LG, Geraghty AWA, Lloyd S, Goodman N, Michaelides DT, Hargood C, et al. Comparing usage of a Web and app stress management intervention: An observational study. Internet Interv 2018 Jun;12:74-82 [FREE Full text] [doi: 10.1016/j.invent.2018.03.006] [Medline: 30135771]
13. Band R, Bradbury K, Morton K, May C, Michie S, Mair FS, et al. Intervention planning for a digital intervention for self-management of hypertension: A theory-, evidence- and person-based approach. Implement Sci 2017 Dec 23;12(1):25 [FREE Full text] [doi: 10.1186/s13012-017-0553-4] [Medline: 28231840]
14. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: Application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
15. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: An integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
16. 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]
17. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From theory-inspired to theory-based interventions: A protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. Ann Behav Med 2018 May 18;52(6):501-512. [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
18. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: Mapping theoretically derived behavioural determinants to behaviour change techniques. Appl Psychol 2008 Jul 08;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
19. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
20. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. J Med Internet Res 2006 Aug 30;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]
21. Paz Castro R, Haug S, Filler A, Kowatsch T, Schaub MP. Engagement within a mobile phone-based smoking cessation intervention for adolescents and its association with participant characteristics and outcomes. J Med Internet Res 2017 Dec 01;19(11):e356 [FREE Full text] [doi: 10.2196/jmir.7928] [Medline: 29092811]
22. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: Improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

23. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: The new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
24. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: Collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 07;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
25. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing user engagement of an mHealth intervention: Development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
26. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: Systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
27. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
28. Arden-Close EJ, Smith E, Bradbury K, Morrison L, Dennison L, Michaelides D, et al. A visualization tool to analyse usage of Web-based interventions: The example of positive online weight reduction (POWeR). JMIR Hum Factors 2015 May 19;2(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.4310] [Medline: 27026372]
29. Morrison C, Doherty G. Analyzing engagement in a Web-based intervention platform through visualizing log-data. J Med Internet Res 2014 Nov 13;16(11):e252 [FREE Full text] [doi: 10.2196/jmir.3575] [Medline: 25406097]
30. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, Manage My Pain: Results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
31. van der Aalst W. Process Mining: Data Science in Action. 2nd edition. Berlin, Germany: Springer; 2016.
32. Little P, Stuart B, Hobbs FDR, Moore M, Barnett J, Popoola D, et al. An Internet-delivered handwashing intervention to modify influenza-like illness and respiratory infection transmission (PRIMIT): A primary care randomised trial. Lancet 2015 Oct 24;386(10004):1631-1639. [doi: 10.1016/S0140-6736(15)60127-1] [Medline: 26256072]
33. Williams S, Yardley L, Weal M, Willis G. Introduction to LifeGuide: Open-source software for creating online interventions for health care, health promotion, and training. In: Proceedings of Med-e-Tel 2010: The International eHealth, Telemedicine, and Health ICT Forum for Education, Networking, and Business. Grimbergen, Belgium: International Society for Telemedicine & eHealth (ISfTeH); 2010 Presented at: Med-e-Tel 2010: The International eHealth, Telemedicine, and Health ICT Forum for Education, Networking, and Business; April 14-16, 2010; Luxembourg p. 460-463 URL: https://www.isfteh.org/files/media/Electronic_Proceedings_of_Med-e-Tel_2010.pdf
34. Little P, Stuart B, Andreou P, McDermott L, Joseph J, Mullee M, et al. Primary care randomised controlled trial of a tailored interactive website for the self-management of respiratory infections (Internet Doctor). BMJ Open 2016 Dec 20;6(4):e009769 [FREE Full text] [doi: 10.1136/bmjopen-2015-009769] [Medline: 27098821]
35. Leventhal H, Brissette I, Leventhal EA. The common-sense model of self-regulation of health and illness. In: Cameron LD, Leventhal H, editors. The Self-Regulation of Health and Illness Behaviour. London, UK: Routledge; 2003:42-65.
36. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY: Worth Publishers; 1997.
37. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
38. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]

Abbreviations

AMUsED: Analyzing and Measuring Usage and Engagement Data
BCT: behavior change technique
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth
GP: general practitioner
MOST: multiphase optimization strategy
PRIMIT: PRImary care trial of a website-based Infection control intervention to Modify Influenza-like illness and respiratory infection Transmission
RCT: randomized controlled trial
RTI: respiratory tract infection
SMART: sequential multiple assignment randomized trial

TIDieR: Template for Intervention Description and Replication

Edited by G Eysenbach; submitted 04.05.18; peer-reviewed by O Perski, C Yeager, L Nelson; comments to author 07.08.18; revised version received 17.10.18; accepted 30.10.18; published 15.02.19.

Please cite as:
Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, Morrison L
A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint
J Med Internet Res 2019;21(2):e10966
URL: http://www.jmir.org/2019/2/e10966/
doi:10.2196/10966
PMID:30767905

Original Paper

Electronic Health Program to Empower Patients in Returning to Normal Activities After General Surgical and Gynecological Procedures: Intervention Mapping as a Useful Method for Further Development

Chantal M den Bakker[1,2], MD; Frederieke G Schaafsma[1], MD, PhD; Eva van der Meij[1,3], MD; Wilhelmus JHJ Meijerink[4], MD, PhD; Baukje van den Heuvel[4], MD, PhD; Astrid H Baan[5], MD; Paul HP Davids[6], MD, PhD; Petrus C Scholten[7], MD, PhD; Suzan van der Meij[8], MD; W Marchien van Baal[9], MD, PhD; Annette D van Dalsen[10], MD; Daniel J Lips[11], MD, PhD; Jan Willem van der Steeg[12], MD, PhD; Wouter KG Leclercq[13], MD; Peggy MAJ Geomini[14], MD, PhD; Esther CJ Consten[15], MD, PhD; Steven E Schraffordt Koops[16], MD, PhD; Steve MM de Castro[17], MD, PhD; Paul JM van Kesteren[18], MD, PhD; Huib A Cense[19], MD, PhD; Hein BAC Stockmann[20], MD, PhD; A Dorien ten Cate[21], MD; Hendrik J Bonjer[2], MD, PhD; Judith AF Huirne[3], MD, PhD; Johannes R Anema[1], MD, PhD

[1]Amsterdam Public Health Research Institute, Department of Occupational and Public Health, VU University Medical Center, Amsterdam, Netherlands
[2]Department of Surgery, VU University Medical Center, Amsterdam, Netherlands
[3]Department of Obstetrics and Gynaecology, VU University Medical Center, Amsterdam, Netherlands
[4]Department of Operation Rooms, Radboud University Medical Center, Nijmegen, Netherlands
[5]Department of Surgery, Amstelland Ziekenhuis, Amstelveen, Netherlands
[6]Department of Surgery, Diakonessenhuis, Utrecht, Netherlands
[7]Department of Obstetrics and Gynaecology, Diakonessenhuis, Utrecht, Netherlands
[8]Department of Surgery, Flevoziekenhuis, Almere, Netherlands
[9]Department of Obstetrics and Gynaecology, Flevoziekenhuis, Almere, Netherlands
[10]Department of Surgery, Isala, Zwolle, Netherlands
[11]Department of Surgery, Jeroen Bosch Ziekenhuis, 's-Hertogenbosch, Netherlands
[12]Department of Obstetrics and Gynaecology, Jeroen Bosch Ziekenhuis, 's-Hertogenbosch, Netherlands
[13]Department of Surgery, Máxima Medisch Centrum, Veldhoven, Netherlands
[14]Department of Obstetrics and Gynaecology, Máxima Medisch Centrum, Veldhoven, Netherlands
[15]Department of Surgery, Meander Medisch Centrum, Amersfoort, Netherlands
[16]Department of Obstetrics and Gynaecology, Meander Medisch Centrum, Amersfoort, Netherlands
[17]Department of Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam, Netherlands
[18]Department of Obstetrics and Gynaecology, Onze Lieve Vrouwe Gasthuis, Amsterdam, Netherlands
[19]Department of Surgery, Rode Kruis Ziekenhuis, Beverwijk, Netherlands
[20]Department of Surgery, Spaarne Gasthuis, Hoofddorp, Netherlands
[21]Department of Obstetrics and Gynaecology, Spaarne Gasthuis, Haarlem, Netherlands

Corresponding Author:
Chantal M den Bakker, MD
Amsterdam Public Health Research Institute
Department of Occupational and Public Health
VU University Medical Center
Van der Boechorststraat 7
Amsterdam, 1081 BT
Netherlands
Phone: 31 204445703
Fax: 31 204448387
Email: c.denbakker@vumc.nl

Abstract

Background: Support for guiding and monitoring postoperative recovery and resumption of activities is usually not provided to patients after discharge from the hospital. Therefore, a perioperative electronic health (eHealth) intervention ("ikherstel" intervention or "I recover" intervention) was developed to empower gynecological patients during the perioperative period. This eHealth intervention requires a need for further development for patients who will undergo various types of general surgical and gynecological procedures.

Objective: This study aimed to further develop the "ikherstel" eHealth intervention using Intervention Mapping (IM) to fit a broader patient population.

Methods: The IM protocol was used to guide further development of the "ikherstel" intervention. First, patients' needs were identified using (1) the information of a process evaluation of the earlier performed "ikherstel" study, (2) a review of the literature, (3) a survey study, and (4) focus group discussions (FGDs) among stakeholders. Next, program outcomes and change objectives were defined. Third, behavior change theories and practical tools were selected for the intervention program. Finally, an implementation and evaluation plan was developed.

Results: The outcome for an eHealth intervention tool for patients recovering from abdominal general surgical and gynecological procedures was redefined as "achieving earlier recovery including return to normal activities and work." The Attitude-Social Influence-Self-Efficacy model was used as a theoretical framework to transform personal and external determinants into change objectives of personal behavior. The knowledge gathered by needs assessment and using the theoretical framework in the preparatory steps of the IM protocol resulted in additional tools. A mobile app, an activity tracker, and an electronic consultation (eConsult) will be incorporated in the further developed eHealth intervention. This intervention will be evaluated in a multicenter, single-blinded randomized controlled trial with 18 departments in 11 participating hospitals in the Netherlands.

Conclusions: The intervention is extended to patients undergoing general surgical procedures and for malignant indications. New intervention tools such as a mobile app, an activity tracker, and an eConsult were developed.

Trial Registration: Netherlands Trial Registry NTR5686; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=5686

(J Med Internet Res 2019;21(2):e9938) doi:10.2196/jmir.9938

KEYWORDS

intervention mapping; eHealth; return to normal activities; return to work; patient reported outcome measures; colectomy; hysterectomy

Introduction

Background

The length of in-hospital stay after general surgical and gynecological procedures has decreased significantly due to a growing trend in day-care surgery, introduction of minimal invasive techniques, and enhanced recovery after surgery programs (ERAS) [1-3]. Due to this shortening of in-hospital stay, perioperative in-hospital care has been reduced and the greater part of the recovery period takes place at home [4,5]. As a result, guiding and monitoring of resumption of normal activities (RNA) including return to work (RTW) and long-term recovery are now transferred to primary care [6]. However, frequently no or conflicting advice is given, resulting in patients being unsure whom to contact for support in case of complaints. Patients often lack the knowledge themselves to determine how and when to resume activities [7-10]. As a consequence, full recovery after surgery takes much longer than expected despite improved surgical treatment. A longer recovery at home could result in diminished general and mental health, higher medical consumption, lower quality of life, and longer sick leave period [10-13].

Electronic health (eHealth) can be a suitable tool to optimize the quality of perioperative care of patients who will undergo general surgical and gynecological procedures. eHealth can provide tailored information, increase patients' self-management, and has interactive communication features [14]. Furthermore, it has the potential to empower patients, to motivate patients, and turn them into more active and effective managers of their own health [15-17]. A recovery-oriented eHealth intervention ("ikherstel" intervention or "I recover" intervention) has already proven to be effective with a significant faster RTW after benign gynecological surgical procedures [18]. This care program requires a need for further development to fit a broader population of patients who will undergo various types of general surgical and gynecological procedures.

Objectives

In this paper, the further development with corresponding process of this "ikherstel" intervention is described [19]. The objective of the eHealth intervention development includes to further optimize (1) empowerment of general surgical and gynecological patients during the perioperative period to RNA and RTW and (2) partial substitution of perioperative care with eHealth. For this, the Intervention Mapping (IM) protocol is used, which is a suitable systematic and scientifically accepted method for the (further) development and implementation of a wide range of eHealth and RTW interventions. This method is based on theory and stakeholders' (including patients') involvement [20,21].

Methods

Overview

The IM protocol for the further development of the "ikherstel" intervention consists of 6 steps (see Figure 1): (1) forming a logic model of the problem; (2) defining program outcomes and

objectives; (3) designing the eHealth program; (4) producing the eHealth program; (5) developing a program implementation plan; and (6) making an evaluation plan [20,21].

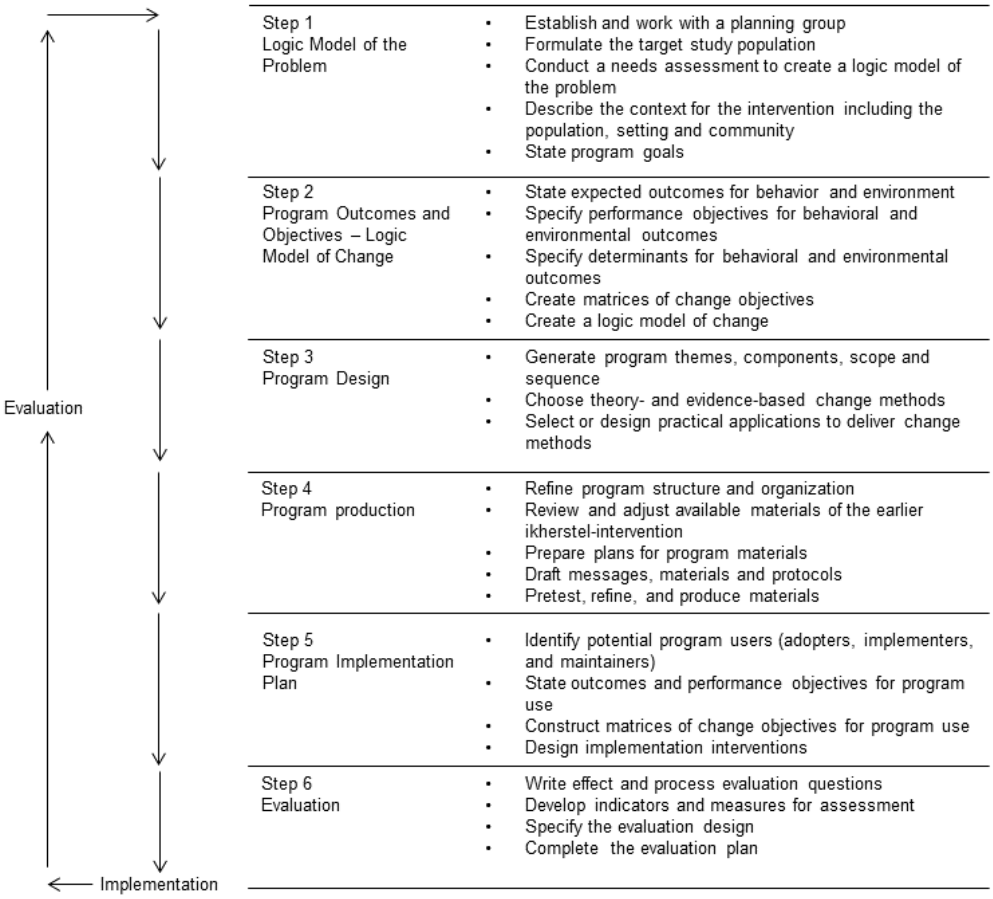
In the previous "ikherstel" study, only women were included as the intervention was only available for patients undergoing gynecological abdominal procedures. In this study, men will also be included because the target population includes patients undergoing general surgical procedures. Currently, the most performed general surgical abdominal procedures in elective setting in the Netherlands are hernia inguinal repair (with 28,232 procedures per year), cholecystectomy (25,203 procedures), and colectomy (14,012 procedures) [22]. Patients undergoing these surgical procedures are chosen as part of the target population, next to those patients undergoing gynecological procedures including hysterectomy and adnexal surgery as the other part of the target population. This way, a broad patient population of general surgical and gynecological patients can be achieved [18,19,23,24]. The study population is divided in (1) the minor abdominal general surgical and gynecological procedures group, which consists of laparoscopic cholecystectomy and adnexal surgery and laparoscopic or open hernia inguinal repair and (2) the major abdominal general surgical and gynecological procedures group, which consists of laparoscopic or open colectomy and hysterectomy.

In the first 3 preparatory steps of further development, the process is described, starting from, and based on, the original developed "ikherstel" intervention [19]. This original intervention is extended, optimized, and made applicable to the broader target population. In the last 3 steps of the intervention development, the implementation and evaluation plan is described for only the major abdominal general surgical and gynecological procedures group, because the protocol of the randomized controlled trial (RCT) of the minor abdominal procedures is already published elsewhere [25].

Step 1: Logic Model of the Problem

First, a planning group was composed to work with during the whole IM process. Next, a multifactorial patients' needs assessment was conducted which included (1) a process evaluation of the earlier performed "ikherstel" study, (2) a review of the literature, (3) a survey study, and (4) focus group discussion (FGDs). Findings from the needs assessment were then compiled into a PRECEDE-PROCEED model (predisposing, reinforcing, and enabling constructs in educational diagnosis and evaluation-policy, regulatory, and organizational constructs in educational and environmental development) to identify the factors the program should address to help improve full recovery including RNA and RTW after surgery at home for abdominal general surgical and gynecological patients in the Netherlands.

Figure 1. Intervention mapping steps.



Process Evaluation

A process evaluation was performed on the previous "ikherstel" intervention to gain more insights into the facilitators and barriers for the acceptance and implementation of this intervention [26]. These results were used for the further development of the new "ikherstel" intervention. We evaluated how the eHealth intervention was delivered to and received by participants and how participants and health care professionals had appreciated the intervention.

Review of the Literature

A systematic review was performed to evaluate the effect of perioperative eHealth interventions on the postoperative course of any surgical procedure [27]. This review provides us important and general information about effective aspects of eHealth interventions for the further development of our eHealth intervention.

Furthermore, a literature search was performed to further investigate and gather more information about the average duration of full recovery including RNA and RTW after abdominal surgical procedures including factors that affect this duration. The literature search of the IM procedure of the previous "ikherstel" intervention was used and broadened to identify behavioral and environmental conditions of prolonged sick leave and delayed RTW among gynecological patients [19]. The search was executed in PubMed with the key words "abdominal surger*" and "recover*" or "rehabilitation" or "return to normal activit*" to select articles describing additional factors for delayed RNA in general surgical patients. This also included a search for behavioral and environmental explanations for delayed full recovery.

Survey Study

A survey study was performed to explore specific needs in the target population regarding the "ikherstel" intervention. Patients who underwent surgery (minor and major abdominal general surgical or gynecological procedures) between August 2013 and August 2014 in the VU University medical center (located in Amsterdam) received a questionnaire. This survey study was performed to (1) evaluate shortcomings in information and guidance supplied to patients in current perioperative care for patients undergoing general surgical and gynecological procedures, (2) investigate whether eHealth may be of assistance in this, and (3) to identify gender-specific needs [28]. On the basis of these results, the previous "ikherstel" intervention could be further adapted to this population.

Focus Group Discussions

In total, 4 FGDs were performed to gather patients' experiences with recovery after (multimodal) cancer treatment and to identify particular needs during the various phases of cancer treatment. The second objective was to evaluate possible solutions for unmet needs by the introduction of eHealth. By gathering this information, the intervention can be adjusted in such a way that it is also applicable to patients undergoing surgery due to a malignant indication. Colon cancer patients were recruited from patient files of 2 Dutch teaching hospitals: Meander Medical Center (located in Amersfoort) and Spaarne Gasthuis hospital (located in Hoofddorp and in Haarlem). Endometrial cancer patients were recruited from the patient files of 3 teaching hospitals: VU University medical center (located in Amsterdam), Antoni van Leeuwenhoek hospital (located in Amsterdam), and the Flevo Hospital (located in Almere).

Step 2: Program Outcomes and Objectives-Logic Model of Change

In the second step, the program outcomes and objectives were developed. Performance objectives were specified to describe in detail the patients' behavioral and environmental outcomes that were considered necessary to reach full recovery after surgery. Determinants for these behavioral and environmental outcomes were selected, and by crossing the performance objectives with these determinants (constructing matrices), the change objectives were identified. These matrices were used to identify behaviors and conditions that result in a sustainable full recovery. This step aimed to determine whether performance objectives, behavioral determinants, and change objectives of the earlier "ikherstel" intervention needed to be modified to better fit the broader target population. A selection of the performance objectives of the IM article of Vonk Noordegraaf et al was further supplemented with new performance objectives based on the findings of the needs assessment and the literature on behavior change [19].

Step 3: Program Design

In the third step, program themes, components, scope, and sequence were specified and theory- and evidence-based change methods were searched in PubMed and applied based on the findings in the first 2 steps. Theory-based methods for change are general techniques or processes that have been shown to enable change in one or more determinants of behavior and have their origins in behavioral and social science theories [20,29]. Choosing the theory- and evidence-based change methods and selecting or designing practical applications to deliver change methods was done by the planning group and guided by the formulated performance and change objectives in the second step of the IM protocol and were also based on the performance and change objectives as described in the earlier performed "ikherstel" study [19].

Step 4: Program Production

In the fourth step, the program structure was refined and organization was planned. All gathered information from the previous steps was synthesized and translated into plans for drafting program materials, that is, tailored tools and information to empower major abdominal general surgical and gynecological patients by an innovative eHealth care program. Participants of the FGDs in step 1 and health care professionals (surgeons, gynecologists, and residents in training) outside the project group pretested the different aspects of our eHealth intervention, whereby possible bugs and shortcomings were identified, refined, and adjusted.

Step 5: Program Implementation Plan

In the fifth step, potential program users were reidentified in line with the broadening of the patient population and the consequences for the implementation and acceptance of the eHealth intervention. With this information, an implementation plan to enable an extensive evaluation of the intervention was

developed including a plan to identify and inform the health care professionals and researchers in the participating hospitals about the different core components of the intervention. Furthermore, details about optimal delivery of the intervention were incorporated herein. Hospitals that participated in previous "ikherstel" studies gave their approval to participate again. Other hospitals were recruited by the PhD candidates. The planning of the adoption and implementation of the intervention was established.

Step 6: Evaluation

In the sixth step, an evaluation plan of the intervention was developed, which involves determining whether behavior outcomes change as a result of the adjusted eHealth intervention. For this evaluation, an RCT was designed to measure the effects of the eHealth intervention on RNA and RTW for patients undergoing major abdominal general surgical and gynecological procedures. This study was approved by the Medical Ethics Committee of the VU University medical center under registration number 2014.301. This study was also registered at the Netherlands Trial Registry under registration number NTR5686.

Results

Step 1: Logic Model of the Problem

The planning group (development team) consisted of 2 PhD candidates, 2 occupational health physicians, 3 surgeons, and 1 gynecologist. This multidisciplinary planning group was established to further develop the intervention by applying the various steps of the IM process. The members were selected to represent various fields of expertise required for the design of the intervention.

Process Evaluation

In the earlier performed "ikherstel" study, 210 patients were included, of which 110 patients were allocated to the intervention group. The implementation score of the intervention was 80%. All patients were given access to the "ikherstel" intervention and 86.4% made a convalescence plan. Reasons for not making or not adhering to the convalescence plan include patients (1) preferred to resume activities when they felt ready for it, (2) found the convalescence plan too optimistic or too conservative, (3) felt pushed by the convalescence plan, and (4) felt that the plan did not apply to their personal situation.

The eHealth intervention was perceived effective by 74% of patients, and 76% of involved gynecologists were satisfied with the Web portal. In total, 95% of the health care professionals would offer the intervention to their patients in the future, and 85.3% of patients would recommend the "ikherstel" eHealth program to a friend.

Required access to internet, the inflexibility of the eHealth intervention in case of complications for patients, and an increased time investment for gynecologists were rated as possible future usage barriers. Suggestions for improvement included an extra section with experiences of other women. All results of the process evaluation are published in detail elsewhere [26].

Review of the Literature

The systematic review included 27 studies that focused on replacing or complementing perioperative usual care with some form of care via information and communication technology (ICT) such as telemonitoring, telerehabilitation, teleconsultation, or an educational or supportive website in various types of surgery. Of these studies, 92% reported at least an equal or positive effect of the eHealth intervention compared with usual care on patient-related outcomes. The results show that eHealth or other forms of ICT and telehealth improve clinical outcomes, knowledge, and satisfaction of patients undergoing various forms of surgery. Furthermore, eHealth improves RTW and daily functioning after surgery. Positive influencing factors on these patient-related outcomes were (1) easy access to the intervention, (2) expectation management, and (3) a combined symptom monitoring by blended care [27].

The literature search in the IM protocol article of Vonk Noordegraaf et al identified that pain and discomfort, feelings of fear, and infections were delaying factors for RTW [19]. In addition, literature showed that the substantial variation in convalescence recommendations given by health care professionals were also of influence on the total duration of sickness absence [9,30]. In the broadened search, a wide variety in convalescence duration after minor general surgical procedures (eg, cholecystectomy and hernia inguinal repairs) was observed [31,32]. For major abdominal procedures, limited research was available to analyze the time of full recovery including RNA (and RTW) [33-36]. However, literature does show that the total duration of convalescence was longer than expected despite the implementation of minimal invasive techniques and ERAS programs applied by major abdominal procedures [1,37,38]. Duration of convalescence depends on pain, complications, or fear of recurrence (in case of cancer of hernia inguinal repairs) [39-41]. Preoperative expectations of convalescence, size of the incision, and fatigue were important contributory factors to explain actual convalescence [39-42]. Furthermore, patients who were physically active after colorectal surgery were more likely to recover faster [43].

Survey Study

In total, 57.2% (207/362) potential participants completed the survey. Mean age of participants was 46.6 years and almost 30% were male. For 87.4% participants, the indication for surgery was benign with an equal distribution between general surgical and gynecological procedures.

A reported shortcoming related to information and guidance provision was the lack of detailed advice about the resumption of activities following surgery. Many participants reported receiving inconsistent recommendations from medical specialists, general practitioners, and occupational physicians. Limited guidance from professionals during the recovery process was also mentioned as a shortcoming by 40% of the participants. Some participants preferred to receive more information or more emotional and mental support after their surgical procedure. A perceived lack of information or support resulted in more nervousness before surgery or insecurity after surgery.

eHealth was expected to be a very useful tool to overcome these above-mentioned shortcomings. If an eHealth intervention had been available before or after their surgical procedure, 71% of the participants reported they would have used it. Most popular rated items of a future eHealth care program were a page containing an overview of important telephone numbers, a list with frequently asked questions, and the possibility to self-evaluate symptoms after surgery. Furthermore, the option of an electronic consultation (eConsult) was rated useful by 57.6%, and almost half of the participants preferred to use the "ikherstel" intervention also via a mobile phone app. The option to give an employer or an occupational physician access to parts of the website and the option of a patient forum were not rated useful. Limited gender differences in preferences were identified in this survey. Women showed a slightly higher need for information and preferred some extra eHealth support. The total results of the survey study are published elsewhere [28].

Focus Group Discussions

For this study, 40% (30/75) potential participants were willing and available to participate in the colon cancer FGDs. For the endometrial cancer FGDs, 35% (17/48) potential participants were willing to participate. A total of 22 patients actually participated in the colon cancer FGDs and 12 in the endometrial cancer FGDs. Most frequently reported unmet needs in the perioperative phase of colon cancer patients were an absence of tailored, dosed, and understandable information and advice regarding RNA. Colon cancer patients who had finished the adjuvant chemotherapy phase would have liked to receive more information about side effects, more mental support, and a longer aftercare period. Endometrial cancer patients evaluated the received information and guidance as very well. This was mainly due to very good guidance from the gynecologist and physician assistant. The participants without adjuvant treatment reported that they would like to have more recommendations regarding resuming normal activities, and they saw a role for eHealth to support this. However, patients who had adjuvant treatment rated the nurses' guidance as sufficient and needed no additional support.

Colon cancer patients treated with multimodal treatment and endometrial cancer patients who did not receive any adjuvant treatment reported that eHealth services could be supportive but not a substitute of personal interaction with health care professionals. There was a preference for a Web-based health care system that is readily available 24/7 in the form of blended care. As there is already a lot of information on the internet about cancer diagnoses and treatment, it was sometimes hard for participants to differentiate which information is correct. Recommendation of the tool by their own health care professional would enhance perception of safety and, therefore, increase usage. In contrast, endometrial cancer patients who received adjuvant treatment did not see an added value for eHealth. This difference might be explained by the age difference in these patient groups. Results of the FGDs will be reported in more detail as separate papers.

On the basis of the results of this needs assessment, the overall desired outcome for further development of the "ikherstel" intervention tool was defined as "achieving earlier recovery including RNA and RTW." Furthermore, the findings from the needs assessment were compiled into a PRECEDE-PROCEED model to identify patients' problems and needs in perioperative surgical care in which eHealth and mobile health (mHealth) can have a (complementary) role and are defined as:

- A lack of clear and simple instructions for the RNA-including work;
- Inconsistent recommendations from different health care providers;
- A lack of information about surgical procedures and the perioperative course, symptoms, and complications;
- Limited mental support in case of patients with a malignant indication;
- Delayed and limited mobilization interventions to realize earlier RNA and RTW; and
- No or limited interaction with their health care professional during postoperative course.

Step 2: Program Outcomes and Objectives-Logic Model of Change

In the previous "ikherstel" intervention, the following overall objectives were formulated. After the needs assessment, these objectives were adjusted to align to patients' problems and needs as identified under step 1:

- To enhance recovery by giving clear and simple instructions for the RNA;
- To stop inconsistent recommendations from different health care providers;
- To take away the insecurity with respect to postoperative course, symptoms, and complications.

The following overall objectives were added to the existing objectives after the needs assessment in step 1:

- To provide extra attention and mental support in case of patients with a malignant indication,
- To encourage patients to a quicker and more intense mobilization and earlier RNA, and
- To let patients have more interaction with their health care professional via a web portal.

Specified performance objectives for the further development of the "ikherstel" intervention are presented in Multimedia Appendix 1. To create a matrix of *performance objectives*, the main personal and external determinants of behavior change for each performance objective were operationalized. The Attitude-Social Influence-Self-Efficacy model was used in the IM procedure of the earlier "ikherstel" intervention and was used for the further development of this intervention as well [19]. This model was still considered by the planning group as the most suitable model for recovery and behavior change and, thereby, used to form change objectives [44-49]. Skills, barriers, and facilitators were considered relevant factors for RNA [50,51]. In Multimedia Appendix 2, an example of the performance objective "mobilize quickly and more intense after operation" is presented.

Step 3: Program Design

The same practical methods and suitable strategies as in the IM article of Vonk Noordegraaf et al were used for the further

development of tools and materials of the "ikherstel" intervention as the methods and strategies used in this intervention already had been proven effective [19,52]. For example, self-monitoring of behavior (awareness) and getting direct feedback will be used to stimulate patients to mobilize quickly and more intensively. An activity tracker will be used in the "ikherstel" intervention to encourage this. Multimedia Appendix 3 presents more examples of these methods with preconditions and final tool/materials of the eHealth intervention.

Step 4: Program Production

The knowledge gathered in the first 3 steps was discussed by the planning group at several meetings to add various appropriate tools to the existing "ikherstel" intervention. A designer/developer specialized in eHealth and mHealth interventions was consulted during some of the meetings. In addition, experienced surgeons, gynecologists, and residents in training outside the planning group were consulted to judge the medical content of the tools. The words "ikherstel.nl" in the internet address of the eHealth intervention means "I am recovering." The mobile app was made available for iOS and Android mobile phones and tablets. The design of the further developed "ikherstel" intervention for patients undergoing major abdominal surgical and gynecological procedures (this includes open or laparoscopic colectomy and open or laparoscopic hysterectomy) is described below. In Table 1, an overview of all tools of the developed eHealth intervention is presented.

Adaptation of Existing "ikherstel" Intervention Tools

Website

The website aims to prepare patients in the best possible manner for their surgery and to offer guidance during their recovery process until full recovery and resumption of all activities are achieved. The following tools on the website support this.

Table 1. Component of the "ikherstel" intervention regarding each target population.

Tool	Content	Target population
Website		
Information by text and animations about the surgical procedure	Enhancing patient preparation including creating expectations	All participants
Personalized convalescence plan	Enhancing patient preparation including creating expectations	All participants
Recovery monitor and recovery report	Monitoring recovery and offering assistance when relevant	All participants
Video	Increasing the information provision by using several ways to provide this	All participants
Glossary	Increasing the information provision by using several ways to provide this	All participants
Frequently asked questions	Increasing the information provision by using several ways to provide this	All participants
Electronic consultation	Increasing access to care and reducing patient uncertainties and fear related to the recovery process and workload	All participants
Information about malignancies	Reducing anxiety and uncertainty and increasing the amount of information provision	All participants with malignant disease
Information about chemotherapy and side effects	Reducing anxiety and uncertainty and increasing the amount of information provision	All participants with malignant disease
Links for supportive care needs	Creating long-term support to reduce anxiety and uncertainty	All participants with malignant disease
Mobile app		
Information by text	Enhancing patient involvement and recovery expectations and reducing anxiety	All participants with a smartphone
Insight into the convalescence plan	Creating recovery expectations and improving recovery	All participants with a smartphone
Recovery monitor and recovery report	Reducing uncertainties and fear related to the recovery process and improving monitoring and transition of postoperative care	All participants with a smartphone
Creating a packing list	Increasing better patient preparation before admission	All participants with a smartphone
Section to make notes	Increasing better patient preparation before admission and during in-hospital stay	All participants with a smartphone
Activity tracker		
Monitoring and giving feedback on recovery	Reducing uncertainties and fear related to the recovery process, which may improve recovery	All participants with a suitable smartphone for the activity tracker

Making a Personalized Convalescence Plan

A personalized and tailored convalescence plan, including advice about resumption of (work) activities is the most important tool on the website. The specific tailored convalescence recommendations were developed for relevant types of abdominal surgical procedures by using a modified Delphi procedure [53,54]. On the website, information will be tailored for each patient, offering the opportunity to enhance patient involvement. This is possible as some data are already prefilled when patients receive their website account (eg, surgical procedure, sex, and hospital). This tool will enhance recovery by giving clear and simple instructions for the RNA and will stop inconsistent recommendations from different health care providers.

Providing Information About the Surgical Procedure and Recovery Process

Information per treatment phase (preoperative, perioperative, and postoperative) will be provided by text and images. These services will contribute perioperative to the patients' awareness and expectations; both factors have proven to be important predictors of the length of recovery. Information will also have a positive effect on anxiety and satisfaction because the patients can prepare themselves better for the surgery. Postoperative information will be offered about the recovery period and common postoperative complaints. This could support patients during this period and may help them with feelings of insecurity. In addition, practical advice about when, how, and whom to contact in case of complaints will be provided. Patients will be helped in deciding whether to contact a health care professional in case of complaints or complications during their recovery. Frequently asked questions will be added as well and these were formulated based on main topics in patients' brochures and Web-based patient discussion forums. This tool will take away the insecurity with respect to the postoperative course, symptoms, and complications.

Getting Feedback on Recovery by a Recovery Monitor and Recovery Report

The recovery monitor and report are tools to identify recovery problems and give patients feedback on their recovery progress. Patients will be asked to indicate in a recovery monitor to what extent they have resumed their activities, which will be subsequently graphically displayed in a recovery report allowing them to track their progress. It also aims to improve monitoring and transition of postoperative care. After the patient has given consent, the Web portal can be accessed by a health care professional in secondary care to monitor patients' recovery and, thus, identify recovery problems.

Developing the Materials and Tools of the Revised "ikherstel" Intervention

Providing Information on the Website

For patients who will receive adjuvant chemotherapy, information about this treatment, including side effects, is provided on the website. This will provide extra attention and mental support in case of patients with a malignant indication. Videos are also added on the website. Videos are considered the most appropriate medium to deliver an informative message to patients because of the influence of modeling behavior on attitude [55]. Videos about the admission day, the surgical procedures, receiving and managing a permanent or definitive stoma, the postoperative period, and a simulation of patients and employers to discuss potential RTW problems are provided.

Postoperative Consult by Electronic Consultation

In case of recovery problems, patients have the option to ask questions to a health care professional from their own hospital by means of an eConsult via the website. Patients will be informed that eConsults are only suitable for nonurgent questions and that these questions will be answered within 2 working days. In case of urgent questions, they receive a phone number for direct contact. The hypothesis is that patients will be more comfortable and less hampered in resuming their activities with the opportunity to ask questions whenever they prefer. This tool will let patients have more interaction with their health care professional via a Web portal.

Mobile Phone App (Mobile Health)

All information which is available on the website is also available on the mobile phone app ("ikherstel" app), which will be synchronized with the website. This includes among others the convalescence plan that patients created on the website. A section to make notes and the option to compose a list of what to pack when being admitted to the hospital will also be available on the app. If patients do not have a smartphone, they will only use the website. This tool will enhance recovery by giving clear and simple instructions for the RNA and will stop inconsistent recommendations from different health care providers (see Multimedia Appendix 4).

Activity Tracker

An activity tracker that measures the daily step count can be linked to the "ikherstel" application for patients with a smartphone. This tracker will be used as a support tool for patients to monitor and to give feedback on their recovery. The average daily step count in the week before surgery will be set as their baseline measurement and, thereby, target postoperative activity level. The daily step count will be postoperative graphically displayed in the app as a percentage of their target activity level, including a target level the patient is expected to reach. Patients will be asked to wear the activity tracker in the first 4 weeks after surgery, and again after 8 weeks as the hypothesis is that baseline activity level should be reached in this week. This tool will encourage patients to a quicker and more intense mobilization and earlier RNA.

Pretest of Materials

In total, 10 patients and a representative sample of health care professionals evaluated the demo version of the eHealth "ikherstel" intervention. Patients got 3 weeks to test the intervention before they were interviewed by the researcher. Semistructured interviews were conducted with patients who also had participated in the FGDs. Health care professionals were asked to judge the demo version on several items (eg, layout, comprehensibility of all informative text). The test patients were satisfied with the content of the information, the way it was delivered, and the messages (source and style). They also found all text provided on the website as very useful.

However, a few remarks for improvement were suggested. These were related to supplying more detailed information about the side effects of chemotherapy, adding more attention to mental support, and less complicated sentences. All test patients would like to use the intervention if they will have surgery again.

Refinement and Production of Materials After Testing

After participants' and health care professionals' feedback, minimal adjustments were made. Mainly textual changes were made to simplify and order sentences. This resulted in the final eHealth intervention that was used to perform the RCT.

Step 5: Program Implementation Plan

Patients undergoing major abdominal, general surgical, or gynecological procedures were identified as new program users. No new professionals were identified as program users as the introduction and implementation of the program would remain in the perioperative phase also for the extended patient group. The 7 hospitals that participated in the earlier performed "ikherstel" study wished to continue the use of the "ikherstel" intervention including broadening the usage of "ikherstel" intervention at the general surgery department of their hospital. Many other general hospitals in different regions in the Netherlands signed the letter of intent to participate in this multicenter study. After a kick-off meeting at the research institute, one meeting with surgeons or gynecologists (depending on the department) and one meeting with nurses per hospital was held to discuss the content and logistics of the study and its implementation. Finally, 10 general surgery practices and 8 gynecology practices (all teaching hospitals) were positive about the further development of "ikherstel" intervention and willing to implement the intervention. They will participate in the evaluation of this intervention through implementation of the eHealth intervention as a supplement to the standard perioperative care given at their hospital.

Step 6: Evaluation

Study Design

The evaluation of the eHealth intervention will be performed by a multicenter single-blinded RCT. The effectiveness and cost-effectiveness of this innovative eHealth care program compared with the usual care given in 11 participating hospitals on RNA and RTW will be evaluated. A process evaluation will also be performed using a mixed-methods design [56].

Eligibility Criteria

Patients who will undergo a laparoscopic or open colectomy or hysterectomy and are aged between 18 and 75 years will be contacted. Exclusion criteria are surgery without a curative intention, deep infiltrating endometriosis, concomitant surgical procedures, not able to use the internet, unable to understand Dutch questionnaires, malignancy (in case of the hysterectomy), and receiving neoadjuvant treatment.

Outcome Measures

Our primary outcome measure is RNA. The Patient-Reported Outcomes Measurement Information System physical functioning item bank version 1.2 will be used to measure limitations in daily activities. A list of 29 most relevant selected activities will be presented to participants before surgery with the goal to select 8 activities, which are most relevant for them in their daily life. They will be asked in the following questionnaires after surgery if they can already perform one of these 8 activities. RNA is defined as the time in calendar days from the day of surgery until a participant has resumed all activities [57,58].

Secondary outcomes are social participation, self-rated health, duration until RTW, physical activity, length of recovery, pain intensity, and patient satisfaction [59-63]. Costs will be measured from a societal and health care perspective and consist of costs of the intervention, health care utilization costs, and costs associated with lost productivity [64,65]. Sociodemographic data and questions regarding expectations about the length of recovery and the amount of anxiety will be assessed at the baseline measurement. Complications will be assessed by reviewing the surgical reports and postoperative notes and scored by using the Clavien-Dindo classification [66]. The outcome measures will be obtained by using questionnaires administered at baseline (approximately 1-2 week preoperative) and at 2, 4, and 6 weeks and 3, 6, 9, and 12 months after surgery. For a total overview of all outcomes per measurement moment of this study, see Multimedia Appendix 5.

Sample Size Calculation

To detect a hazard ratio of 1.4 for RNA (corresponding to a decrease in median time to RNA from 10 weeks to 7.14 weeks as a result of the intervention) with 80% power while testing using a 2-sided log-rank test at a significance level of 5%, a total of 318 events need to be observed. The total sample size is set at 354 (177 per arm) to account for an anticipated proportion of 2.5% of patients not returning to daily activities within the 12 month follow-up period and a dropout rate of 10%.

Recruitment, Inclusion, Allocation, and Blinding of Patients

Patients will be recruited for study participation when they are on the waiting list in one of the participating hospitals and will receive a study information letter on behalf of their doctor. Contact will be made by phone to check their willingness to participate and to access eligibility. Eligible patients willing to participate will be included. After the patient completes the baseline questionnaire T(0) within 2 week before surgery and has signed informed consent, randomization will be executed by an (independent) research assistant. A computer-generated randomization in a 1:1 ratio will be performed on individual level stratified regarding hospital, sex, and surgical procedure using permuted blocks of size 2. Patients will be blinded to the intervention, as they do not know which program is developed as a nonintervention or intervention care program. The researchers involved in the analyses will be blinded to the allocation throughout the analyses. Health care professionals cannot be blinded to the intervention because it is highly likely that they will be notified of the allocation either by the patient or the patients' medical file.

Data Analyses Plan

All analyses will be performed in IBM SPSS. Baseline characteristics will be summarized using descriptive statistics and compared between the intervention and control group using *t* tests by normal distributions of variables and Mann-Whitney *U* tests, chi-square tests, or Fisher exact tests by non-normal distributions of variables. Our primary outcome, time until full RNA, will be analyzed by both crude and adjusted survival analyses where hospital, surgical procedure, and sex will be taken into account as covariates in the adjusted analyses due to stratification. To describe the distribution of the duration until RNA in both groups, the Kaplan Meier method will be used. The Cox proportional hazard model will be applied to calculate hazard ratios. Adjustments will be made if there are clinically relevant differences between the intervention and the control group in the baseline characteristics or if other potential confounding factors are observed. For the longitudinal secondary outcomes, mixed models and multilevel logistic regression models will be performed. For the cross-sectional secondary outcomes, *t* tests, Mann-Whitney *U* tests, chi-square tests, or Fisher exact tests will be used to compare differences. Intention-to-treat analyses will be compared with per-protocol analyses to identify whether there are differences if patients used the intervention as intended. Subgroup analyses will be performed regarding the surgical procedure (colectomy and hysterectomy) and indication (benign or malignant disease). A post hoc analysis will be carried out on patients without major complications.

Discussion

Principal Findings

The purpose of this study was to describe the systematic process of the further development of the "ikherstel" intervention following the 6 steps of the IM protocol [20,21]. This eHealth intervention was adjusted to target different behavioral determinants relevant for the overarching program goal to achieve earlier full recovery for general surgical and gynecological patients. From the beginning, patients and health care professionals were involved in the developmental process and patients' needs were taken into account by offering a user-friendly eHealth and mHealth intervention for our target group [67]. By involving hospitals in an early stage of the development process, the expectation is that a supportive basis for the intervention was created and that the implementation followed by evaluation will go as planned [68].

Comparison With Other Studies

The IM protocol has proven to be suitable to systematically and scientifically develop an eHealth intervention for various health issues [69,70]. In addition, a few studies have also used the IM protocol to further develop, adapt, or adjust an existing intervention [71,72]. In these studies, the IM protocol was considered suitable and enabled researchers to reconsider points of view and to integrate new information into existing tools. These studies used the IM protocol in the same manner as we did in this study. They decided to further build on the methodological choices made in the previously developed intervention by adjusting and improving the existing intervention for a new study population [71,72]. Methodological choices made in steps 2 and 3 of the IM process of developing the original "ikherstel" intervention, as described in the IM article of Vonk Noordegraaf et al, are considered effective and, therefore, used again in this IM procedure [19].

Comparison with other studies focusing on the development of eHealth recovery interventions is limited due to a low number of studies [18,23]. When comparing this developed intervention to the eHealth interventions included in the systematic review conducted by our planning group in step 1, it can be concluded that our intervention consists of multiple components on a diversity of functions that makes this a comprehensive intervention. In addition, most RCTs regarding enhancing the postoperative course are performed in cardiac surgery, whereas there is a lack of RCTs evaluating postoperative recovery after general surgical and gynecological procedures [73-80]. The studies of Vonk et al and Bouwsma et al are most comparable [18,23]. However, our study will be the only RCT that will be performed in patients undergoing major abdominal surgery focusing on RNA, which also includes procedures performed due to a malignant indication [81]. By determining the time until RNA as primary outcome, all patients aged between 18 and 75 years can participate in this study allowing more people to be reached.

The intervention was further developed and improved with respect to the previous "ikherstel" intervention. The shortcomings identified in the needs assessment are reflected, and new tools have been introduced. On the basis of the input from the FGDs, the accessibility has been improved by developing a mobile app and by the possibility of an eConsult. This facilitates direct access and communication with health care professionals and increases the reach of the "ikherstel" intervention [82]. By using the activity tracker, patients can monitor themselves, resulting in more involvement in their recovery process and motivating them to mobilize quicker and more intensively. Given the "ikherstel" intervention is based on the concept of computer-driven tailoring, the information regarding recovery is more personally relevant, which will increase the likelihood of behavior change and maintenance [83,84].

Strengths and Limitations

One of the strengths of our study is that by involving patients and different health care professionals in the development process of the intervention, experiences from multiple fields of expertise were included. We are convinced that this is necessary because this recovery-focused eHealth intervention is designed to include a multidisciplinary approach. Another strength is that all information gathered in the development (IM) process of the original intervention was based on findings of qualitative and quantitative studies and is used and extended with additional information obtained during this IM process. By adjusting and improving the "ikherstel" intervention, a more complete eHealth intervention for the broadened target population was created.

A limitation is that nurses were not involved in this process of development of the intervention, whereas they are linked to a part of the intervention itself. However, we did have meetings

with nurses in step 5 of the IM protocol, which attributes to the acceptability.

Clinical Relevance

Involving different program users and using a theory- and evidence-based systematic approach in the development of the intervention as noted in the IM procedure results in the best opportunity on effectiveness and implementation. If the further developed "ikherstel" intervention is proven effective, the content can be extended to other surgical procedures. In addition, if it proves to be cost-effective for a broad surgical group of patients, an implementation plan for future nationwide implementation has to be generated. It will also provide insight into the question of whether a systematically further developed version of an effective intervention is still effective for a different target group.

Conclusions

This study showed that with the use of IM, we were able to optimize and further develop the original "ikherstel" intervention. The intervention is extended to patients undergoing general surgical procedures and for malignant indications. New intervention tools such as a mobile app, an activity tracker, and an eConsult were developed. Consequently, with these tools, there is an increase in accessibility coupled with provision of monitoring and interactive feedback. The further developed "ikherstel" intervention will be evaluated in a multicenter single-blinded RCT.

Acknowledgments

This study is carried out with funding of ZonMw (project number 837002409), an organization for health research and development in the Netherlands.

Authors' Contributions

CMdB and FGS performed the literature search in step 1. HBACS, ECJC, ADtC, JAFH, and WMvB invited patients to participate in the FGDs, and they contributed to the design of this study and interpreted the analyses. CMdB and FGS analyzed the FGDs. AHB, PHPD, PCS, SvdM, WMvB, ADvD, DJL, JWvdS, WKGL, PMAJG, ECJC, SESK, SMMdC, PJMvK, HAC, HBACS, and ADtC contributed to steps 2, 5, and 6. CMdB, FGS, EvdM, WJHJM, BvdH, JAFH, HJB, and JRA contributed to all the steps. In addition, all authors participated in the drafting and revising of this paper, and all approved this version to be published.

Conflicts of Interest

WJHJM, BvdH, AHB, PHPD, PCS, SvdM, WMvB, ADvD, DJL, JWvdS, WKGL, PMAJG, ECJC, SESK, SMMdC, PJMvK, HAC, HBACS, and ADtC have no conflicts of interest. CMdB, FGS, EvdM, JAFH, HJB, and JRA are the developers of the care program under study. JRA and JAFH intend to set up a spin-off company concerning the implementation of a mobile app concerning the "ikherstel" intervention in the Netherlands. JAFH received grants from ZonMw-NWO, and Samsung during the conduct of the study and received a fee from Olympus, outside the submitted work. HJB received personal fees from Olympus, Stryker, Medtronic, and Applied Medical, outside the submitted work. JRA holds a chair in Insurance Medicine paid by the Dutch Social Security Agency, is a stockholder of Evalua, and received grants from ZonMw-NWO, Instituut Gak, UWV, SZW, VWS, Pfizer, Achmea, and CVZ/Zorg Instituut outside the submitted work.

Multimedia Appendix 1

Performance objectives to empower patients to return to normal activities including work. Asterisk indicates that the performance objective is based on the intervention mapping study of Vonk Noordegraaf et al (2014).

[XLSX File (Microsoft Excel File), 12KB - jmir_v21i2e9938_app1.xlsx]

Multimedia Appendix 2

Example of performance objective of patients.

[XLSX File (Microsoft Excel File), 11KB - jmir_v21i2e9938_app2.xlsx]

Multimedia Appendix 3

Determinants, methods, preconditions and strategies tools.

[XLSX File (Microsoft Excel File), 14KB - jmir_v21i2e9938_app3.xlsx]

Multimedia Appendix 4

The layout of the mobile app (directly translated from Dutch).

[PNG File, 465KB - jmir_v21i2e9938_app4.png]

Multimedia Appendix 5

Primary and secondary outcomes per measurement moment.

[XLSX File (Microsoft Excel File), 14KB - jmir_v21i2e9938_app5.xlsx]

References

1. Spanjersberg WR, Reurings J, Keus F, van Laarhoven CJ. Fast track surgery versus conventional recovery strategies for colorectal surgery. Cochrane Database Syst Rev 2011 Feb 16(2):CD007635. [doi: 10.1002/14651858.CD007635.pub2] [Medline: 21328298]
2. Bonjer HJ, Deijen CL, Abis GA, Cuesta MA, van der Pas MH, de Lange-de Klerk ES, COLOR II Study Group. A randomized trial of laparoscopic versus open surgery for rectal cancer. N Engl J Med 2015 Apr 02;372(14):1324-1332. [doi: 10.1056/NEJMoa1414882] [Medline: 25830422]
3. Dutch Hospital data, Utrecht. Dutch Hospital data, Utrecht. 2013 Dec. Kengetallen Nederlandse Ziekenhuizen 2012 URL: https://www.nvz-ziekenhuizen.nl/_library/13152/Kengetallen%20Nederlandse%20Ziekenhuizen%202012.pdf [accessed 2018-11-28] [WebCite Cache ID 74HacBjfo]
4. Henderson M, Glozier N, Holland EK. Long term sickness absence. Br Med J 2005 Apr 09;330(7495):802-803 [FREE Full text] [doi: 10.1136/bmj.330.7495.802] [Medline: 15817531]
5. Lee L, Tran T, Mayo NE, Carli F, Feldman LS. What does it really mean to “recover” from an operation? Surgery 2014 Feb;155(2):211-216. [doi: 10.1016/j.surg.2013.10.002] [Medline: 24331759]
6. Everett TR, Crawford RA. Laparoscopic assisted vaginal hysterectomy and bilateral salpingo-oophorectomy as a day surgery procedure: a promising option. J Obstet Gynaecol 2010;30(7):697-700. [doi: 10.3109/01443615.2010.504866] [Medline: 20925613]
7. Bachoo P, Duncan JL. Prolonged convalescence following inguinal hernia repair: an unnecessary trend. Health Bull (Edinb) 1995 Jul;53(4):209-212. [Medline: 7665348]
8. Baker DM, Rider MA, Fawcett AN. When to return to work following a routine inguinal hernia repair: are doctors giving the correct advice? J R Coll Surg Edinb 1994 Feb;39(1):31-33. [Medline: 7515427]
9. Clayton M, Verow P. Advice given to patients about return to work and driving following surgery. Occup Med (Lond) 2007 Oct;57(7):488-491. [doi: 10.1093/occmed/kqm063] [Medline: 17906266]
10. Ottesen M, Møller C, Kehlet H, Ottesen B. Substantial variability in postoperative treatment, and convalescence recommendations following vaginal repair. A nationwide questionnaire study. Acta Obstet Gynecol Scand 2001 Nov;80(11):1062-1068. [Medline: 11703209]
11. Dirksz A, Bruinvels D, de Vet HCW, Brölmann H. Werkhervatting na gynaecologische ingrepen; Pilot study. 2008.
12. Waddell G, Burton K, Aylward M. Work and common health problems. J Insur Med 2007;39(2):109-120. [Medline: 17941336]
13. Kikuchi I, Takeuchi H, Shimanuki H, Kitade M, Kumakiri J, Kuroda K, et al. Questionnaire analysis of recovery of activities of daily living after laparoscopic surgery. J Minim Invasive Gynecol 2008;15(1):16-19. [doi: 10.1016/j.jmig.2007.08.606] [Medline: 18262138]
14. Lewis J, Ray P, Liaw S. Recent worldwide developments in eHealth and mHealth to more effectively manage cancer and other chronic diseases - a systematic review. Yearb Med Inform 2016 Nov 10(1):93-108 [FREE Full text] [doi: 10.15265/IY-2016-020] [Medline: 27830236]
15. 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]
16. Tang C, Lorenzi N, Harle CA, Zhou X, Chen Y. Interactive systems for patient-centered care to enhance patient engagement. J Am Med Inform Assoc 2016 Jan 14;23(1):2-4. [doi: 10.1093/jamia/ocv198]
17. Dedding C, van Doorn R, Winkler L, Reis R. How will e-health affect patient participation in the clinic? A review of e-health studies and the current evidence for changes in the relationship between medical professionals and patients. Soc Sci Med 2011 Jan;72(1):49-53. [doi: 10.1016/j.socscimed.2010.10.017] [Medline: 21129832]
18. Vonk Noordegraaf A, Anema JR, van Mechelen W, Knol DL, van Baal WM, van Kesteren PJ, et al. A personalised eHealth programme reduces the duration until return to work after gynaecological surgery: results of a multicentre randomised trial. BJOG 2014 Aug;121(9):1127-35; discussion 1136 [FREE Full text] [doi: 10.1111/1471-0528.12661] [Medline: 24511914]
19. Vonk Noordegraaf A, Huirne JA, Pittens CA, van Mechelen W, Broerse JE, Brölmann HA, et al. eHealth program to empower patients in returning to normal activities and work after gynecological surgery: intervention mapping as a useful method for development. J Med Internet Res 2012;14(5):e124 [FREE Full text] [doi: 10.2196/jmir.1915] [Medline: 23086834]
20. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [Medline: 9768376]
21. Bartholomew LK, Parcel GS, Kok G, Gottlieb N. Planning health promotion programs: an intervention mapping approach. San Francisco: Jossey-Bass; 2006.
22. Centraal Bureau voor de Statistiek. Operaties in het ziekenhuis; soort opname, leeftijd en geslacht, 1995-2010 URL: http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=80386NED&LA=NL

23. Bouwsma EV, Anema JR, Vonk Noordegraaf A, Knol DL, Bosmans JE, Schraffordt Koops SE, et al. The cost effectiveness of a tailored, web-based care program to enhance postoperative recovery in gynecologic patients in comparison with usual care: protocol of a stepped wedge cluster randomized controlled trial. JMIR Res Protoc 2014 Jun 18;3(2):e30 [FREE Full text] [doi: 10.2196/resprot.3236] [Medline: 24943277]
24. Stichting Informatie Gezondheidszorg. Prismant. 2009 Jan 01. Landelijke Medische Registratie LMR - Verrichtingen URL: https://nvma.nl/magazine/actueel/registratie-landelijke-medische-registratie-lmr/ [accessed 2018-11-29] [WebCite Cache ID 74HcBZWig]
25. van der Meij E, Huirne JA, Bouwsma EV, van Dongen JM, Terwee CB, van der Ven PM, et al. Substitution of usual perioperative care by eHealth to enhance postoperative recovery in patients undergoing general surgical or gynecological procedures: study protocol of a randomized controlled trial. JMIR Res Protoc 2016 Dec 21;5(4):e245 [FREE Full text] [doi: 10.2196/resprot.6580] [Medline: 28003177]
26. Bouwsma EV, Vonk Noordegraaf A, Szlávik Z, Brölmann HA, Emanuel MH, Lips JP, et al. Process evaluation of a multidisciplinary care program for patients undergoing gynaecological surgery. J Occup Rehabil 2014 Sep;24(3):425-438 [FREE Full text] [doi: 10.1007/s10926-013-9475-4] [Medline: 24057871]
27. van der Meij E, Anema JR, Otten RH, Huirne JA, Schaafsma FG. The effect of perioperative e-health interventions on the postoperative course: a systematic review of randomised and non-randomised controlled trials. PLoS One 2016;11(7):e0158612 [FREE Full text] [doi: 10.1371/journal.pone.0158612] [Medline: 27383239]
28. van der Meij E, Bouwsma EV, van den Heuvel B, Bonjer HJ, Anema JR, Huirne JA. Using e-health in perioperative care: a survey study investigating shortcomings in current perioperative care and possible future solutions. BMC Surg 2017 May 23;17(1):61 [FREE Full text] [doi: 10.1186/s12893-017-0254-6] [Medline: 28535763]
29. van Achterberg T, Huisman-de Waal GG, Ketelaar NA, Oostendorp RA, Jacobs JE, Wollersheim HC. How to promote healthy behaviours in patients? An overview of evidence for behaviour change techniques. Health Promot Int 2011 Jun;26(2):148-162 [FREE Full text] [doi: 10.1093/heapro/daq050] [Medline: 20739325]
30. Møller C, Ottesen M, Kehlet H, Ottesen BS. Convalescence recommendations after hysterectomy. A study of opinions among Danish physicians. Ugeskr Laeger 2001 Dec 10;163(50):7043-7047. [Medline: 11794035]
31. Tolver MA, Rosenberg J, Bisgaard T. Convalescence after laparoscopic inguinal hernia repair: a qualitative systematic review. Surg Endosc 2016 Dec;30(12):5165-5172. [doi: 10.1007/s00464-016-4863-4] [Medline: 27059966]
32. Neville A, Lee L, Antonescu I, Mayo NE, Vassiliou MC, Fried GM, et al. Systematic review of outcomes used to evaluate enhanced recovery after surgery. Br J Surg 2014 Feb;101(3):159-170. [doi: 10.1002/bjs.9324] [Medline: 24469616]
33. Hamaker ME, Prins MC, Schiphorst AH, van Tuyl SA, Pronk A, van den Bos F. Long-term changes in physical capacity after colorectal cancer treatment. J Geriatr Oncol 2015 Mar;6(2):153-164. [doi: 10.1016/j.jgo.2014.10.001] [Medline: 25454769]
34. Miller TE, Mythen M. Successful recovery after major surgery: moving beyond length of stay. Perioper Med (Lond) 2014;3:4 [FREE Full text] [doi: 10.1186/2047-0525-3-4] [Medline: 25018877]
35. Christensen T, Kehlet H. Postoperative fatigue. World J Surg 1993;17(2):220-225. [Medline: 8511917]
36. Carli F, Mayo N, Klubien K, Schricker T, Trudel J, Belliveau P. Epidural analgesia enhances functional exercise capacity and health-related quality of life after colonic surgery: results of a randomized trial. Anesthesiology 2002 Sep;97(3):540-549. [Medline: 12218518]
37. Brölmann HA, Vonk Noordegraaf A, Bruinvels DJ, de Vet RH, Dirksz AA, Huirne JA. Can prolonged sick leave after gynecologic surgery be predicted? An observational study in The Netherlands. Surg Endosc 2009 Oct;23(10):2237-2241 [FREE Full text] [doi: 10.1007/s00464-008-0287-0] [Medline: 19118421]
38. Johansen P, Al-Khafagi SK, Thøstesen LM, Lauszus FF, Rasmussen KL. Analysis of need for sick leave after hysterectomy. Ugeskr Laeger 2008 Apr 21;170(17):1465-1468. [Medline: 18462627]
39. Bisgaard T, Klarskov B, Rosenberg J, Kehlet H. Factors determining convalescence after uncomplicated laparoscopic cholecystectomy. Arch Surg 2001 Aug;136(8):917-921. [Medline: 11485527]
40. Callesen T, Klarskov B, Bech K, Kehlet H. Short convalescence after inguinal herniorrhaphy with standardised recommendations: duration and reasons for delayed return to work. Eur J Surg 1999 Mar;165(3):236-241. [doi: 10.1080/110241599750007108] [Medline: 10231657]
41. van Muijen P, Weevers NL, Snels IA, Duijts SF, Bruinvels DJ, Schellart AJ, et al. Predictors of return to work and employment in cancer survivors: a systematic review. Eur J Cancer Care (Engl) 2013 Mar;22(2):144-160. [doi: 10.1111/ecc.12033] [Medline: 23279195]
42. Rodriguez-Bigas MA, Chang GJ, Skibber JM. Barriers to rehabilitation of colorectal cancer patients. J Surg Oncol 2007 Apr 01;95(5):400-408. [doi: 10.1002/jso.20778] [Medline: 17345598]
43. van Zutphen M, Winkels RM, van Duijnhoven FJ, van Harten-Gerritsen SA, Kok DE, van Duijvendijk P, et al. An increase in physical activity after colorectal cancer surgery is associated with improved recovery of physical functioning: a prospective cohort study. BMC Cancer 2017 Dec 25;17(1):74 [FREE Full text] [doi: 10.1186/s12885-017-3066-2] [Medline: 28122534]
44. Shaw L, Segal R, Polatajko H, Harburn K. Understanding return to work behaviours: promoting the importance of individual perceptions in the study of return to work. Disabil Rehabil 2002 Mar 10;24(4):185-195. [Medline: 11926259]

45. Rosenberger PH, Jokl P, Ickovics J. Psychosocial factors and surgical outcomes: an evidence-based literature review. J Am Acad Orthop Surg 2006 Jul;14(7):397-405. [Medline: 16822887]
46. Celestin J, Edwards RR, Jamison RN. Pretreatment psychosocial variables as predictors of outcomes following lumbar surgery and spinal cord stimulation: a systematic review and literature synthesis. Pain Med 2009;10(4):639-653. [doi: 10.1111/j.1526-4637.2009.00632.x] [Medline: 19638142]
47. Shehmar M, Gupta JK. The influence of psychological factors on recovery from hysterectomy. J R Soc Med 2010 Feb;103(2):56-59 [FREE Full text] [doi: 10.1258/jrsm.2009.090198] [Medline: 20118335]
48. Brouwer S, Krol B, Reneman MF, Bültmann U, Franche R, van der Klink JJL, et al. Behavioral determinants as predictors of return to work after long-term sickness absence: an application of the theory of planned behavior. J Occup Rehabil 2009 Jun;19(2):166-174. [doi: 10.1007/s10926-009-9172-5] [Medline: 19333738]
49. Oyeflaten I, Hysing M, Eriksen HR. Prognostic factors associated with return to work following multidisciplinary vocational rehabilitation. J Rehabil Med 2008 Jul;40(7):548-554 [FREE Full text] [doi: 10.2340/16501977-0202] [Medline: 18758672]
50. Svajger A, Winding K. Perceptions of possibilities of returning to work with chronic musculoskeletal disorders. Work 2009;32(4):443-454. [doi: 10.3233/WOR-2009-0855] [Medline: 19478415]
51. Lambeek LC, van Mechelen W, Knol DL, Loisel P, Anema JR. Randomised controlled trial of integrated care to reduce disability from chronic low back pain in working and private life. Br Med J 2010;340:c1035 [FREE Full text] [Medline: 20234040]
52. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an Intervention Mapping approach. Health Psychol Rev 2015 Oct 15:1-16. [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
53. Vonk NA, Huirne JA, Brölmann HA, van Mechelen W, Anema JR. Multidisciplinary convalescence recommendations after gynaecological surgery: a modified Delphi method among experts. BJOG 2011 Dec;118(13):1557-1567 [FREE Full text] [doi: 10.1111/j.1471-0528.2011.03091.x] [Medline: 21895950]
54. van Vliet DC, van der Meij E, Bouwsma EV, Vonk Noordegraaf A, van den Heuvel B, Meijerink WJ, et al. A modified Delphi method toward multidisciplinary consensus on functional convalescence recommendations after abdominal surgery. Surg Endosc 2016 Dec;30(12):5583-5595 [FREE Full text] [doi: 10.1007/s00464-016-4931-9] [Medline: 27139706]
55. Meppelink CS, van Weert JC, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
56. Linnan L, Steckler A. Process evaluation for public health interventions and research. 1st edition. San Francisco: Jossey-Bass; 2002:1-23.
57. Terwee CB, Roorda LD, de Vet HC, Dekker J, Westhovens R, van Leeuwen J, et al. Dutch-Flemish translation of 17 item banks from the patient-reported outcomes measurement information system (PROMIS). Qual Life Res 2014 Aug;23(6):1733-1741. [doi: 10.1007/s11136-013-0611-6] [Medline: 24402179]
58. Dutch Flemish Promis. Welkom bij de Dutch-Flemish PROMIS groep URL: http://www.dutchflemishpromis.nl [accessed 2018-11-27] [WebCite Cache ID 74FXky119]
59. Bode RK, Hahn EA, DeVellis R, Cella D, Patient-Reported Outcomes Measurement Information System Social Domain Working Group. Measuring participation: the patient-reported outcomes measurement information system experience. Arch Phys Med Rehabil 2010 Sep;91(9 Suppl):S60-S65 [FREE Full text] [doi: 10.1016/j.apmr.2009.10.035] [Medline: 20801282]
60. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997 Nov;35(11):1095-1108. [Medline: 9366889]
61. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
62. Kluivers KB, Hendriks JC, Mol BW, Bongers MY, Vierhout ME, Brölmann HA, et al. Clinimetric properties of 3 instruments measuring postoperative recovery in a gynecologic surgical population. Surgery 2008 Jul;144(1):12-21. [doi: 10.1016/j.surg.2008.03.027] [Medline: 18571580]
63. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992 Aug;50(2):133-149. [Medline: 1408309]
64. Zorginstituutnederland. Richtlijn voor het uitvoeren van economische valuaties in de gezondheidszorg URL: https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg [accessed 2018-11-27] [WebCite Cache ID 74FXRNBZk]
65. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Hakkaart-van Roijen L. The iMTA productivity cost questionnaire: a standardized instrument for measuring and valuing health-related productivity losses. Value Health 2015 Sep;18(6):753-758 [FREE Full text] [doi: 10.1016/j.jval.2015.05.009] [Medline: 26409601]
66. Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg 2009 Aug;250(2):187-196. [doi: 10.1097/SLA.0b013e3181b13ca2] [Medline: 19638912]

67. van Dulmen S, Driesenaar JA, van Weert JC, van Osch M, Noordman J. PatientVOICE: development of a preparatory, pre-chemotherapy online communication tool for older patients with cancer. JMIR Res Protoc 2017 May 10;6(5):e85 [FREE Full text] [doi: 10.2196/resprot.6979] [Medline: 28490421]
68. Bartholomew L, Parcel G, Kok G, Gottlieb N. Intervention Mapping: designing theory and evidence-based health promotion programs. 1st ed. Mountain View, California: Mayfield Publishing Company; 2001.
69. Garba RM, Gadanya MA. The role of intervention mapping in designing disease prevention interventions: a systematic review of the literature. PLoS One 2017;12(3):e0174438 [FREE Full text] [doi: 10.1371/journal.pone.0174438] [Medline: 28358821]
70. Lamort-Bouché M, Sarnin P, Kok G, Rouat S, Péron J, Letrilliart L, et al. Interventions developed with the Intervention Mapping protocol in the field of cancer: a systematic review. Psychooncology 2018 Apr;27(4):1138-1149. [doi: 10.1002/pon.4611] [Medline: 29247578]
71. de Kroon ML, Bulthuis J, Mulder W, Schaafsma FG, Anema JR. Reducing sick leave of Dutch vocational school students: adaptation of a sick leave protocol using the intervention mapping process. Int J Public Health 2016 Dec;61(9):1039-1047. [doi: 10.1007/s00038-016-0840-x] [Medline: 27318528]
72. Boekhout JM, Peels DA, Berendsen BA, Bolman CA, Lechner L. An eHealth intervention to promote physical activity and social network of single, chronically impaired older adults: adaptation of an existing intervention using intervention mapping. JMIR Res Protoc 2017 Nov 23;6(11):e230 [FREE Full text] [doi: 10.2196/resprot.8093] [Medline: 29170146]
73. Körtke H, Stromeyer H, Zittermann A, Buhr N, Zimmermann E, Wienecke E, et al. New East-Westfalian Postoperative Therapy Concept: a telemedicine guide for the study of ambulatory rehabilitation of patients after cardiac surgery. Telemed J E Health 2006 Aug;12(4):475-483. [doi: 10.1089/tmj.2006.12.475] [Medline: 16942420]
74. Barnason S, Zimmerman L, Nieveen J, Hertzog M. Impact of a telehealth intervention to augment home health care on functional and recovery outcomes of elderly patients undergoing coronary artery bypass grafting. Heart Lung 2006;35(4):225-233. [doi: 10.1016/j.hrtlng.2005.10.003] [Medline: 16863894]
75. Barnason S, Zimmerman L, Nieveen J, Schmaderer M, Carranza B, Reilly S. Impact of a home communication intervention for coronary artery bypass graft patients with ischemic heart failure on self-efficacy, coronary disease risk factor modification, and functioning. Heart Lung 2003;32(3):147-158. [Medline: 12827099]
76. Barnason S, Zimmerman L, Nieveen J, Schulz P, Miller C, Hertzog M, et al. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. Heart Lung 2009;38(5):364-376 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.01.005] [Medline: 19755186]
77. Martorella G, Côté J, Racine M, Choinière M. Web-based nursing intervention for self-management of pain after cardiac surgery: pilot randomized controlled trial. J Med Internet Res 2012;14(6):e177 [FREE Full text] [doi: 10.2196/jmir.2070] [Medline: 23241361]
78. Miller C, Zimmerman L, Barnason S, Nieveen J. Impact of an early recovery management intervention on functioning in postoperative coronary artery bypass patients with diabetes. Heart Lung 2007;36(6):418-430. [doi: 10.1016/j.hrtlng.2007.02.011] [Medline: 18005803]
79. Zimmerman L, Barnason S, Nieveen J, Schmaderer M. Symptom management intervention in elderly coronary artery bypass graft patients. Outcomes Manag 2004;8(1):5-12. [Medline: 14740578]
80. Keeping-Burke L, Purden M, Frasure-Smith N, Cossette S, McCarthy F, Amsel R. Bridging the transition from hospital to home: effects of the VITAL telehealth program on recovery for CABG surgery patients and their caregivers. Res Nurs Health 2013 Dec;36(6):540-553. [doi: 10.1002/nur.21571] [Medline: 24242195]
81. Bolle S, Romijn G, Smets EM, Loos EF, Kunneman M, van Weert JC. Older cancer patients' user experiences with Web-based health information tools: a think-aloud study. J Med Internet Res 2016 Jul 25;18(7):e208 [FREE Full text] [doi: 10.2196/jmir.5618] [Medline: 27457709]
82. Nasi G, Cucciniello M, Guerrazzi C. The role of mobile technologies in health care processes: the case of cancer supportive care. J Med Internet Res 2015;17(2):e26 [FREE Full text] [doi: 10.2196/jmir.3757] [Medline: 25679446]
83. 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]
84. Quintiliani LM, Campbell MK, Bowling JM, Steck S, Haines PS, DeVellis BM. Results of a randomized trial testing messages tailored to participant-selected topics among female college students: physical activity outcomes. J Phys Act Health 2010 Jul;7(4):517-526 [FREE Full text] [Medline: 20683094]

Abbreviations

eConsult: electronic consultation
eHealth: electronic health
ERAS: enhanced recovery after surgery programs
FGD: focus group discussion
ICT: information and communication technology

IM: Intervention Mapping
mHealth: mobile health
PRECEDE: predisposing, reinforcing, and enabling constructs in educational diagnosis and evaluation
PROCEED: policy, regulatory, and organizational constructs in educational and environmental development
RCT: randomized controlled trial
RNA: resumption of normal activities
RTW: return to work

Edited by C Gibbons; submitted 25.01.18; peer-reviewed by G Kok, J Scheerman; comments to author 21.05.18; revised version received 16.07.18; accepted 30.08.18; published 06.02.19.

Please cite as:
den Bakker CM, Schaafsma FG, van der Meij E, Meijerink WJHJ, van den Heuvel B, Baan AH, Davids PHP, Scholten PC, van der Meij S, van Baal WM, van Dalsen AD, Lips DJ, van der Steeg JW, Leclercq WKG, Geomini PMAJ, Consten ECJ, Schraffordt Koops SE, de Castro SMM, van Kesteren PJM, Cense HA, Stockmann HBAC, ten Cate AD, Bonjer HJ, Huirne JAF, Anema JR
Electronic Health Program to Empower Patients in Returning to Normal Activities After General Surgical and Gynecological Procedures: Intervention Mapping as a Useful Method for Further Development
J Med Internet Res 2019;21(2):e9938
URL: https://www.jmir.org/2019/2/e9938/
doi:10.2196/jmir.9938
PMID:30724740

Review

Weight Management in Young Adults: Systematic Review of Electronic Health Intervention Components and Outcomes

Taylor Jade Willmott[1], BBus (Ec&Marketing), BBus (Hons); Bo Pang[1], PhD; Sharyn Rundle-Thiele[1], PhD; Abi Badejo[1], PhD

Social Marketing @ Griffith, Griffith Business School, Griffith University, Nathan, Australia

Corresponding Author:
Taylor Jade Willmott, BBus (Ec&Marketing), BBus (Hons)
Social Marketing @ Griffith
Griffith Business School
Griffith University
170 Kessels Road
Nathan, 4111
Australia
Phone: 61 737358433
Email: t.willmott@griffith.edu.au

Abstract

Background: Young adulthood is a vulnerable period for unhealthy lifestyle adoption and excess weight gain. Scant attention has been focused on developing and evaluating effective weight gain prevention strategies for this age group. Electronic health (eHealth) offers potential as a cost-effective means of delivering convenient, individually-tailored, and contextually-meaningful interventions at scale.

Objective: The primary aim of this systematic review was to locate and synthesize the evidence on eHealth weight management interventions targeting young adults, with a particular focus on (eHealth) intervention components and outcomes.

Methods: A systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. The search strategy was executed across the following electronic databases: Cumulative Index to Nursing and Allied Health Literature, Cochrane Library, EBSCO, EMBASE, Emerald, Education Resources Information Center, Medical Literature Analysis and Retrieval System Online, Ovid, ProQuest, PsycINFO, PubMed, Science Direct, Scopus, and Web of Science. Furthermore, 2 reviewers independently assessed records for eligibility: peer-reviewed, published in English, and report evaluations of eHealth weight management interventions targeting healthy young adults (aged 18-35 years). Data were then extracted from studies that met the criteria for inclusion. The methodological quality of studies was independently assessed by 2 reviewers using the Effective Public Health Practice Project's (EPHPP) quality assessment tool. A comprehensive narrative evidence synthesis was then completed.

Results: Out of the 1301 studies assessed for eligibility, 24 met the criteria for inclusion. According to the EPHPP quality assessment tool, overall, 19 studies were as rated weak, 5 as moderate, and none as strong. The narrative synthesis of intervention outcomes found 8 studies reported positive weight-related outcomes, 4 reported mixed outcomes, and 12 did not report any significant changes in weight-related outcomes. The narrative synthesis of (eHealth) intervention components led to 3 levels of classification. A total of 14 studies were classified as *Web-based*, 3 as *mobile-based*, and 7 as *multicomponent* interventions. Following the narrative synthesis, 5 key strategies were thematically identified: self-regulation (goal setting and self-monitoring), tailored or personalized feedback, contact with an interventionist, social support, and behavioral prompts (nudges and reminders) and booster messages.

Conclusions: Findings highlight the limited evidence base for eHealth weight management interventions targeting young adults. The complex nature of weight management presents an ongoing challenge for interventionists to identify *what works, for whom, how, and when*. The quality of the evidence in this review was generally assessed as weak; however, assessment tools such as the EPHPP are principally concerned with *what should be* and this is seldom equivalent to *what works*. Thus, while sampling, study design and retention rates will remain key determining factors of reliability and validity, further research attention directed toward the development of guiding tools for community trials is warranted.

(J Med Internet Res 2019;21(2):e10265) doi:10.2196/10265

KEYWORDS

body weight maintenance; eHealth; health behavior; obesity; overweight; review; technology; weight gain; young adult

Introduction

Background

Nearly one-third of the global population is overweight or obese, that is, more than 2.1 billion people [1]. The prevalence of obesity is rising rapidly throughout both the developed and the developing world, creating a substantial social, economic, and health burden on society [2]. If current trends continue, it is estimated that by 2030 almost half of the world's adult population will be overweight or obese [3]. Such a scenario would have devastating consequences for the global burden of noncommunicable diseases, with increasing body mass index (BMI) associated with an elevated risk of developing a chronic disease such as cardiovascular disease, diabetes, respiratory disease, and certain cancers [4]. The magnitude of the obesity epidemic has led to a shift in focus from the clinical treatment of obesity to the development of prevention strategies that address the economic, environmental, sociocultural, and lifestyle-related causes of population weight gain [5-7]. The prevention of weight gain and the maintenance of a healthy weight are considered less challenging, less expensive, and potentially more effective than the treatment of obesity after it has fully developed [8]. Once established, obesity is difficult and costly to treat [9,10]. Owing to the projected increases in obesity prevalence, the challenges faced in delivering effective treatment, and the costs associated with treatment, it will not be possible to deliver care for all individuals in need [11]. Therefore, the prevention of obesity and its comorbidities are, and must continue to be, a foremost public health priority.

Targeting high-risk groups with prevention interventions is hypothesized to have the greatest impact on the rising incidence of overweight and obesity [4]. Efforts to prevent obesity have mainly focused on children and adolescents, whereas other important age groups have been overlooked [12]. The most rapid weight gain in the life course has been observed during the early twenties to midthirties [12,13], with incident obesity at a younger age associated with increased risk of chronic disease and mortality in later adult life [12,14,15].

Young adulthood is a transitional life stage in which young people experience significant life changes, increasing independence, and adopt lasting health behavior patterns [16]. The cause of age-related weight gain during young adulthood appears to be lifestyle-based, resulting from marked declines in physical activity (PA), increases in sedentary behavior, and poor dietary habits [17-22]. These changes in PA and diet-related behaviors likely result from the significant life transitions that occur during young adulthood, such as moving out of the family home, relocating to new environments, beginning full-time work or tertiary study, and establishing financial, residential, and employment stability [16]. Among this demographic, barriers to healthy weight maintenance exceed enablers [23], with healthful eating and regular PA not considered high priorities [24]. Perceived time constraints, lack of discipline, inadequate self-regulation skills, and a lack of environmental support for healthy eating and PA have all been cited as common barriers to healthy weight maintenance among young adults [23-26]. Common enablers to healthy weight maintenance include education and awareness (eg, what to eat and what not to eat), self-regulation skills (eg, practicing moderation and portion control), and positive social and environmental support [23,24,26]. Importantly, the adoption of healthier lifestyle behaviors in young adulthood has been associated with a lower risk of developing chronic disease in later adult life [27]. Given obesity is entirely preventable, the establishment and maintenance of healthy behavioral patterns in young adulthood would deliver long-term health benefits to individuals as well as cost benefits to society. Therefore, a more fine-grained understanding of the means that can be reliably used to effectively assist young adults in managing their weight is needed.

Review Rationale and Aim

Previous reviews [28-33] of lifestyle interventions for obesity prevention and weight management have highlighted the limited evidence base for successful interventions among this age group. Findings from these reviews were inconclusive owing to the small number of studies available [30], small sample sizes [30,32], heterogeneity across intervention designs [30,31], differences in participant characteristics [30], gender biases [32], and short intervention durations [30,32]. Traditional weight management interventions (ie, face-to-face sessions with a trained interventionist) may not meet the needs of many young adults, as evidenced by lower recruitment and retention rates, inferior attendance and compliance, and poor weight-related outcomes relative to older adult participants [34]. Traditional interventions are resource intensive in terms of the commitment required by participants and intervention providers, which can create barriers for full participation and adherence [29]. Moreover, the resources required to deliver face-to-face interventions (individual or group-based) prevent large-scale deployment to the wider community [28].

Young adulthood is a developmentally unique life stage [16]. Therefore, weight management interventions aimed at this demographic must have a specific focus on the distinct challenges faced by young adults that are known to contribute to weight gain, including rapidly shifting life circumstances related to home, work, family, and other relationships [16]; examples of the challenges faced during this developmental period include juggling the many responsibilities that come with being an 'adult' [16], continuing cognitive development through the midtwenties (eg, impulse control, regulation of emotions, and rational decision making) [35], and learning the skills needed to sustain oneself, such as home food preparation and meal planning [36,37]. Technology may offer a cost-effective means of engaging young adults in weight management, with the current generation of young adults among the highest users of digital technologies such as social media, mobile phones, and wireless information sharing platforms [38].

Electronic health (eHealth), defined as the use of information and communication technologies (ICTs) for health [39], offers a feasible alternative to traditional weight management interventions and has the potential to be delivered at scale. Telemedicine, first used in the 1920s, is the oldest form of eHealth. The introduction of broadband internet in the 1990s, followed by wireless technologies, precipitated an explosion of eHealth and mobile health apps within the health care field [40]. Interventions that encompass ICTs (eg, internet-enabled mobile and tablet devices, wearable monitors) permit the efficient delivery of individually-tailored, context-specific health behavior change programs, with time-unlimited feedback, coaching, and support [41]. The popularity, mobility, and capability of modern ICTs allow temporal synchronization of intervention delivery and allow the intervention to be delivered at a convenient time and place [41]. For example, young adults may be sent a short message service (SMS) text message in the morning to remind them that having a nutritious breakfast is important for healthy weight maintenance [42], with a link to healthy breakfast recipes based on items commonly available at home. eHealth-based interventions have previously demonstrated the potential to promote healthful changes in both diet and PA behaviors [43] and have been used as a treatment option for obesity in adults [44]. However, there is limited evidence on the effectiveness of eHealth-based approaches for weight loss maintenance and weight gain prevention [44], especially among young adult populations. As such, the primary aim of this review was to locate and synthesize the evidence on eHealth weight management interventions targeting young adults, with a particular focus on (eHealth) intervention components and outcomes.

Methods

Review Protocol

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [45]. Refer to Multimedia Appendix 1 for the PRISMA checklist used in this review.

Data Sources and Search Methods

The systematic literature search was completed in September 2018. The search strategy was executed across the following electronic databases: Cumulative Index to Nursing and Allied Health Literature, Cochrane Library, EBSCO, EMBASE, Emerald, Education Resources Information Center, Medical Literature Analysis and Retrieval System Online, Ovid, ProQuest, PsycINFO, PubMed, Science Direct, Scopus, and Web of Science. As outlined by Smith et al (2011) in a systematic review of individual studies, the search should be as wide as possible to maximize the likelihood of capturing all relevant data and minimizing the effects of reporting biases. As such, a search of a wide variety of electronic databases relevant to the topic of interest is recommended as a best practice [46]. The predetermined search strategy was designed by combining relevant search terms related to eHealth, weight management, and young adults. Search terms were divided into 4 groups: (1) intervention type (ie, eHealth variations), (2) outcome (ie, weight-related and behavioral variations), (3) study design (ie, study type variations), and (4) participants (ie, young adult variations). The full search strategy and database results are provided in Multimedia Appendix 2. The reference lists of all included papers (backward search) and pertinent systematic reviews [28-33] were also hand searched to identify additional studies for inclusion. Google Scholar was used to screen papers citing included studies (forward search).

Study Screening and Selection

All records were downloaded to Endnote Version X8 (Clarivate Analytics), duplicates were removed, and the remaining studies were assessed for eligibility via title and abstract by 2 independent reviewers. The results were categorized by title and abstract into (1) papers appearing to meet study selection criteria, (2) papers that should be retrieved for further examination, and (3) excluded papers. In cases where there were several publications from the same cohort, the study with the longest follow-up was selected; if the follow-up was equivalent, the most recent study was included. The full-text of potentially relevant papers was then obtained and assessed by 2 independent reviewers. These papers were further categorized. At all stages, any discrepancies were discussed and resolved by consensus. Where consensus could not be reached, a third independent reviewer acting as an arbitrator was consulted.

Eligibility Criteria

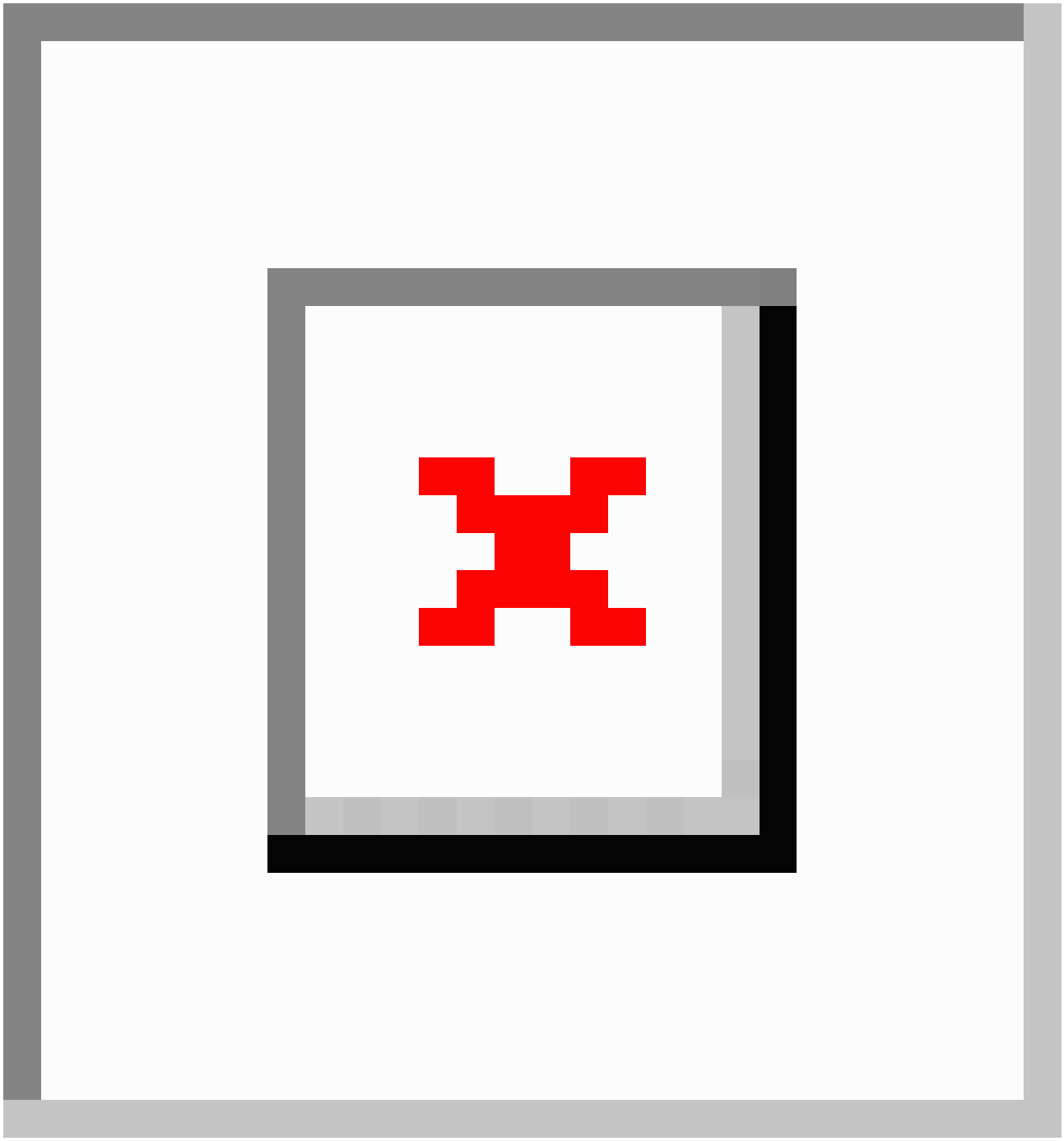
The eligibility criteria adopted in the present review are as follows. To be included in the review, studies had to (1) be peer-reviewed, (2) be published in the English language, (3) report evaluations of eHealth weight management interventions targeting young adults (aged 18-35 years old), including randomized controlled trials (RCTs), controlled clinical trials (CCTs), and cohort studies (pretest-posttest and posttest only), (4) include participants who were healthy and free of acute illness or chronic disease, and (5) report a measure of weight pre and postintervention.

For the purposes of this review, eHealth referred to behavior change interventions, which were operationalized and transformed for delivery via ICTs including computers, tablets, mobile phones, wearable and nonwearable tracking devices, and digital games. For studies to be eligible for review, eHealth had to form the primary means of intervention delivery in *at least one* treatment arm. The technology could be used as both a tool to enable a process, function or service, or as the embodiment of eHealth itself [47]. Weight management was defined as the prevention of weight gain via the maintenance of a healthy body weight or the reversal of small gains to maintain a healthy body weight [8]. Studies that purposively recruited and subsequently evaluated weight loss or weight loss maintenance interventions among the obese (mean BMI>30 kg/m^2) were excluded as the prevention of weight gain (ie, management) was the focus of this study; participants who have lost a significant amount of weight do not represent the general young adult population [48]. The age range of 18 to 35 years was selected based on the protocol included in the *National Heart, Lung, and Blood Institute's* Early Adult Reduction of Weight through Lifestyle Intervention trials [49]. Weight gain is most rapid during these years [12,13], and increasing BMI

in young adulthood increases the risk of developing metabolic syndrome over the subsequent 15 years almost 20-fold [50].

Studies were excluded on the basis of the following criteria: (1) not peer-reviewed, (2) not in English, (3) not related to eHealth *and* weight management, (4) not an intervention evaluation, (5) included participants who were not healthy and free from acute and chronic disease, or were pregnant, (6) did not report a measure of weight pre and postintervention, or (7) did not specifically target young adults (aged 18-35 years). Studies that did not report an age range, the mean age of the sample, or the percentage of the sample who were within a given age range, were also excluded. Numbers and reasons for exclusions are reported in Figure 1.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of study selection process.



Data Extraction and Management

A data extraction form informed by the PRISMA statement was developed for abstracting study characteristics [45]. Data extracted included the following: study details (author, year of publication, and country), study design, participants (sample size, characteristics, setting, retention, and blinding), intervention and comparator details, duration, and data collection methods, measures, outcomes, and conclusions (refer to Multimedia Appendix 3). Following this, summary tables were thoroughly and independently reviewed by all authors for accuracy and relevance. Any inconsistencies were resolved through discussion.

Quality Assessment

The Effective Public Health Practice Project's (EPHPP) quality assessment tool [51] for quantitative studies was used to assess the methodological quality of included studies. The tool requires the assessment of 6 individual quality components (selection bias, study design, confounders, blinding, data collection methods, and withdrawals and dropouts) before assigning an overall quality rating (strong, moderate, or weak) based on a 3-point scale. The tool has been judged suitable for use in systematic reviews of effectiveness [52] and has been reported to have content and construct validity [51,53]. Moreover, a study comparing the EPHPP quality assessment tool with the Cochrane Collaboration Risk of Bias (CCRB) tool found the EPHPP tool to have fair interrater agreement for individual domains and excellent agreement for the final grade. In contrast, the CCRB tool had only slight interrater agreement for individual domains and fair interrater agreement for the final grade. Of note, no agreement between the 2 tools was evident in their final grade assigned to each study. The authors concluded that although both tools were developed to assess *quality of the evidence*, they appear to measure different constructs [54]. In the present review, 2 independent reviewers completed assessments of methodological quality according to the EPHPP tool. Any discrepancies were resolved through discussion with a third independent reviewer, acting as an arbitrator, when required.

Data Synthesis

In line with the primary aim of this review, a comprehensive narrative evidence synthesis was completed. Each study was of intrinsic interest on its own and combining such complex interventions was likely to yield a meaningless result that would not provide actionable insights for improving the design of future interventions [55]. As such, the reviewers sought to describe the variation in study findings by qualitatively examining (eHealth) intervention components and outcomes rather than attempting to combine findings into 1 overall estimate of effectiveness [51].

Studies were categorized into 3 groups based on (eHealth) intervention components: *Web*-based, *mobile*-based, and *multi* component. Web-based refers to interventions that were predominantly delivered through the use of internet-enabled functions such as e-learning platforms, websites, and email. Mobile-based denotes interventions that were primarily delivered through mobile-enabled functions including SMS text messages and mobile phone apps. Multicomponent represents interventions that used a combination of the above technologies to deliver the intervention. Behavioral change strategies were thematically identified using the Coventry, Aberdeen, and London-Refined taxonomy [56].

Outcomes were classified as *positive* if there was a significant desired change in the weight-related measure postintervention delivery, for example, decrease or maintenance in body weight, BMI, or %body fat. Outcomes were classified as *mixed* if not *all* weight-related changes were statistically significant in *all* intervention arms. Outcomes were classified as having *no change* if no statistically significant differences in the weight-related measure were reported postintervention (*and* when compared with control groups, if applicable).

Results

Search Results

The initial database search located 3280 records, 1979 duplicates were removed, and the remaining 1301 records were retained for title and abstract screening. Of these, 1237 were excluded as they did not meet the criteria for inclusion. The most common reasons for exclusion were the following: (1) not related to eHealth and weight management, (2) not an intervention, (3) participants had an acute illness or chronic disease, or were pregnant, or (4) the age range was too broad (eg, 18-65 years). A total of 75 full-text papers were retained and assessed for inclusion. Of these, 24 studies met the criteria for inclusion and were included in the narrative evidence synthesis. Figure 1 illustrates the PRISMA study selection process employed.

Study Characteristics

Of the 24 studies included, over 92% (n=22) were published from 2010 onward, and all were conducted in developed countries: a total of 17 in the United Sates [57-72], 4 in Australia [73-76], 2 in the United Kingdom [77,78], and 1 in Belgium [79]. The majority employed either a CCT design (n=7) or an RCT design (n=15). All interventions addressed weight management; however, the behavioral focus of each intervention differed: a total of 10 focused on both healthy eating and PA [61,62,64-67,69,71,73,74], 7 focused on multiple behaviors (eg, healthy eating, PA, stress management, and sleep) [57,59,60,72,76-78], 3 focused on self-weighing [63,70], 3 focused on PA only [58,68,79], and 1 focused on healthy eating only [75]. The number of participants within each of the studies ranged from 12 to 2621, with a mean sample size of 468 participants. The majority of studies (n=20) recruited participants from colleges or universities, with only 4 studies extending their recruitment beyond an academic setting [74-76,79]. The duration of interventions ranged from 6 weeks to 24 months (mean=22 weeks), with an average retention rate at the final point of data collection of 79%. In terms of outcomes, 12 out of the 24 studies did not report any statistically significant changes in the weight-related measure(s) [57,58,61,65,66,68,71-73,77-79], 8 reported significant positive weight-related changes (eg, maintenance of a healthy weight or reversal of small gains) [60,62-64,70,74,76], and 4 reported mixed outcomes [59,67,69,75]. Refer to Multimedia Appendix 3 for a summary of individual study characteristics.

Intervention (Electronic Health) Components

Of the 24 studies included this review, 14 evaluated Web-based interventions [59-70,78], 3 evaluated mobile-based interventions [58,75,79], and 7 evaluated multicomponent interventions [57,71-74,76,77]. The following section provides a narrative synthesis of the different eHealth components employed in these studies.

Web-Based Interventions

Among the 14 studies evaluating Web-based interventions, 4 [65-67,78] comprised a Web-based education (e-learning)

program, 8 [59-64,68,69] used a combination of internet-enabled functions (eg, e-learning, website, email, e-counselor, e-newsletter, and Wi-Fi enabled scale), and 2 [70] used email as the sole method of intervention delivery. Typically, in the e-learning-based programs, participants (college or university students) were required to enroll in the program and complete the required modules to receive course credit for their participation [59,64,65,67,69,72]. The main behavioral change strategies employed in the e-learning-based studies were knowledge shaping, self-monitoring, social support, and contact with an interventionist (see Multimedia Appendix 3). For instance, in the study conducted by Gow et al (2010), students were randomized to either the (1) *internet* intervention arm (6 intensive e-learning sessions delivered via Blackboard), (2) *feedback* intervention arm (encompassing feedback from interventionists and using Blackboard for self-weighing) or (3) the *combined* intervention arm (e-learning sessions plus feedback) [69]. Similar programs were designed and evaluated by Dennis et al (2012), Greene et al (2012), LaChausse et al (2012), Kattelmann et al (2014), and Nikolaou et al (2015), whereby students randomized to the intervention arm(s) completed a semester long e-learning program accessible via a centralized website [57,65-67]. Specifically, Nikolaou et al (2015) used the Web-based e-learning program *Moodle* to deliver intervention content, with weekly email reminders sent to alert participants of new materials and mailboxes used to encourage communication between participants and interventionists [62]. Conversely, the study conducted by Harvey-Berino et al (2012) used a Web-based e-learning platform to facilitate weekly Web-based synchronous *group chats* led by a trained interventionist and supported by materials accessible via the intervention website [64].

The remaining Web-based interventions integrated multiple internet-enabled functions including e-newsletters, social network sites (SNSs), and email. The main behavioral change strategies employed in these studies were knowledge shaping, goal setting, self-monitoring, and contact with an interventionist (see Multimedia Appendix 3). For example, the study conducted by West et al (2016) was delivered via weekly e-newsletters and a private Facebook page, with participants also receiving a Wi-Fi scale and PA tracker (Fitbit Zip) for self-monitoring [72]. Similarly, the study conducted by Schweitzer et al (2016) comprised an adapted eHealth intervention where participants received weekly emails with tips for achieving set goals and weblinks to their personal accounts for viewing progress and accessing additional material [61]. In contrast, Bertz et al (2015) used Wi-Fi scales and email to facilitate the implementation of the caloric titration method, which involves daily self-weighing and visual feedback to promote weight management [63]. The final 2 studies combined both Web-based and offline components. The intervention evaluated by Dennis et al (2012) integrated both Web-based modules and biweekly in-class sessions with an expert instructor in nutrition and exercise science [67]. Similarly, the *Choosing Healthy Options in College Environments and Settings* (CHOICES) trial evaluated by Lytle et al (2017) offered an academic course with e-learning, face-to-face, and hybrid options for program delivery. The Web-based program included e-learning modules, an SNS, and a support website [59].

Mobile-Based Interventions

The 3 studies evaluating mobile-based interventions [58,75,79] delivered intervention content primarily via SMS text messages and mobile phone apps. The main behavioral change strategies employed in the mobile-based studies were goal setting, self-monitoring, and behavioral prompts (see Multimedia Appendix 3). For example, in the study conducted by Munoz et al (2014), participants used a pedometer to track PA, with brief SMS text messages (2-3 per week) sent throughout the intervention period to encourage the adoption of healthy behaviors [58]. Similarly, in the study conducted by Kerr et al (2016), dietary intake was monitored using a mobile food record app and tailored dietary feedback was sent weekly via SMS text messages to nudge healthy eating habits [75]. The study conducted by Simons et al (2018) comprised an investigator-designed mobile phone app (Active Coach) and a wearable device (Fitbit Charge) for tracking PA. The app included goal-setting functionalities, practical tips, and educational facts [79].

Multicomponent Interventions

The 7 studies [57,71-74,76,77] that were categorized as multicomponent used various eHealth technologies to deliver or support the intervention. The multicomponent studies employed a larger number of behavioral change strategies including knowledge shaping, barrier identification, goal setting, outcome expectation setting, behavioral prompts, self-monitoring, graded tasks, skill development, personalized feedback, social support, and contact with an interventionist (see Multimedia Appendix 3). The more complex interventions such as *TXT2BFiT* [73,74] and *HEYMAN* [76] incorporated multiple (eHealth) intervention components and associated change strategies. The pilot TXT2BFiT study comprised short SMS text messages, emails, mobile phone apps, and an internet forum [73]. The pilot was later refined and trialed in a larger RCT, which encompassed coaching calls by a dietician skilled in motivational interviewing, personalized SMS text messages tailored to participants' stage of change to prompt behavior change, a website (resource bank) for knowledge shaping, and 4 designer mobile phone apps for goal setting and self-monitoring. Following the completion of the 12-week intervention, booster SMS text messages, emails, and coaching calls were used to promote long-term behavioral change [74]. Similarly, the HEYMAN study included a website (resource bank) for knowledge shaping, wearable PA tracker (Jawbone) for goal setting and self-monitoring, weekly face-to-face sessions (60 min), personalized feedback reports, private Facebook group to facilitate social support and engagement (reminders and notifications), Gymstick resistance band to facilitate home-based strength training, and finally a TEMPlate dinner disc to guide main meal portion size [76]. In contrast, the study conducted by West et al (2016) had an educational focus and comprised 8 sessions delivered weekly via electronic newsletters and a (private) Facebook group [72]. The intervention encouraged frequent self-weighing, regular PA, and healthy eating. Participants received a Wi-Fi enabled scale and a wearable PA tracker (Fitbit Zip) to facilitate self-monitoring and weight maintenance. Similarly, the *Tweeting to Health* intervention used a Twitter account to deliver

education-based content. Participants also received a wearable PA tracker (Fitbit Zip) to facilitate self-monitoring [71].

Quality Assessment

According to the EPHPP quality assessment tool, overall, 19 out of the 24 included studies were rated as weak [57-59,61-72,75,77,78], 5 as moderate [60,73,74,76,79], and none as strong. A summary of the individual component ratings and overall quality ratings is provided in Multimedia Appendix 4. In terms of selection bias, no study reported representative sampling, with the majority using convenience sampling to recruit eligible participants from university or college-based settings. Participation rates were difficult to determine in most studies as details on consent throughout the recruitment, screening, and randomization stages were not clearly reported. Therefore, all 24 studies were classified as weak for *Component A*. With respect to study design, 92% (n=22) of the studies employed a CCT [63,65-68,70] or RCT [57-62,69,72-79] design. The 2 remaining studies used cohort designs: 1 employed a 1-group pretest-posttest design [64] and the other employed a 1-group posttest only design [71]. Consequently, 22 studies were rated as strong and 2 as moderate for *Component B*. In terms of confounders, 13 studies [57-63,69,72-75,78] reported no significant differences between intervention and comparison (control) groups at baseline; 3 [76,77,79] reported significant differences among groups but controlled for these in analyses; 3 [68,70] did not report any potential confounders; 3 [65-67] reported significant differences between groups at baseline but did not report whether these differences were controlled for; 2 [64,71] did not include a comparison group and therefore group differences were not applicable. Consequently, 16 studies [57-63,69,72-79] were rated as strong and 8 [62,65-68,70,74] as weak for *Component C*.

In terms of blinding, 16 studies [57-59,61,63,65-70,72,75,77-79] did not describe blinding of outcome assessors or participants. Owing to the recruitment methods employed and the nature of the interventions (ie, behavioral modification), participants in these studies were assessed as being aware of the study's research question unless explicitly stated otherwise. Of the remaining studies, 2 studies [62,74] were double blinded; 1 [73] did not report blinding of outcome assessors but participants were reportedly blinded; 2 reported outcome assessors as blinded but participants as not [60,76]; blinding was not applicable in 2 studies [64,71]. Subsequently, 18 studies [57-59,61,63-72,75,77,78] were assessed as weak, 4 [60,73,76,79] as moderate, and 2 [62,74] as strong for *Component D*. With regard to data collection methods, 20 studies [57-63,65-69,72-79] reported some evidence of reliability (eg, Cronbach alpha) and validity (eg, reference to a validation study) for measures used to assess primary outcomes; 1 [64] reported measures to be valid but not reliable; 3 [70,71] did not report any evidence of the reliability and validity of the measures used. Consequently, 20 studies were assessed as strong [57-63,65-69,72-79], 1 as moderate [64], and 3 [70,71] as weak for *Component E*.

With respect to withdrawals and dropouts, 12 studies [57,58,61,65,66,68-70,72,77,78] reported the number of dropouts but not the reasons for this attrition; 9 [59,60,63,67,73-76,79] reported both the numbers and reasons (eg, medical reasons, life changes could no longer commit, no contact); 1 study [64] did not report numbers or reasons. For the 2 remaining studies, withdrawals and dropouts were unclear [62] and not applicable [71]. On the basis of study completion rates, 10 studies [59,60,65,67,72-76,79] were assessed as strong (80%-100% retention), 8 [61,63,66,68-70,77] as moderate (60%-79% retention), and 6 [57,58,62,64,71,78] as weak (<60% retention) for *Component F*.

Discussion

Principal Findings

This systematic review provides a comprehensive narrative evidence synthesis of eHealth weight management interventions targeting young adults, with a particular focus on (eHealth) intervention components and outcomes. A total of 24 studies were identified and included in the review. A majority were published from 2010 onward, conducted in developed countries, and used convenience sampling to recruit young adults from university- or college-based settings. There was large variation in the behavioral focus, intervention design and duration, sample size, and outcomes reported across the included studies. The variability across intervention outcomes highlights that additional research is warranted to extend our understanding of *what works, for whom, how, and when?* The following discussion provides further commentary on review findings, along with recommendations for future research.

Intervention (Electronic Health) Components

Technology as a means to communicate content in eHealth interventions is often overlooked [80]. Frequently, technology is seen as a black box, a mere tool that has no effect or value and serves only as a vehicle to deliver intervention content [81]. However, recent research suggests that technology should be seen as a vital and inseparable aspect of interventions [82] and should be examined from a more holistic perspective [81,83]. With differences in persuasive technology elements and user interaction shown to be significant predictors of adherence [80], the design of *persuasive technology* should be an important consideration in the development of any eHealth intervention seeking sustained adherence [84]. Nonadherence is an issue that continues to plague the effectiveness of eHealth interventions [80,82,85], with many participants failing to sustain their use of the intervention in the desired way [81]. Given that nonoptimal exposure to an intervention has been shown to lessen intervention effect [86], examining technology and user interaction from a more holistic perspective is necessary for improving adherence and in turn the effectiveness of eHealth interventions.

In this review, the use of eHealth components and behavioral change strategies varied, with some studies only utilizing 1 technological function (eg, SMS text message or email) and others employing a range of internet- and mobile-enabled functions (eg, website, mobile phone apps, email, and SMS text message). Earlier studies (published 2006-2012) were generally more basic by design with the majority employing an e-learning-based approach to deliver a didactic education-focused weight management program, usually as part

of a college or university-based course. With a focus on education and raising awareness, these interventions offered limited opportunities for participants to interact and actively engage with the technology, and as a result, exposure to intervention content was likely suboptimal. Interaction has been shown to be a significant predictor of adherence; therefore, eHealth interventions that fail to promote user interaction are unlikely to achieve the intended usage target [81]. Furthermore, research indicates that focusing on education (ie, knowledge shaping) alone is unlikely to achieve the level of behavior change necessary to address weight status [87,88]. Moreover, participants enrolled in the e-learning-based interventions were likely to be more motivated by the course credit on offer rather than learning new skills for healthy lifestyle adoption and weight maintenance. Later studies (published 2013-2017) became more sophisticated in their use of technology and associated behavioral change strategies, leveraging modern technological innovations. For example, using algorithms for content tailoring (eg, knowledge shaping), mobile phone apps and wearable devices for tracking behavior (eg, goal setting and self-monitoring) and relaying real-time feedback (eg, prompt review and reinforcement) to improve the capability, interactivity, and mobility of the intervention. Research has found eHealth interventions, which are enhanced by a range of features (eg, personalized e-feedback, chat rooms, and goal-setting and self-monitoring tools), support greater retention and usage of the intervention than standard (or basic) eHealth interventions [89]. It should be noted that 5 studies [59,60,65,67,76] also incorporated face-to-face (individual or group-based) sessions in 1 or more of the intervention (treatment) arms, further highlighting the limited evidence base for eHealth weight management interventions targeting young adults.

Recommended Intervention Strategies Delivered Via Electronic Health

Although the evidence for successful eHealth weight management interventions targeting young adults (aged 18 to 35 years) was limited, common behavioral change strategies and techniques were able to be thematically identified, with an emphasis placed on the studies categorized as having positive or mixed weight-related outcomes. The 5 strategies identified included the following: self-regulation (goal setting and self-monitoring), tailored or personalized feedback, contact with an interventionist, social support, and behavioral prompts (nudges and reminders) and booster messages.

Self-Regulation (Via Goal Setting and Self-Monitoring)

Most weight management interventions promote goal setting along with some form of self-monitoring, usually recommending that participants should record details pertaining to their behavioral patterns (eg, dietary intake and PA) and weight (eg, BMI) and review tracking data in line with their goals or recommended guidelines to evaluate progress and identify where further changes are needed [90]. The premise of self-regulation for changing finely ingrained habits is that monitoring of one's behavior will lead to self-evaluation of progress made toward previously set goals, with ensuing self-reinforcement following this evaluation. Thus, the process of changing habits requires well-developed self-regulatory skills [91,92]. Self-monitoring and goal setting are central to this process [93]. Self-monitoring requires deliberate attention to one's own actions, as well as the conditions under which they occur, and their immediate and long-term effects [93]. Research indicates that self-monitoring of key behaviors has been associated with successful weight maintenance [92,94-96]. In particular, the use of technology for self-monitoring has been suggested as a way of lessening the burden of self-monitoring and enhancing adherence [97]. In this review, all studies reporting positive weight-related outcomes implemented some form of self-monitoring (eg, frequent self-weighing, monitoring PA, or dietary intake). For instance, in the HEYMAN intervention [76], participants received a Jawbone wearable PA tracker with an associated mobile phone app (UP app) to assist in goal setting and self-monitoring and a TEMPlate dinner disc to guide main meal portion size. The findings from this review suggest that improving self-regulation skills should be a central focus of future eHealth weight management interventions, particularly given young adults often lack such skills [24].

Tailored or Personalized Feedback

Tailoring has been shown to enhance the effectiveness of behavior change interventions, including eHealth-based interventions [94,95]. Tailoring involves gathering and assessing personal data to determine the most effective strategy to meet the specific needs of an individual [96]. Collecting data for tailoring intervention content enables personalized feedback, commands greater attention and is processed more deeply by the recipient and is perceived as more likable than a generic message [96,98]. With ready access to data provision and retrieval, the internet provides a powerful tool for tailoring interventions [96]. Interactive and responsive tailoring enhances the user's experience with and understanding of intervention content [94,95]. Tailoring can range from simple Web-based assessments and feedback to highly sophisticated interventions that are completely customized [95]. Of the studies reporting tailoring in this review, most only employed simple tailoring based on either Web-based or in-person health assessments. For example, in the trial conducted by Bertz et al (2015), participants weighed themselves daily using Wi-Fi scales and immediately received an email containing their weight plotted over time with a horizontal reference line indicating their target weight [63]. A few studies employed more sophisticated levels of tailoring. The TXT2BFiT intervention [74] used a staging algorithm based on the Transtheoretical/Stages of Change Model to generate a personalized set of SMS text messages, which were tailored to whether the participant was in precontemplation, contemplation, preparation, action, or maintenance stages of change for each of the 4 behaviors addressed. More cognitive messages were included if a participant was in 1 of the early stages of change, and the messages were more behavioral if the participant was in the action or maintenance stages of change. We recommend that future studies experiment with more sophisticated methods of tailoring to empirically test which aspects of the tailored messages promote adherence and in turn enhance effectiveness in this context.

Contact With an Interventionist

Several studies included in this review incorporated in-person support from an interventionist. Human support has been shown to enhance the effectiveness of and adherence to eHealth interventions via accountability to a coach who is seen as trustworthy, benevolent, and having expertise [99]. However, intervention designs incorporating in-person support are resource intensive. A trained specialist is needed to deliver intervention content and monitor participants to ensure the correct treatment dose is received and the fidelity of the intervention is maintained. In addition, the facilities and equipment required to deliver the intervention must be procured. Furthermore, the effort that is required on the part of participants to commit to and attend in-person counseling sessions can create barriers (eg, cost of travel, lack of parking at venues, and limited availability) to full participation and adherence [100]. The high cost and inability of these interventions to reach diverse demographic and socioeconomic groups thwart large-scale deployment to the wider community [28]. More economical methods that may provide similar outcomes to face-to-face contact, while reducing the costs associated with intervention delivery, include coaching calls via telephone, email, chat forums, and SNSs [101]. For example, in the TXT2BFiT intervention [74], participants received 5 coaching calls focused on goal setting and a review from a dietician skilled in motivation interviewing. Similarly, in the CHOICES intervention [59], a study specific SNS facilitated participant engagement with peers and the intervention staff. Future research should consider the potential benefits and disadvantages of different communication mediums to deliver expert support at scale.

Social Support

Social support has been identified as an important factor in the provision of lifestyle-focused weight management interventions [102-104], including those supported by technology [105]. In particular, SNSs provide an ideal platform for facilitating social support with access to large existing (or new) networks of influencers [106]. The studies incorporating social support in this review typically facilitated peer support via online chat forums or SNSs. For example, in the CHOICES intervention conducted by Lytle et al (2017), a study-specific SNS was created to encourage discussion and interaction among participants [59]. Similarly, in the HEYMAN intervention, a combination of in-person (via group-based sessions) and Web-based social support (via a private Facebook group) was employed to facilitate interaction among participants [76]. Given that there is research to show that social contacts and normative beliefs influence weight status and intentions for weight control in young adults [106], mediums for delivering social support should be a key consideration in future research.

Behavioral Prompts (Nudges and Reminders) and Booster Messages

Maintenance of behavior change presents an ongoing challenge for behavior change research, with very little actually known about the process of behavioral maintenance [107]. The evidence supporting the use of behavioral nudges, reminders, and booster sessions for behavioral change and maintenance is mixed [108]. However, findings from this review indicate that booster emails, SMS text messages, and coaching calls may help promote behavioral maintenance over the longer term. For example, in the TXT2BFiT intervention [74], a low dose maintenance phase was implemented following the completion of the initial 12-week intervention. In this maintenance phase, participants received monthly SMS text messages and emails and 2 booster coaching calls at 5 and 8 months. Technology offers a feasible means of delivering strategies that promote behavioral maintenance; however, further research is needed to better understand the process of behavioral maintenance.

Quality of Included Studies

A majority of studies included in this review were of weak methodological quality according to EPHPP quality assessment ratings. The main weaknesses identified were the following: a lack of studies employing representative sampling, not clearly reporting participation rates, not blinding assessors and participants to group allocation, and low completion rates. Future research should aim to address these issues to improve the methodological quality of the evidence for eHealth weight management interventions targeting young adults.

Representativeness in eHealth-based research is crucial for ensuring interventions are capable of reaching large, representative numbers of the target population, particularly those who are most in need of treatment [109]. According to Glasgow (2007), reach is a function of both participation rate and the representativeness of participants compared with nonparticipants based upon a set of key characteristics including race, ethnicity, socioeconomic status, computer experience, and health literacy [109]. Participant characteristics across the studies included in this review were similar, with the majority of participants recruited from large western universities or colleges using convenience sampling procedures. Consequently, the results obtained from these studies are limited to a very small, homogenous (ie, high socioeconomic status, education level, and health literacy level) subgroup of the target population, and they are unlikely to generalize to the larger target population, including those most in need (ie, socioeconomically disadvantaged groups, low education and health literacy, ethnic minorities, and rural and remote communities) [28]. The study conducted by Simons et al (2018) is the exception as investigators specifically recruited lower-educated working young adults. Although difficulties in recruiting young adults are acknowledged [110,111], to improve representativeness, future research should aim to employ probability sampling methods, maintain a careful record of recruitment strategies and results, and collect data on both participant and nonparticipant characteristics.

The design of most studies included in this review was rated as strong; however, very few reported blinding of outcome assessors and participants. As a consequence, findings from these studies were likely influenced by detection and reporting bias [55]. Importantly, the 2 studies [62,74] that were reportedly double blinded reported positive weight-related outcomes, and 2 out of the 3 studies [60,73,76] reporting some level of blinding (either outcome assessors or participants) reported positive weight-related outcomes. Therefore, where practical and

feasible, future research should aim for double-blinded allocations.

Finally, intervention durations and completion rates varied significantly across studies (6 weeks-24 months for duration; 42%-98% for retention), which affects the veracity of study findings and the ability to compare outcomes. Coupled with the lack of studies reporting details on effective recruitment strategies and reasons for attrition, the current understandings on how best to recruit and engage young adults in weight management studies is limited [28,112,113]. To improve participation rates, retention, and resource allocation efficiency, future research should keep a careful record of recruitment strategies, participation rates, and reasons for attrition by following up with withdrawals and dropouts.

Limitations

A number of limitations, many of which represent opportunities for future research, are acknowledged. First, the search parameters employed were specific to the review's research aim, thereby limiting the number of studies identified. For example, grey literature, nonpeer-reviewed research, and studies not published in English were excluded. Future research may therefore extend this review by including grey literature, nonpeer-reviewed research, and studies not published in English. Moreover, all studies included in this review were conducted in developed countries; however, obesity is not isolated to the developed world [114]. Thus, extending the current review to include research from developing countries could provide valuable insight into the generalizability of study findings in different geographic contexts. Furthermore, future research may adopt the Patient, Problem or Population, Intervention, Comparison, Control or Comparator, and Outcomes framework to inform the search strategy and eligibility criteria and compare whether this approach yields the same or different results as in this review. Second, the highly complex nature of the interventions included in this review limited our ability to confidently isolate the active drivers of intervention outcomes [112]. Although some potential behavioral change strategies were thematically identified, definitive conclusions as to which intervention components were contributing most to outcomes (or lack thereof) were not able to be made. Future research should follow published guidelines on developing and evaluating complex interventions to permit critical appraisal [112], and research is called for to expand the evidence base. As the evidence base grows, we recommend that narrower age ranges should be set to extend understanding. Third, the EPHPP quality assessment tool, although deemed appropriate for the purposes of this review, is one of several tools that can be used to evaluate the quality of quantitative studies. As such, overall quality ratings should be interpreted with the specific characteristics of this tool in mind, given different assessment outcomes can arise from different tools [54].

Furthermore, obtaining representative samples and blinding participants to group allocation in interventions attempting to modify behavior(s) are not often practical or feasible. Assessment tools such as the EPHPP are principally concerned with *what should be* and this is seldom equivalent to *what works* in the field. Therefore, although the quality of the evidence in this review was generally assessed as weak, results should be interpreted tentatively. Sampling, study design, and retention rates will remain key determining factors of reliability and validity; however, further research attention should be directed toward the applicability, generalizability, and impact potential of studies. Given a large proportion of weight management interventions are delivered in-field, with varying budget amounts, expecting study designs to conform to the standards set within quality assessment tools arising from controlled clinical settings may not be realistic. As such, additional research is needed to better understand which metrics can be reliably applied within different research designs. For instance, future reviews may consider incorporating the Grading of Recommendations Assessment, Development and Evaluation [113] assessment tool to provide an overall judgement of the evidence base and in turn guide future practice. Future research may opt to narrow the scope of review to 1 specific behavior or study design to permit metaanalytic comparisons; however, the results of this review suggest such a narrow scope would significantly limit the number of eligible studies available for quantitative comparison at this point in time. Finally, we recommend that future studies publish a review protocol (researchprotocols.org) to establish an early scientific record, promote transparency, solicit early feedback, and enhance review methods and processes.

Conclusions

The prevention of unhealthy weight gain in young adults provides a new target for reducing the rising prevalence of obesity, and it is one that could offer an effective transgenerational approach to obesity prevention. Consequently, there is a need to develop effective weight management programs that are capable of engaging a large number of young adults in healthy lifestyle adoption over the longer term. An eHealth-based approach offers potential, with young adults among the highest users of digital technologies. However, at present, there is limited high-quality, peer-reviewed evidence available. Future research must be directed toward improving the methodological quality of the evidence and establishing which specific elements of eHealth weight management interventions are most effective in achieving the desired outcomes, thereby answering the *for whom, how, and when* question.

Acknowledgments

This work was supported by the Commonwealth Government of Australia via a Research Training Program Scholarship to the first author (TJW).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist completed during the review.

[PDF File (Adobe PDF File), 50KB - jmir_v21i2e10265_app1.pdf]

Multimedia Appendix 2

Search strategy and database results.

[PDF File (Adobe PDF File), 40KB - jmir_v21i2e10265_app2.pdf]

Multimedia Appendix 3

Summary of individual study characteristics.

[PDF File (Adobe PDF File), 262KB - jmir_v21i2e10265_app3.pdf]

Multimedia Appendix 4

Summary of Effective Public Health Practice Project quality assessment results.

[PDF File (Adobe PDF File), 33KB - jmir_v21i2e10265_app4.pdf]

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781. [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
2. Dobbs R, Sawers C, Thompson F. McKinsey Global Institute. 2014. Overcoming obesity: An initial economic analysis URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Economic%20Studies%20TEMP/Our%20Insights/How%20the%20world%20could%20better%20fight%20obesity/MGI_Overcoming_obesity_Full_report.ashx [accessed 2018-02-28] [WebCite Cache ID 6xa1dhP4b]
3. Kelly T, Yang W, Chen C, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. Int J Obes (Lond) 2008 Sep;32(9):1431-1437. [doi: 10.1038/ijo.2008.102] [Medline: 18607383]
4. World Health Organization. 2010. Global status report on noncommunicable diseases 2010 URL: http://apps.who.int/iris/bitstream/10665/44579/6/9789244564226_rus.pdf [accessed 2017-08-28] [WebCite Cache ID 6xa1vmPyq]
5. Müller MJ, Mast M, Asbeck I, Langnäse K, Grund A. Prevention of obesity--is it possible? Obes Rev 2001 Feb;2(1):15-28. [doi: 10.1046/j.1467-789x.2001.00012.x] [Medline: 12119633]
6. Kumanyika SK, Obarzanek E, Stettler N, Bell R, Field AE, Fortmann SP, American Heart Association Council on Epidemiology Prevention, Interdisciplinary Committee for Prevention. Population-based prevention of obesity: the need for comprehensive promotion of healthful eating, physical activity, and energy balance: a scientific statement from American Heart Association Council on Epidemiology and Prevention, Interdisciplinary Committee for Prevention (formerly the expert panel on population and prevention science). Circulation 2008 Jul 22;118(4):428-464 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.189702] [Medline: 18591433]
7. King L, Gill T, Allender S, Swinburn B. Best practice principles for community-based obesity prevention: development, content and application. Obes Rev 2011 May;12(5):329-338. [doi: 10.1111/j.1467-789X.2010.00798.x] [Medline: 20880111]
8. Gill T, King L, Caterson I. Obesity prevention: necessary and possible. A structured approach for effective planning. Proc Nutr Soc 2005 May;64(2):255-261. [doi: 10.1079/PNS2005425] [Medline: 15960870]
9. Pi-Sunyer X. A clinical view of the obesity problem. Science 2003 Feb 07;299(5608):859-860. [doi: 10.1126/science.1082319] [Medline: 12574620]
10. Proietto J. Why is treating obesity so difficult? Justification for the role of bariatric surgery. Med J Aust 2011 Aug 01;195(3):144-146. [doi: 10.5694/j.1326-5377.2011.tb03242.x] [Medline: 21806533]
11. Dietz WH. Obesity and excessive weight gain in young adults: new targets for prevention. J Am Med Assoc 2017 Dec 18;318(3):241-242. [doi: 10.1001/jama.2017.6119] [Medline: 28719674]
12. Zheng Y, Manson JE, Yuan C, Liang MH, Grodstein F, Stampfer MJ, et al. Associations of weight gain from early to middle adulthood with major health outcomes later in life. J Am Med Assoc 2017 Jul 18;318(3):255-269. [doi: 10.1001/jama.2017.7092] [Medline: 28719691]
13. Flegal KM, Kruszon-Moran D, Carroll MD, Fryar CD, Ogden CL. Trends in obesity among adults in the United States, 2005 to 2014. J Am Med Assoc 2016 Jun 07;315(21):2284-2291. [doi: 10.1001/jama.2016.6458] [Medline: 27272580]

14. de Mutsert R, Sun Q, Willett WC, Hu FB, van Dam RM. Overweight in early adulthood, adult weight change, and risk of type 2 diabetes, cardiovascular diseases, and certain cancers in men: a cohort study. Am J Epidemiol 2014 Jun 1;179(11):1353-1365 [FREE Full text] [doi: 10.1093/aje/kwu052] [Medline: 24786797]
15. Adams KF, Leitzmann MF, Ballard-Barbash R, Albanes D, Harris TB, Hollenbeck A, et al. Body mass and weight change in adults in relation to mortality risk. Am J Epidemiol 2014 Jan 15;179(2):135-144 [FREE Full text] [doi: 10.1093/aje/kwt254] [Medline: 24173550]
16. Nelson MC, Story M, Larson NI, Neumark-Sztainer D, Lytle LA. Emerging adulthood and college-aged youth: an overlooked age for weight-related behavior change. Obesity (Silver Spring) 2008 Oct;16(10):2205-2211. [doi: 10.1038/oby.2008.365] [Medline: 18719665]
17. Hankinson AL, Daviglus ML, Bouchard C, Carnethon M, Lewis CE, Schreiner PJ, et al. Maintaining a high physical activity level over 20 years and weight gain. J Am Med Assoc 2010 Dec 15;304(23):2603-2610 [FREE Full text] [doi: 10.1001/jama.2010.1843] [Medline: 21156948]
18. De Cocker KA, van Uffelen JG, Brown WJ. Associations between sitting time and weight in young adult Australian women. Prev Med 2010 Nov;51(5):361-367. [doi: 10.1016/j.ypmed.2010.07.009] [Medline: 20638408]
19. Larson N, Neumark-Sztainer D, Laska MN, Story M. Young adults and eating away from home: associations with dietary intake patterns and weight status differ by choice of restaurant. J Am Diet Assoc 2011 Nov;111(11):1696-1703 [FREE Full text] [doi: 10.1016/j.jada.2011.08.007] [Medline: 22027052]
20. Duffey KJ, Gordon-Larsen P, Steffen LM, Jacobs DR, Popkin BM. Drinking caloric beverages increases the risk of adverse cardiometabolic outcomes in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. Am J Clin Nutr 2010 Oct;92(4):954-959 [FREE Full text] [doi: 10.3945/ajcn.2010.29478] [Medline: 20702604]
21. Sherwood NE, Jeffery RW, French SA, Hannan PJ, Murray DM. Predictors of weight gain in the Pound of Prevention study. Int J Obes Relat Metab Disord 2000 Apr;24(4):395-403. [doi: 10.1038/sj.ijo.0801169] [Medline: 10805494]
22. Kimmons J, Gillespie C, Seymour J, Serdula M, Blanck HM. Fruit and vegetable intake among adolescents and adults in the United States: percentage meeting individualized recommendations. Medscape J Med 2009;11(1):26. [Medline: 19295947]
23. Greaney ML, Less FD, White AA, Dayton SF, Riebe D, Blissmer B, et al. College students' barriers and enablers for healthful weight management: a qualitative study. J Nutr Educ Behav 2009;41(4):281-286. [doi: 10.1016/j.jneb.2008.04.354] [Medline: 19508934]
24. Strong KA, Parks SL, Anderson E, Winett R, Davy BM. Weight gain prevention: identifying theory-based targets for health behavior change in young adults. J Am Diet Assoc 2008 Oct;108(10):1708-1715 [FREE Full text] [doi: 10.1016/j.jada.2008.07.007] [Medline: 18926139]
25. Andajani-Sutjahjo S, Ball K, Warren N, Inglis V, Crawford D. Perceived personal, social and environmental barriers to weight maintenance among young women: a community survey. Int J Behav Nutr Phys Act 2004 Dec 5;1(1):15 [FREE Full text] [doi: 10.1186/1479-5868-1-15] [Medline: 15462679]
26. Sand A, Emaus N, Lian OS. Motivation and obstacles for weight management among young women-a qualitative study with a public health focus - the Tromsø study: Fit Futures. BMC Public Health 2017 Dec 08;17(1):417 [FREE Full text] [doi: 10.1186/s12889-017-4321-9] [Medline: 28482855]
27. Spring B, Moller AC, Colangelo LA, Siddique J, Roehrig M, Daviglus ML, et al. Healthy lifestyle change and subclinical atherosclerosis in young adults: Coronary Artery Risk Development in Young Adults (CARDIA) study. Circulation 2014 Jul 1;130(1):10-17. [doi: 10.1161/CIRCULATIONAHA.113.005445] [Medline: 24982115]
28. Partridge SR, Juan SJ, McGeechan K, Bauman A, Allman-Farinelli M. Poor quality of external validity reporting limits generalizability of overweight and/or obesity lifestyle prevention interventions in young adults: a systematic review. Obes Rev 2015 Jan;16(1):13-31. [doi: 10.1111/obr.12233] [Medline: 25407633]
29. Hutchesson MJ, Hulst J, Collins CE. Weight management interventions targeting young women: a systematic review. J Acad Nutr Diet 2013 Jun;113(6):795-802. [doi: 10.1016/j.jand.2013.01.015] [Medline: 23473986]
30. Hebden L, Chey T, Allman-Farinelli M. Lifestyle intervention for preventing weight gain in young adults: a systematic review and meta-analysis of RCTs. Obes Rev 2012 Aug;13(8):692-710. [doi: 10.1111/j.1467-789X.2012.00990.x] [Medline: 22413804]
31. Laska MN, Pelletier JE, Larson NI, Story M. Interventions for weight gain prevention during the transition to young adulthood: a review of the literature. J Adolesc Health 2012 Apr;50(4):324-333 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.01.016] [Medline: 22443834]
32. Poobalan AS, Aucott LS, Precious E, Crombie IK, Smith WC. Weight loss interventions in young people (18 to 25 year olds): a systematic review. Obes Rev 2010 Aug;11(8):580-592. [doi: 10.1111/j.1467-789X.2009.00673.x] [Medline: 19874531]
33. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. Prev Med 2017 Jun;99:197-206. [doi: 10.1016/j.ypmed.2017.01.009] [Medline: 28130046]
34. Gokee-LaRose J, Gorin AA, Raynor HA, Laska MN, Jeffery RW, Levy RL, et al. Are standard behavioral weight loss programs effective for young adults? Int J Obes (Lond) 2009 Dec;33(12):1374-1380 [FREE Full text] [doi: 10.1038/ijo.2009.185] [Medline: 19786967]

35. Giedd JN. Structural magnetic resonance imaging of the adolescent brain. Ann N Y Acad Sci 2004 Jun;1021:77-85. [doi: 10.1196/annals.1308.009] [Medline: 15251877]
36. Larson NI, Perry CL, Story M, Neumark-Sztainer D. Food preparation by young adults is associated with better diet quality. J Am Diet Assoc 2006 Dec;106(12):2001-2007. [doi: 10.1016/j.jada.2006.09.008] [Medline: 17126631]
37. Laska MN, Graham D, Moe SG, Lytle L, Fulkerson J. Situational characteristics of young adults' eating occasions: a real-time data collection using Personal Digital Assistants. Public Health Nutr 2011 Mar;14(3):472-479. [doi: 10.1017/S1368980010003186] [Medline: 21138611]
38. Lenhart A, Purcell K, Smith A, Zickuhr K. Pew Internet. Social Media & Mobile Internet Use among Teens and Young Adults URL: https://files.eric.ed.gov/fulltext/ED525056.pdf [accessed 2018-03-01] [WebCite Cache ID 6xa4ECJNi]
39. World Health Organization. 2017. eHealth at WHO URL: http://www.who.int/ehealth/about/en/ [accessed 2018-03-01] [WebCite Cache ID 6xa4K5lTl]
40. Rooij TV, Marsh S. eHealth: past and future perspectives. Per Med 2016 Jan;13(1):57-70. [doi: 10.2217/pme.15.40] [Medline: 29749870]
41. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
42. Deshmukh-Taskar PR, Nicklas TA, O'Neil CE, Keast DR, Radcliffe JD, Cho S. The relationship of breakfast skipping and type of breakfast consumption with nutrient intake and weight status in children and adolescents: the National Health and Nutrition Examination Survey 1999-2006. J Am Diet Assoc 2010 Jun;110(6):869-878. [doi: 10.1016/j.jada.2010.03.023] [Medline: 20497776]
43. 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]
44. 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]
45. 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]
46. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011;11(1):15 [FREE Full text] [doi: 10.1186/1471-2288-11-15] [Medline: 21291558]
47. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
48. Australian Institute of Health and Welfare. Australian Institute of Health and Welfare (AIHW). Overweight and Obesity URL: https://www.aihw.gov.au/reports-data/behaviours-risk-factors/overweight-obesity/data [accessed 2018-12-14] [WebCite Cache ID 74eEWX8Zb]
49. Lytle LA, Svetkey LP, Patrick K, Belle SH, Fernandez ID, Jakicic JM, et al. The EARLY trials: a consortium of studies targeting weight control in young adults. Transl Behav Med 2014 Sep;4(3):304-313 [FREE Full text] [doi: 10.1007/s13142-014-0252-5] [Medline: 25264469]
50. Lloyd-Jones DM, Liu K, Colangelo LA, Yan LL, Klein L, Loria CM, et al. Consistently stable or decreased body mass index in young adulthood and longitudinal changes in metabolic syndrome components: the Coronary Artery Risk Development in Young Adults Study. Circulation 2007 Feb 27;115(8):1004-1011. [doi: 10.1161/CIRCULATIONAHA.106.648642] [Medline: 17283263]
51. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004;1(3):176-184. [doi: 10.1111/j.1524-475X.2004.04006.x] [Medline: 17163895]
52. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, International Stroke Trial Collaborative Group, European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7(27):iii-x, 1 [FREE Full text] [doi: 10.3310/hta7270] [Medline: 14499048]
53. Jackson N, Waters E, Guidelines for Systematic Reviews in Health Promotion Public Health Taskforce. Criteria for the systematic review of health promotion and public health interventions. Health Promot Int 2005 Dec;20(4):367-374. [doi: 10.1093/heapro/dai022] [Medline: 16169885]
54. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
55. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. West Sussex, England: Wiley, John & Sons, Incorporated; 2008.

56. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
57. Kattelmann KK, Bredbenner CB, White AA, Greene GW, Hoerr SL, Kidd T, et al. The effects of Young Adults Eating and Active for Health (YEAH): a theory-based web-delivered intervention. J Nutr Educ Behav 2014;46(6):S27-S41. [doi: 10.1016/j.jneb.2014.08.007] [Medline: 25457733]
58. Muñoz L, La FK, Dominguez D. Text messaging as a tool to increase physical activity in college students. Physical Educator 2014;71(3):442 [FREE Full text]
59. Lytle LA, Laska MN, Linde JA, Moe SG, Nanney MS, Hannan PJ, et al. Weight-gain reduction among 2-year college students: the CHOICES RCT. Am J Prev Med 2017 Feb;52(2):183-191 [FREE Full text] [doi: 10.1016/j.amepre.2016.10.012] [Medline: 27939237]
60. Wing RR, Tate DF, Espeland MA, Lewis CE, LaRose JG, Gorin AA, Study of Novel Approaches to Weight Gain Prevention (SNAP) Research Group. Innovative self-regulation strategies to reduce weight gain in young adults: the Study of Novel Approaches to Weight Gain Prevention (SNAP) randomized clinical trial. JAMA Intern Med 2016 Dec 01;176(6):755-762 [FREE Full text] [doi: 10.1001/jamainternmed.2016.1236] [Medline: 27136493]
61. Schweitzer AL, Ross JT, Klein CJ, Lei KY, Mackey ER. n electronic wellness program to improve diet and exercise in college students: a pilot study. JMIR Res Protoc 2016;5(1):e29 [FREE Full text] [doi: 10.2196/resprot.4855] [Medline: 26929118]
62. Nikolaou CK, Hankey CR, Lean ME. Elearning approaches to prevent weight gain in young adults: a randomized controlled study. Obesity (Silver Spring) 2015 Dec;23(12):2377-2384 [FREE Full text] [doi: 10.1002/oby.21237] [Medline: 26538383]
63. Bertz F, Pacanowski CR, Levitsky DA. Frequent self-weighing with electronic graphic feedback to prevent age-related weight gain in young adults. Obesity (Silver Spring) 2015 Oct;23(10):2009-2014 [FREE Full text] [doi: 10.1002/oby.21211] [Medline: 26414563]
64. Harvey-Berino J, Pope L, Gold BC, Leonard H, Belliveau C. Undergrad and overweight: an online behavioral weight management program for college students. J Nutr Educ Behav 2012;44(6):604-608. [doi: 10.1016/j.jneb.2012.04.016] [Medline: 23140565]
65. Lachausse RG. My student body: effects of an internet-based prevention program to decrease obesity among college students. J Am Coll Health 2012;60(4):324-330. [doi: 10.1080/07448481.2011.623333] [Medline: 22559092]
66. Greene GW, White AA, Hoerr SL, Lohse B, Schembre SM, Riebe D, et al. Impact of an online healthful eating and physical activity program for college students. Am J Health Promot 2012;27(2):e47-e58. [doi: 10.4278/ajhp.110606-QUAN-239] [Medline: 23113786]
67. Dennis EA, Potter KL, Estabrooks PA, Davy BM. Weight gain prevention for college freshmen: comparing two social cognitive theory-based interventions with and without explicit self-regulation training. J Obes 2012;2012:803769 [FREE Full text] [doi: 10.1155/2012/803769] [Medline: 22778919]
68. Wadsworth DD, Hallam JS. Effect of a web site intervention on physical activity of college females. Am J Health Behav 2010;34(1):60-69. [doi: 10.5993/AJHB.34.1.8] [Medline: 19663753]
69. Gow RW, Trace SE, Mazzeo SE. Preventing weight gain in first year college students: an online intervention to prevent the "freshman fifteen". Eat Behav 2010 Jan;11(1):33-39 [FREE Full text] [doi: 10.1016/j.eatbeh.2009.08.005] [Medline: 19962118]
70. Levitsky DA, Garay J, Nausbaum M, Neighbors L, Dellavalle DM. Monitoring weight daily blocks the freshman weight gain: a model for combating the epidemic of obesity. Int J Obes (Lond) 2006 Jun;30(6):1003-1010. [doi: 10.1038/sj.ijo.0803221] [Medline: 16446748]
71. Chung AE, Skinner AC, Hasty SE, Perrin EM. Tweeting to Health: a novel mHealth intervention using Fitbits and Twitter to foster healthy lifestyles. Clin Pediatr (Phila) 2016 Jun 16;56(1):26-32. [doi: 10.1177/0009922816653385] [Medline: 27317609]
72. West DS, Monroe CM, Turner-McGrievy G, Sundstrom B, Larsen C, Magradey K, et al. A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study. J Med Internet Res 2016 Jun 13;18(6):e133 [FREE Full text] [doi: 10.2196/jmir.5474] [Medline: 27296086]
73. Hebden L, Cook A, van der Ploeg HP, King L, Bauman A, Allman-Farinelli M. A mobile health intervention for weight management among young adults: a pilot randomised controlled trial. J Hum Nutr Diet 2014 Aug;27(4):322-332. [doi: 10.1111/jhn.12155] [Medline: 23992038]
74. Allman-Farinelli M, Partridge SR, McGeechan K, Balestracci K, Hebden L, Wong A, et al. A mobile health lifestyle program for prevention of weight gain in young adults (TXT2BFiT): nine-month outcomes of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jun 22;4(2):e78 [FREE Full text] [doi: 10.2196/mhealth.5768] [Medline: 27335237]
75. Kerr DA, Harray AJ, Pollard CM, Dhaliwal SS, Delp EJ, Howat PA, et al. The connecting health and technology study: a 6-month randomized controlled trial to improve nutrition behaviours using a mobile food record and text messaging support in young adults. Int J Behav Nutr Phys Act 2016 Apr 21;13:52 [FREE Full text] [doi: 10.1186/s12966-016-0376-8] [Medline: 27098449]

76. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the 'HEYMAN' healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J 2017 Jan 13;16(1):2 [FREE Full text] [doi: 10.1186/s12937-017-0227-8] [Medline: 28086890]
77. Epton T, Norman P, Dadzie A, Harris PR, Webb TL, Sheeran P, et al. A theory-based online health behaviour intervention for new university students (U@Uni): results from a randomised controlled trial. BMC Public Health 2014;14:563 [FREE Full text] [doi: 10.1186/1471-2458-14-563] [Medline: 24903620]
78. Cameron D, Epton T, Norman P, Sheeran P, Harris PR, Webb TL, et al. A theory-based online health behaviour intervention for new university students (U@Uni:LifeGuide): results from a repeat randomized controlled trial. Trials 2015;16(1):555 [FREE Full text] [doi: 10.1186/s13063-015-1092-4] [Medline: 26643917]
79. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. Effect and process evaluation of a smartphone app to promote an active lifestyle in lower educated working young adults: cluster randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10003 [FREE Full text] [doi: 10.2196/10003] [Medline: 30143477]
80. 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]
81. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen EW. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
82. 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;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
83. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
84. Fogg B. Persuasive Technology: Using Computers to Change What We Think and Do. Burlington: Morgan Kaufmann; Dec 01, 2002:2.
85. 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]
86. 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]
87. Gill T, Boylan S. Public health messages: why are they ineffective and what can be done? Curr Obes Rep 2012;1(1):50-58 [FREE Full text] [doi: 10.1007/s13679-011-0003-6]
88. Gill T, King L, Bauman A. The University of Sydney. A "state of the knowledge" assessment of comprehensive interventions that address the drivers of obesity URL: https://pdfsecret.com/download/a-state-of-the-knowledge-assessment-of-comprehensive_5a132869d64ab24772a6c3f3_pdf [accessed 2018-03-01] [WebCite Cache ID 6xa6bVf3V]
89. Collins CE, Morgan PJ, Hutchesson MJ, Callister R. Efficacy of standard versus enhanced features in a Web-based commercial weight-loss program for obese adults, part 2: randomized controlled trial. J Med Internet Res 2013;15(7):e140 [FREE Full text] [doi: 10.2196/jmir.2626] [Medline: 23876832]
90. Van Dorsten B, Lindley EM. Cognitive and behavioral approaches in the treatment of obesity. Endocrinol Metab Clin North Am 2008 Dec;37(4):905-922. [doi: 10.1016/j.ecl.2008.08.003] [Medline: 19026939]
91. Butryn ML, Phelan S, Hill JO, Wing RR. Consistent self-monitoring of weight: a key component of successful weight loss maintenance. Obesity (Silver Spring) 2007 Dec;15(12):3091-3096. [doi: 10.1038/oby.2007.368] [Medline: 18198319]
92. Linde JA, Jeffery RW, French SA, Pronk NP, Boyle RG. Self-weighing in weight gain prevention and weight loss trials. Ann Behav Med 2005 Dec;30(3):210-216. [doi: 10.1207/s15324796abm3003_5] [Medline: 16336072]
93. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998;13(4):623-649 [FREE Full text] [doi: 10.1080/08870449808407422]
94. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
95. 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]
96. Kreuter MW, Farrell DW, Olevitch LR, Brennan LK. Tailoring Health Messages: Customizing Communication With Computer Technology. Abingdon, UK: Routledge; 2013.
97. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
98. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003 Dec;27(Suppl 3):S227-S232. [doi: 10.5993/AJHB.27.1.s3.6] [Medline: 14672383]

99. 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]
100. Dombrowski SU, Sniehotta FF, Johnston M, Broom I, Kulkarni U, Brown J, et al. Optimizing acceptability and feasibility of an evidence-based behavioral intervention for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: an open-pilot intervention study in secondary care. Patient Educ Couns 2012 Apr;87(1):108-119. [doi: 10.1016/j.pec.2011.08.003] [Medline: 21907528]
101. 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;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
102. Elfhag K, Rössner S. Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. Obes Rev 2005 Feb;6(1):67-85. [doi: 10.1111/j.1467-789X.2005.00170.x] [Medline: 15655039]
103. Leahey TM, Gokee LJ, Fava JL, Wing RR. Social influences are associated with BMI and weight loss intentions in young adults. Obesity (Silver Spring) 2011 Jun;19(6):1157-1162 [FREE Full text] [doi: 10.1038/oby.2010.301] [Medline: 21164501]
104. Verheijden MW, Bakx JC, van Weel C, Koelen MA, van Staveren WA. Role of social support in lifestyle-focused weight management interventions. Eur J Clin Nutr 2005 Aug;59(Suppl 1):S179-S186. [doi: 10.1038/sj.ejcn.1602194] [Medline: 16052189]
105. Hamel LM, Robbins LB. Computer- and web-based interventions to promote healthy eating among children and adolescents: a systematic review. J Adv Nurs 2013 Jan;69(1):16-30. [doi: 10.1111/j.1365-2648.2012.06086.x] [Medline: 22757605]
106. Laranjo L, Arguel A, Neves AL, 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. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
107. Wing RR. Cross-cutting themes in maintenance of behavior change. Health Psychol 2000 Jan;19(1S):84-88. [doi: 10.1037/0278-6133.19] [Medline: 10709952]
108. House of Lords. Parliament of the United Kingdom. Science and Technology Committee-Second Report URL: https://publications.parliament.uk/pa/ld201012/ldselect/ldsctech/179/17902.htm [accessed 2017-02-24] [WebCite Cache ID 6xa7bH6Uk]
109. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
110. Corsino L, Lin P, Batch BC, Intille S, Grambow SC, Bosworth HB, et al. Recruiting young adults into a weight loss trial: report of protocol development and recruitment results. Contemp Clin Trials 2013 Jul;35(2):1-7 [FREE Full text] [doi: 10.1016/j.cct.2013.04.002] [Medline: 23591327]
111. Partridge SR, Balestracci K, Wong AT, Hebden L, McGeechan K, Denney-Wilson E, et al. Effective strategies to recruit young adults into the TXT2BFiT mHealth randomized controlled trial for weight gain prevention. JMIR Res Protoc 2015;4(2):e66 [FREE Full text] [doi: 10.2196/resprot.4268] [Medline: 26048581]
112. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
113. Dijkers M. Icahn School of Medicine at Mount Sinai. 2013. Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development URL: http://www.mitaeroa.com/pdf/c101.pdf [accessed 2019-01-24] [WebCite Cache ID 75etIFcKT]
114. World Health Organization. 2017. Obesity and overweight URL: http://www.who.int/mediacentre/factsheets/fs311/en/ [WebCite Cache ID 6xa7tt9yD]

Abbreviations

BMI: body mass index
CCT: controlled clinical trial
CCRB: Cochrane Collaboration Risk of Bias
CHOICES: Choosing Healthy Options in College Environments and Settings
eHealth: electronic health
EPHPP: Effective Public Health Practice Project
ICT: information and communication technology
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis
RCT: randomized controlled trial
SMS: short message service
SNS: social network site

Edited by G Eysenbach; submitted 28.02.18; peer-reviewed by M Allman-Farinelli, L Ashton, L Brennan, S Partridge, J Graff; comments to author 30.07.18; revised version received 23.11.18; accepted 10.12.18; published 06.02.19.

Please cite as:
Willmott TJ, Pang B, Rundle-Thiele S, Badejo A
Weight Management in Young Adults: Systematic Review of Electronic Health Intervention Components and Outcomes
J Med Internet Res 2019;21(2):e10265
URL: http://www.jmir.org/2019/2/e10265/
doi:10.2196/10265
PMID:30724736

Original Paper

An Internet-Based Childhood Obesity Prevention Program (Time2bHealthy) for Parents of Preschool-Aged Children: Randomized Controlled Trial

Megan L Hammersley[1], PhD; Anthony D Okely[1], EdD; Marijka J Batterham[2], PhD; Rachel A Jones[1], PhD

[1]Early Start, Faculty of Social Sciences, University of Wollongong, Wollongong, Australia

[2]Statistical Consulting Service, National Institute for Applied Statistics Research Australia, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, Australia

Corresponding Author:
Megan L Hammersley, PhD
Early Start
Faculty of Social Sciences
University of Wollongong
Northfields Avenue
Wollongong, 2522
Australia
Phone: 61 242215670
Email: megan.hammersley1@uowmail.edu.au

Abstract

Background: Electronic health (eHealth) obesity programs offer benefits to traditionally delivered programs and have shown promise in improving obesity-related behaviors in children.

Objective: This study aimed to assess the efficacy of a parent-focused, internet-based healthy lifestyle program for preschool-aged children, who are overweight or at or above the fiftieth percentile for body mass index (BMI) for their age and sex, on child BMI, obesity-related behaviors, parent modeling, and parent self-efficacy.

Methods: The *Time2bHealthy* randomized controlled trial was conducted in Australia, during 2016 to 2017. Participants were recruited both online and through more traditional means within the community. Parent or carer, and child (aged 2-5 years) dyads were randomized into an intervention or comparison group. Intervention participants received an 11-week internet-based healthy lifestyle program, underpinned by social cognitive theory, followed by fortnightly emails for 3 months thereafter. Intervention participants set goals and received individual feedback from a dietitian. They were also encouraged to access and contribute to a closed Facebook group to communicate with other participants and the dietitian. Comparison participants received email communication only. Objectively measured child BMI was the primary outcome. Secondary outcomes included objectively measured physical activity, parent-measured and objectively measured sleep habits, and parent-reported dietary intake, screen time, child feeding, parent modeling, and parent self-efficacy. All data were collected at face-to-face appointments at baseline, 3 months, and 6 months by blinded data collectors. Randomization was conducted using a computerized random number generator post baseline data collection.

Results: A total of 86 dyads were recruited, with 42 randomized to the intervention group and 44 to the comparison group. Moreover, 78 dyads attended the 3- and 6-month follow-ups, with 7 lost to follow-up and 1 withdrawing. Mean child age was 3.46 years and 91% (78/86) were in the healthy weight range. Overall, 69% (29/42) of participants completed at least 5 of the 6 modules. Intention-to-treat analyses found no significant outcomes for change in BMI between groups. Compared with children in the comparison group, those in the intervention group showed a reduced frequency of discretionary food intake (estimate −1.36, 95% CI −2.27 to −0.45; *P*=.004), and parents showed improvement in child feeding pressure to eat practices (−0.30, 95% CI 0.06 to −0.00; *P*=.048) and nutrition self-efficacy (0.43, 95% CI 0.10 to 0.76; *P*=.01). No significant time by group interaction was found for other outcomes.

Conclusions: The trial demonstrated that a parent-focused eHealth childhood obesity prevention program can provide support to improve dietary-related practices and self-efficacy but was not successful in reducing BMI. The target sample size was not achieved, which would have affected statistical power.

Trial Registration: Australian New Zealand Clinical Trials Registry ANZCTR12616000119493; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=370030 (Archived by WebCite at http://www.webcitation.org/74Se4S7ZZ).

(J Med Internet Res 2019;21(2):e11964) doi:10.2196/11964

KEYWORDS

internet; eHealth; food intake; physical activity; screen time; sleep; self efficacy; body mass index

Introduction

Background

The World Health Organization (WHO) has described childhood obesity as one of the most significant public health issues [1]. Around 23% of children and adolescents in developed countries and 13% in developing countries are overweight or obese [2]. One of the main influences on the development of childhood obesity is parental guidance and role modeling around obesity-related behaviors [3-5], particularly in the early years of life up to 5 years of age [5]. Health behaviors become more difficult to change with age [5] and tend to track into adulthood [6], but are quite malleable in the early years [5]. Therefore, early childhood is an opportune time to intervene, and involving parents in interventions appears to be crucial [7].

Targeted interventions have the potential to alter the trajectory of childhood overweight and obesity continuing into adulthood, and interventions that involve parents are the most successful [8-10]. However, barriers to traditional face-to-face interventions such as scheduling of appointments [10], stigma, parental denial [11], childcare for other siblings [12], travel [13], and cost [10] can prevent sustained parental involvement and commitment and, therefore, potentially impact the success of interventions.

Overweight and obesity interventions, which use an electronic health (eHealth) delivery method, offer many advantages compared with traditional delivery methods, particularly around convenience and accessibility. Most interventions using eHealth delivery methods have been conducted in older children and have not involved parents [14]. In a recent meta-analysis of parent-focused eHealth obesity interventions for 0- to 18-year-olds, around half of the included studies showed significant improvements in the dietary intake or physical activity when compared with a control group, but there was no significant change in the body mass index (BMI)/BMI z-score. In this review, no studies targeting children aged under 5 years were included, and it was recommended that larger, higher-quality parent-focused eHealth studies be conducted, with a particular focus on younger age groups [14]. There is also a lack of studies that focus on obesity-related behaviors beyond dietary intake and physical activity. It is important that interventions focus on total movement throughout the day and incorporate strategies to improve sleep and reduce sedentary behavior, aligning with the recommendations of newly released 24-hour movement guidelines [15,16]. Furthermore, although some studies have been underpinned by social cognitive theory [17-21], few have assessed change in parent self-efficacy, a key construct of social cognitive theory.

Objectives

This paper reports the outcomes of a randomized controlled trial (RCT) evaluating the efficacy of a parent-focused, internet-based program in facilitating behavior change in preschool-aged children who are overweight or at risk of becoming overweight. We hypothesized that children in the intervention group would achieve significantly greater reductions in BMI compared with those in the comparison group at 6-month follow-up. It was also hypothesized that the intervention group would achieve significantly greater improvements in child dietary intake, physical activity, screen time, sleep, child feeding, and parent self-efficacy and role modeling.

Methods

Study Design

The protocol for this study has been published [22]. Briefly, the *Time2bHealthy* study was based on formative research with parents of preschool-aged children [23] and was piloted [24] before this trial. This study was a 2-arm parallel RCT involving parent-child dyads recruited into 6 cohorts. The trial was conducted between January 2016 and December 2017 in the Illawarra, Southern and South-Western Sydney, Southern Highlands, and Shoalhaven areas of New South Wales and Melbourne, Victoria, Australia. Measures were collected at baseline, 3 months post baseline, and 6 months post baseline. The primary outcome was change in BMI 6-months post baseline. The 6-month time point was selected as it was not expected that the 3-month time point would provide adequate time to detect changes in BMI. Secondary outcomes included child dietary intake, physical activity, screen time, sleep, child-feeding practices, and parent self-efficacy and role modeling.

The Consolidated Standards of Reporting Trials statement was used to guide the reporting of this study [25]. The study was registered with the Australian and New Zealand Clinical Trials Registry (12616000119493) and approved by the University of Wollongong Human Research Ethics Committee (HE15/354).

Participant Recruitment and Eligibility Criteria

Potential participants were informed about the study through flyers distributed at early childhood education and care centers, general practices/primary health care centers, early childhood health centers, playgroups, and local sporting groups. Flyers were also displayed on community notice boards (eg, libraries, shopping centers, children's activity centers), and articles were placed in the University of Wollongong and Local Health District newsletters and posted on Facebook. Media releases were also sent to local media outlets.

As the focus of the program was prevention of childhood obesity, healthy weight children as well as overweight children were included in the sample. Participants were eligible if they had access to the internet, if their child was 2 to 5 years old (and not yet attending school), and was at or above the WHO fiftieth percentile for BMI for their age and sex [26,27], a criterion used in other similar studies [17-19]. Parents also needed to have a Facebook account or agreed to create one.

Child participants were excluded if they were taking medications or had a medical condition with the potential to affect weight or restrict age-appropriate play. Children with conditions that required the restriction of certain foods (eg, celiac disease or food allergies) were deemed eligible to participate, but parents were informed that parts of the program would not be completely appropriate and that they would need to make some adaptations to the material provided to match their child's individual dietary/health needs.

Informed written consent was provided by the parents/guardians after reading a participant information sheet. Provisional eligibility was determined through contact with participants via phone or email and was confirmed at the face-to-face baseline data collection visit when the child's height and weight were measured to confirm if the child's BMI was at or above the WHO fiftieth percentile for age and sex. Participants below the fiftieth percentile were excluded.

Randomization and Blinding

Participants were randomized into the intervention or comparison group following the collection of baseline measures. Randomization was performed in a 1:1 ratio using a computerized random number generator. A data manager with no other involvement in the study conducted the randomization. The researcher responsible for implementing the intervention was the only person who was informed about group allocation. At the follow-up data collection time points, height and weight measurements were taken by trained data collectors blinded to group allocation.

Time2bHealthy Intervention

Participants randomized to the intervention group were provided with an individual log-in to access the *Time2bHealthy* program. The development, content, and theoretical framework for this intervention have been previously published [22]. Briefly, the intervention was guided by Bandura's social cognitive theory [28] and was designed using a backwards intervention mapping process [29,30]. The intervention targeted multiple behaviors and consisted of 6 modules including an introduction, nutrition (n=2), physical activity, screen time, and sleep module, which were completed by the participants over an 11-week period. Each module comprised reading material, videos, activities, quizzes, and a goal-setting component. Participants received feedback on their goals at the end of each module by a dietitian and were provided with advice to improve their goals using the SMART (Specific, Measurable, Attainable, Realistic, Timely) goal framework [31]. Participants also received weekly emails reminding them to log on to the website and participate in the activities. Participants were informed that they could make contact via email or phone if they had questions or concerns at any time. Participants in each of the cohorts were also encouraged to access and contribute to a closed (secret) Facebook group to communicate with other members of the cohort and the dietitian. There was a separate group for each cohort, and they were regularly monitored and moderated by the dietitian. Participants were asked to post photos, recipes, and personal experiences and ideas that they had found helpful for behavior change, which were relevant to each module. If the dietitian could not answer a question raised, advice was sought from another member of the research team, which included experts in physical activity. An incentive to post to the group was provided, with 1 post being selected from each module (2 to 6) to receive a gift card.

Participants continued to receive emails fortnightly at the end of the program until the 6-month follow-up. Infographics summarizing the key points from each of the modules were provided in these emails, and participants were also encouraged to log back into the website to revise the material and review their progress with their goals.

Comparison Condition

Participants randomized to the comparison group received fortnightly emails, which contained links to the *Raising Children Network* website (an Australian government-funded parenting website). The topics were similar to *Time2bHealthy* (nutrition, physical activity, screen time, and sleep) and also included other general health information. There were no interactive components available to this group. After the final data collection point at 6 months, participants from this group were provided access to *Time2bHealthy*, but they did not receive access to a Facebook group or to the regular emails.

Outcome Measures

Measurements were taken at baseline and 3 and 6 months post baseline. Participant measures were collected at the University of Wollongong, in the participant's home, or in a community setting. Questionnaires were completed by the parents on an iPad during these sessions, which took approximately 30 to 45 min. Demographic information was also collected from parents at the baseline data collection point. Participants in the intervention group were asked to complete a process evaluation questionnaire at the end of the Web-based program, which assessed user acceptability of the program content, length, goal setting, Facebook discussion group, and the modality used.

Primary Outcome Measure

Child height and weight were measured using a standardized method [32] to calculate BMI. A stadiometer was used to measure height to the nearest 0.1 mm. Weight was measured (with no shoes and minimal clothing) to the nearest 0.1 kg using a Seca scale. Both height and weight were measured twice. The mean of these 2 measurements was used to calculate BMI. A third measurement was taken when height measurements differed by more than 0.5 cm and weight measurements differed by more than 0.5 kg.

Secondary Outcome Measures

Dietary intake was assessed using both a parent-reported food questionnaire (modified from the Eating and Physical Activity

Questionnaire) [33] and a parent-reported 24-hour recall of child dietary intake (using the "Easy Diet Diary" app [Xyris Software, Australia, Pty Ltd]). The section of the food questionnaire, which asked about the frequency of intake of discretionary foods, was expanded to include additional discretionary food categories, which used the same scale as the existing question. Cronbach alpha=.68 for these discretionary food questions. Data from the 24-hour recall was used to calculate kJ per kg of body weight, percentage of kJ from sugar, and percentage of kJ from saturated fat. Data from the food questionnaire were used to assess the daily fruit intake, daily vegetable intake, and frequency of fruit juice and sugary drinks intake. A discretionary food score was calculated based on responses to questions on the frequency of intake of takeaway or fast food; sugary cereals; potato chips or other salty snacks; sweets; cakes, doughnuts, and sweet cookies, or muffins.

Physical activity intensity and duration were measured using an ActiGraph GT3X+ accelerometer (ActiGraph Corporation, Pensacola, FL), which was worn on an elasticized belt around the child's waist for 7 days. Accelerometer data were analyzed in ActiLife version 6 (ActiGraph Corporation, Pensacola, FL). A sampling frequency of 30 Hz was used, with the files then reintegrated into 15-second epochs. Nonwear time was defined as 20 min or more of 0 counts. Accelerometer data used for the physical activity analysis were considered valid based on wear time of at least 6 hours per day on 3 days, which has been found to be reliable in previous research [34]. The following cut points appropriate for preschool-aged children were used to categorize physical activity intensity: sedentary, <100 counts/min; low light–intensity physical activity, 101 to 800 counts/min; high light–intensity physical activity, 801 to 1679 counts/min; moderate-intensity physical activity, 1680 to 3367 count/min; and vigorous-intensity physical activity, ≥3368 count/min [35].

Sleep habits were assessed using 4 questions assessing sleep latency, sleep reluctance, difficulty sleeping, and difficulty falling to sleep in own bed based on questions from the Children's Sleep Habits Questionnaire [36] (Cronbach alpha=.63 for the 3 scaled questions relating to sleep reluctance, difficulty falling asleep, and difficulty falling to sleep in own bed) and questions about the child's usual sleep and wake times and an Actigraph GT3X+ accelerometer. Sleep accelerometer data were analyzed in ActiLife using the Sadeh algorithm, which is appropriate for use in children [37]. Sleep accelerometer data were considered valid based on a wear time of at least 3 nights [38].

Parent-reported questionnaires were used to assess child feeding (from the Child Feeding Questionnaire predefined subscales of "restriction" and "pressure to eat" [39]), screen time (based on the studies by Downing et al and Hinkley et al [40,41] and additional questions relating to screen entertainment rules, presence of a television in the child's bedroom and frequency of watching television while eating a meal), parent modeling (developed after reviewing the studies by Palfreyman et al and Gattshall et al [42,43]; Cronbach alpha=.63), and parent self-efficacy in nutrition, physical activity, screen time, and sleep (modified from Bohman et al [44] by adding 6 additional questions and making small changes to some existing questions to align the questionnaire to the program content; Cronbach alpha=.89).

Power and Sample Size

On the basis of the results of the pilot study [24], we expected a BMI effect size of approximately 0.4 for this trial. To detect a statistically significant difference between groups (alpha=.05 and power=.8), 136 participants were required (68 per group), and based on an estimated attrition rate of 15%, we aimed to recruit 160 participants (80 per group).

Statistical Analyses

Differences in changes over time between the intervention and comparison groups were assessed for each outcome. Linear mixed models were used to determine differences between groups over time (baseline, 3 months, and 6 months) with adjustment for potential covariates. Intention-to-treat (ITT) principles were used for parametric data, with all participants analyzed in the group to which they were randomized regardless of whether they attended all data collection time points or completed the intervention. Covariates included baseline values, age, and cohort. Due to nonparametric distributions for some variables, Freidman tests and Wilcoxon signed rank tests were used followed by Mann-Whitney tests to analyze nonparametric data using completed cases. Generalized estimating equations were considered; however, the analyses would not converge.

Post hoc analysis of covariance (ANCOVA) analyses were used to detect changes between groups at individual time points, which included the baseline value, age, and cohort as covariates. Within-group changes were analyzed using repeated measures analysis of variance (ANOVA), which included age and cohort as covariates. These were complete case analyses. Analyses were performed using IBM SPSS Statistics for Windows, version 25 (IBM Corp, Armonk, NY, USA).

Results

Overview

Figure 1 shows the flow of participants through the study. Recruitment was conducted between January 2016 and June 2017. Enquiries were received from 372 parents initially. After viewing the information sheet, 159 parents remained interested in the study and were screened via phone or email, with 104 being potentially eligible. Of the 93 parent-child dyads who attended the initial visit, 86 were eligible and enrolled in the study. A total of 42 participant dyads were randomized to the intervention group and 44 to the comparison group. The mean number of participants per cohort was 14 (range 8-22), and the mean number of participants in each Facebook group was 6 (range 3-10). Follow-up was conducted between July 2016 and December 2017. Moreover, 78 participants (91%) attended the 3- and 6-month follow-ups, with 7 (8%) lost to follow-up and 1 participant (1%) withdrawing from the intervention group due to problems accessing the internet. Figure 2 shows the completion of each of the intervention program modules. At least 5 of the 6 modules were completed by 29 participants (69%).

Figure 1. CONSORT flow diagram for *Time2bHealthy* randomized controlled trial.

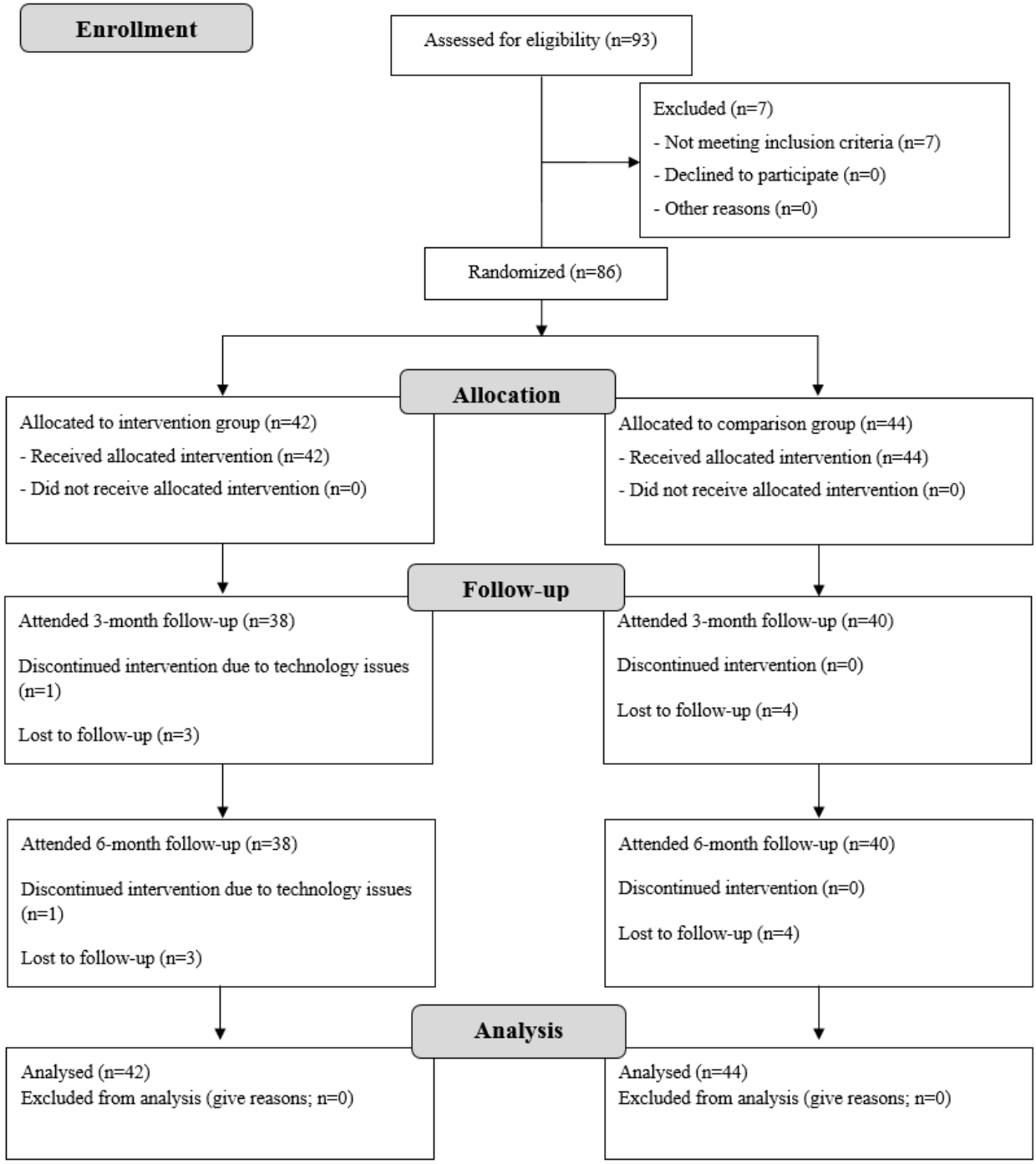
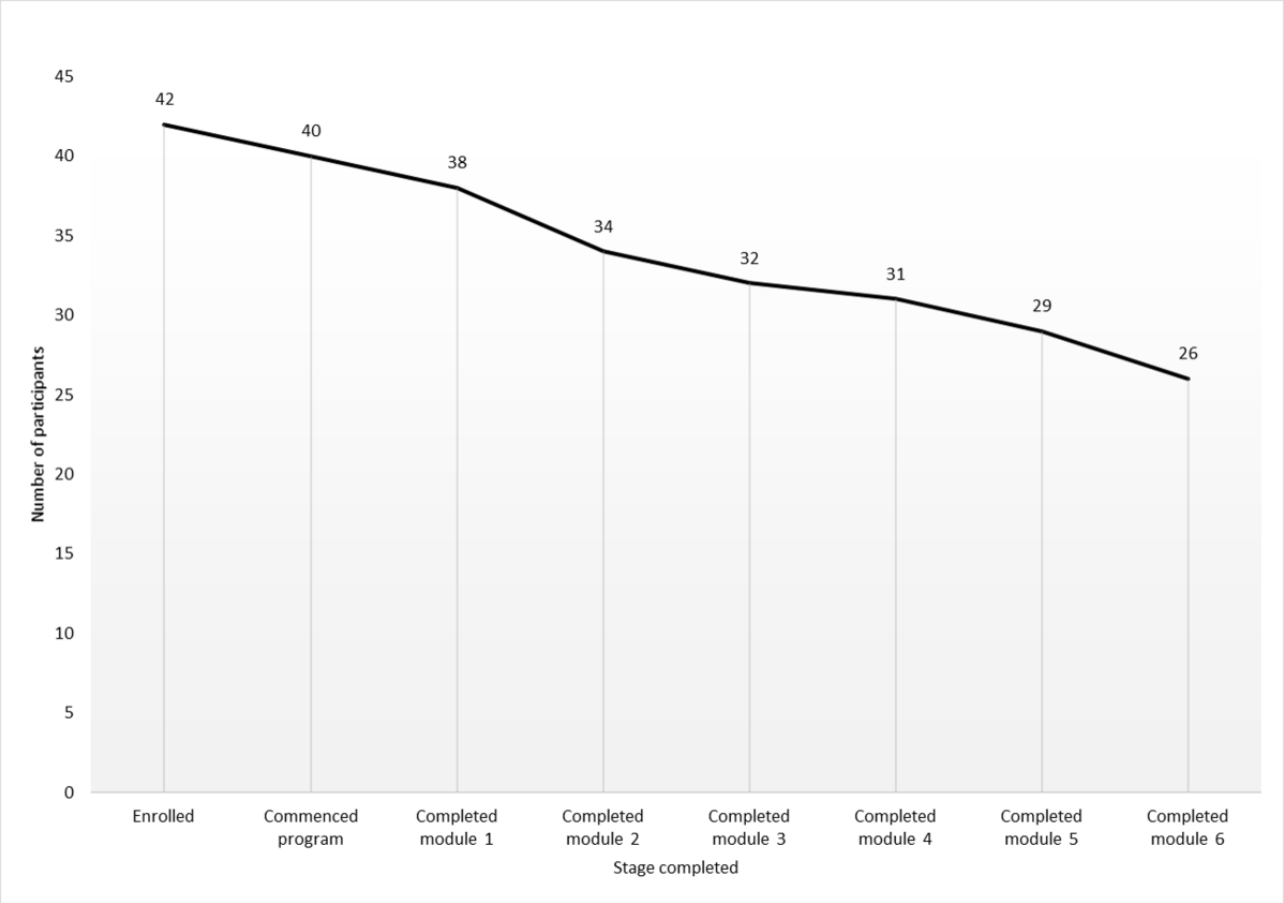


Figure 2. Participant completion of *Time2bHealthy* modules.



Participant Characteristics

The baseline characteristics of participants are displayed in Table 1. The mean age (SD) of the participating children was 3.46 (0.92) years, and 50% of child participants were female. The mean age (SD) of the participating parents was 35.17 (4.80) years and 97% (83/86) were female, 63% (54/86) had a university degree, 50% (43/86) had an after-tax income of at least Aus $580/week, and 85% (73/86) were married or had a partner. The majority of children were in the healthy weight range (78/86, 91%) according to the WHO criteria [1]. The mean (SD) BMI of the participating children was 17.01 (1.24). The mean (SD) BMI of participating parents was 26.08 (5.97), and 45% (39/86) were overweight or obese.

Primary Outcome

Table 2 displays the baseline, 3-month, and 6-month BMI results. The results of the ITT, displayed in Table 3, indicated that there was no group-by-time effect for BMI. The ANCOVA analyses (shown in Tables 4 and 5) also found no significant differences between groups at each time point. When considering changes within groups, the repeated-measures ANOVA found a significant change in BMI within the intervention group at both the 3-month (adjusted mean difference −.26, 95% CI −0.51 to −0.02; *P*=.03) and 6-month time points (adjusted mean difference −0.22, 95% CI −0.40 to −0.03; *P*=.02) and no significant changes within the comparison group.

Secondary Outcomes

Table 2 displays the baseline, 3-month, and 6-month results for parametric secondary outcomes. The linear mixed model analyses (displayed in Table 3) found a significant group-by-time interaction for frequency of consumption of discretionary foods (estimate −1.36, 95% CI −2.27 to −0.45; *P* ≤.01), nutrition parent self-efficacy (estimate 0.43, 95% CI 0.10 to 0.76, *P*=.01), and child feeding–pressure to eat (estimate −0.30, 95% CI 0.61 to −0.00, *P*=.048). No group-by-time interaction effects for any other secondary outcomes were observed.

The posthoc ANCOVA analyses results (displayed in Table 4 and 5) showed a significant difference between groups in frequency of consumption of discretionary foods at 3 months (adjusted mean difference −1.45, 95% CI −2.42 to −0.43; *P*=.01) and 6 months (adjusted mean difference −1.30, 95% CI −2.34 to −0.26; *P*=.02), nutrition parent self-efficacy at 6 months (adjusted mean difference 0.53, 95% CI 0.13 to 0.93; *P*=.01), child feeding–pressure to eat at 6 months (adjusted mean difference −0.35, 95% CI −0.68 to −0.02; *P*=.04), and accelerometer-measured sleep duration (in the nonhypothesized direction) at 6 months (adjusted mean difference −0.55, 95% CI −1.01 to −0.03; *P*=.04). The results of the Mann-Whitney tests for the nonparametric data showed that there were no significant differences between groups for any parameter (at Bonferroni adjusted *P*<.008).

Table 1. Baseline characteristics of participants.

Variable	Comparison group (n=44)	Intervention group (n=42)	All (n=86)
Child (sex), n (%)			
Male	19 (43)	24 (57)	43 (50)
Female	25 (57)	18 (43)	43 (50)
Child age (months), mean (SD)	43 (12.26)	40 (9.65)	42 (11.05)
Child age (years), mean (SD)	3.55 (1.02)	3.36 (0.80)	3.46 (0.92)
Child body mass index (BMI), mean (SD)	16.72 (0.92)	17.28 (1.44)	17.01 (1.24)
Child weight status[a], n (%)			
Healthy weight	38 (86)	40 (95)	78 (91)
Overweight	5 (11)	2 (5)	7 (8)
Obese	1 (2)	0 (0)	1 (1)
Median BMI percentile range	85 to ≤95	75 to ≤85	75 to ≤85
Child aboriginal/Torres Strait Islander status, n (%)			
Aboriginal	4 (9)	1 (2)	5 (6)
Torres Strait Islander	0 (0)	0 (0)	0 (0)
No	39 (89)	40 (95)	79 (92)
Not answered	1 (2)	1 (2)	2 (2)
Participating parent's sex, n (%)			
Male	1 (2)	2 (5)	3 (3)
Female	43 (98)	40 (95)	83 (97)
Participating parent's age, mean (SD)	34.91 (4.68)	35.45 (4.95)	35.17 (4.80)
Highest level of education of the participating parent, n (%)			
Not university qualified	22 (50)	8 (19)	30 (35)
University qualified	22 (50)	32 (76)	54 (63)
Currently studying	0 (0)	2 (5)	2 (2)
Participating parent's income after tax (Aus $), n (%)			
<$580/week	23 (52)	20 (48)	43 (50)
$580-$1240/week	15 (34)	16 (38)	31 (36)
>$1240/week	6 (14)	6 (14)	12 (14)
BMI of participating parent, mean (SD)	27.38 (21.61)	24.81 (4.64)	26.08 (5.97)
Weight status of participating parent, n (%)			
Underweight	1 (2)	1 (2)	2 (2)
Healthy weight	15 (34)	26 (62)	42 (49)
Overweight	13 (30)	9 (21)	22 (26)
Obese	11 (25)	6 (14)	17 (20)
Not answered	3 (7)	0 (0)	3 (3)
Aboriginal status of participating parent, n (%)			
Aboriginal	2 (4.55)	1 (2.38)	3 (3.49)
No	41 (88.64)	40 (95.24)	81 (94.19)
Not answered	1 (2.27)	1 (2.38)	2 (2.33)
Participating parent's relationship with child, n (%)			
Biological mother	41 (93)	39 (93)	80 (93)
Biological father	2 (5)	2 (5)	4 (5)

Variable	Comparison group (n=44)	Intervention group (n=42)	All (n=86)
Other	1 (2)	1 (2)	2 (2)
Marital status of participating parent, n (%)			
Single/divorced/separated/widowed	10 (23)	3 (7)	13 (15)
Married/with partner	34 (77)	39 (93)	73 (85)
BMI of other parent, mean (SD)	27.61 (4.51)	28.24 (6.72)	27.95 (5.76)
Weight status of other parent, n (%)			
Underweight	0 (0)	0 (0)	0 (0)
Healthy weight	9 (20)	15 (36)	24 (28)
Overweight	13 (30)	9 (21)	22 (26)
Obese	9 (20)	11 (26)	20 (23)
No answer/not applicable	13 (30)	7 (17)	19 (22)
Income of other parent (Aus $), n (%)			
<$580/week	6 (14)	5 (12)	11 (13)
$580-$1240/week	20 (45)	19 (45)	39 (45)
>$1240/week	9 (20)	15 (36)	24 (28)
No answer/not applicable	9 (20)	3 (7)	12 (14)
Language spoken at home, n (%)			
English	40 (91)	37 (88)	77 (90)
Other	4 (9)	5 (12)	9 (10)
Found out about the program, n (%)			
Early childhood education center	18 (41)	16 (38)	34 (40)
Flyer	5 (11)	7 (17)	12 (14)
Early childhood nurse/center	2 (5)	5 (12)	7 (8)
Email	0 (0)	4 (10)	4 (5)
School newsletter	2 (5)	1 (2)	3 (3)
Media (print, television, and radio)	2 (5)	1 (2)	3 (3)
Social media	5 (11)	4 (10)	9 (10)
Playgroup	3 (7)	0 (0)	3 (3)
Other	7 (16)	4 (10)	11 (13)

[a]World Health Organization definition [1].

Table 2. Mean (SD) values for primary and secondary outcomes at each time point.

Variable	Baseline, mean (SD)		3 months, mean (SD)		6 months, mean (SD)	
	Comparison (n=44)	Intervention (n=42)	Comparison (n=40)	Intervention (n=38)	Comparison (n=40)	Intervention (n=38)
Body mass index (BMI)	17.28 (1.44)	16.72 (0.92)	16.99 (1.25)	16.46 (0.80)	16.87 (1.24)	16.51 (0.75)
Median BMI percentile range	85 to ≤95	75 to ≤85	85 to ≤95	75 to ≤85	85 to ≤95	75 to ≤85
kJ/kg of body weight[a]	330.43 (125.08)	343.64 (112.01)	296.24 (114.64)	303.75 (120.15)	296.20 (82.05)	327.60 (104.06)[b]
Percentage of kJ from sugar[a]	22.24 (6.75	20.14 (7.01)	21.15 (7.30)	20.83 (6.02)	19.29 (7.01)	19.54 (6.95)[b]
Percentage of kJ from saturated fat[a]	12.52 (4.77)	11.74 (3.95)	11.58 (3.84)	11.37 (3.91)	12.50 (3.74)	11.00 (3.90)[b]
Servings of fruit[c]	2.91 (1.03)	2.52 (0.92)	2.95 (0.96)	2.47 (0.83)	2.88 (1.04)	2.53 (0.86)
Servings of vegetables[c]	2.34 (1.08)	2.62 (1.27)	2.53 (1.22)	2.84 (1.22)	2.65 (1.05)	2.97 (1.28)
Discretionary food frequency score[d]	11.73 (2.86)	11.21 (3.82)	11.60 (2.73)	9.82 (3.21)	11.90 (2.29)	10.40 (3.22)
Nutrition self-efficacy[e]	7.94 (1.13)	8.19 (1.36)	8.28 (1.19)	8.69 (0.97)	8.30 (1.22)	8.89 (0.89)
Child feeding–restriction[f]	3.60 (0.92)	3.630 (0.78)	3.73 (0.84)	3.69 (0.75)	3.58 (0.89)	3.66 (0.79)
Child feeding–pressure[f]	2.34 (0.98)	2.52 (0.99)	2.34 (1.09)	2.17 (1.08)	2.43 (1.04)	2.14 (0.99)
Parent modeling[g]	3.95 (0.76)	3.98 (0.79)	3.93 (0.85)	4.18 (0.55)	4.16 (0.73)	4.36 (0.54)
Sleep reluctance[h]	3.00 (1.24)	2.36 (1.06)	2.65 (1.00)	2.13 (0.99)	2.68 (0.97)	2.24 (1.14)[b]
Sleep	**n=34**	**n=34**	**n=19**	**n=28**	**n=20**	**n=21**
Duration (hours)[i]	9.59 (0.93)	9.85 (0.78)	9.74 (0.72)	9.91 (0.62)	9.78 (0.96)	9.54 (0.64)
Latency (minutes)[i]	19.92 (16.55)	20.98 (14.41)	19.97 (18.05)	16.44 (11.91)	22.19 (11.85)	25.00 (18.03)
Screen time	**n=44**	**n=41**	**n=40**	**n=38**	**n=40**	**n=38**
Week day (hours)[j]	2.52 (2.55)	2.82 (3.87)	1.37 (1.06)	1.73 (2.47)	2.20 (2.91)	1.26 (0.99)
Weekend day (hours)[j]	2.94 (1.98)	3.15 (2.95)	2.31 (1.56)	1.84 (1.43)	2.68 (2.33)	2.04 (1.39)
Percentage activity	**n=34**	**n=35**	**n=27**	**n=31**	**n=26**	**n=27**
Sedentary time[i]	46.28 (7.98)	47.44 (11.09)	48.28 (7.87)	49.17 (4.03)	46.45 (6.21)	49.47 (5.56)
Light, moderate, and vigorous physical activity[i]	27.74 (7.40)	25.82 (6.24)	26.18 (6.16)	25.61 (4.38)	27.73 (5.42)	25.44 (4.93)
Moderate-to-vigorous physical activity[i]	13.88 (5.04)	12.02 (3.60)	13.56 (4.43)	12.91 (3.70)	14.38 (4.11)	13.01 (3.77)

[a]Calculated from 24-hour diet recall using Easy Diet Diary/Foodworks.

[b]n=37.

[c]From food questionnaire.

[d]Scored from food questionnaire questions on frequency of intake of takeaway or fast food; sugary cereals; potato chips or other salty foods; sweets; and cakes, doughnuts, sweet cookies, or muffins. Responses of never or rarely, 1 to 3 times per month, 1 to 2 times per week, 3 to 4 times per week, 5 to 6 times per week, once per day, and 2 or more times per day were coded as 1 to 6, respectively, and summed to obtain a discretionary food score.

[e]Self-efficacy questionnaire.

[f]Child-feeding questionnaire.

[g]Parent modeling questionnaire.

[h]From sleep questionnaire.

[i]Accelerometer measures.

[j]From screen time questionnaire.

Table 3. Results of intention-to-treat analyses for primary and secondary outcomes; linear mixed model group × time interaction (random intercept and compound symmetry covariance structure). Age, cohort, and baseline values included as covariates in the model (n=86).

Variable	Estimate	95% CI	*P* value[a]
Body mass index	−0.11	−0.34 to 0.12	.35
kJ/kg of body weight[b]	10.89	−29.94 to 51.73	.60
Percentage of kJ from sugar[b]	−0.09	−2.44 to 2.25	.94
Percentage of kJ from saturated fat[b]	−0.61	−3.09 to 1.87	.63
Servings of fruit[c]	−0.24	−0.58 to 0.10	.17
Servings of vegetables[c]	0.17	−0.15 to 0.49	.24
Discretionary food frequency score[d]	−1.36	−2.27 to −0.45	<.01
Nutrition self-efficacy[e]	0.43	0.10 to 0.76	*.01*[f]
Child feeding–restriction[g]	0.04	−0.21 to 0.29	.76
Child feeding–pressure[g]	−0.30	−0.61 to −0.00	*.048*
Parent modeling[h]	0.21	−0.02 to 0.44	.08
Sleep duration (hours)[i]	−0.22	−0.57 to 0.13	.21
Sleep latency (minutes)[i]	−0.25	−0.79 to 0.74	.95
Sleep reluctance[j]	−0.36	−0.77 to 0.06	.09
Screen time–week day (hours)[k]	−0.20	−0.87 to 0.47	.56
Screen time–weekend day (hours)[k]	−0.40	−0.90 to 0.10	.11
Percentage sedentary time[i]	0.84	−1.60 to −3.27	.49
Percentage light, moderate, and vigorous intensity physical activity[i,k]	−0.99	−2.20 to 2.01	.93
Percentage moderate-to-vigorous intensity physical activity[i]	0.54	−0.94 to 2.01	.47

[a]Significant at $P<.05$.

[b]Calculated from 24-hour diet recall using Easy Diet Diary/Foodworks.

[c]From Food Questionnaire.

[d]Scored from food questionnaire questions on the frequency of intake of takeaway or fast food; sugary cereals; potato chips or other salty foods; sweets; and cakes, doughnuts, sweet cookies, or muffins. Responses of never or rarely, 1 to 3 times per month, 1 to 2 times per week, 3 to 4 times per week, 5 to 6 times per week, once per day, and 2 or more times per day were coded as 1 to 6, respectively, and summed to obtain a discretionary food score.

[e]Self-efficacy questionnaire.

[f]Italicized text: statistically significant result.

[g]Child feeding questionnaire.

[h]Parent modeling questionnaire.

[i]Accelerometer measures.

[j]From sleep questionnaire.

[k]From screen time questionnaire.

Table 4. Adjusted mean differences (and 95% CI) for primary and secondary outcomes at 3 months (complete case analyses). Analysis of covariance (ANCOVA) analyses, with baseline value, age, and cohort as covariates (n=78).

Variable	Intervention comparison, adjusted mean difference (95% CI)	*P* value[a]
Body mass index	−0.23 (−0.50 to 0.04)	.09
kJ/kg of body weight[b]	−0.57 (−57.71 to 46.26)	.83
Percentage of kJ from sugar[b]	−0.23 (−3.29 to 2.83)	.88
Percentage of kJ from saturated fat[b]	−0.15 (−1.94 to 1.63)	.87
Servings of fruit[c]	−0.31 (−0.69 to 0.07)	.11
Servings of vegetables[c]	0.19 (−0.23 to 0.60)	.37
Frequency discretionary foods[d]	−1.45 (−2.47 to −0.43)	.01
Nutrition self-efficacy[e]	0.33 (−0.03 to 0.69)	.07
Child feeding–restriction[f]	0.01 (−0.28 to 0.29)	.96
Child feeding–pressure[f]	−0.27 (−0.61 to 0.07)	.12
Parent modeling[g]	0.24 (0.06 to 0.53)	.12
Sleep duration[h]	0.04 (−0.35 to 0.43)	.84
Sleep latency[h]	−4.46 (−13.91 to 4.98)	.35
Sleep reluctance[i]	−0.36 (−0.82 to 0.09)	.11
Screen time–weekday[j]	0.45 (−0.36 to 1.27)	.27
Screen time–weekend[j]	−0.30 (−0.86 to 0.26)	.29
Percentage sedentary time[h]	0.14 (−2.76 to 3.04)	.92
Percentage light, moderate, and vigorous intensity physical activity[h]	0.92 (−1.60 to 3.44)	.47
Percentage moderate-to-vigorous intensity physical activity[h]	1.10 (−0.65 to 2.84)	.21

[a]Significant at $P<.05$.

[b]Calculated from 24-hour diet recall using Easy Diet Diary/Foodworks.

[c]From food questionnaire.

[d]Scored from food questionnaire questions on frequency of intake of takeaway or fast food; sugary cereals; potato chips or other salty foods; sweets; and cakes, doughnuts, sweet cookies, or muffins. Responses of never or rarely, 1 to 3 times per month, 1 to 2 times per week, 3 to 4 times per week, 5 to 6 times per week, once per day, and 2 or more times per day were coded as 1-6, respectively, and summed to obtain a discretionary food score.

[e]Self-efficacy questionnaire.

[f]Child feeding questionnaire.

[g]Parent modeling questionnaire.

[h]Accelerometer measures.

[i]From sleep questionnaire.

[j]From screen time questionnaire.

Table 5. Adjusted mean differences (and 95% CI) for primary and secondary outcomes at 6 months (complete case analyses). Analysis of covariance (ANCOVA) analyses, with baseline value, age, and cohort as covariates (n=78).

Variable	Intervention comparison, adjusted mean difference (95% CI)	*P* value[a]
Body mass index	0.01 (−0.27 to −0.29)	.95
kJ/kg of body weight[b]	24.80 (−17.75 to 67.35)	.25
Percentage of kJ from sugar[b]	0.05 (−3.18 to 3.29)	.97
Percentage of kJ from saturated fat[b]	−1.41 (−3.19 to 0.37)	.12
Servings of fruit[c]	−0.17 (−0.57 to 0.23)	.39
Servings of vegetables[c]	0.16 (−0.24 to 0.56)	.44
Frequency discretionary foods[d]	−1.30 (−2.34 to −0.26)	.02
Nutrition self-efficacy[e]	0.53 (0.13 to 0.93)	.01
Child feeding–restriction[f]	0.10 (−0.18 to 0.37)	.48
Child feeding–pressure[f]	−0.35 (−0.68 to −0.02)	.04
Parent modeling[g]	0.18 (−0.05 to 0.41)	.12
Sleep duration[h]	−0.55 (−1.01 to −0.03)	.04
Sleep latency[h]	6.00 (−4.09 to 16.09)	.24
Sleep reluctance[i]	−0.33 (−0.82 to 0.15)	.18
Screen time–weekday[j]	−0.84 (−1.76 to 0.07)	.07
Screen time–weekend[j]	−0.49 (−1.14 to 0.15)	.13
Percentage sedentary time[k]	1.590 (−1.415 to 4.60)	.29
Percentage light, moderate, and vigorous intensity physical activity[h]	−1.106 (−3.601 to 1.40)	.38
Percentage moderate-to-vigorous intensity physical activity[h]	−1.110 (−1.912 to 1.69)	.90

[a]Significant at $P<.05$.

[b]Calculated from 24-hour diet recall using Easy Diet Diary/Foodworks.

[c]From food questionnaire.

[d]Scored from food questionnaire questions on frequency of intake of takeaway or fast food; sugary cereals; potato chips or other salty foods; sweets; and cakes, doughnuts, sweet cookies, or muffins. Responses of never or rarely, 1 to 3 times per month, 1 to 2 times per week, 3 to 4 times per week, 5 to 6 times per week, once per day, and 2 or more times per day were coded as 1-6, respectively, and summed to obtain a discretionary food score.

[e]Self-efficacy questionnaire.

[f]Child feeding questionnaire.

[g]Parent modeling questionnaire.

[h]Accelerometer measures.

[i]From sleep questionnaire.

[j]From screen time questionnaire.

Table 6. *Time2bHealthy* intervention process evaluation (n=38).

Question	Strongly agree, n (%)	Agree, n (%)	Neutral, n (%)	Disagree, n (%)	Strongly disagree, n (%)	Not applicable, n (%)
The program content was interesting	21 (55)	15 (39)	2 (5)	0 (0)	0 (0)	0 (0)
The program content was easy to understand	28 (74)	10 (26)	0 (0)	0 (0)	0 (0)	0 (0)
The program content was relevant	22 (58)	15 (39)	1 (3)	0 (0)	0 (0)	0 (0)
The length of the program was appropriate	15 (39)	18 (47)	2 (5)	3 (8)	0 (0)	0 (0)
One module every 2 weeks was appropriate	11 (29)	23 (61)	4 (11)	0 (0)	0 (0)	0 (0)
The tips and tricks for parents was helpful	20 (53)	17 (45)	1 (3)	0 (0)	0 (0)	0 (0)
The information about meals was helpful	22 (58)	13 (34)	2 (5)	1 (3)	0 (0)	0 (0)
There was enough information in the module about meals	14 (37)	20 (53)	0 (0)	4 (11)	0 (0)	0 (0)
The information on snacks and drinks was helpful	21 (55)	15 (39)	1 (3)	1 (3)	0 (0)	0 (0)
There was enough information in the module about snacks and drinks	15 (39)	19 (50)	2 (5)	2 (5)	0 (0)	0 (0)
The information about physical activity was helpful	20 (53)	16 (42)	2 (5)	0 (0)	0 (0)	0 (0)
There was enough information in the module about physical activity	19 (50)	17 (45)	2 (5)	0 (0)	0 (0)	0 (0)
The information on screen time was helpful	20 (53)	13 (34)	4 (11)	1 (3)	0 (0)	0 (0)
There was enough information in the module about screen time	14 (37)	20 (53)	4 (11)	0 (0)	0 (0)	0 (0)
The information about sleep was helpful	11 (29)	19 (50)	6 (16)	0 (0)	1 (3)	1 (3)
There was enough information about sleep	8 (21)	24 (63)	4 (11)	0 (0)	1 (3)	1 (3)
The goal setting was helpful	12 (32)	18 (47)	7 (18)	1 (3)	0 (0)	0 (0)
The number of goals set was appropriate	12 (32)	18 (47)	7 (18)	0 (0)	1 (3)	0 (0)
The health consultants were helpful and knowledgeable	20 (53)	15 (39)	1 (3)	1 (3)	0 (0)	1 (3)
The time the health consultants responded in was appropriate	21 (55)	16 (42)	0 (0)	0 (0)	0 (0)	1 (3)
The online delivery mode was suitable	19 (50)	18 (47)	1 (3)	0 (0)	0 (0)	0 (0)
The Facebook group component was useful	3 (8)	12 (32)	18 (47)	4 (11)	1 (3)	0 (0)

Process Evaluation

Overall, 38 participants from the intervention group (38/42, 90%) completed the process evaluation questionnaire. The results are displayed in Table 6. Most participants agreed or strongly agreed that the program content was interesting (36/38, 95%), easy to understand (38/38, 100%), and relevant (37/38, 97%). Most also agreed or strongly agreed that the length of the program was appropriate (33/38, 87%), the goal-setting component was helpful (30/38, 79%), and that the dietitian was helpful and knowledgeable (35/38, 92%). Most participants discussed the program with extended family members (28/38, 74%). The internet-based delivery mode of the program was suitable for the majority of participants (37/38, 97%); however, 6 participants stated that they would have preferred a different mode of delivery such as a mobile-optimized website (2) mobile phone app (2), face-to-face (2), or hard copy (2). Only 15 participants (15/38, 39%) agreed or strongly agreed that the Facebook component was useful.

Discussion

Principal Findings

In this RCT, we found no significant difference in the BMI change between the 2 groups at 6 months post baseline. There were no significant differences in physical activity, screen time, or sleep outcomes between groups. The intervention did, however, demonstrate some positive group-by-time outcomes in relation to dietary intake, child feeding, and nutrition parent self-efficacy. To the best of our knowledge, *Time2bHealthy* is the first RCT to assess the efficacy of a parent-focused healthy lifestyle intervention on BMI in preschool-aged children, which is delivered entirely Web-based.

Our null finding regarding BMI change at 6 months aligns with similar eHealth obesity prevention studies conducted in young [45] and older children [17,21,46] and a recent mobile health study in preschool-aged children that measured fat mass index [47]. Due to a lack of eHealth studies in this age group, we have also compared our findings with studies delivered by more traditional methods. Mixed results have been reported from traditionally delivered parent-focused obesity prevention studies in young children, with a recent meta-analysis finding a short-term, but not a long-term, effect [48]. This meta-analysis

also found that interventions targeting only overweight and obese children were more effective than those that included children in the healthy weight range [48]. Given that more than 90% (78/86) of children recruited in our study were in the healthy weight range, significant changes may have been unrealistic. Superior outcomes may have been achieved had our study included only overweight and obese children. Healthy weight children were included in this study as prevention is key to impacting childhood obesity rates, and it is critical to design interventions that facilitate establishment of healthy behaviors and maintenance of healthy weight in all children at an early age [49]. There was a significant within-group difference in BMI in the intervention group. Had the target sample size been achieved, it is possible that a difference between groups would have been found.

Other eHealth parent-focused studies have demonstrated similar improvements in dietary outcomes, such as energy dense food consumption [50,51]. The discretionary food group-by-time outcomes in this study most closely align with Williamson et al's [52] internet-based study targeting adolescent overweight girls, which demonstrated a reduction in "eating fattening foods." Contrary to this study, previous eHealth studies have also shown improvements in fruit and vegetable intake, including Chen et al's internet-based study on adolescents [21] and Knowlden and Conrad's internet-based study for mothers of 4- to 6-year-old children [53]. Reduction in sugar-sweetened beverage intake was also reported in an internet-based parent-focused study for children aged 18 to 24 months [54]. Some traditionally delivered parent-focused interventions in preschool-aged children have also demonstrated improvements in fruit and vegetable consumption [55] and reductions in mean energy intake [56].

Our null findings in regard to kJ/kg body weight and kJ from sugar and saturated fat were perhaps due to the fact that (due to resource constraints) the 24-hour recall was administered on 1 single weekday at each time point and was not sufficient to capture regular and weekend consumption patterns. It is also possible that the intervention effects on each of the obesity-related behaviors could have been diluted due to the multi-behavior focus and breadth of the content covered compared with previous studies that have focused on fewer behaviors.

Similar eHealth parent-focused studies in a range of age groups have shown mixed physical activity outcomes [21,46,47,52,57,58]. One successful internet-based study of adolescents used pedometers to self-monitor activity [21], which may have enhanced motivation. Few traditionally delivered parent-focused studies have demonstrated an improvement in physical activity [59]. Accelerometry compliance was not optimal in our study (n=53 to 68), and therefore, the results may not be indicative of the whole sample. Night-time accelerometry compliance was even lower (n=41 to 68). To the best of our knowledge, no similar eHealth studies have assessed sleep outcomes; however, a traditionally delivered program found a significant increase in parent-reported sleep duration [60]. Further studies are needed, which objectively measure sleep duration and explore strategies to improve night-time accelerometry compliance, such as the use of wrist-worn monitors [61], incentives, or phone calls/email reminders [62]. Screen time behavior has also not been a focus of many parent-focused childhood obesity studies. One eHealth study in young children [45] and 2 in older children found null screen time outcomes [20,46], which align with our findings. Similar to our study, Knowlden et al [53] found improvements in both groups and an improvement in screen time parent self-efficacy in the intervention group; perhaps, a minimal intervention can effect change in this area.

There was a significant group-by-time interaction for nutrition parent self-efficacy but no significant differences between the intervention and comparison groups for parent self-efficacy in relation to physical activity, screen time, or sleep. The reason why positive outcomes were achieved for nutrition parent self-efficacy and not for the other behaviors is unclear, but it may be due to the higher proportion of program time dedicated to healthy eating and nutrition (2 modules compared with only 1 module for the other behaviors) and the larger number of videos, providing a greater opportunity for vicarious learning.

It is established that parent self-efficacy is crucial for implementing obesity-related behavior change in children [44]. Positive relationships have been reported between high parental (or maternal) self-efficacy and fruit and vegetable intake [63-66] and moderate-to-vigorous intensity physical activity [65], and an inverse relationship has been reported with consumption of unhealthy food [63,65]. Although social cognitive theory has been used as a basis for other similar studies [17,20], parent self-efficacy has rarely been assessed, despite this being key in parent-focused interventions.

There was a significant group-by-time interaction for "pressure-to-eat" child feeding practices, but there was no significant difference between groups for "restriction" of child-feeding practices. Despite the body of evidence regarding child feeding practices and risk of overweight and obesity, there are limited studies that have used child feeding as an outcome measure. No other eHealth study to the best of our knowledge has assessed child feeding practices; hence, the outcomes of this study will be compared with traditionally delivered programs in preschool-aged children. Similar to this study, a significant improvement in "pressure-to-eat" child feeding practices was reported in a group that received a regular newsletter (compared with a group that received a single booklet), but no significant changes in other child-feeding practices were reported in a study of mothers of African American preschool-aged children [67]. Conversely, Harvey-Berino et al [50] found a significant reduction in "restriction" child-feeding practices, but not for other child-feeding practices in their childhood obesity prevention study in Native American preschool children, which was delivered in the home [50]. As most studies have reported a significant change in only 1 child-feeding practice, it is possible that in this study as well as others, parents may find it difficult to focus on changing more than 1 of the practices simultaneously.

Research clearly demonstrates the need to intervene early to establish healthy behaviors [68], and the role of parents at this stage is instrumental in achieving change [4,7,49,69]. The results

of this RCT suggest that an internet-based program can be effective in facilitating change, particularly for dietary-related behaviors, and weight status range of children in this sample demonstrates that the intervention can be applied to both healthy weight and overweight/obese children. The positive dietary-related outcomes may be a reflection of a higher proportion of the program being focused on healthy eating and the activities in these modules being more intensive and involving more practical application. The dietary-related modules were also completed first, with 32 (76.19%) participants completing these 2 modules. Participation (and perhaps motivation) dropped off as participants worked through the modules, with 26 participants (61.90%) completing all the 6 modules.

A cost-effectiveness analysis was not within the scope of this study. Although it is generally perceived that eHealth interventions are more cost-effective than traditionally delivered programs, more research is needed [70].

Recruitment for this study was challenging, despite the expansion of the recruitment area and extension of the recruitment period, and we are not able to determine with certainty the factors involved in the lower than anticipated sample size without further investigation. Further work is required to explore optimal avenues to access at-risk and hard-to-reach populations. The program was marketed as a "healthy lifestyle program" and appeared to be more successful in recruiting parents of children in the healthy weight range than overweight or obese ranges. Parental awareness of their child's weight status may have been a factor in the low enrollment rates in the overweight and obese ranges. Previous research has found that the majority of parents do not recognize that their child is overweight [71], and therefore, parents may not have recognized the need for the program. Education and monitoring initiatives may, therefore, be useful to enhance parent awareness. Feedback from participants who initially enquired about the study indicated that the need to attend face-to-face appointments for data collection was a deterrent. As the intervention is solely internet-based, it could be easily translated to a real-world setting, given that most developed countries [72-74] have a high proportion of internet users. In a real-world setting, data collection could be Web-based, which could improve participant recruitment and retention, but lack of objectively measured data may create bias issues. The requirement for participants to have a Facebook account may also have been a factor if potential participants did not have an interest in engaging with social media or felt uncomfortable sharing information online with people they did not know. It is recommended that further studies with a longer follow-up period and those that translate programs into primary health care be conducted to demonstrate long-term effectiveness.

Strengths and Limitations

This study used a randomized controlled design, applying a backwards intervention mapping exercise to align the intervention with social cognitive theory [29,30]. Multiple health behaviors were targeted, and outcome measures were based on objective and valid methods where possible. There was a low attrition rate, and the mode of delivery, content, and format of the program demonstrated a high rate of user acceptability.

There are several limitations of this study. Although it was intentional to include healthy weight children in this study, there were a higher than anticipated proportion of children (over 90%, 78/86) in the healthy weight range. Therefore, the effect on BMI may have been diluted. Due to the small number of children in the overweight and obese ranges, it was not possible to conduct a subanalysis of these participants. Statistical power would have been affected by the fact that the target sample size was not achieved despite measures to enhance participant recruitment, including expanding the recruitment area and extending the recruitment period. It is also possible that a longer follow-up period may have been required to demonstrate differences in BMI change between groups. As there were multiple outcomes assessed, there is a risk that there may have been a type 1 error. Questionnaire-based measures and the 24-hour recall used for secondary outcomes, involving self-reporting of data, were used, and therefore, it may have been possible that parents misreported this information (either intentionally or unintentionally) and such misreporting would probably have occurred in both groups. This is a familiar challenge to researchers assessing behavioral outcomes [75,76]. A height measure could not be obtained at the data collection appointment for 2 participants. Parent-provided measures were used in these instances.

In conclusion, *Time2bHealthy* led to a significant improvement in the frequency of discretionary food intake, nutrition parent self-efficacy, and pressure-to-eat child-feeding practices, but no improvement in BMI. The program has the potential for scalability and wide reach. Future studies with a larger sample size and longer follow-up period and those that translate effective eHealth childhood obesity prevention programs into primary health care are needed.

Acknowledgments

The authors would like to thank the data collectors who assisted with this study and Penny Cross for her assistance with data management and randomization. The authors would also like to thank the parents and carers and children who participated in the study and the individuals and organizations who disseminated information about the study to families. This research has been conducted with the support of the Australian Government Research Training Program Scholarship. The study was also supported by funding from the Telstra Foundation and Australian Health Management. The funding body was not involved in the design, data collection, analysis, interpretation, or writing.

Conflicts of Interest

The University of Wollongong developed the *Time2bHealthy* program.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 568KB - jmir_v21i2e11964_app1.pdf]

References

1. World Health Organization. 2016. Report of the Commission on Ending Childhood Obesity URL: http://apps.who.int/iris/bitstream/handle/10665/204176/9789241510066_eng.pdf;jsessionid=1899C78BCBF2B1DA7D75167FA0896531?sequence=1 [accessed 2018-12-07] [WebCite Cache ID 74TcgGfRt]
2. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781. [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
3. Golley RK, Hendrie GA, Slater A, Corsini N. Interventions that involve parents to improve children's weight-related nutrition intake and activity patterns - what nutrition and activity targets and behaviour change techniques are associated with intervention effectiveness? Obes Rev 2011 Feb;12(2):114-130. [doi: 10.1111/j.1467-789X.2010.00745.x] [Medline: 20406416]
4. Niemeier BS, Hektner JM, Enger KB. Parent participation in weight-related health interventions for children and adolescents: a systematic review and meta-analysis. Prev Med 2012 Jul;55(1):3-13. [doi: 10.1016/j.ypmed.2012.04.021] [Medline: 22575353]
5. Natale RA, Messiah SE, Asfour L, Uhlhorn SB, Delamater A, Arheart KL. Role modeling as an early childhood obesity prevention strategy: effect of parents and teachers on preschool children's healthy lifestyle habits. J Dev Behav Pediatr 2014;35(6):378-387. [doi: 10.1097/DBP.0000000000000074] [Medline: 25007060]
6. Oude Luttikhuis H, Baur L, Jansen H, Shrewsbury VA, O'Malley C, Stolk RP, et al. Interventions for treating obesity in children. Cochrane Database Syst Rev 2009(1):CD001872. [doi: 10.1002/14651858.CD001872.pub2] [Medline: 19160202]
7. Sung-Chan P, Sung YW, Zhao X, Brownson RC. Family-based models for childhood-obesity intervention: a systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):265-278. [doi: 10.1111/obr.12000] [Medline: 23136914]
8. Ho M, Garnett SP, Baur L, Burrows T, Stewart L, Neve M, et al. Effectiveness of lifestyle interventions in child obesity: systematic review with meta-analysis. Pediatrics 2012 Dec;130(6):e1647-e1671 [FREE Full text] [doi: 10.1542/peds.2012-1176] [Medline: 23166346]
9. Young KM, Northern JJ, Lister KM, Drummond JA, O'Brien WH. A meta-analysis of family-behavioral weight-loss treatments for children. Clin Psychol Rev 2007 Mar;27(2):240-249. [doi: 10.1016/j.cpr.2006.08.003] [Medline: 17070638]
10. Grimes-Robison C, Evans RR. Benefits and barriers to medically supervised pediatric weight-management programs. J Child Health Care 2008 Dec;12(4):329-343. [doi: 10.1177/1367493508096319] [Medline: 19052190]
11. Kelleher E, Davoren MP, Harrington JM, Shiely F, Perry IJ, McHugh SM. Barriers and facilitators to initial and continued attendance at community-based lifestyle programmes among families of overweight and obese children: a systematic review. Obes Rev 2017 Dec;18(2):183-194 [FREE Full text] [doi: 10.1111/obr.12478] [Medline: 27862851]
12. Warren JM, Golley RK, Collins CE, Okely AD, Jones RA, Morgan PJ, et al. Randomised controlled trials in overweight children: practicalities and realities. Int J Pediatr Obes 2007;2(2):73-85. [doi: 10.1080/17477160601133671] [Medline: 17763014]
13. Fitch A, Fox C, Bauerly K, Gross A, Heim C. Institute for Clinical Systems Improvement. 2013. Prevention and Management of Obesity for Children and Adolescents URL: https://www.ohcoop.org/wp-content/uploads/Clinical-Guidelines-Prevention-and-Management-Obesity-in-Children-and-Adolscent.pdf [accessed 2018-12-07] [WebCite Cache ID 74TdEOw3n]
14. Hammersley ML, Jones RA, Okely AD. Parent-focused childhood and adolescent overweight and obesity eHealth interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jul 21;18(7):e203 [FREE Full text] [doi: 10.2196/jmir.5893] [Medline: 27443862]
15. Australian Government, Department of Health. 2017. Australian 24-Hour Movement Guidelines for the Early Years (birth to 5 years): An Integration of Physical Activity, Sedentary Behaviour, and Sleep URL: http://www.health.gov.au/internet/main/publishing.nsf/content/F01F92328EDADA5BCA257BF0001E720D/$File/Birthto5years_24hrGuidelines_Brochure.pdf [accessed 2018-12-07] [WebCite Cache ID 74TdM45Kb]
16. Tremblay MS, Chaput JP, Adamo KB, Aubert S, Barnes JD, Choquette L, et al. Canadian 24-Hour Movement Guidelines for the Early Years (0-4 years): an integration of physical activity, sedentary behaviour, and sleep. BMC Public Health 2017 Nov 20;17(Suppl 5):874 [FREE Full text] [doi: 10.1186/s12889-017-4859-6] [Medline: 29219102]
17. Baranowski T, Baranowski JC, Cullen KW, Thompson DI, Nicklas T, Zakeri IE, et al. The Fun, Food, and Fitness Project (FFFP): the Baylor GEMS pilot study. Ethn Dis 2003;13(1 Suppl 1):S30-S39. [Medline: 12713209]
18. Sherwood NE, JaKa MM, Crain AL, Martinson BC, Hayes MG, Anderson JD. Pediatric primary care-based obesity prevention for parents of preschool children: a pilot study. Child Obes 2015 Dec;11(6):674-682 [FREE Full text] [doi: 10.1089/chi.2015.0009] [Medline: 26478951]

19. Baranowski T, Baranowski J, Thompson D, Buday R, Jago R, Griffith MJ, et al. Video game play, child diet, and physical activity behavior change a randomized clinical trial. Am J Prev Med 2011 Jan;40(1):33-38 [FREE Full text] [doi: 10.1016/j.amepre.2010.09.029] [Medline: 21146765]
20. Wright JA, Phillips BD, Watson BL, Newby PK, Norman GJ, Adams WG. Randomized trial of a family-based, automated, conversational obesity treatment program for underserved populations. Obesity (Silver Spring) 2013 Sep;21(9):E369-E378 [FREE Full text] [doi: 10.1002/oby.20388] [Medline: 23512915]
21. Chen JL, Weiss S, Heyman MB, Cooper B, Lustig RH. The efficacy of the web-based childhood obesity prevention program in Chinese American adolescents (Web ABC study). J Adolesc Health 2011 Aug;49(2):148-154 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.11.243] [Medline: 21783046]
22. Hammersley ML, Jones RA, Okely AD. Time2bHealthy - an online childhood obesity prevention program for preschool-aged children: a randomised controlled trial protocol. Contemp Clin Trials 2017 Oct;61:73-80 [FREE Full text] [doi: 10.1016/j.cct.2017.07.022] [Medline: 28739536]
23. Jones RA, Price N, Okely AD, Lockyer L. Developing an online program to prevent obesity in preschool-aged children: what do parents recommend? Nutr Diet 2009;66(3):151-157. [doi: 10.1111/j.1747-0080.2009.01359.x]
24. Jones R, Wells M, Okely A, Lockyer L, Walton K. Is an online healthy lifestyles program acceptable for parents of preschool children? Nutr Diet 2011;68:149-154. [doi: 10.1111/j.1747-0080.2011.01514.x]
25. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Br Med J 2010;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
26. World Health Organization. Geneva; 2006. WHO BMI-for-age (2-5 years) percentiles tables URL: https://www.who.int/childgrowth/standards/bmi_for_age/en/ [accessed 2018-12-07] [WebCite Cache ID 74TdlfpCz]
27. World Health Organization. Geneva; 2007. WHO BMI-for-age (5-19 years) percentiles tables URL: https://www.who.int/growthref/who2007_bmi_for_age/en/ [accessed 2018-12-07] [WebCite Cache ID 74TdvBpw1]
28. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, London: Prentice Hall; 1986.
29. Cornelius M, Armel KC, Hoffman K, Allen L, Bryson SW, Desai M, et al. Increasing energy- and greenhouse gas-saving behaviors among adolescents: a school-based cluster-randomized controlled trial. Energ Effic 2013 Jul 20;7(2):217-242. [doi: 10.1007/s12053-013-9219-5]
30. Robinson TN, Borzekowski DL. Effects of the SMART classroom curriculum to reduce child and family screen time. J Communication 2006 Mar;56(1):1-26. [doi: 10.1111/j.1460-2466.2006.00001.x]
31. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [Medline: 12237980]
32. National Health and Medical Research Council of Australia. 2013. Clinical Practice Guidelines for the Management of Overweight and Obesity in Adults, Adolescents and Children URL: https://nhmrc.gov.au/about-us/publications/clinical-practice-guidelines-management-overweight-and-obesity [accessed 2018-12-07] [WebCite Cache ID 74TePdlsk]
33. Bennett C, de Silva-Sanigorski AM, Nichols M, Bell AC, Swinburn BA. Assessing the intake of obesity-related foods and beverages in young children: comparison of a simple population survey with 24 hr-recall. Int J Behav Nutr Phys Act 2009 Oct 26;6:71 [FREE Full text] [doi: 10.1186/1479-5868-6-71] [Medline: 19857247]
34. Bingham DD, Costa S, Clemes SA, Routen AC, Moore HJ, Barber SE. Accelerometer data requirements for reliable estimation of habitual physical activity and sedentary time of children during the early years - a worked example following a stepped approach. J Sports Sci 2016 Oct;34(20):2005-2010. [doi: 10.1080/02640414.2016.1149605] [Medline: 26920123]
35. Pate RR, Almeida MJ, McIver KL, Pfeiffer KA, Dowda M. Validation and calibration of an accelerometer in preschool children. Obesity (Silver Spring) 2006 Nov;14(11):2000-2006 [FREE Full text] [doi: 10.1038/oby.2006.234] [Medline: 17135617]
36. Sneddon P, Peacock GG, Crowley SL. Assessment of sleep problems in preschool aged children: an adaptation of the children's sleep habits questionnaire. Behav Sleep Med 2013;11(4):283-296. [doi: 10.1080/15402002.2012.707158] [Medline: 23394069]
37. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 1994 Apr;17(3):201-207. [Medline: 7939118]
38. Bagley EJ, El-Sheikh M. Familial risk moderates the association between sleep and zBMI in children. J Pediatr Psychol 2013 Aug;38(7):775-784 [FREE Full text] [doi: 10.1093/jpepsy/jst031] [Medline: 23699749]
39. Birch LL, Fisher JO, Grimm-Thomas K, Markey CN, Sawyer R, Johnson SL. Confirmatory factor analysis of the Child Feeding Questionnaire: a measure of parental attitudes, beliefs and practices about child feeding and obesity proneness. Appetite 2001 Jun;36(3):201-210. [doi: 10.1006/appe.2001.0398] [Medline: 11358344]
40. Downing KL, Hinkley T, Hesketh KD. Associations of parental rules and socioeconomic position with preschool children's sedentary behaviour and screen time. J Phys Act Health 2015 Apr;12(4):515-521 [FREE Full text] [doi: 10.1123/jpah.2013-0427] [Medline: 25155232]
41. Hinkley T, Salmon J, Okely AD, Crawford D, Hesketh K. Preschoolers' physical activity, screen time, and compliance with recommendations. Med Sci Sports Exerc 2012 Mar;44(3):458-465. [doi: 10.1249/MSS.0b013e318233763b] [Medline: 21900847]

42. Palfreyman Z, Haycraft E, Meyer C. Development of the Parental Modelling of Eating Behaviours Scale (PARM): links with food intake among children and their mothers. Matern Child Nutr 2014 Oct;10(4):617-629. [doi: 10.1111/j.1740-8709.2012.00438.x] [Medline: 22906242]
43. Gattshall ML, Shoup JA, Marshall JA, Crane LA, Estabrooks PA. Validation of a survey instrument to assess home environments for physical activity and healthy eating in overweight children. Int J Behav Nutr Phys Act 2008 Jan 11;5:3 [FREE Full text] [doi: 10.1186/1479-5868-5-3] [Medline: 18190709]
44. Bohman B, Ghaderi A, Rasmussen F. Psychometric properties of a new measure of parental self-efficacy for promoting healthy physical activity and dietary behaviors in children. Eur J Psychol Assess 2013;29(4):291-298. [doi: 10.1027/1015-5759/a000159]
45. Wald ER, Ewing LJ, Moyer SC, Eickhoff JC. An interactive web-based intervention to achieve healthy weight in young children. Clin Pediatr (Phila) 2018 May;57(5):547-557. [doi: 10.1177/0009922817733703] [Medline: 29067819]
46. Paineau DL, Beaufils F, Boulier A, Cassuto DA, Chwalow J, Combris P, et al. Family dietary coaching to improve nutritional intakes and body weight control: a randomized controlled trial. Arch Pediatr Adolesc Med 2008 Jan;162(1):34-43. [doi: 10.1001/archpediatrics.2007.2] [Medline: 18180410]
47. Nyström CD, Sandin S, Henriksson P, Henriksson H, Trolle-Lagerros Y, Larsson C, et al. Mobile-based intervention intended to stop obesity in preschool-aged children: the MINISTOP randomized controlled trial. Am J Clin Nutr 2017 Jun;105(6):1327-1335. [doi: 10.3945/ajcn.116.150995] [Medline: 28446496]
48. Yavuz HM, van Ijzendoorn MH, Mesman J, van der Veek S. Interventions aimed at reducing obesity in early childhood: a meta-analysis of programs that involve parents. J Child Psychol Psychiatry 2014 Oct 8;56(6):677-692. [doi: 10.1111/jcpp.12330] [Medline: 25292319]
49. Gruber KJ, Haldeman LA. Using the family to combat childhood and adult obesity. Prev Chronic Dis 2009 Jul;6(3):A106 [FREE Full text] [Medline: 19527578]
50. Harvey-Berino J, Rourke J. Obesity prevention in preschool native-american children: a pilot study using home visiting. Obes Res 2003 May;11(5):606-611 [FREE Full text] [doi: 10.1038/oby.2003.87] [Medline: 12740449]
51. Louzada ML, Campagnolo PD, Rauber F, Vitolo MR. Long-term effectiveness of maternal dietary counseling in a low-income population: a randomized field trial. Pediatrics 2012 Jun;129(6):e1477-e1484. [doi: 10.1542/peds.2011-3063] [Medline: 22566413]
52. Williamson DA, Walden HM, White MA, York-Crowe E, Newton RL, Alfonso A, et al. Two-year internet-based randomized controlled trial for weight loss in African-American girls. Obesity (Silver Spring) 2006 Jul;14(7):1231-1243. [doi: 10.1038/oby.2006.140] [Medline: 16899804]
53. Knowlden AP, Conrad E. Two-year outcomes of the Enabling Mothers to Prevent Pediatric Obesity Through Web-Based Education and Reciprocal Determinism (EMPOWER) randomized control trial. Health Educ Behav 2018 Apr;45(2):262-276. [doi: 10.1177/1090198117732604] [Medline: 28954544]
54. van Grieken A, Vlasblom E, Wang L, Beltman M, Boere-Boonekamp MM, L'Hoir MP, et al. Personalized web-based advice in combination with well-child visits to prevent overweight in young children: cluster randomized controlled trial. J Med Internet Res 2017 Jul 27;19(7):e268 [FREE Full text] [doi: 10.2196/jmir.7115] [Medline: 28751299]
55. Haire-Joshu D, Elliott MB, Caito NM, Hessler K, Nanney MS, Hale N, et al. High 5 for Kids: the impact of a home visiting program on fruit and vegetable intake of parents and their preschool children. Prev Med 2008 Jul;47(1):77-82 [FREE Full text] [doi: 10.1016/j.ypmed.2008.03.016] [Medline: 18486203]
56. Shelton D, le Gros K, Norton L, Stanton-Cook S, Morgan J, Masterman P. Randomised controlled trial: a parent-based group education programme for overweight children. J Paediatr Child Health 2007 Dec;43(12):799-805. [doi: 10.1111/j.1440-1754.2007.01150.x] [Medline: 17854421]
57. Haerens L, Deforche B, Maes L, Stevens V, Cardon G, de Bourdeaudhuij I. Body mass effects of a physical activity and healthy food intervention in middle schools. Obesity (Silver Spring) 2006 May;14(5):847-854. [doi: 10.1038/oby.2006.98] [Medline: 16855194]
58. Williamson DA, Martin PD, White MA, Newton R, Walden H, York-Crowe E, et al. Efficacy of an internet-based behavioral weight loss program for overweight adolescent African-American girls. Eat Weight Disord 2005 Sep;10(3):193-203. [Medline: 16277142]
59. Skouteris H, McCabe M, Swinburn B, Newgreen V, Sacher P, Chadwick P. Parental influence and obesity prevention in pre-schoolers: a systematic review of interventions. Obes Rev 2011 May;12(5):315-328. [doi: 10.1111/j.1467-789X.2010.00751.x] [Medline: 20492538]
60. Haines J, McDonald J, O'Brien A, Sherry B, Bottino CJ, Schmidt ME, et al. Healthy habits, happy homes: randomized trial to improve household routines for obesity prevention among preschool-aged children. JAMA Pediatr 2013 Nov;167(11):1072-1079. [doi: 10.1001/jamapediatrics.2013.2356] [Medline: 24019074]
61. Fairclough SJ, Noonan R, Rowlands AV, van Hees V, Knowles Z, Boddy LM. Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers. Med Sci Sports Exerc 2016 Feb;48(2):245-253. [doi: 10.1249/MSS.0000000000000771] [Medline: 26375253]
62. Tudor-Locke C, Barreira TV, Schuna JM, Mire EF, Chaput JP, Fogelholm M, ISCOLE Research Group. Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International Study of Childhood Obesity, Lifestyle

and the Environment (ISCOLE). Int J Behav Nutr Phys Act 2015 Feb 11;12:11 [FREE Full text] [doi: 10.1186/s12966-015-0172-x] [Medline: 25881074]
63. Campbell K, Hesketh K, Silverii A, Abbott G. Maternal self-efficacy regarding children's eating and sedentary behaviours in the early years: associations with children's food intake and sedentary behaviours. Int J Pediatr Obes 2010 Dec;5(6):501-508. [doi: 10.3109/17477161003777425] [Medline: 20429735]
64. Jago R, Sebire SJ, Edwards MJ, Thompson JL. Parental TV viewing, parental self-efficacy, media equipment and TV viewing among preschool children. Eur J Pediatr 2013 Nov;172(11):1543-1545. [doi: 10.1007/s00431-013-2077-5] [Medline: 23812514]
65. Rohde JF, Bohman B, Berglind D, Hansson LM, Frederiksen P, Mortensen EL, et al. Cross-sectional associations between maternal self-efficacy and dietary intake and physical activity in four-year-old children of first-time Swedish mothers. Appetite 2018 Jun 01;125:131-138. [doi: 10.1016/j.appet.2018.01.026] [Medline: 29408332]
66. Koh GA, Scott JA, Woodman RJ, Kim SW, Daniels LA, Magarey AM. Maternal feeding self-efficacy and fruit and vegetable intakes in infants. Results from the SAIDI study. Appetite 2014 Oct;81:44-51. [doi: 10.1016/j.appet.2014.06.008] [Medline: 24911620]
67. Essery EV, DiMarco NM, Rich SS, Nichols DL. Mothers of preschoolers report using less pressure in child feeding situations following a newsletter intervention. J Nutr Educ Behav 2008;40(2):110-115. [doi: 10.1016/j.jneb.2007.02.008] [Medline: 18314087]
68. Campbell KJ, Hesketh KD. Strategies which aim to positively impact on weight, physical activity, diet and sedentary behaviours in children from zero to five years. A systematic review of the literature. Obes Rev 2007 Jul;8(4):327-338. [doi: 10.1111/j.1467-789X.2006.00305.x] [Medline: 17578382]
69. Ventura AK, Birch LL. Does parenting affect children's eating and weight status? Int J Behav Nutr Phys Act 2008;5:15 [FREE Full text] [doi: 10.1186/1479-5868-5-15] [Medline: 18346282]
70. Jelalian E, Rancourt D, Sato AF. Innovative interventions in pediatric obesity: commentary and future directions. J Pediatr Psychol 2013 Oct;38(9):1030-1036. [doi: 10.1093/jpepsy/jst072] [Medline: 24039241]
71. Lundahl A, Kidwell KM, Nelson TD. Parental underestimates of child weight: a meta-analysis. Pediatrics 2014 Mar;133(3):e689-e703 [FREE Full text] [doi: 10.1542/peds.2013-2690] [Medline: 24488736]
72. Australian Bureau of Statistics. Household Use of Information Technology, Australia, 2016-17 URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0 [accessed 2018-12-07] [WebCite Cache ID 74Teccwox]
73. Pew Research Center. 2018. Internet/Broadband Fact Sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2018-08-15] [WebCite Cache ID 71gkd1Xjx]
74. Office for National Statistics. Statistical bulletin: Internet users, UK: 2018 URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2018 [accessed 2018-11-29] [WebCite Cache ID 71g9ZPsaM]
75. Gemming L, Jiang Y, Swinburn B, Utter J, Mhurchu CN. Under-reporting remains a key limitation of self-reported dietary intake: an analysis of the 2008/09 New Zealand Adult Nutrition Survey. Eur J Clin Nutr 2014 Feb;68(2):259-264. [doi: 10.1038/ejcn.2013.242] [Medline: 24300904]
76. Poslusna K, Ruprich J, de Vries JH, Jakubikova M, van't Veer P. Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. Br J Nutr 2009 Jul;101(Suppl 2):S73-S85. [doi: 10.1017/S0007114509990602] [Medline: 19594967]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance
BMI: body mass index
eHealth: electronic health
ITT: intention-to-treat
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.08.18; peer-reviewed by E Hohman, A Beleigoli; comments to author 14.10.18; revised version received 31.10.18; accepted 09.11.18; published 08.02.19.

Please cite as:
Hammersley ML, Okely AD, Batterham MJ, Jones RA
An Internet-Based Childhood Obesity Prevention Program (Time2bHealthy) for Parents of Preschool-Aged Children: Randomized Controlled Trial
J Med Internet Res 2019;21(2):e11964
URL: http://www.jmir.org/2019/2/e11964/
doi:10.2196/11964
PMID:30735139

Original Paper

Effectiveness of a Technology-Based Supportive Educational Parenting Program on Parental Outcomes (Part 1): Randomized Controlled Trial

Shefaly Shorey[1], PhD; Yvonne Peng Mei Ng[2], MBBS; Esperanza Debby Ng[3], BA; An Ling Siew[3], BSc (Hons); Evalotte Mörelius[4], PhD; Joanne Yoong[1], PhD; Mihir Gandhi[5], PhD

[1]Clinical Research Centre, Alice Lee Centre for Nursing Studies, National University of Singapore, Singapore, Singapore

[2]National University Hospital, Singapore, Singapore

[3]National University of Singapore, Singapore, Singapore

[4]Linkoping University, Linkoping, Sweden

[5]Singapore Clinical Research Institute, Singapore, Singapore

Corresponding Author:
Shefaly Shorey, PhD
Clinical Research Centre
Alice Lee Centre for Nursing Studies
National University of Singapore
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011294
Email: nurssh@nus.edu.sg

Related Article:

Companion article: https://www.jmir.org/2019/2/e11065/

Abstract

Background: Transitioning into parenthood can be stressful for new parents, especially with the lack of continuity of care from health care professionals during the postpartum period. Short hospital stays limit the availability of support and time parents need to be well equipped with parenting and infant care skills. Poor parental adjustment may, in turn, lead to negative parental outcomes and adversely affect the child's development. For the family's future well-being, and to facilitate a smoother transition into parenthood, there is a need for easily accessible, technology-based educational programs to support parents during the crucial perinatal period.

Objective: This study aimed to examine the effectiveness of a technology-based supportive educational parenting program (SEPP) on parenting outcomes during the perinatal period in couples.

Methods: A randomized, single-blinded, parallel-armed, controlled trial was conducted. The study recruited 236 parents (118 couples) from an antenatal clinic of a tertiary hospital in Singapore. Eligible parents were randomly assigned to the intervention group (n=118) or the control group (n=118). The SEPP is based on Bandura's self-efficacy theory and Bowlby's theory of attachment. Components of the intervention include 2 telephone-based educational sessions (1 antenatal and 1 immediately postnatal) and a mobile health app follow-up for 1 month. The control group only received routine perinatal care provided by the hospital. Outcome measures including parenting self-efficacy (PSE), parental bonding, perceived social support, parenting satisfaction, postnatal depression (PND), and anxiety were measured using reliable and valid instruments. Data were collected over 6 months at 4 time points: during pregnancy (third trimester), 2 days postpartum, 1 month postpartum, and 3 months postpartum. Outcomes were standardized using baseline means and SDs. Linear mixed models were used to compare the groups for postpartum changes in the outcome variables.

Results: The intervention group showed significantly better outcome scores than the control group from baseline to 3 months postpartum for PSE (mean difference, MD, 0.37; 95% CI 0.06 to 0.68; *P*=.02), parental bonding (MD −1.32; 95% CI −1.89 to

−0.75; *P*<.001), self-perceived social support (MD 0.69; 95% CI 0.18 to 1.19; *P*=.01), parenting satisfaction (MD 1.40; 95% CI 0.86 to 1.93; *P*<.001), and PND (MD −0.91; 95% CI −1.34 to −0.49; *P*<.001). Postnatal anxiety (PNA) scores of the intervention group were only significantly better after adjusting for covariates (MD −0.82; 95% CI −1.15 to −0.49; *P*<.001).

Conclusions: The technology-based SEPP is effective in enhancing parental bonding, PSE, perceived social support and parental satisfaction, and in reducing PND and PNA. Health care professionals could incorporate it with existing hands-on infant care classes and routine care to better meet parents' needs and create positive childbirth experiences, which may in turn encourage parents to have more children.

Trial Registration: ISRCTN Registry ISRCTN48536064; http://www.isrctn.com/ISRCTN48536064 (Archived by WebCite at http://www.webcitation.org/6wMuEysiO).

(J Med Internet Res 2019;21(2):e10816) doi:10.2196/10816

KEYWORDS

parents; social support

Introduction

Background

Singapore's fertility rate has been declining over the years despite the government's desperate attempts to incentivize married couples to have more children. Career prioritization [1], previous negative childbirth experiences, and unmet parental expectations were the main reasons for this declining trend [2,3]. To mitigate negative childbirth experiences and prepare parents for parenthood, perinatal educational classes have been made available in Singapore's hospitals. However, owing to unawareness, time, and financial constraints, only few parents attend these classes [4]. The dissemination of overwhelming infant care information through pamphlets or in a didactic style during the short hospital stay also tends to cause information overload for parents [4-6]. Despite the growing interest and involvement of fathers in parenting [5], perinatal care by Singapore's hospitals still focuses primarily on breastfeeding and the physical health of the mother and child, failing to consider paternal involvement and the importance of parent-child bonding [5,7]. Recently, under a smart nation initiative, Singapore aimed to deliver holistic health care through technological innovations [8]. Given the increasing number of parents relying on Web-based information and online support communities [9,10], and considering the unreliability of sources and lack of professional moderation of Web-based information [11], there is a need for an improved technology and a theory-based perinatal educational program for parents.

Among all parental outcomes, parenting self-efficacy (PSE) is a major determinant of a positive parenting experience. According to Bandura, self-efficacy refers to one's feeling of effectiveness in accomplishing required tasks and activities [12]. For better PSE, Bandura emphasized that parents must have confidence in their ability to perform specific skills and believe that their actions will have the desired outcomes to ensure successful parenting [13]. Self-efficacy can be developed through the mastery of experiences, vicarious experiences, social persuasion, and affective and physiological factors [12]. Especially for first-time parents, PSE is highly associated with better coping responses and parenthood adjustment and positive psychological and developmental outcomes for parents and children [14,15]. In addition, Bowlby's attachment theory [16] also theorizes that PSE, social support, and parental emotional well-being are essential to establishing early parent-infant bonding, which is the foundation of a positive development of social relationships in infants. Therefore, it is important to investigate the relationships among PSE, social support, parent-infant bonding, parents' psychological health (ie, postnatal depression [PND] and anxiety), and parenting satisfaction.

During the postpartum period, the stressful adaptation to new parental roles, additional infant care responsibilities, and a lack of social support from one's partner [17-19] can adversely affect parental bonding with the infant, which may give rise to child developmental and attachment issues [20]. Poor parental adjustment also increases the risks of postpartum psychological disorders in both mothers and fathers [21,22]. On the contrary, high levels of perceived PSE [23] and social support help to facilitate smoother transition to parenthood [24], leading to increased parental bonding, parenting satisfaction, and parenting competence [25], and lowered risks of PND [23] and postnatal anxiety (PNA) [26]. This is evident in the vital role of PSE in promoting a positive parenting experience. Hence, intervention programs should emphasize on increasing PSE among parents.

In a recent review, educational intervention for parents was shown to be effective in enhancing and sustaining PSE long-term [27]. This corresponds with previous technology-based intervention studies, which were shown to not only boost PSE [28-31] but also increase parenting satisfaction [28,30,31], parental bonding [32,33], and perceived social support [29,31,34-36] and reduce postpartum psychological disorders [32,35,37]. However, most of these interventions cater only to mothers [32,35,37,38] and are only introduced during the postnatal period [28,30,31,33,34,38]. This study addresses the lack of continuity of care during the perinatal period [4,39] and the lack of inclusion of fathers' involvement in parenting [28,31] by providing new insights with a technology- and couple-based parenting educational program made available during the perinatal period.

Aim and Hypotheses

This study aimed to examine the effectiveness of a technology-based supportive educational parenting program (SEPP) on parental outcomes in terms of PSE, parental bonding, perceived social support, parenting satisfaction, PND, and PNA during the perinatal period in couples.

Compared with the control group receiving standard care, we hypothesized that the intervention group will have significantly better scores for PSE, perceived social support, parental bonding, and parental satisfaction, and lower scores for PND and PNA from baseline to 3 months postpartum.

Methods

Study Design

This was a randomized, single-blinded, parallel-armed controlled trial. The research assistant responsible for data collection was blinded to the intervention assignment of participants. Before recruitment, an independent statistician generated a randomization list using a permuted block randomization method (no stratification factor) with a 1:1 allocation ratio using Research Randomizer [40]. The block size was blinded to the study team. Couples were randomized into 2 groups (59 couples in each group) using an opaque envelope containing nonduplicated numbers (1-118).

Eligibility Criteria

Eligible participants were heterosexual married couples aged 21 years and older (individuals aged 21 years and below are considered minors in Singapore), were proficient in spoken and written English, owned a mobile phone with internet access, and planned to stay in Singapore for the first 3 months post delivery. Only mothers who had a low-risk singleton pregnancy with more than 28 weeks gestation were included. Fathers and mothers were excluded if they were a single parent, had self-reported physical or mental disorders that would interfere with their ability to participate in the study "and" or "or" if the mother had a high-risk pregnancy (eg, placenta-previa major, preeclampsia, or pregnancy-induced hypertension), had assisted delivery such as vacuum or forceps with a fourth degree perineal tear, and/or had given birth to a stillborn or newborn with congenital abnormalities and/or medical complications. Mothers with high-risk pregnancy were excluded to reduce confounding influences on the parental outcome scores. Upon recruitment, couples were informed of possible exclusion from the study if mothers were to experience complications during pregnancy and/or delivery.

Intervention

Parents assigned to the control group received routine perinatal care provided by the hospital, which includes antenatal checkups with an obstetrician, optional antenatal educational classes and postnatal parent-craft educational classes, and regular follow-ups with doctors from 10 days to 6 weeks postpartum.

Parents in the intervention group received the SEPP in addition to the standard routine perinatal hospital care. The SEPP adopted a 3-step approach, including (1) a 30-min telephone-based antenatal educational session, (2) a 60-min telephone-based immediate postnatal educational session, and (3) a mobile health (mHealth) app follow-up educational session made available for 4 weeks postpartum. Individual usernames, masking the parents' identities, and passwords were issued to the parents for access to the mHealth app. Details of the SEPP are summarized in the protocol [41].

The mHealth app contained knowledge-based content that addressed issues on breastfeeding, maternal self-care, newborn care tasks, dealing with emotional challenges, and enhancing parental efficacy and bonding, besides providing insights to new parents to facilitate their transition into parenthood. In addition, parental queries could be posted in the app's discussion forum, which were answered daily by a trained midwife for the first 4 weeks post childbirth. Parents were also highly encouraged to share their personal insights and experiences in response to such queries. The mHealth app issued daily push notifications regarding important milestones on parenting. Further specifications on the intervention can be found in the study protocol [41].

Procedure

The study took place in the antenatal clinic of a tertiary hospital, National University Hospital, in Singapore from December 2016 to December 2017. Participants were recruited as a couple (father and mother dyad) when they went for their routine antenatal checkup at the antenatal clinic. With the support and referral of nurse managers and clinicians at the antenatal clinic, a research assistant (RA1) approached referred couples to explain the purpose and details of the study. After being screened for eligibility, interested couples gave their informed consent and had to complete a demographics form and baseline questionnaire. They were then randomized into either the intervention group or the control group. For the SEPP intervention group, the RA1 proceeded to deliver a 30-min telephone-based antenatal educational session to participants. After childbirth, the couples were reapproached by the RA1 in the postnatal wards to finish another set of questionnaires. They then received a 60-min telephone-based postnatal educational session conducted by the RA1. Before discharge from the hospital, the couples were required to download the supportive parenting educational mHealth app. The RA1 guided couples through the app's functions on the spot. Individual usernames and passwords, which would expire in 4 weeks, were provided for access to the mHealth app.

For couples in the control group, only routine perinatal care by the hospital was provided. Subsequent postbaseline data collection was done through telephone calls by another research assistant (RA2) who was blinded to the group allocation. Data collection took place at the following time points for all parents: (1) during pregnancy (baseline), (2) 2 days postpartum (immediate), (3) 4 weeks (1 month) postpartum, and (4) 3 months postpartum.

Outcome Measures

The primary outcome (PSE) and secondary outcomes (parental bonding, PND, PNA, perceived social support, and parenting satisfaction) were measured using validated and reliable self-report questionnaires. PSE was measured using the 10-item Parenting Efficacy Scale (PES) [42], with a score range of 10 to 40. A high PES score indicates a high level of perceived self-efficacy [43]. Internal consistency of the PES was high across all time points (baseline, immediate postpartum, 1 month postpartum, and 3 months postpartum) with standardized Cronbach alphas of .935, .928, .925, and .868, respectively. Parental bonding was measured using the 8-item Parent-to-Infant

Bonding Questionnaire (PIBQ) [44], which has a 4-point Likert scale. As the eighth item on aggression toward newborns was found to be poorly correlated with other items, this item was dropped to improve internal consistency among items and to help in calculating the total score. Total scores range from 0 to 21, and a score of 2 and above for each item suggests poor and ineffective parental bonding [30,45,46]. Standardized Cronbach alphas for the PIBQ were .704, .585, .663, and .624 across each time point. The 10-item Edinburgh Postnatal Depression Scale (EPDS), with a score range of 0 to 30, was used to measure PND [47]. A higher score indicates a higher risk of PND. The recommended cutoff score for PND screening in mothers is 12 or 13 [48], whereas the recommended cutoff score for PND in fathers is above 10 [49]. The EPDS had high internal consistency across each time point (.811, .834, .853, and .834). The 40-item State Trait Anxiety Inventory (STAI) [50] is widely used to measure PNA and has a score range of 40 to 160, with a higher score indicating a higher level of parental anxiety [31,51]. The internal consistencies of STAI were .957, .962, .964, and .961 across each time point. The 8-item Perceived Social Support for Parenting (PSSP) [52] scale constitutes 2 subparts (4 items each) that are used to measure parents' perceived social support received from their partner and others. It has a total score range of 0 to 40, with a higher score indicating higher perceived social support [52] . The internal consistencies of the PSSP scale were .936, .875, .923, and .936 across each time point. Finally, parenting satisfaction was measured with an evaluation subscale of the What Being a Parent of a Baby Is Like (WPBL) scale [46]. It consists of 11 items, each with a 10-point semantic differential scale ranging from 0 to 9. A higher score indicates higher parenting satisfaction [53,54]. The WPBL scale has high internal consistencies of .956, .929, .958, and .954 across each time point. Further details on the psychometric properties of each outcome measure are mentioned in the protocol paper [41].

Data Analysis

The sample size was calculated based on a repeated measure analysis accounting for intracluster (within couple) to examine the differences between the 2 groups (intervention and control) at 3 months. Assuming a medium effect size of 0.3, an intracluster correlation of .05, and a correlation between repeated measurements of .5 at a power of 85% with a significance level of 5% (2-sided), 88 participants (44 couples) in each group were required. Factoring a 25% attrition rate, 236 participants (118 couples), with 118 participants (59 couples) in each group, were required.

The analysis was performed on the intention-to-treat population. All outcome data were treated continuously. As the total scores for the outcome variables (PSE, PIBQ, EPDS, STAI, PSSP, and WPBL) were on different scales, they were standardized to a z-score using their baseline mean and SD. This will enable interpreting all the outcome variables on the same scale, equivalent to the standardized effect size. In this study, each couple should be considered as a cluster of 2 individuals whose outcomes could be correlated. To account for intracluster correlation within clusters, linear mixed-effect models were used to compare the 2 groups on change in PSE, PIBQ, EPDS, STAI, PSSP, and WPBL z-scores at immediate, 1 month, and 3 months postpartum. The unadjusted model included couple-specific random intercepts, and baseline value, indicator variable for the intervention (reference—control), indicator variables for time points: immediate, 1 month, 3 months postpartum (reference—baseline), and interaction between indicator variables for intervention and 3 time points as fixed effects. A sensitivity analysis was performed using the same model adjusted for age, gender, ethnicity, education, employment status, household income, length of marriage, antenatal class attendance, confinement period, maternal/paternal leave, and mode of feeding. Similar models were performed separately for mothers and fathers to understand the intervention effect in each of these subgroups. Finally, the unadjusted and adjusted analyses were performed using the complete case data and multiple imputed data based on the Markov Chain Monte Carlo method (50 imputations) to assess the robustness of the model in the presence of missing data. A *P* value (*P*) of less than .05 was considered statistically significant. The analysis was performed using IBM SPSS 24.0.

Ethical Considerations

This study received ethics approval from the National Health Group Domain Specific Review Board (Ref. No: NHG DSRB: 2016/00651) before recruitment of participants commenced. All participants were given a participant information sheet and were briefed thoroughly on the purpose of the study and procedures before their consent was obtained. Participation was strictly voluntary, and anonymity was guaranteed. Participants were also informed of the right to withdraw at any point of the study without consequences.

Results

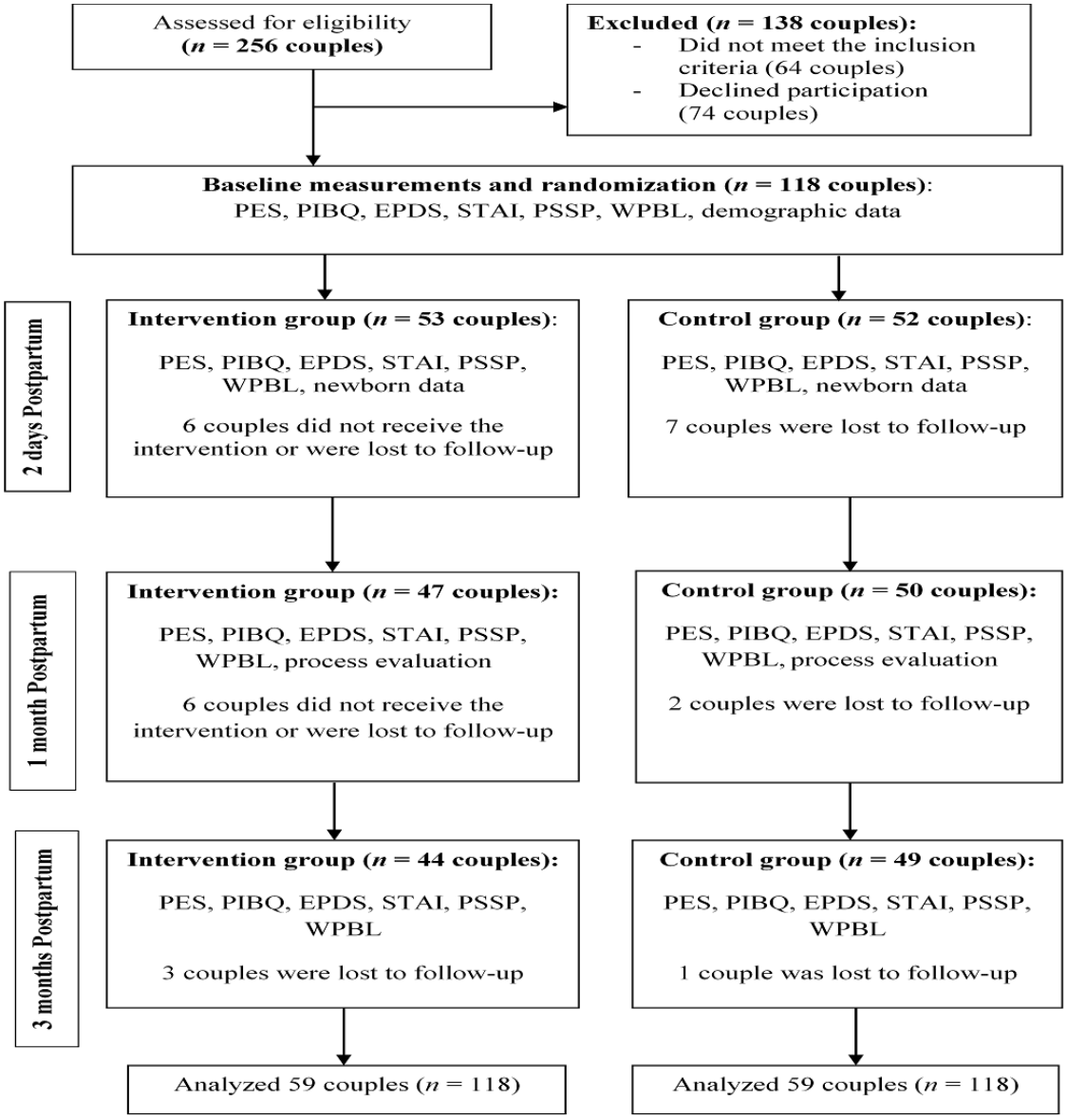
Participants' Details

A total of 236 (118 couples) participants were recruited and randomized into the SEPP intervention group (n=59) and control group (n=59). The baseline sociodemographic and pregnancy-related characteristics of the study participants are presented in Table 1. Participants had a mean age of 32 years (SD 4.81, range 22-51). All participants were married with an average marriage length of 3.5 years (SD 2.67, range 1-10). The majority of the participants were Chinese (109/236, 46.2%), university graduates (173/236, 73.3%), and employed (215/236, 91.1%) with a household income of more than SGD $5000 (138/236, 58.5%). The majority of the participants did not attend antenatal classes; most mothers had a normal vaginal delivery and followed a confinement period. There were no statistically significant differences between the control and intervention groups on demographic characteristics except for age (*P*=.026) and length of marriage (*P*=.008). Figure 1 shows the Consolidated Standards of Reporting Trial flowchart of the study.

Table 1. Couples (mothers and fathers) population: summary of sociodemographic and pregnancy-related characteristics at the baseline.

Characteristics	Intervention (n=118)	Control (n=118)
Age (years), mean (SD; min, max)		
Couples	31.3 (4.6; 22, 51)	32.6 (5.0; 24, 51)
Fathers	32.1 (4.6; 25, 51)	33.9 (5.1; 26, 51)
Mothers	30.4 (4.4; 22, 42)	31.4 (4.6; 24, 45)
Ethnicity, n (%)		
Chinese	54 (45.8)	55 (46.6)
Malay	28 (23.7)	32 (27.1)
Indian	22 (18.6)	17 (14.4)
Others	14 (11.9)	14 (11.9)
Marriage length (years), mean (SD; min, max)	3.1 (2.5; 1, 10)	4.0 (2.8; 1, 10)
Educational level, n (%)		
Primary/secondary school	7 (5.9)	1 (0.9)
Diploma/polytechnic	28 (23.7)	26 (22.2)
University graduates	83 (70.3)	90 (76.3)
Employed participants, n (%)	106 (89.8)	110 (93.2)
Monthly household income, n (%)		
<SGD $3000	18 (15.3)	12 (10.2)
SGD $3000-$5000	36 (30.5)	30 (25.4)
>SGD $5000	62 (52.5)	76 (64.4)
Planned pregnancy, n (%)	80 (67.8)	80 (67.8)
Attended antenatal class, n (%)	40 (33.9)	31 (26.3)
Mode of delivery, n (%)		
Normal vaginal delivery/water birth	72 (61.0)	76 (64.4)
Instrumental delivery	8 (6.8)	4 (3.4)
Cesarean section	32 (27.1)	32 (27.1)
Female babies, n (%)	48 (40.7)	48 (40.7)
Birth order, n (%)		
First	84 (71.2)	74 (62.7)
Second	22 (18.6)	36 (30.5)
Third and above	10 (8.5)	4 (3.4)
Paternal/maternal leave, n (%)		
No leave	3 (2.5)	2 (1.7)
≤12 weeks	69 (58.5)	61 (51.7)
>12 weeks	30 (25.4)	40 (33.9)
Confinement period, n (%)	98 (83.1)	94 (79.7)
Mode of feeding, n (%)		
Breastfeeding	74 (62.7)	70 (59.3)
Formula feeding	2 (1.7)	4 (3.4)
Breastfeeding and formula feeding	34 (28.8)	42 (35.6)

Figure 1. Consolidated standards of reporting trial flowchart of the study. EPDS: Edinburgh Postnatal Depression Scale; PES: Parenting Efficacy Scale; PIBQ: Parent-to-Infant Bonding Questionnaire; PSSP: Perceived Social Support for Parenting; STAI: State Trait Anxiety Inventory; WPBL: What Being a Parent of a Baby Is Like.



Follow-up assessments at 1 month postpartum were completed for 47 couples (47/59, 80%) in the intervention group and 50 couples (50/59, 85%) in the control group. At 3 months postpartum, follow-up assessments were completed for 44 couples (44/59, 75%) in the intervention group and 49 couples (49/59, 83%) in the control group. However, as the intention-to-treat analysis was adopted, data were analyzed for all 59 couples in both control and intervention groups. The overall attrition rate was 21.8%. Outcome scores were adjusted for ethnicity, maternal/paternal leave, confinement period, infant feeding mode, age, length of marriage, household income, employment status, and education. Table 2 summarizes the mean scores for parental outcomes of PSE, parental bonding, PND, PNA, perceived social support, and parenting satisfaction in the intervention and control groups at each time point. Multimedia Appendix 1 shows key baseline characteristics of participants who provided all data (complete case analysis) and those who provided partial data. As compared with participants who provided complete data, participants excluded from the complete case analysis were married slightly longer, had lower education, had less monthly household income, and fewer of them practiced the confinement period; therefore, the results of the unadjusted complete case analysis may or may not be extrapolated to this group of parents.

Table 2. Couples (mothers and fathers) population: summary of parental outcomes in the control (n=118) and intervention groups (n=118) at baseline, 2 days, 1 month, and 3 months postpartum.

Outcomes	Baseline, mean (SD)		2 days postpartum, mean (SD)		1 month postpartum, mean (SD)		3 months postpartum, mean (SD)	
	Control	Intervention	Control	Intervention	Control	Intervention	Control	Intervention
Parenting self-efficacy (10-40)	28.45 (6.2)	29.91 (5.8)	27.37 (5.9)	27.27 (6.3)	29.46 (6.2)	29.07 (5.9)	32.10 (4.8)	31.95 (4.4)
Parental bonding (PIBQ[a], 0-21)	2.73 (3.1)	2.54 (3.1)	1.84 (2.2)	1.58 (1.6)	14.05 (2.8)	1.30 (1.7)	5.64 (6.0)	1.37 (1.7)
Postnatal depression (EPDS[b], 0-30)	6.02 (4.1)	6.39 (4.1)	5.96 (4.6)	6.09 (3.8)	4.76 (4.9)	5.65 (4.2)	4.16 (4.3)	5.37 (4.1)
Postnatal anxiety (STAI[c], 40-160)	67.46 (17.2)	68.23 (17.8)	66.47 (19.2)	69.29 (18.1)	64.34 (21.9)	66.97 (17.8)	61.28 (16.7)	62.29 (18.2)
Perceived social support (PSSP[d], 0-40)	33.38 (5.8)	32.14 (7.5)	35.86 (4.4)	35.41 (7.7)	34.53 (6.0)	33.27 (6.5)	35.72 (4.3)	33.36 (7.2)
Parenting satisfaction (WP-BL[e], 0-99)	85.32 (12.2)	82.14 (12.1)	83.32 (12.1)	81.18 (12.7)	83.87 (13.5)	82.8 (12.2)	88.85 (8.7)	87.61 (10.5)

[a]PIBQ: Parent-to-Infant Bonding Questionnaire.

[b]EPDS: Edinburgh Postnatal Depression Scale.

[c]STAI: State Trait Anxiety Inventory.

[d]PSSP: Perceived Social Support for Parenting.

[e]WPBL: What Being a Parent of a Baby Is Like.

Main Analysis

The mean difference of the standardized scores between the control and intervention groups immediately postpartum from the baseline (ie, intervention × time effects) was not significant for all parental outcomes (PSE, parental bonding, PND, PNA, social support, and parenting satisfaction; Table 3). At 1 month postpartum, the mean difference of the standardized scores between the control and intervention groups from the baseline was significant for all parental outcomes except PNA (difference [*d*]=−3.30; 95% CI −8.17 to 1.57; *P*=.07). However, after adjusting for covariates, the mean difference of the standardized score between the groups for PNA was significant (*d*=−3.25; 95% CI −3.65 to −2.85; *P*<.001). At 3 months postpartum, all mean difference scores between the groups were significant for all parental outcomes. The mean difference between the 2 groups at the baseline (ie, main effect of intervention) was close to 0 for all outcomes (each *P*>.05; results not shown). A comparison of the mean outcome scores between the control and intervention groups across all time points is shown in Figure 2. Sensitivity analyses based on the complete case and multiple imputed data showed results similar to the main analysis (Multimedia Appendices 2 and 3).

Subgroup Analysis

For the analysis of parental outcome scores for mothers, the mean difference of scores between groups immediately postpartum was not significant for all outcomes, but the mean difference of scores between groups (ie, intervention × time effects) was significant for all parental outcomes at 1 month postpartum. At 3 months postpartum, the mean difference of the scores was only significant for parental bonding (*d*=−1.33; 95% CI −1.92 to −0.74; *P*<.001), PND (*d*=−0.84; 95% CI −1.24 to −0.44; *P*<.001), PNA (*d*=−0.55; 95% CI −0.93 to −0.17; *P*=.01), and parenting satisfaction (*d*=1.31; 95% CI 0.72 to 1.90; *P*<.001). After adjusting for covariates, the mean difference of the scores for social support became significant (*d*=0.69; 95% CI 0.06 to 1.33; *P*=.03), whereas the mean difference of the scores for PSE remained as not significant (*d*=0.33; 95% CI −0.20 to 0.85; *P*=.22).

Similarly, for the analysis of parental outcome scores for fathers, the mean difference of the scores between the groups was not significant immediately postpartum. However, the mean difference of the scores for all parental outcomes was significant at 1 and 3 months postpartum. The mean difference between the 2 groups at the baseline (ie, main effect of intervention) was close to 0 for all outcomes (each *P*>.05) in models for mothers and fathers (results not shown). A summary of the mean difference of standardized parental outcome scores between the groups for mothers and fathers is shown in Tables 4 and 5, respectively.

Table 3. Couple (mothers and fathers) population: estimated differences between the intervention and control groups for changes in standardized parental outcomes at 2 days, 1 month, and 3 months postpartum.

Standardized outcomes	2 days postpartum				1 month postpartum				3 months postpartum			
	Unadjusted difference[a]		Adjusted difference[b]		Unadjusted difference		Adjusted difference		Unadjusted difference		Adjusted difference	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Parenting self-efficacy	−0.20 (−0.53 to 0.12)	0.21	−0.29 (−0.75 to 0.17)	0.21	2.23 (1.92 to 2.54)	<.001	2.36 (1.94 to 2.79)	<.001	0.37 (0.06 to 0.68)	0.021	0.45 (0.03 to 0.86)	0.034
Parental bonding	−0.03 (−0.36 to 0.31)	0.88	−0.08 (−0.47 to 0.31)	0.67	−4.00 (−4.39 to −3.60)	<.001	−4.07 (−4.50 to −3.64)	0.001	−1.32 (−1.89 to −0.75)	<.001	−1.53 (−2.17 to −0.89)	<.001
Postnatal depression	−0.05 (−0.33 to 0.22)	0.69	0.05 (−0.26 to 0.35)	0.76	−3.40 (−3.97 to −3.22)	<.001	−3.54 (−3.93 to −3.15)	<.001	−0.91 (−1.34 to −0.49)	<.001	−0.92 (−1.38 to −0.47)	<.001
Postnatal anxiety	0.11 (−0.14 to 0.36)	0.37	0.15 (−0.12 to 0.43)	0.26	−3.30 (−8.17 to 1.57)	0.07	−3.25 (−3.65 to −2.85)	<.001	−0.71 (−1.01 to −0.42)	<.001	−0.82 (−1.15 to −0.49)	<.001
Perceived social support	−0.13 (−0.42 to 0.16)	0.37	−0.18 (−0.47 to 0.10)	0.19	3.14 (2.75 to 3.53)	<.001	3.13 (2.78 to 3.47)	<.001	0.69 (0.18 to 1.19)	0.008	0.76 (0.36 to 1.16)	<.001
Parenting Satisfaction	0.14 (−0.21 to 0.49)	0.44	0.03 (−0.37 to 0.34)	0.83	3.31 (2.92 to 3.69)	<.001	3.48 (3.14 to 3.83)	<.001	1.40 (0.86 to 1.93)	<.001	1.44 (1.05 to 1.82)	<.001

[a]Unadjusted differences were estimated using a linear mixed model adjusted for baseline values.

[b]Adjusted differences were estimated using the same model with additions of covariates: ethnicity, maternal/paternal leave, confinement period, infant feeding mode, age, length of marriage, household income, employment status, and education. See Methods for outcome definitions.

[c]OR: odds ratio.

Figure 2. Couples' (mothers and fathers) population: changes in the standardized estimated mean scores of parental outcomes at 2 days, 1 month, and 3 months postpartum in control and intervention groups.

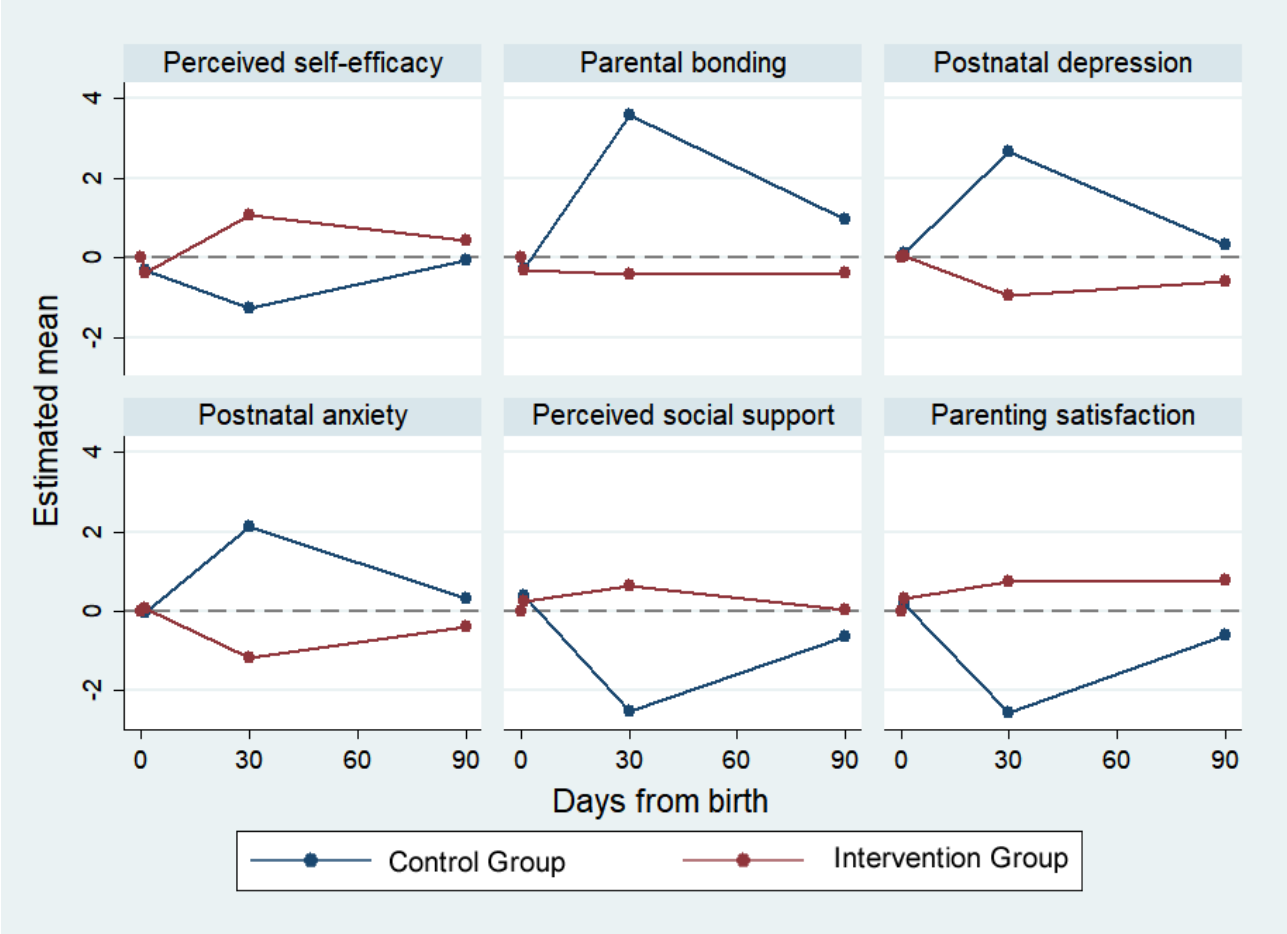


Table 4. Mothers-only subgroup: estimated differences between intervention and control groups for changes in standardized parental outcomes at immediately, 1 month, and 3 months postpartum.

Standardized outcomes	Immediately postpartum				1 month postpartum				3 months postpartum			
	Unadjusted difference[a]		Adjusted difference[b]		Unadjusted difference		Adjusted difference		Unadjusted difference		Adjusted difference	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Parenting self-efficacy	−0.12 (−0.49 to 0.26)	.54	−0.22 (−0.76 to 0.31)	.41	2.29 (1.92 to 2.67)	<.001	2.38 (1.89 to 2.88)	<.001	0.29 (−0.09 to 0.68)	.13	0.33 (−0.20 to 0.85)	.22
Parental bonding	−0.04 (−0.41 to 0.33)	.82	−0.19 (−0.62 to 0.25)	.39	−3.98 (−4.44 to −3.51)	<.001	−4.25 (−4.79 to −3.70)	<.001	−1.33 (−1.92 to −0.74)	<.001	−1.50 (−2.26 to −0.74)	<.001
Postnatal depression	−0.33 (−0.68 to 0.03)	.07	−0.15 (−0.54 to 0.25)	.46	−3.69 (−4.08 to −3.30)	<.001	−3.57 (−4.00 to −3.15)	<.001	−0.84 (−1.24 to −0.44)	<.001	−0.81 (−1.25 to −0.37)	<.001
Postnatal anxiety	−0.02 (−0.39 to 0.35)	.91	0.17 (−0.22 to 0.57)	.39	−3.26 (−3.72 to −2.79)	<.001	−3.32 (−3.80 to −2.84)	<.001	−0.55 (−0.93 to −0.17)	.005	−0.55 (−0.94 to −0.15)	.007
Perceived social support	−0.03 (−0.40 to 0.33)	.85	−0.10 (−0.52 to 0.33)	.65	3.36 (2.85 to 3.86)	<.001	3.40 (2.86 to 3.95)	<.001	0.57 (−0.06 to 1.20)	.075	0.69 (0.06 to 1.33)	.033
Parenting satisfaction	0.02 (−0.42 to 0.46)	.92	−0.15 (−0.61 to 0.32)	.53	3.64 (2.98 to 4.30)	<.001	3.77 (2.92 to 4.63)	<.001	1.31 (0.72 to 1.90)	<.001	1.31 (0.61 to 2.02)	<.001

[a]Unadjusted differences were estimated using a linear mixed model adjusted for baseline values.

[b]Adjusted differences were estimated using the same model with additions of covariates: ethnicity, maternal/paternal leave, confinement period, infant feeding mode, age, length of marriage, household income, employment status, and education. See Methods for outcome definitions.

[c]OR: odds ratio.

Table 5. Fathers-only subgroup: estimated differences between the intervention and control groups for changes in standardized parental outcomes at immediately, 1 month, and 3 months postpartum.

Standardized outcomes	Immediately postpartum				1 month postpartum				3 months postpartum			
	Unadjusted difference[a]		Adjusted difference[b]		Unadjusted difference		Adjusted difference		Unadjusted difference		Adjusted difference	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Parenting self-efficacy	−0.28 (−0.77 to 0.21)	.26	−0.40 (−0.91 to 0.11)	.12	2.15 (1.64 to 2.66)	<.001	2.13 (1.59 to 2.68)	<.001	0.48 (0.01 to 0.95)	.044	0.51 (0.03 to 0.99)	.036
Parental bonding	0.01 (−0.37 to 0.39)	.94	−0.04 (−0.49 to 0.41)	.84	−3.99 (−4.40 to −3.59)	<.001	−4.31 (−4.78 to −3.84)	<.001	−1.29 (−1.80 to −0.77)	<.001	−1.48 (−2.22 to −0.75)	<.001
Postnatal depression	0.11 (−0.24 to 0.45)	.53	0.13 (−0.23 to 0.49)	.46	−3.56 (−3.94 to −3.17)	<.001	−3.58 (−4.07 to −3.10)	<.001	−0.92 (−1.33 to −0.50)	<.001	−1.09 (−1.68 to −0.49)	<.001
Postnatal anxiety	0.17 (−0.14 to 0.48)	.26	0.13 (−0.18 to 0.44)	.4	−3.37 (−3.74 to −3.00)	<.001	−3.40 (−3.93 to −2.86)	<.001	−0.90 (−1.26 to −0.54)	<.001	−1.09 (−1.57 to −0.61)	<.001
Perceived social support	−0.22 (−0.57 to 0.13)	.22	−0.22 (−0.62 to 0.18)	.28	3.00 (2.61 to 3.39)	<.001	2.98 (2.52 to 3.44)	<.001	0.67 (0.18 to 1.15)	.007	0.81 (0.16 to 1.45)	.015
Parenting satisfaction	0.26 (−0.09 to 0.61)	.14	0.21 (−0.23 to 0.64)	.35	3.45 (3.08 to 3.81)	<.001	3.45 (3.01 to 3.89)	<.001	1.45 (1.02 to 1.88)	<.001	1.55 (0.95 to 2.16)	<.001

[a]Unadjusted differences were estimated using a linear mixed model adjusted for baseline values.

[b]Adjusted differences were estimated using the same model with additions of covariates: ethnicity, maternal/paternal leave, confinement period, infant feeding mode, age, length of marriage, household income, employment status, and education. See Methods for outcome definitions.

[c]OR: odds ratio.

Discussion

Overview

This study examined the effectiveness of a technology-based SEPP among parents in Singapore. As the first 2 days postpartum is a tumultuous time for parents, those in the intervention group did not show enhanced parental outcomes during the immediate postpartum period. However, they had significantly better parental outcome scores than those in the control group at 1 and 3 months postpartum, which suggests the effectiveness of the intervention. In terms of score trends for parental outcomes between 1 and 3 months postpartum, there are noticeable increases in the outcome scores for PSE, social support, and parenting satisfaction and decreases in the outcome scores for PND, parental bonding, and PNA for the control group. For the intervention group, scores for PSE, social support, and parenting satisfaction decreased, whereas scores for PND, PNA, and parental bonding increased. This implied that the cessation of the mHealth app usage at 1 month caused a decrease in parental outcomes.

Parenting Self-Efficacy

The significantly better PSE scores for the intervention group suggest that the SEPP is effective in enhancing parents' confidence in infant care skills and capability as a parent. Adhering to Bandura's self-efficacy theory [12], components of this technology-based SEPP allowed parents to obtain self-efficacy through mastery experiences (via telephone-based educational sessions and mobile phone app content information), vicarious experiences (via learning from other parents on the forum), and verbal persuasion (via constructive feedback and encouragement from the midwife).

The positive results for parental PSE in the intervention group were similar to a Finnish study by Salonen et al where parents who received a Web-based educational intervention reported significant increases in PSE during the postpartum period [30]. However, another study by Bartholomew et al that evaluated the frequency of Facebook usage and communication with Facebook friends by parents during the postpartum period found no significant increase in PSE [55]. This suggests that apart from social support, educational information is essential in enhancing PSE among parents. The effectiveness of the technology-based SEPP in boosting PSE in fathers and mothers also corresponds with other parent-specific Web-based intervention studies (eg, Home-but-not-alone [31], Mom Mood Booster [56], and New Fathers Network [28]), which showed significantly better PSE scores in the intervention group around 1 to 6 months postpartum.

After the cessation of the mHealth app usage, the PSE of parents in the intervention group decreased slightly, whereas PSE in the control group increased, resulting in similar scores. This suggested that parents in the control group needed 3 months of mastery experience to reach the same level of PSE as parents

in the intervention group, who were able to achieve high levels of PSE within 1 month. These results corresponded with Porter and Hsu's study on first-time mothers, which reported a significant increase in maternal self-efficacy at 3 months postpartum even without an educational intervention [57]. A pilot study in Iran discovered an improvement in maternal self-efficacy after 6 weeks of an intervention and no change in PSE even 1 month after the withdrawal of the intervention [58]. Along with our results, this supports Bandura's theory [12] that childcare experience increases maternal PSE regardless of the intervention received and that the SEPP is only effective in helping parents gain PSE in a shorter time.

Parental Bonding

There were significant differences in parental bonding scores between the control and intervention groups at 1 month and 3 months postpartum. According to Bowlby's attachment theory, good parental bonding is dependent on high PSE, good psychological health, and good social support [16]. This corresponds with a recent review by Edward et al who reported PND and poor maternal social networks as predictors of impaired maternal bonding [49]. As the SEPP was effective in increasing PSE, parents were more confident in infant care skills such as breastfeeding, bathing, and swaddling the baby. These interactions and increased skin-to-skin contact [59] helped to release the oxytocin hormone, which is essential for parent-infant bonding [60].

Results from existing studies on the effectiveness of educational interventions on parental bonding are still inconclusive. Some studies have found increased parental bonding after receiving an intervention [61-63], whereas other studies reported no change in parental bonding [53,54,64]. Parental bonding plays an important role in promoting the healthy psychosocial well-being of parents and social development of the child [61,63]; therefore, more rigorous testing is required to determine the effectiveness of such educational interventions on parental bonding.

Social Support

The intervention group scored significantly higher than the control group for perceived social support at 1 month postpartum. This implied that the rich knowledge-based content and video and audio recordings of infant care were effective in providing professional informational support to both parents. The online discussion forum provided emotional support by creating a social support system among parents that fostered a sense of belonging within the community. This result corresponded with a local study, which found that the use of an informational mobile app significantly improved perceived social support of mothers during the postpartum period [31]. Two reviews further revealed that such online-based forums are common sources of peer support, helping to reduce stigma and promoting help-seeking behavior, which reduces PND and PNA [9,65]. An increasing number of young parents are also turning to the internet for social support rather than to their family and friends [10]. A combination of professional and peer support was also found to increase the efficacy of educational programs [66].

On the contrary, the decrease in scores for both groups at 3 months postpartum could be attributed to the cessation of the mHealth app usage (for the intervention group), which removed an essential source of social support, or it could be due to other external factors such as the intrusiveness of family members/in-laws [67-70] or the 1-month confinement period itself [4,5]. However, further studies are required to validate these findings.

Postnatal Depression and Postnatal Anxiety

For PND, there were significant differences between the control and intervention groups at 1 month and 3 months postpartum. However, in the intervention group, PND scores increased after the termination of SEPP, indicating more depressive symptoms. This suggested that the SEPP is effective in mitigating PND and the lack of it increases the risk of PND. These results were contrary to a local study where an educational mHealth app administered for 1 month postpartum did not show significant results in mitigating PND at 3 months postpartum [31]. However, a study by Danaher et al involving a 6- to 12-week-long Web-based intervention for PND prevention among mothers found significant decrease in PND scores at 3 and 6 months postpartum [56]. Despite a similar outcome to the present results, more research is required to determine the effects of length of intervention and follow-up time points on the effectiveness of technology-based psychoeducational programs on PND prevention.

For PNA also, there were significant differences between intervention and control groups at 1 and 3 months postpartum. A study by Seymour et al reported that postpartum maternal anxiety was associated with poor partner relationship quality, need for social support, low involvement, low efficacy, and low parenting satisfaction [26]. However, comorbid depression and anxiety were more strongly correlated with these negative experiences than anxiety alone [26]. Studies often investigate PNA and depression together because of their high comorbidity rates [71]. According to a study in the Netherlands, in 57% of the comorbid cases, anxiety preceded depression, and in 18% of cases, depression preceded anxiety [72]. In this study, PND and PNA showed similar trends in both groups, with PNA scores being slightly lower than PND at all time points. Similarly, O'Mahen et al reported a significant decline in both PND and PNA scores for mothers at 3 months postpartum using an 11-week technology-based intervention involving weekly phone calls from coaches and Web-based information [73]. Although few studies focused on the comorbidity of depression and anxiety in fathers [74], previous studies have mentioned the spillover effects of maternal psychological health on paternal psychological health [75], which can adversely affect the development of the child [74]. However, coparenting alliance and paternal self-efficacy were found to reduce the paternal stress and overall psychological health of fathers [76,77]. Therefore, couple-based educational programs are important for facilitating partner support, ensuring each parent's well-being, and increasing parenting dynamics [78-80].

Parenting Satisfaction

There was an increase in parenting satisfaction scores in the intervention group, whereas there was a large decrease in

parenting satisfaction scores in the control group. These results were congruent with Salonen et al's study in which mothers who received a Web-based educational intervention reported higher parenting satisfaction, but not fathers [30]. However, Hudson et al reported an increase in paternal parenting satisfaction after an online fathers' network intervention [28]. In both studies, PSE was found to be positively correlated with parenting satisfaction as parents gained more confidence in caring for their baby [28,30]. This suggests that higher PSE scores predict higher levels of parenting satisfaction, which corresponds to our findings.

Strengths and Limitations

Although most studies focused on technology-based interventions in the postpartum period [28,30,31], this study administered its intervention during the perinatal period, which may increase the effectiveness of improving parental outcomes. This study adds on to existing literature on maternal well-being and fills knowledge gaps on paternal well-being during the perinatal period. Given that most parents today are turning to the internet for information and social support, the technology-based SEPP is an ideal time- and cost-effective method to meet parental needs and reduce dependency on midwives who may not be available.

A major limitation of this study is that as the SEPP is a technology-based intervention, there is no practical hands-on skill incorporated into the program. This reduces the effectiveness of parenting skill-based learning, especially for experiential/hands-on learners. However, the short telephone-based educational classes (30 min and 1 hour) in the SEPP may suit working parents who lack the time to attend antenatal classes, which may last up to 16 hours. Therefore, future studies should investigate the difference in effectiveness between technology-based classes and traditional face-to-face didactic perinatal classes.

Owing to a lack of available data, we were unable to analyze the cost-effectiveness of this program; future research may consider evaluating cost-effectiveness as an outcome. This study lacks infant outcome data, thus restricting the evaluation of actual benefits of intervention on infants. Moreover, this is a single-site study that only included married English-speaking couples. As we believe, the SEPP will be equally beneficial to parents from the minority group (ie, single parents, other races, etc); future research should be more inclusive and consider a multisite study with a minority population. Future studies should also investigate the long-term effectiveness (beyond 3 months) of the SEPP on parental outcomes.

The internal validity of PIBQ was not optimal to measure parental bonding among multiracial parents in this study, although it has been validated in previous studies [44,81,82]. This could be due to cultural influence on parental bonding and that some questionnaire items may require further evaluation. Nevertheless, the precision of the intervention effect based on the PIBQ was not compromised because of the adequately large sample size of the study. However, we strongly recommend further testing of this instrument among various cultural populations and the testing of its psychometric properties.

Conclusions

This study demonstrated the effectiveness of a technology-based SEPP in improving PSE, parental bonding, perceived social support, and parenting satisfaction, while reducing PND and PNA. Such educational programs are vital to equip parents with the necessary parenting skills to facilitate a smoother transition into parenthood by increasing their self-efficacy and enhancing parental bonding. They also serve as a reliable source of social support to promote good psychological health in both fathers and mothers. We hope that the clinical implementation of the SEPP will meet parents' needs and create positive childbirth experiences, which may in turn encourage parents to have more children and alleviate Singapore's declining birth rate.

Acknowledgments

The authors thank the National University of Singapore for the National University Health System Clinician Research grant FY2015 (Ref No. NUHSRO/2016/023/CRG/02) and the National University Health System, Medical Publications Support Unit, for assistance in language editing of the manuscript.

Conflicts of Interest

None declared.

Editorial Notice: This is a companion article to "Evaluation of Mothers' Perceptions of a Technology-Based Supportive Educational Parenting Program (Part 2): Qualitative Study" by Shorey and Ng (J Med Internet Res 2018;21(2):e11065) which can be found at https://www.jmir.org/2019/2/e11065/.

Multimedia Appendix 1
Comparison of sociodemographic and pregnancy-related characteristics of participants at the baseline who were included and excluded from the complete case analysis.
[DOCX File , 16 KB - jmir_v21i2e10816_app1.docx]

Multimedia Appendix 2

Complete cases analysis for couple population (mothers and fathers): estimated differences between the intervention and control groups for changes in standardized parental outcomes at postpartum time-points from the baseline.
[PDF File (Adobe PDF File), 30 KB - jmir_v21i2e10816_app2.pdf]

Multimedia Appendix 3
Multiple imputation analyses for couple population (mothers and fathers): estimated differences between the intervention and control groups for changes in standardized parental outcomes at postpartum time points from the baseline.
[PDF File (Adobe PDF File), 30 KB - jmir_v21i2e10816_app3.pdf]

Multimedia Appendix 4
CONSORT‐EHEALTH checklist (V.1.6.1).
[PDF File (Adobe PDF File), 1376 KB - jmir_v21i2e10816_app4.pdf]

References

1. Jones G. The Japanese Journal of Population. 2012 Mar. Late marriage and low fertility in Singapore: the limits of policy URL: https://pdfs.semanticscholar.org/0505/ddda55ac9f93082a62a93f877220afb0bbd5.pdf [accessed 2018-12-07] [WebCite Cache ID 74TxIYZM9]
2. Gottvall K, Waldenström U. Does a traumatic birth experience have an impact on future reproduction? BJOG 2002 Mar;109(3):254-260 [FREE Full text] [doi: 10.1111/j.1471-0528.2002.01200.x] [Medline: 11950179]
3. Waldenström U, Hildingsson I, Rubertsson C, Rådestad I. A negative birth experience: prevalence and risk factors in a national sample. Birth 2004 Mar;31(1):17-27. [doi: 10.1111/j.0730-7659.2004.0270.x] [Medline: 15015989]
4. Ong SF, Chan WC, 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]
5. Shorey S, Dennis CL, Bridge S, Chong YS, Holroyd E, He HG. First-time fathers' postnatal experiences and support needs: a descriptive qualitative study. J Adv Nurs 2017 Dec;73(12):2987-2996. [doi: 10.1111/jan.13349] [Medline: 28557020]
6. Melnyk BM, Crean HF, Feinstein NF, Fairbanks E. Maternal anxiety and depression after a premature infant's discharge from the neonatal intensive care unit: explanatory effects of the creating opportunities for parent empowerment program. Nurs Res 2008;57(6):383-394 [FREE Full text] [doi: 10.1097/NNR.0b013e3181906f59] [Medline: 19018213]
7. Ngai FW, Chan SW, Ip WY. The effects of a childbirth psychoeducation program on learned resourcefulness, maternal role competence and perinatal depression: a quasi-experiment. Int J Nurs Stud 2009 Oct;46(10):1298-1306. [doi: 10.1016/j.ijnurstu.2009.03.007] [Medline: 19361801]
8. Smart Nation Singapore. 2017. Transforming Singapore URL: https://www.smartnation.sg/about-smart-nation [accessed 2017-07-06] [WebCite Cache ID 6rkbI9Ekw]
9. Gulliver A, Griffiths KM, Christensen H, Brewer JL. A systematic review of help-seeking interventions for depression, anxiety and general psychological distress. BMC Psychiatry 2012 Jul 16;12:81 [FREE Full text] [doi: 10.1186/1471-244X-12-81] [Medline: 22799879]
10. Plantin L, Daneback K. Parenthood, information and support on the internet. A literature review of research on parents and professionals online. BMC Fam Pract 2009 May 18;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
11. Salonen AH, Kaunonen M, Astedt-Kurki P, Järvenpää AL, Isoaho H, Tarkka MT. Parenting self-efficacy after childbirth. J Adv Nurs 2009 Nov;65(11):2324-2336. [doi: 10.1111/j.1365-2648.2009.05113.x] [Medline: 19761456]
12. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191] [Medline: 847061]
13. Bandura A. Self-efficacy: the exercise of control. New York: WH Freeman and Company; 1997.
14. Biehle SN, Mickelson KD. Personal and co-parent predictors of parenting efficacy across the transition to parenthood. J Soc Clin Psychol 2011 Nov;30(9):985-1010. [doi: 10.1521/jscp.2011.30.9.985]
15. Leerkes EM, Burney RV. The development of parenting efficacy among new mothers and fathers. Infancy 2007 Feb 03;12(1):45-67. [doi: 10.1111/j.1532-7078.2007.tb00233.x]
16. Bowlby J. Attachment and Loss, Volume 1: Attachment. New York: Basic Books; 1969.
17. Doss BD, Rhoades GK, Stanley SM, Markman HJ. The effect of the transition to parenthood on relationship quality: an 8-year prospective study. J Pers Soc Psychol 2009 Mar;96(3):601-619 [FREE Full text] [doi: 10.1037/a0013969] [Medline: 19254107]
18. Lawrence E, Rothman AD, Cobb RJ, Rothman MT, Bradbury TN. Marital satisfaction across the transition to parenthood. J Fam Psychol 2008 Feb;22(1):41-50 [FREE Full text] [doi: 10.1037/0893-3200.22.1.41] [Medline: 18266531]
19. Biehle SN, Mickelson KD. First-time parents' expectations about the division of childcare and play. J Fam Psychol 2012 Feb;26(1):36-45. [doi: 10.1037/a0026608] [Medline: 22182336]

20. Feldman R. Parents' convergence on sharing and marital satisfaction, father involvement, and parent–child relationship at the transition to parenthood. Infant Ment Health J 2000 Jul 14;21(3):176-191. [doi: 10.1002/1097-0355(200007)21:3<176::AID-IMHJ3>3.0.CO;2-4]
21. Bielawska-Batorowicz E, Kossakowska-Petrycka K. Depressive mood in men after the birth of their offspring in relation to a partner's depression, social support, fathers' personality and prenatal expectations. J Reprod Infant Psychol 2006 Feb;24(1):21-29. [doi: 10.1080/02646830500475179]
22. Harwood K, McLean N, Durkin K. First-time mothers' expectations of parenthood: what happens when optimistic expectations are not matched by later experiences? Dev Psychol 2007 Jan;43(1):1-12. [doi: 10.1037/0012-1649.43.1.1] [Medline: 17201504]
23. 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]
24. Negron R, Martin A, Almog M, Balbierz A, Howell EA. Social support during the postpartum period: mothers' views on needs, expectations, and mobilization of support. Matern Child Health J 2013 May;17(4):616-623 [FREE Full text] [doi: 10.1007/s10995-012-1037-4] [Medline: 22581378]
25. Brown SG, Hudson DB, Campbell-Grossman C, Kupzyk KA, Yates BC, Hanna KM. Social support, parenting competence, and parenting satisfaction among adolescent, African American, mothers. West J Nurs Res 2018 Apr;40(4):502-519. [doi: 10.1177/0193945916682724] [Medline: 28322635]
26. Seymour M, Giallo R, Cooklin A, Dunning M. Maternal anxiety, risk factors and parenting in the first post-natal year. Child Care Health Dev 2015 Mar;41(2):314-323. [doi: 10.1111/cch.12178] [Medline: 25074519]
27. Liyana Amin NA, Tam WW, Shorey S. Enhancing first-time parents' self-efficacy: a systematic review and meta-analysis of universal parent education interventions' efficacy. Int J Nurs Stud 2018 Jun;82:149-162. [doi: 10.1016/j.ijnurstu.2018.03.021] [Medline: 29656206]
28. Hudson DB, Campbell-Grossman C, Fleck MO, Elek SM, Shipman A. Effects of the New Fathers Network on first-time fathers' parenting self-efficacy and parenting satisfaction during the transition to parenthood. Issues Compr Pediatr Nurs 2003;26(4):217-229. [doi: 10.1080/01460860390246687] [Medline: 14630529]
29. Lindberg I, Christensson K, Ohrling K. Parents' experiences of using videoconferencing as a support in early discharge after childbirth. Midwifery 2009 Aug;25(4):357-365. [doi: 10.1016/j.midw.2007.06.002] [Medline: 17935844]
30. Salonen AH, Kaunonen M, Astedt-Kurki P, Järvenpää AL, Isoaho H, Tarkka MT. Effectiveness of an internet-based intervention enhancing Finnish parents' parenting satisfaction and parenting self-efficacy during the postpartum period. Midwifery 2011 Dec;27(6):832-841. [doi: 10.1016/j.midw.2010.08.010] [Medline: 20932612]
31. Shorey S, Lau Y, Dennis CL, 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 Mar 09;73(9):2103-2117. [doi: 10.1111/jan.13293] [Medline: 28276086]
32. Haga SM, Drozd F, Brendryen H, Slinning K. Mamma mia: a feasibility study of a web-based intervention to reduce the risk of postpartum depression and enhance subjective well-being. JMIR Res Protoc 2013 Aug 12;2(2):e29 [FREE Full text] [doi: 10.2196/resprot.2659] [Medline: 23939459]
33. Luu TM, Xie LF, Peckre P, Cote S, Karsenti T, Walker CD, et al. Web-based intervention to teach developmentally supportive care to parents of preterm infants: feasibility and acceptability study. JMIR Res Protoc 2017 Nov 30;6(11):e236 [FREE Full text] [doi: 10.2196/resprot.8289] [Medline: 29191797]
34. 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]
35. Dunham PJ, Hurshman A, Litwin E, Gusella J, Ellsworth C, Dodd PW. Computer-mediated social support: single young mothers as a model system. Am J Community Psychol 1998 Apr;26(2):281-306. [doi: 10.1023/A:1022132720104] [Medline: 9693693]
36. Slomian J, Bruyère O, Reginster JY, Emonts P. The internet as a source of information used by women after childbirth to meet their need for information: a web-based survey. Midwifery 2017 May;48:46-52. [doi: 10.1016/j.midw.2017.03.005] [Medline: 28324809]
37. Evans M, Donelle L, Hume-Loveland L. Social support and online postpartum depression discussion groups: a content analysis. Patient Educ Couns 2012 Jun;87(3):405-410. [doi: 10.1016/j.pec.2011.09.011] [Medline: 22019021]
38. Lee EW, Denison FC, Hor K, Reynolds RM. Web-based interventions for prevention and treatment of perinatal mood disorders: a systematic review. BMC Pregnancy Childbirth 2016 Feb 29;16:38 [FREE Full text] [doi: 10.1186/s12884-016-0831-1] [Medline: 26928898]
39. Slomian J, Emonts P, Vigneron L, Acconcia A, Reginster JY, Oumourgh M, et al. Meeting the needs of mothers during the postpartum period: using co-creation workshops to find technological solutions. JMIR Res Protoc 2017 May 03;6(5):e76 [FREE Full text] [doi: 10.2196/resprot.6831] [Medline: 28468746]
40. Urbaniak G, Plous S. Research Randomizer. URL: https://www.randomizer.org/ [accessed 2018-04-17] [WebCite Cache ID 6yjd5HGA9]

41. Shorey S, Ng YP, Siew AL, Yoong J, Mörelius E. Effectiveness of a technology-based supportive educational parenting program on parental outcomes in Singapore: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Jan 10;7(1):e4 [FREE Full text] [doi: 10.2196/resprot.8062] [Medline: 29321127]
42. Leerkes EM, Crockenberg SC. The impact of maternal characteristics and sensitivity on the concordance between maternal reports and laboratory observations of infant negative emotionality. Infancy 2003 Oct 01;4(4):517-539. [doi: 10.1207/S15327078IN0404_07]
43. Berle J, Mykletun A, Daltveit AK, Rasmussen S, Holsten F, Dahl AA. Neonatal outcomes in offspring of women with anxiety and depression during pregnancy. A linkage study from the Nord-Trøndelag Health Study (HUNT) and Medical Birth Registry of Norway. Arch Womens Ment Health 2005 Sep;8(3):181-189. [doi: 10.1007/s00737-005-0090-z] [Medline: 15959623]
44. Taylor A, Atkins R, Kumar R, Adams D, Glover V. A new mother-to-infant bonding scale: links with early maternal mood. Arch Womens Ment Health 2005 May;8(1):45-51. [doi: 10.1007/s00737-005-0074-z] [Medline: 15868385]
45. Bienfait M, Maury M, Haquet A, Faillie JL, Franc N, Combes C, et al. Pertinence of the self-report mother-to-infant bonding scale in the neonatal unit of a maternity ward. Early Hum Dev 2011 Apr;87(4):281-287. [doi: 10.1016/j.earlhumdev.2011.01.031] [Medline: 21316876]
46. Pridham KF, Chang AS. What being the parent of a new baby is like: revision of an instrument. Res Nurs Health 1989 Oct;12(5):323-329. [doi: 10.1002/nur.4770120508] [Medline: 2798953]
47. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [Medline: 3651732]
48. Edwards H, Phillips C, Esterman A, Buisman-Pijlman F, Gordon A. Risk factors and assessment tools for mother-infant bonding: a scoping review to assist future research. Evidence Based Midwifery 2017 Dec 01;15(4):128-134.
49. Edmondson OJ, Psychogiou L, Vlachos H, Netsi E, Ramchandani PG. Depression in fathers in the postnatal period: assessment of the Edinburgh Postnatal Depression Scale as a screening measure. J Affect Disord 2010 Sep;125(1-3):365-368 [FREE Full text] [doi: 10.1016/j.jad.2010.01.069] [Medline: 20163873]
50. Spielberger CD. Manual for the state-trait anxiety inventory. Palo Alto, California: Consulting Psychologists Press; Jan 1983.
51. Julian LJ. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care Res (Hoboken) 2011 Nov;63(Suppl 11):S467-S472 [FREE Full text] [doi: 10.1002/acr.20561] [Medline: 22588767]
52. Leerkes EM, Crockenberg SC. The development of maternal self-efficacy and its impact on maternal behavior. Infancy 2002 Apr 01;3(2):227-247. [doi: 10.1207/S15327078IN0302_7]
53. Deave T, Johnson D. The transition to parenthood: what does it mean for fathers? J Adv Nurs 2008 Sep;63(6):626-633. [doi: 10.1111/j.1365-2648.2008.04748.x] [Medline: 18808584]
54. Figueroa Leigh F, Binda Vergara V, Pía Santelices M. Enhancing early attachment: design and pilot study of an intervention for primary health care dyads. J Child Health Care 2013 Mar;17(1):82-91. [doi: 10.1177/1367493512447533] [Medline: 23162050]
55. Bartholomew MK, Schoppe-Sullivan SJ, Glassman M, Kamp Dush CM, Sullivan JM. New parents' Facebook use at the transition to parenthood. Fam Relat 2012 Jul;61(3):455-469 [FREE Full text] [doi: 10.1111/j.1741-3729.2012.00708.x] [Medline: 23671354]
56. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. MomMoodBooster web-based intervention for postpartum depression: feasibility trial results. J Med Internet Res 2013 Nov 04;15(11):e242 [FREE Full text] [doi: 10.2196/jmir.2876] [Medline: 24191345]
57. Porter CL, Hsu HC. First-time mothers' perceptions of efficacy during the transition to motherhood: links to infant temperament. J Fam Psychol 2003 Mar;17(1):54-64. [doi: 10.1037/0893-3200.17.1.54] [Medline: 12666463]
58. Abarashi Z, Tahmassian K, Mazaheri MA, Panaghi L, Mansoori N. Parental self-efficacy as a determining factor in healthy mother-child interaction: a pilot study in Iran. Iran J Psychiatry Behav Sci 2014;8(1):19-25 [FREE Full text] [Medline: 24995026]
59. Shorey S, He HG, Morelius E. Skin-to-skin contact by fathers and the impact on infant and paternal outcomes: an integrative review. Midwifery 2016 Sep;40:207-217. [doi: 10.1016/j.midw.2016.07.007] [Medline: 27476026]
60. Feldman R, Gordon I, Schneiderman I, Weisman O, Zagoory-Sharon O. Natural variations in maternal and paternal care are associated with systematic changes in oxytocin following parent-infant contact. Psychoneuroendocrinology 2010 Sep;35(8):1133-1141. [doi: 10.1016/j.psyneuen.2010.01.013] [Medline: 20153585]
61. Gao LL, Chan SW, Sun K. Effects of an interpersonal-psychotherapy-oriented childbirth education programme for Chinese first-time childbearing women at 3-month follow up: randomised controlled trial. Int J Nurs Stud 2012 Mar;49(3):274-281. [doi: 10.1016/j.ijnurstu.2011.09.010] [Medline: 21962336]
62. Razurel C, Antonietti JP, Rulfi F, Pasquier N, Domingues-Montanari S, Darwiche J. The impact of pre- and post-natal psycho-educational intervention on the construction of parenthood. Arch Womens Ment Health 2017 Jun;20(3):469-472. [doi: 10.1007/s00737-017-0720-2] [Medline: 28357527]

63. Doherty WJ, Erickson MF, LaRossa R. An intervention to increase father involvement and skills with infants during the transition to parenthood. J Fam Psychol 2006 Sep;20(3):438-447. [doi: 10.1037/0893-3200.20.3.438] [Medline: 16938002]
64. Serçekus P, Baskale H. Effects of antenatal education on fear of childbirth, maternal self-efficacy and parental attachment. Midwifery 2016 Mar;34:166-172. [doi: 10.1016/j.midw.2015.11.016] [Medline: 26656473]
65. Niela-Vilén H, Axelin A, Salanterä S, Melender HL. Internet-based peer support for parents: a systematic integrative review. Int J Nurs Stud 2014 Nov;51(11):1524-1537. [doi: 10.1016/j.ijnurstu.2014.06.009] [Medline: 24998788]
66. Hannula L, Kaunonen M, Tarkka MT. A systematic review of professional support interventions for breastfeeding. J Clin Nurs 2008 May;17(9):1132-1143. [doi: 10.1111/j.1365-2702.2007.02239.x] [Medline: 18416790]
67. Chan SW, Levy V, Chung TK, Lee D. A qualitative study of the experiences of a group of Hong Kong Chinese women diagnosed with postnatal depression. J Adv Nurs 2002 Sep;39(6):571-579. [doi: 10.1046/j.1365-2648.2002.02326.x] [Medline: 12207755]
68. Leung SS, Martinson IM, Arthur D. Postpartum depression and related psychosocial variables in Hong Kong Chinese women: findings from a prospective study. Res Nurs Health 2005 Feb;28(1):27-38. [doi: 10.1002/nur.20053] [Medline: 15625707]
69. Burke CT, Goren J. Self-evaluative consequences of social support receipt: the role of context self-relevance. Pers Relatsh 2014 Jul 25;21(3):433-450. [doi: 10.1111/pere.12039]
70. Burke CT, Perndorfer CC. Negative emotional responses to motherhood-related support receipt during pregnancy predict postpartum depressive symptoms. Anxiety Stress Coping 2016 Sep;29(5):580-588. [doi: 10.1080/10615806.2015.1092023] [Medline: 26361232]
71. Falah-Hassani K, Shiri R, Dennis CL. Prevalence and risk factors for comorbid postpartum depressive symptomatology and anxiety. J Affect Disord 2016 Jul 01;198:142-147. [doi: 10.1016/j.jad.2016.03.010] [Medline: 27016657]
72. Lamers F, van Oppen P, Comijs HC, Smit JH, Spinhoven P, van Balkom AJ, et al. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). J Clin Psychiatry 2011 Mar;72(3):341-348. [doi: 10.4088/JCP.10m06176blu] [Medline: 21294994]
73. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided internet behavioural activation treatment for postpartum depression. Psychol Med 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: 10.1017/S0033291713002092] [Medline: 24148703]
74. O'Brien AP, McNeil KA, Fletcher R, Conrad A, Wilson AJ, Jones D, et al. New fathers' perinatal depression and anxiety-treatment options: an integrative review. Am J Mens Health 2017 Jul;11(4):863-876 [FREE Full text] [doi: 10.1177/1557988316669047] [Medline: 27694550]
75. Giallo R, D'Esposito F, Cooklin A, Mensah F, Lucas N, Wade C, et al. Psychosocial risk factors associated with fathers' mental health in the postnatal period: results from a population-based study. Soc Psychiatry Psychiatr Epidemiol 2013 Apr;48(4):563-573. [doi: 10.1007/s00127-012-0568-8] [Medline: 22898826]
76. Kwan RW, Kwok SY, Ling CC. The moderating roles of parenting self-efficacy and co-parenting alliance on marital satisfaction among Chinese fathers and mothers. J Child Fam Stud 2015 Feb 13;24(12):3506-3515. [doi: 10.1007/s10826-015-0152-4]
77. Pinto TM, Figueiredo B, Pinheiro LL, Canário C. Fathers' parenting self-efficacy during the transition to parenthood. J Reprod Infant Psychol 2016 Jun 09;34(4):343-355. [doi: 10.1080/02646838.2016.1178853]
78. Shapiro AF, Nahm EY, Gottman JM, Content K. Bringing baby home together: examining the impact of a couple-focused intervention on the dynamics within family play. Am J Orthopsychiatry 2011 Jul;81(3):337-350. [doi: 10.1111/j.1939-0025.2011.01102.x] [Medline: 21729014]
79. Midmer D, Wilson L, Cummings S. A randomized, controlled trial of the influence of prenatal parenting education on postpartum anxiety and marital adjustment. Fam Med 1995 Mar;27(3):200-205. [Medline: 7774781]
80. Feinberg ME, Kan ML. Establishing family foundations: intervention effects on coparenting, parent/infant well-being, and parent-child relations. J Fam Psychol 2008 Apr;22(2):253-263 [FREE Full text] [doi: 10.1037/0893-3200.22.2.253] [Medline: 18410212]
81. van Bussel JC, Spitz B, Demyttenaere K. Three self-report questionnaires of the early mother-to-infant bond: reliability and validity of the Dutch version of the MPAS, PBQ and MIBS. Arch Womens Ment Health 2010 Oct;13(5):373-384. [doi: 10.1007/s00737-009-0140-z] [Medline: 20127128]
82. Yoshida K, Yamashita H, Conroy S, Marks M, Kumar C. A Japanese version of mother-to-infant bonding scale: factor structure, longitudinal changes and links with maternal mood during the early postnatal period in Japanese mothers. Arch Womens Ment Health 2012 Oct;15(5):343-352 [FREE Full text] [doi: 10.1007/s00737-012-0291-1] [Medline: 22733162]

Abbreviations

EPDS: Edinburgh Postnatal Depression Scale
mHealth: mobile health
OR: odds ratio
PES: Parenting Efficacy Scale

PIBQ: Parent-to-Infant Bonding Questionnaire
PNA: postnatal anxiety
PND: postnatal depression
PSE: parenting self-efficacy
PSSP: Perceived Social Support for Parenting
RA: research assistant
SEPP: supportive educational parenting program
STAI: State Trait Anxiety Inventory
WPBL: What Being a Parent of a Baby Is Like

Edited by G Eysenbach; submitted 19.04.18; peer-reviewed by F Drozd, L McGoron; comments to author 14.06.18; revised version received 17.07.18; accepted 25.11.18; published 13.02.19.

Please cite as:
Shorey S, Ng YPM, Ng ED, Siew AL, Mörelius E, Yoong J, Gandhi M
Effectiveness of a Technology-Based Supportive Educational Parenting Program on Parental Outcomes (Part 1): Randomized Controlled Trial
J Med Internet Res 2019;21(2):e10816
URL: http://www.jmir.org/2019/2/e10816/
doi:10.2196/10816
PMID:30758289

Original Paper

Evaluation of Mothers' Perceptions of a Technology-Based Supportive Educational Parenting Program (Part 2): Qualitative Study

Shefaly Shorey[1], PhD; Esperanza Debby Ng[1], BA

National University of Singapore, Singapore, Singapore

Corresponding Author:
Shefaly Shorey, PhD
National University of Singapore
Level 2 Clinical Research Centre, 10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011294
Fax: 65 67767135
Email: nurssh@nus.edu.sg

Related Article:

Companion article: https://www.jmir.org/2019/2/e10816/

Abstract

Background: Transitioning into parenthood can be stressful as parents struggle to cope with new parenting responsibilities. Although perinatal care in hospitals aims to improve parental outcomes, there is a general consensus that it is suboptimal and insufficient. Therefore, many studies have designed intervention methods to supplement support for parents during this stressful period. However, studies often focus on parental outcomes as indicators of their interventions' success and effectiveness. Studies evaluating participants' experiences and feedback are limited.

Objective: This study aimed to examine the experiences and perceptions of participants who participated in a supportive education parenting program intervention study.

Methods: A qualitative semistructured interview was conducted with 16 mothers (6 control and 10 intervention) from a randomized controlled trial. The supportive education parenting program received by the intervention group included 2 phone-based perinatal educational sessions, a phone-based educational session after childbirth, and a 1-month postpartum access to a mobile health app. The interviews were approximately 30- to 60-min long, audiotaped and transcribed verbatim, and analyzed using thematic analysis. Study findings were reported according to the Consolidated Criteria for Reporting Qualitative Research checklist.

Results: The 3 main themes evaluating mothers' experiences and perceptions were generated: (1) changed perspective toward parenthood, (2) journey from pregnancy to after birth, and (3) a way forward. Mothers from the intervention group mostly had good perinatal experiences with sufficient support received, which elevated their emotional well-being and increased parenting involvement. Mothers in the control group, although satisfied with the hospital care received, were more stressed and shared a need for professional advice and extra support. Apart from technical enhancements, mothers also requested extended social support during early pregnancy up to 1 year postpartum, taking into consideration Asian cultural practices.

Conclusions: Mothers who received the intervention were overall satisfied with the support provided by the technology-based supportive educational parenting program. The success of the educational program in this study highlights the need to supplement standard care in hospitals with technology-based educational programs. Future research should include fathers' perceptions to attain an in-depth understanding of overall participants' experiences and needs in the future development of supportive and educational programs.

(J Med Internet Res 2019;21(2):e11065) doi:10.2196/11065

KEYWORDS

education; mothers; parenting; technology

Introduction

Background

The birth of a child is typically a joyous occasion, but it can elicit an amalgamation of emotions from parents, ranging from elation to anxiety. Transitioning into parenthood can be stressful as parents often feel overwhelmed while struggling to achieve a new balance between their work lives and infant care [1,2]. This is especially challenging for mothers who are the sole caregivers and lack social support in the postpartum period [3,4]. Mothers have to constantly monitor their babies' health, engage in baby care tasks, breastfeed, and experience varied sleep patterns, which cause high stress levels and increase their susceptibility to postpartum affective disorders such as postpartum anxiety and postpartum depression [5-7].

To this day, specific causes of postpartum psychiatric disorders remain unknown, although many studies suggest a multifactorial etiology [3,6,8]. However, certain protective factors such as social support received [9] and parental self-efficacy [10,11] are known to elevate maternal moods and keep postpartum psychiatric disorders at bay [10-12]. Maternal moods also adversely affect paternal psychological well-being [13], which subsequently disrupts parent-child bonding [14-16] and jeopardizes marital relationships [15] as negative emotions resonate off each other, creating a complex snowballing cycle. The resulting poor family dynamics and unconducive home environment, in turn, affect the child's physical, behavioral, and cognitive development [16,17].

To avoid negative social consequences inflicted by high maternal morbidity and poor child development, hospitals have implemented perinatal classes to prepare mothers for parenthood. However, there is a lack of continuity of care after hospital discharge [18], and the current perinatal care provided by hospitals is reportedly suboptimal and insufficient in meeting mothers' needs [19].

Postnatal follow-ups often only focus on breastfeeding and the physical well-being of mothers, often neglecting mothers' psychological and emotional well-being [20]. Moreover, in 1 study by Ong et al [19], because of short hospital stays, unawareness, lack of time, or financial constraints, very few mothers attended perinatal educational classes by hospitals. This can result in a lack of preparedness and poor parenting self-efficacy (PSE), which not only affect parental bonding and maternal emotional well-being but may also affect future childbearing decisions [21]. However, the Asian practice of a confinement period, which typically lasts for a month and mainly involves a confinement lady or the new mother's own mother or mother-in-law taking care of the special dietary needs of the new mother and infant care tasks, was shown to provide instrumental support and improve maternal outcomes during the postpartum period [22].

In the effort to help parents tide over challenging times in the perinatal period, many preventive and treatment programs were designed to improve parental outcomes and reduce postpartum affective disorders, which were mostly successful [10,23-25]. Successful intervention methods have included home visits [26], telephone call follow-ups [27], and face-to-face educational programs [10,28], but they required significant resource costs because of practitioner involvement. An ideal solution is to harness widely available and advanced technology to provide sustainable and cost-effective health care [29]. However, there is a lack of available literature on multimodal technology-based interventions administered during the perinatal period, with most available studies adopting a single modal approach [23,24]. The majority of existing literature [9,30,31] tends to focus on at-risk parents, often neglecting healthy groups of parents who are also going through this stressful postpartum period. Therefore, there is a need for a universal intervention that is also applicable to the healthy population. In addition, most randomized controlled trials (RCTs) only focused on the quantitative analysis of parental outcomes; hence, the qualitative evaluations of participants' perceptions of such perinatal technology-based interventions are limited. Considering that similar interventions may have different results depending on the setting and the population group involved, it is important to incorporate detailed insights on users' experiences and feedback on the intervention when gauging its success and effectiveness on parental outcomes [32]. A qualitative analysis also provides descriptive and exploratory insights that can better capture complex social needs and allow intervention methods to be tailored according to the population's needs [33]. Process-oriented studies evaluating parenting programs have increased substantially, but the programs are often home-based or face-to-face [34-38], which differ in administration and intervention components compared with technology-based programs. The heterogeneity of the studies in terms of findings or themes identified [39] and interview sample population (eg, fathers [40], immigrants [35], health care professionals [37], and low-income mothers [38]), warrants a need for a qualitative study specific to the current educational program.

Aim of the Study

Therefore, the aim of this study was to explore the experiences and perceptions of participants in a technology-based supportive educational parenting program (SEPP). The perinatal experiences of the control group were included to show the contrast in experiences from those who received the intervention.

Methods

Design and Setting

This study adopted a descriptive qualitative study design. It is a follow-up of an RCT that examined the effectiveness of a technology-based SEPP on parental outcomes during the perinatal period [41]. The original study took place in a tertiary hospital in Singapore, where a convenience sample of 118 couples was recruited. Participants in the control group only received routine hospital perinatal care, whereas those in the intervention group received additional phone-based educational

sessions during the perinatal period and access to a mobile health (mHealth) app for 1 month post delivery. The 2 antenatal and postnatal phone-based education sessions were administered by a professionally trained research assistant. The details of each SEPP component can be found in the study protocol [41]. Despite both parents being involved in the initial trial [41], the unavailability of fathers because of their work commitments limited the interviewees to mothers only. Mothers from the control group were also included in this study to provide a comparison of perinatal experiences and further insight into unmet maternal needs. Before the intervention study, participants were already informed of this optional extended qualitative arm of the study and that they would be further reimbursed for participation.

Recruitment

Recruitment took place 1 month postpartum (immediately after the completion of the intervention). Participants in this study were a subset of mothers involved in the initial intervention study [41]. Convenience sampling was conducted through the blasting of messages to all 118 participants (from both intervention and control groups) until data saturation was reached. Overall, 30 mothers volunteered to be interviewed, but data saturation was reached at the 14th participant, when no new findings emerged. Moreover, 2 additional interviews were performed to confirm the findings, resulting in a total of 16 participants (6 from the control group and 10 from the intervention group). The remaining 14 mothers agreed to withdraw from the study. The mothers were notified of the estimated length of the interview (approximately 30-60 min) and that all interviews would be audio-recorded for research purposes only.

Data Collection

A research assistant trained in qualitative interviewing techniques conducted the face-to-face interviews with the mothers at a time and location of each mother's convenience, which was typically during their postnatal follow-ups in the hospital. A semistructured interview guide was adopted with probing questions to attain a more comprehensive view of the effectiveness of the SEPP. The interview guides for both groups are presented in Textboxes 1 and 2. The interviews lasted approximately 30 min and were subsequently transcribed verbatim. Field notes were taken during the interviews to note nonverbal cues that were used to supplement the transcript.

Textbox 1. Qualitative evaluation semistructured guide for the control group.

Probing question
1. How did you prepare yourself during the pregnancy? Did you seek any support?
2. How did you bond with your baby? Did you think it was effective?
3. How did you provide newborn care? Did you find it efficient?
4. Did you seek any social support regarding newborn care or postdelivery care?
5. How did you feel emotionally during the pregnancy and post delivery?
6. How did you gain knowledge on self-care and newborn care during your pregnancy or post delivery?
7. Did you think any nurse's/midwife's expert advice will be beneficial during the pregnancy or post delivery?
8. Do you think it is worthwhile to spend extra time receiving educational programs provided by the hospital? Or any peer support groups?
9. How did you find the perinatal care provided to you by the hospital throughout your pregnancy, during labor, and after delivery?
10. Do you have any suggestions for the improvement of care or support for example during pregnancy, labor, and after delivery, provided to you as a parent?

Textbox 2. Qualitative evaluation semistructured guide for the intervention group.

Probing question
1. How did you feel participating in this research study?
2. Was this supportive educational parenting program helpful?
3. Did you find the supportive educational parenting program useful in improving your bond with your newborn?
4. Did you find the supportive educational parenting program useful in improving your self-efficacy in newborn care?
5. Did you find the supportive educational parenting program useful in improving your social support seeking behavior?
6. Did you find the supportive educational parenting program useful in improving your mood and decreasing your negative emotions?
7. Did you find the supportive educational parenting program useful in improving your knowledge about self-care and newborn care post delivery?
8. Did you find the nurse's/midwife's expert advice beneficial?
9. Do you think it was worthwhile spending extra time to receive this supportive educational parenting program intervention?
10. What were the strengths and weaknesses of the intervention?
11. How did you find the perinatal care provided to you by the hospital?
12. Do you have any suggestion for the improvement of perinatal support provided to you as a parent?

Data Analysis

Thematic analysis was conducted by 2 authors independently according to Braun and Clark's [42] 6 phases of analysis. The authors read the 16 transcribed interviews multiple times to familiarize themselves with the data. A manual color-coding method was employed to highlight different concepts and generate the initial codes. Related codes from all excerpts were collated to generate subthemes and overarching themes, which were reviewed comprehensively for homogeneity by both authors. Field notes were also constantly referred to as supplementary materials. Any discrepancies were discussed and clarified between the 2 authors until consensus was achieved. Upon further discussion, prominent themes constituting frequently reported overlapping data were selected from the authors' independent analysis, renamed, and included in the final analysis.

Ethical Considerations

Ethics approval was obtained from the National Health Group Domain Specific Review Board (Reference Number: NHG DSRB: 2016/00651) of the participating hospital. Relevant information about the research study was comprehensively explained to the participants, and the participants were informed of their rights to withdraw at any time during the study. After which, a written informed consent was obtained from each participant. Participation in this research was strictly voluntary and confidentiality was adhered to.

Results

Study findings were reported according to the Consolidated Criteria for Reporting Qualitative Research checklist [43].

Participant Characteristics

A total of 16 mothers were interviewed (6 control and 10 intervention). They had an age range of 23 to 41 years. The majority of the mothers were Chinese (n=9), followed by Malay (n=3), Indian (n=3), and Pakistani (n=1) (Table 1).

Themes and Subthemes

Overall, 3 major themes were identified (Figure 1): (1) changed perspective toward parenthood, (2) journey from pregnancy to after birth, and (3) a way forward.

Table 1. Description of the interviewed mothers (n=16).

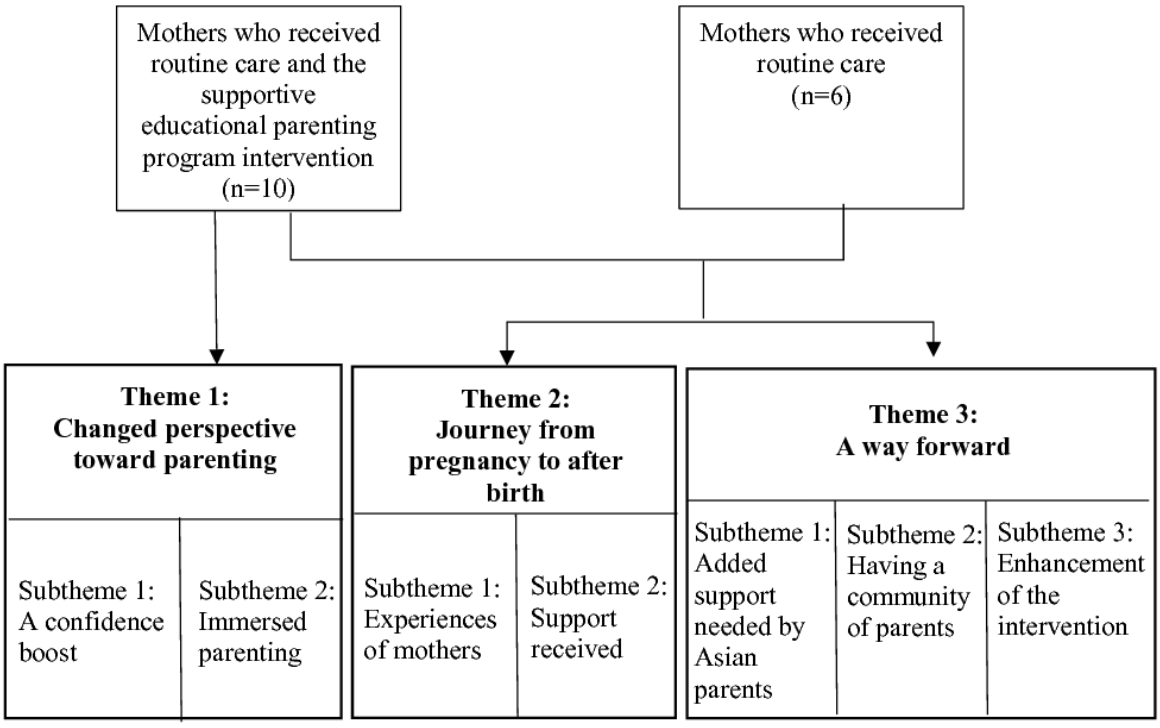
ID	Group	Age (years)	Ethnicity	Employment	Antenatal class	Type of delivery	Duration of maternity leave	Confinement period	Child birth order
C1	Control	31	Indian	Employed	No	Assisted	>12 weeks	No	First
C2	Control	27	Malay	Employed	Yes	NVD[a]	>12 weeks	Yes	Second
C3	Control	41	Malay	Unemployed	No	NVD	NS[b]	Yes	Second
C4	Control	34	Chinese	Employed	No	NVD	>12 weeks	Yes	First
C5	Control	34	Chinese	Employed	No	NVD	NS	Yes	Second
C6	Control	30	Chinese	Employed	Yes	C-sect[c]	≤12 weeks	Yes	Second
T1	Intervention	33	Chinese	Employed	No	C-sect	>12 weeks	Yes	First
T2	Intervention	34	Indian	Employed	No	NVD	≤12 weeks	No	First
T3	Intervention	28	Chinese	Employed	No	NVD	≤12 weeks	Yes	First
T4	Intervention	34	Chinese	Employed	No	NVD	>12 weeks	Yes	≥ third
T5	Intervention	23	Pakistani	Unemployed	Yes	C-sect	NS	Yes	First
T6	Intervention	25	Chinese	Unemployed	No	NVD	>12 weeks	Yes	First
T7	Intervention	28	Chinese	Unemployed	No	NVD	NS	Yes	First
T8	Intervention	29	Indian	Employed	No	C-sect	≤12 weeks	No	First
T9	Intervention	32	Malay	Employed	No	NVD	>12 weeks	Yes	≥ third
T10	Intervention	30	Chinese	Employed	Yes	NVD	≤12 weeks	Yes	First

[a]NVD: normal vaginal delivery.

[b]NS: not stated.

[c]C-sect: cesarean section.

Figure 1. Themes and subthemes.



Theme 1: Changed Perspective Toward Parenthood

Overall, mothers who received the intervention had improved emotional well-being and confidence levels in handling baby care tasks. This, in turn, led to more involved parenting and effective parental bonding, which encouraged their partners' active involvement as well.

Subtheme 1: A Confidence Boost

The majority of the mothers from the intervention group felt more relaxed, encouraged to seek help, and confident in taking care of their babies. One mother shared how the postnatal education phone call helped her to manage her emotions better and built her confidence:

> *The starting period when I was having difficulties, the phone call supported me a lot...to understand my own emotions because I was depressed at the start, and I was restless not having enough sleep...Also, I got some confidence in understanding my baby...* [T11]

A different mother shared how the mHealth app improved her baby care skills, which increased her confidence:

> *Through this application, there was a page on how we should take care of the baby's bathing and I have also seen a video...Like, from 0%, at least, I have improved to 70%. I can confidently say that I can take care of the baby, but I am still learning...* [T8]

In addition, the majority of the mothers stated how the forum component of the app encouraged help-seeking behaviors and motivated them. According to 1 mother:

> *Sometimes, mothers don't dare to ask questions...So, if you ask through such a platform, no one else knows that you ask questions and you still get the answers you need.* [T4]

Apart from the given anonymity and the reduced stigma, 2 mothers also shared how the forum:

> *encourages you, like you don't have to be shy to ask.* [T9]

And that:

> *it makes you more willing to speak up...they are also facing the same issues and, then, they are overcoming them well, which gives you more confidence and motivation to persevere.* [T3]

Subtheme 2: Immersed Parenting

On receiving this intervention, most mothers reported improved bonds with their babies. Moreover, 1 mother shared that the self-care tips and infant care information available on the app helped with maternal bonding:

> *For the first few weeks, I was not able to have a proper [maternal] attachment with the baby, but I read through how I have to take care of myself after my C-section, and, then, I had bonding with my baby and...I read through the positions on how to hold my baby...So, it helped me bond with my baby...* [T8]

Another mother shared that the information on the app allowed her to understand her baby better, which enhanced her bonding:

> *I found [out] how to take care of the newborn in terms of basic needs and stuff, making the baby more pleasant and happy. So, when he is happy, it is easier to bond...* [T1]

She also added:

> *Because of the video, my husband is more knowledgeable about breastfeeding...So, it is like a guidebook to him too...He is more supportive towards breastfeeding now.* [T1]

A first-time mother and an experienced mother also gave feedback on increased paternal involvement because of educational information and reminders provided by the app:

> *It helped my husband to bond with my baby...He hasn't carried any baby in his lifetime...So, the application...helped him a lot.* [T8]

> *Because she is already the third child, so bonding time with the father is not as much as with me...So, I guess with this app, it reminds us [that] the father needs to still be involved because, I mean, the father honestly spent a lot of time dealing with the other two [kids]. Sometimes, I will take over and, then, he comes around.* [T4]

Theme 2: Journey From Pregnancy to After Birth

This theme encapsulates the journeys and experiences of mothers during their last trimester to 1 month postpartum, including the support they received during this period. Overall, mothers from both groups (intervention and control) had positive pregnancy experiences; however, mothers in the control group were more stressed than those in the intervention group after delivery. Most mothers from both groups were also satisfied with the care and support received, but mothers from the control group highlighted the need for professional advice.

Subtheme 1: Experiences of Mothers

During pregnancy, a first-time mother from the control group reported being emotionally scared because she did not know what to expect. In contrast, first-time mothers in the intervention group felt good and excited. One of them shared:

> *My mood was very good...as they [phone call education] told me what to expect...I was excited to see my baby.* [T10]

After delivery, although some mothers from the control group reported that they were managing well, they also frequently used the words *stressed*, *anxious*, and *challenging*. A mother who had undergone a cesarean section shared her postnatal experiences:

> *We were quite helpless sometimes because we are new parents...I felt incompetent...It was more on anxiety and, sometimes, anger at myself for not being able to recover in time...I didn't really know whether, with my wound [still recovering], I can actually care for her in a right, responsible way.* [C4]

On the contrary, mothers from the intervention group reported feeling less anxious:

> *Actually, I feel more at ease, not worried. I am not that worried that I am feeding wrongly and stuff...as my queries have been answered in a timely manner.* [T2]

Subtheme 2: Support Received

Mothers from both groups were overall satisfied with the hospital care received and the social support received from their loved ones:

> *I think the perinatal care itself...was very good according to me. Even after the delivery, we did get visits to see a better doctor and so on. So, I didn't feel that anything was lacking, so I was quite happy with the service that I got overall.* [C5]

> *I am really pleased to be here. For once, they responded very quickly to the request...and every nurse was very willing to do things for you...and they are very nice when approaching you and teaching you...so I feel like I am in good hands.* [T4]

However, despite the availability of external platforms (ie, the internet, Facebook groups, books, or confinement ladies) on infant and self-care information, several mothers from the control group still shared the need for advice and emotional support from a health care professional whom they could trust. One mother justified her need for added professional advice despite having support from a confinement lady, an experienced family member, and the internet:

> *If we have other information, like more on scientific ones...I think it is more helpful...would be better if we have somebody else...an expert...to tell us exactly what is the correct one because...whatever we have, we actually learnt from the confinement lady, so we are not sure whether that is really, like, the correct thing to do, la...and if I tell my sister-in-law the situation, she might come across [different problems] because every baby is different, so it [her solution] might not be applicable to my baby...Sometimes, we share...but it might not actually be the solution for the problem...Sometimes, it is reassuring to know that it comes from a nurse or midwife rather than any Google or website.* [C4]

On the other hand, mothers who received the intervention felt more prepared, reassured, and encouraged by the professional advice and existing external support they got:

> *It definitely helps us to get answers [from the midwife], not spend time worrying, and, then, you Google and, sometimes, we are even more scared. Because we got answers from professionals and they also always replied in a very positive tone, which really helped us to be very confident and just not worry too much...* [T4]

An experienced mother shared how the app served as a *refresher course*:

> *The gap with my third one is quite far, about 5 years. I actually forgot what I must do. What is good for the baby? When somebody replies to you, "Oh, okay, I know this." It [the app] is something like a refresher course.* [T9]

Theme 3: A Way Forward

This theme explores the feedback and opinions from mothers in both groups, which primarily acknowledges the need for additional social support in the Asian context, the importance of being part of a community of new parents, and suggested improvements to this intervention study.

Subtheme 1: Added Support Needed by Asian Parents

Despite the Asian culture being known for its collectivistic culture and its strong emphasis on family-oriented values, certain cultural taboos restrict help-seeking behaviors during early pregnancy. According to a mother:

> *For the Chinese, the first month of the first trimester is a hush-hush kind of thing because we are a bit pantang [superstitious], you know, they [are] afraid that if you tell everybody that you are pregnant, then...[something bad will happen].* [C4]

In addition, although confinement practices were thought to provide more support to mothers during the first month postpartum, 1 mother confessed that having a confinement lady only delayed parenting stress by 1 month. Another mother shared the need for more support, especially after 1 month postpartum:

> *After the confinement lady leaves, then that's when most of the mothers are left on their own and when the fathers go back to work, and they [are] left alone with the baby...[There] will be more things that you [are] worried about, questions they may ask...[but] the people are not around anymore...* [T10]

Subtheme 2: Having a Community of Parents

Mothers from both groups indicated strong preferences for chat groups and interactions with other new parents as it provided communal support, mutual understanding, and encouraged active parenting:

> *[When] my husband sees more daddies asking questions, he gets more involved. He also reads through and, then, he also asks questions himself* [T3]

> *It will be really helpful to have added support from the hospital...because someone who understands you in terms of whatever you say will really make a difference compared to, you know, you speak to your husband and he is not able to understand what you are trying to say.* [C2]

> *Let's say if you are lonely and a struggling mum, you want somebody to listen to you. So, I feel like you need somebody to talk during that crucial time...I thought with this app...somebody is over there, you know, like, watching me, so, okay, I can get through this, you know.* [T6]

Subtheme 3: Enhancement of the Intervention

With regard to the app's features, 1 mother suggested appointment checking for the baby, such as when vaccinations were due. Another mother suggested a feeding reminder:

> *When we feed, then we get very busy, then we really forget when the last feed was...When you are passing the baby to another caregiver, then the caregiver doesn't know when the last time the baby was fed.* [C4]

A few mothers also reported the need for a segregation and categorization of forum queries so that it would be easier for them to view as well as receive notifications whenever someone replied.

Content-wise, 1 mother suggested adding prenatal information on in vitro fertilization–assisted reproduction to facilitate informed decision making and increase the preparedness of couples undergoing it. She shared her experience:

> *We underwent the assisted reproduction programme. For the first time, we were quite clueless...because not a lot of people know what it is exactly. So, I think it would be helpful to actually have some sort of guideline to actually know what to expect and so it is not scary...* [C4]

Although the majority of the mothers in the control group suggested a need for an informational app, most of the mothers in the intervention group requested for extended access to the app for approximately 3 months up to 1 year. Upon termination of the app, 1 mother felt that she was out of it and could not access the app, so it was disappointing to her. Mothers from both the intervention and control groups shared their needs for prolonged use of the app to address different baby issues during the critical first year:

> *One year is a very crucial time for the baby. I thought I can go into the application, but I realize I do not have the application access. I don't have enough duration.* [T8]

> *If there are queries about the infants at least half a year [postpartum]...I think it will be good if there is some sort of, like, platform or support instead of running to the hospital every single time...An app would be a good option.* [C4]

The same mother from the control group added:

> *If it [the app] extended to the prenatal period, it will be good, because, for me, I actually had bleeding in the first trimester, so we were a bit stressed and very scared.* [C4]

A different mother also suggested the extension of the availability of the app to prenatal stages to increase maternal preparedness before the mother gets too busy with the baby:

> *Once when the baby comes, then you are...engaged in being with the baby, doing things, you are sleepless also...When you are pregnant, you have more time, so you can go through the educational videos or seek expert advice.* [T2]

In addition, several mothers from the intervention group requested to have faster responses from the health care professional or to have a live chat as they were usually desperate for answers and the slow responses made them anxious. Overall, 3 mothers shared their experiences:

> *Our query was answered a bit late...We were really anxious and waiting...So, maybe, you know, this process is faster instead of one person, if [there are] two to three people [replying]. If one is busy, the other one can reply.* [T2]

> *When I asked any question, then it takes a day [for it] to be answered. So, if I have some emergency for my baby or anything, I am very much in trouble...* [T5]

> *I thought it's...online when somebody is over there to answer your questions, like instantly...Times like that, my baby was crying, I can't wait for 24 hours...You are so desperate to get the answers really, really fast...* [T6]

Discussion

Principal Findings

In general, participants in the intervention group were more positive, calm, and well adjusted in the postpartum period. On the other hand, mothers in the control group were most likely to report feeling anxious, lost, and incompetent during the postpartum period. The effectiveness of the technology-based educational intervention in reinforcing parental self-efficacy and improving the emotional well-being of parents during the perinatal stage was similarly reflected in previous studies [24,44-46]. Aligning parents' expectations with reality and supplying them with parenting information also helped to ease their transition into parenthood [47]. According to Bandura [48], self-efficacy is the optimistic self-belief in our competence or chances of successfully accomplishing a task and producing a favorable outcome. Compared with mothers in the control group, those in the intervention group reported increased parenting confidence and self-efficacy, which also led to other positive parental outcomes such as better parental bonding and improved emotional well-being. This corresponds with previous intervention studies that reported positive intercorrelations between parental outcomes, including social support received [24,49].

With regard to social support, mothers in the intervention group who used the forum felt more reassured and less alone knowing that other parents encountered the same problems. Although sharing with a professional would generate a clinical or scientific reasoning for their physical experiences, sharing with other mothers produced communal responses to the emotional experiences themselves [50]. Therefore, a cohesive community of parents with shared experiences is essential for providing emotional support during the perinatal period [33,34].

The majority of the mothers in this study followed a confinement period and were generally satisfied with the instrumental support received; however, this support was only temporary and delayed mothers' actual *assumption of duty*, thus causing stress to a

minority of the mothers after the confinement period. A plausible reason is the formation of dependence on caregivers during the 1-month confinement period, which hinders the development of parental self-efficacy; however, further studies are needed to verify this claim. Although most studies agree that confinement practices generally improve parental outcomes, it must not be assumed that the confinement period is available to or practiced by all Asians because of the additional financial burden it imposes [22]. Without anyone to turn to, the forum provided a safe space for parents to share their experiences and queries. The anonymity feature of the forum provided a safe haven for mothers to seek answers by reducing the stigma and fear of appearing as incompetent parents [9,51]. Similarly, in other studies [52,53], the anonymity guaranteed by the technology-based intervention greatly reduced help-seeking barriers across the population regardless of medical conditions [54]. In addition, the Chinese superstition of not announcing one's pregnancy during the first trimester to prevent bad luck serves as a help-seeking barrier for parents in the prenatal stages. This highlights the need for added support by Asian parents in the prenatal period and postconfinement period, which can be met by increasing the duration of accessibility to the app and professional support.

On receiving the intervention, all mothers reported more confidence in their infant care abilities. This increased sense of PSE promoted more immersive parenting, whereby parents were more involved in infant care tasks and actively bonded with their children. Corresponding with Bandura's social learning theory [48], parental self-efficacy is achieved through mastery experience, verbal persuasion, vicarious experience, and physiological and emotional states. Verbal persuasion from other mothers and the midwife through the app's forum also served as a form of appraisal support, which motivated mothers to persevere. Being continually assured and reminded that they had the abilities to be proficient at infant care tasks encouraged mothers to work toward achieving better infant care skills [55].

In addition, medical and parenting advice from health care professionals was deemed trustworthy and necessary by parents to ease their worries and anxiety in the perinatal period. Mothers who lacked professional advice reported poorer self-esteem and emotional well-being, which negatively influenced mothers' sense of self-efficacy. This is congruent to the emotional aspect of Bandura's learning theory [48], whereby a poor emotional or physical state adversely affects PSE. Therefore, receiving guidance from a health care professional is a vital facet to parents developing a sense of security during the postnatal period [24,56].

Various features of the intervention also enhanced mothers' vicarious experiences, which boosted their confidence and self-efficacy. First, the forum established a platform for parents to post queries, share experiences, and provide advice that was moderated by the midwife. Mothers were better equipped with parenting skills through the observation and imitation of other parents' techniques. In addition, there was increased paternal involvement when fathers noticed other fathers' active involvement in the forum. The imitation bias of mimicking a similar individual increased the self-relevance and adaptive value of the information learned [48,57]. A study further discovered that own-gender imitation behavior actually elicits a reward response in the brain [58]. This emphasizes the need for the extension of technology-based supportive community networks to encourage active and immersive parenting.

In consideration of parental preferences for the dissemination of information through multimedia [59], the audio and video recordings on baby care skills proved a useful reference guide for parents to model after. Studies have also reported higher memory retention through visual imagery than auditory or written contents [60], which suggests the effectiveness of the app in boosting parental learning and retention compared with didactic perinatal lectures provided by the hospital. Mothers also recommended a categorization of information on the app for ease of reference and learning. Likewise, Bandura stated that organized information better enhances memory than fragmented information [48].

However, the technology-based SEPP failed to incorporate the mastery experience facet of Bandura's social learning theory [48] as only mothers who had given birth previously could achieve mastery experience. The technology-based intervention did not provide a direct mastery of infant and self-care skills, which is equally vital for promoting self-efficacy. In Saunders et al's study [61], mothers who received hands-on practice improved faster and better incorporated those strategies into their parenting behaviors. Therefore, there is still a necessity for hands-on practice and positive live guidance from health care professionals, suggesting that technology-based interventions should be used as a supplementary tool to existing hospital care and not as a replacement [62].

Overall, the SEPP was generally well received, with certain feedback regarding the expansion of content database and information coverage (eg, in vitro fertilization) as well as the app's availability during the antenatal period. With such improvements, the SEPP can potentially address the common issue of the lack of informed decision making during the antenatal period [63]. Suggestions to include features such as appointment reminders and the categorization of forum queries will also enhance user friendliness and increase usability, creating an all-in-one go-to app for parents. In many studies [24,64], asynchronous communication with health care professionals was reported to be of utmost importance to mothers who see such apps as a lifeline. Although our study has adopted a similar approach through the Web-based forum, mothers gave feedback that the responses from the midwife were too slow, which was not reported in Danjborg et al's study [24], which used Web-based chat instead. Therefore, future studies can consider including an instant messaging function by using an Web-based chat to contact health care professionals directly.

Implication for Future Research and Clinical Practice

Given the rapid advancement of technology and the various benefits technology provides, such as cost-effectiveness, flexibility, accessibility, and its potentially extensive outreach to the wider community, this should serve as an impetus for health care sectors to dedicate more resources to further develop such technology-based interventions.

In a collectivist Asian society that strongly places emphasis on family-oriented values, one's family is often the immediate source of emotional support, and the lack of it may adversely affect a mother's emotional well-being [65]. Along with the rise of paternal involvement in parenting, future interventions should tailor features to encourage involvement and support from other family members such as grandparents. With an increasing dependency on Web-based communities for support, future studies should also emphasize on building more interactive or live communities for parents to interact and exchange experiences.

Strengths and Limitations

Unique to this study is the added insight on Asian cultures such as taboos and traditions during the perinatal period. To the best of our knowledge, this is also the first qualitative study to provide in-depth insight into participants' views on a multimodal technology-based SEPP that was implemented during the perinatal period. A major limitation of this study was the unavailability of fathers during the follow-up interviews, which impeded our understanding of paternal experiences and perceptions of the SEPP intervention. However, mothers who were interviewed mostly reported positive receptivity by fathers and increased fathers' involvement in infant care. The once-off nature of this study disallowed the comparison of pre- and posttest perceptions of mothers, making it difficult to determine the effectiveness of the intervention. In addition, using the convenience sampling method and voluntary participation did not provide a good representation of mothers in the control group. Therefore, future studies should consider employing random or purposive sampling methods when recruiting participants for interviews. The mobile app was a novel approach and the last intervention phase of the SEPP; hence, because of possible recency effect, participants' responses were more inclined to the mobile app. Future qualitative studies should, therefore, consider constructing component-specific interview questions to have a more holistic evaluation of the entire intervention program.

Conclusions

This study examined the perceptions of participants who received routine care and those who received the technology-based SEPP intervention in addition to routine care in the perinatal period. A comparison of mothers' perceptions in both the groups allowed a deeper insight into the needs of mothers in the perinatal period, with the intervention group being generally satisfied with the SEPP. The multifeature technology-based intervention was effective in improving parental outcomes and was well received by parents. This calls for a higher allocation of resources to further develop and tailor technology-based interventions that can supplement existing hospital perinatal care and be implemented in a wider community of parents.

Acknowledgments

The authors would like to thank the funding agency, National University of Singapore, for the National University Health System Clinician Research grant FY2015 (Reference Number: NUHSRO/2016/023/CRG/02). The authors would also like to thank the National University Health System, Medical Publications Support Unit, for assistance in the language editing of this manuscript.

Conflicts of Interest

None declared.

Editorial notice: This is a companion article to "Effectiveness of a Technology-Based Supportive Educational Parenting Program on Parental Outcomes (Part 1): Randomized Controlled Trial" by Shorey et al (J Med Internet Res 2019;21(2):e10816) which can be found at https://www.jmir.org/2019/2/e10816/.

Multimedia Appendix 1

COREQ Checklist.

[PDF File (Adobe PDF File), 486KB - jmir_v21i2e11065_fig.pdf]

References

1. Perry-Jenkins M, Goldberg AE, Pierce CP, Sayer AG. Shift work, role overload, and the transition to parenthood. J Marriage Fam 2007;69(1):123-138 [FREE Full text] [doi: 10.1111/j.1741-3737.2006.00349.x] [Medline: 20216932]
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. Beck CT. Predictors of postpartum depression: an update. Nurs Res 2001;50(5):275-285. [Medline: 11570712]
4. Sepa A, Frodi A, Ludvigsson J. Psychosocial correlates of parenting stress, lack of support and lack of confidence/security. Scand J Psychol 2004 Apr;45(2):169-179. [Medline: 15016271]
5. Vesga-López O, Blanco C, Keyes K, Olfson M, Grant BF, Hasin DS. Psychiatric disorders in pregnant and postpartum women in the United States. Arch Gen Psychiatry 2008 Jul;65(7):805-815 [FREE Full text] [doi: 10.1001/archpsyc.65.7.805] [Medline: 18606953]

6. 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]
7. 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]
8. Brummelte S, Galea LA. Postpartum depression: etiology, treatment and consequences for maternal care. Horm Behav 2016 Jan;77:153-166. [doi: 10.1016/j.yhbeh.2015.08.008] [Medline: 26319224]
9. Negron R, Martin A, Almog M, Balbierz A, Howell EA. Social support during the postpartum period: mothers' views on needs, expectations, and mobilization of support. Matern Child Health J 2013 May;17(4):616-623 [FREE Full text] [doi: 10.1007/s10995-012-1037-4] [Medline: 22581378]
10. Shorey S, Chan SW, Chong YS, He H. 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]
11. Leahy-Warren P, McCarthy G, Corcoran P. First-time mothers: social support, maternal parental self-efficacy and postnatal depression. J Clin Nurs 2012 Feb;21(3-4):388-397. [doi: 10.1111/j.1365-2702.2011.03701.x] [Medline: 21435059]
12. Seymour M, Giallo R, Cooklin A, Dunning M. Maternal anxiety, risk factors and parenting in the first post-natal year. Child Care Health Dev 2015 Mar;41(2):314-323. [doi: 10.1111/cch.12178] [Medline: 25074519]
13. 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]
14. Feldman R. Parents' convergence on sharing and marital satisfaction, father involvement, and parent-child relationship at the transition to parenthood. Infant Ment Health J 2000;21(3):176-191. [doi: 10.1002/1097-0355(200007)21:3<176::AID-IMHJ3>3.0.CO;2-4]
15. Davey SJ, Dziurawiec S, O'Brien-Malone A. Men's voices: postnatal depression from the perspective of male partners. Qual Health Res 2006 Feb;16(2):206-220. [doi: 10.1177/1049732305281950] [Medline: 16394210]
16. Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, et al. Effects of perinatal mental disorders on the fetus and child. Lancet 2014 Nov 15;384(9956):1800-1819. [doi: 10.1016/S0140-6736(14)61277-0] [Medline: 25455250]
17. Letourneau NL, Dennis C, Benzies K, Duffett-Leger L, Stewart M, Tryphonopoulos PD, et al. Postpartum depression is a family affair: addressing the impact on mothers, fathers, and children. Issues Ment Health Nurs 2012 Jul;33(7):445-457. [doi: 10.3109/01612840.2012.673054] [Medline: 22757597]
18. Fogel N. The inadequacies in postnatal health care. Curr Drug Res Pract 2017 Jan;7(1):16-17. [doi: 10.1016/j.cmrp.2016.12.006]
19. 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]
20. Ngai FW, Chan SW, Ip WY. The effects of a childbirth psychoeducation program on learned resourcefulness, maternal role competence and perinatal depression: a quasi-experiment. Int J Nurs Stud 2009 Oct;46(10):1298-1306. [doi: 10.1016/j.ijnurstu.2009.03.007] [Medline: 19361801]
21. Waldenström U, Hildingsson I, Rubertsson C, Rådestad I. A negative birth experience: prevalence and risk factors in a national sample. Birth 2004 Mar;31(1):17-27. [Medline: 15015989]
22. Wong J, Fisher J. The role of traditional confinement practices in determining postpartum depression in women in Chinese cultures: a systematic review of the English language evidence. J Affect Disord 2009 Aug;116(3):161-169. [doi: 10.1016/j.jad.2008.11.002] [Medline: 19135261]
23. 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 Mar 9;73(9):2103-2117. [doi: 10.1111/jan.13293] [Medline: 28276086]
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. Doherty WJ, Erickson MF, LaRossa R. An intervention to increase father involvement and skills with infants during the transition to parenthood. J Fam Psychol 2006 Sep;20(3):438-447. [doi: 10.1037/0893-3200.20.3.438] [Medline: 16938002]
26. Shorey S, Chan SW, Chong YS, He HG. Perceptions of primiparas on a postnatal psychoeducation programme: the process evaluation. Midwifery 2015 Jan;31(1):155-163. [doi: 10.1016/j.midw.2014.08.001] [Medline: 25174540]
27. Downes MJ, Mervin MC, Byrnes JM, Scuffham PA. Telephone consultations for general practice: a systematic review. Syst Rev 2017 Jul 3;6(1):128 [FREE Full text] [doi: 10.1186/s13643-017-0529-0] [Medline: 28673333]
28. Shapiro AF, Nahm EY, Gottman JM, Content K. Bringing baby home together: examining the impact of a couple-focused intervention on the dynamics within family play. Am J Orthopsychiatry 2011 Jul;81(3):337-350. [doi: 10.1111/j.1939-0025.2011.01102.x] [Medline: 21729014]
29. 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]

30. Dennis CL, Hodnett E, Kenton L, Weston J, Zupancic J, Stewart DE, et al. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. Br Med J 2009 Jan 16;338:a3064. [doi: 10.1136/bmj.a3064]
31. Barlow J, Smailagic N, Huband N, Roloff V, Bennett C. Group-based parent training programmes for improving parental psychosocial health. Cochrane Database Syst Rev 2012 Jun 13(6):CD002020. [doi: 10.1002/14651858.CD002020.pub3] [Medline: 22696327]
32. Moore G, Audrey S, Barker M, Bond L, Bonell C, Cooper C, et al. Process evaluation in complex public health intervention studies: the need for guidance. J Epidemiol Community Health 2014 Feb;68(2):101-102 [FREE Full text] [doi: 10.1136/jech-2013-202869] [Medline: 24022816]
33. Haq M. A comparative analysis of qualitative and quantitative research methods and a justification for adopting mixed methods in social research. 2014 Jun Presented at: Annual PhD Conference; 2014 June; University of Bradford Business School of Management, UK p. 1-22. [doi: 10.13140/RG.2.1.1945.8640]
34. Bertoni A, Donato S, Morgano A, Iafrate R, Rosnati R. A qualitative evaluation of a preventive intervention for parents: the groups for family enrichment_parent version (GFE_P). J Prev Interv Community 2017;45(3):215-229. [doi: 10.1080/10852352.2016.1198135] [Medline: 28641064]
35. Yuen LH. New immigrant parents' experiences in a parent education programme. Int J Early Years Educ 2017 Oct 17;1:1-14. [doi: 10.1080/09669760.2017.1390445]
36. Norling-Gustafsson A, Skaghammar K, Adolfsson A. Expectant parents' experiences of parental education within the antenatal health service. Psychol Res Behav Manag 2011;4:159-167 [FREE Full text] [doi: 10.2147/PRBM.S22861] [Medline: 22241955]
37. Wager KA, Lee FW, Bradford WD, Jones W, Kilpatrick AO. Qualitative evaluation of South Carolina's Postpartum/Infant Home Visit program. Public Health Nurs 2004;21(6):541-546. [doi: 10.1111/j.0737-1209.2004.21606.x] [Medline: 15566559]
38. Hartwig SA, Robinson LR, Comeau DL, Claussen AH, Perou R. Maternal perceptions of parenting following an evidence-based parenting program: a qualitative study of legacy for children. Infant Ment Health J 2017 Dec;38(4):499-513 [FREE Full text] [doi: 10.1002/imhj.21657] [Medline: 28658506]
39. Shorey S, Yang YY, Dennis C. A mobile health app-based postnatal educational program (Home-but not Alone): descriptive qualitative study. J Med Internet Res 2018 Apr 19;20(4):e119 [FREE Full text] [doi: 10.2196/jmir.9188] [Medline: 29674314]
40. Scourfield J, Nasiruddin Q. Religious adaptation of a parenting programme: process evaluation of the family links Islamic values course for muslim fathers. Child Care Health Dev 2015 Sep;41(5):697-703 [FREE Full text] [doi: 10.1111/cch.12228] [Medline: 25649634]
41. Shorey S, Ng YP, Siew AL, Yoong J, Mörelius E. Effectiveness of a technology-based supportive educational parenting program on parental outcomes in Singapore: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Jan 10;7(1):e4 [FREE Full text] [doi: 10.2196/resprot.8062] [Medline: 29321127]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Res Psychol 2006;3:77-101. [doi: 10.1191/1478088706qp063oa]
43. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar 23;340:c869 [FREE Full text] [Medline: 20332511]
44. Niela-Vilén H, Axelin A, Salanterä S, Melender HL. Internet-based peer support for parents: a systematic integrative review. Int J Nurs Stud 2014 Nov;51(11):1524-1537. [doi: 10.1016/j.ijnurstu.2014.06.009] [Medline: 24998788]
45. Hudson DB, Campbell-Grossman C, Fleck MO, Elek SM, Shipman A. Effects of the new fathers network on first-time fathers' parenting self-efficacy and parenting satisfaction during the transition to parenthood. Issues Compr Pediatr Nurs 2003;26(4):217-229. [doi: 10.1080/01460860390246687] [Medline: 14630529]
46. Salonen AH, Kaunonen M, Astedt-Kurki P, Järvenpää AL, Isoaho H, Tarkka MT. Effectiveness of an internet-based intervention enhancing Finnish parents' parenting satisfaction and parenting self-efficacy during the postpartum period. Midwifery 2011 Dec;27(6):832-841. [doi: 10.1016/j.midw.2010.08.010] [Medline: 20932612]
47. Barimani M, Vikström A, Rosander M, Forslund FK, Berlin A. Facilitating and inhibiting factors in transition to parenthood - ways in which health professionals can support parents. Scand J Caring Sci 2017 Sep;31(3):537-546. [doi: 10.1111/scs.12367] [Medline: 28144992]
48. Bandura A. Social learning theory. Englewood Cliffs, NJ: Prentice Hall; Nov 11, 1976:256.
49. 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]
50. 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 Apr 19;20(4):e119 [FREE Full text] [doi: 10.2196/jmir.9188] [Medline: 29674314]
51. 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]
52. Senn TE, Braksmajer A, Coury-Doniger P, Urban MA, Carey MP. Mobile technology use and desired technology-based intervention characteristics among HIV+ Black men who have sex with men. AIDS Care 2017 Apr;29(4):423-427. [doi: 10.1080/09540121.2016.1220479] [Medline: 27535069]

53. Raaijmakers LC, Pouwels S, Berghuis KA, Nienhuijs SW. Technology-based interventions in the treatment of overweight and obesity: a systematic review. Appetite 2015 Dec;95:138-151. [doi: 10.1016/j.appet.2015.07.008] [Medline: 26165415]
54. Baum A, Revenson TA, Singer JE. Handbook of health psychology. New York: Psychology Press; 2012:882.
55. 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]
56. 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]
57. Henrich J, McElreath R. The evolution of cultural evolution. Evol Anthropol 2003 May 19;12(3):123-135. [doi: 10.1002/evan.10110]
58. Losin EA, Iacoboni M, Martin A, Dapretto M. Own-gender imitation activates the brain's reward circuitry. Soc Cogn Affect Neurosci 2012 Oct;7(7):804-810 [FREE Full text] [doi: 10.1093/scan/nsr055] [Medline: 22383803]
59. Shorey S, Dennis CL, Bridge S, Chong YS, Holroyd E, He HG. First-time fathers' postnatal experiences and support needs: a descriptive qualitative study. J Adv Nurs 2017 Dec;73(12):2987-2996. [doi: 10.1111/jan.13349] [Medline: 28557020]
60. Cohen MA, Horowitz TS, Wolfe JM. Auditory recognition memory is inferior to visual recognition memory. Proc Natl Acad Sci U S A 2009 Apr 7;106(14):6008-6010 [FREE Full text] [doi: 10.1073/pnas.0811884106] [Medline: 19307569]
61. Saunders R, McFarland-Piazza L, Jacobvitz D, Hazen-Swann N, Burton R. Maternal knowledge and behaviors regarding discipline: the effectiveness of a hands-on education program in positive guidance. J Child Fam Stud 2012 Apr 8;22(3):322-334. [doi: 10.1007/s10826-012-9581-5]
62. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: in review, in use, and into the future. Prof Psychol Res Pr 2003;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
63. Goldberg H. Informed decision making in maternity care. J Perinat Educ 2009;18(1):32-40 [FREE Full text] [doi: 10.1624/105812409X396219] [Medline: 19436598]
64. Bjoernes CD, Laursen BS, Delmar C, Cummings E, Nøhr C. A dialogue-based web application enhances personalized access to healthcare professionals--an intervention study. BMC Med Inform Decis Mak 2012 Sep 4;12:96 [FREE Full text] [doi: 10.1186/1472-6947-12-96] [Medline: 22947231]
65. Jones E, Coast E. Social relationships and postpartum depression in South Asia: a systematic review. Int J Soc Psychiatry 2013 Nov;59(7):690-700. [doi: 10.1177/0020764012453675] [Medline: 22851133]

Abbreviations

C-sect: cesarean section
mHealth: mobile health
NHG DSRB: National Health Group Domain Specific Review Board
NS: not stated
NVD: normal vaginal delivery
PSE: parenting self-efficacy
RCT: randomized controlled trial
SEPP: supportive educational parenting program

Edited by G Eysenbach; submitted 16.05.18; peer-reviewed by J Day, G Hickey; comments to author 02.09.18; revised version received 26.10.18; accepted 12.11.18; published 13.02.19.

Please cite as:
Shorey S, Ng ED
Evaluation of Mothers' Perceptions of a Technology-Based Supportive Educational Parenting Program (Part 2): Qualitative Study
J Med Internet Res 2019;21(2):e11065
URL: https://www.jmir.org/2019/2/e11065/
doi:10.2196/11065
PMID:30758295

Original Paper

Resilience@Work Mindfulness Program: Results From a Cluster Randomized Controlled Trial With First Responders

Sadhbh Joyce[1,2], MClinNeuroPsych; Fiona Shand[2], PhD; Tara J Lal[3], BSc (Hons); Brendan Mott[3], BExScRehab (Hons); Richard A Bryant[4], PhD; Samuel B Harvey[1,2], PhD

[1]School of Psychiatry, Faculty of Medicine, University of New South Wales, Randwick, Australia

[2]The Black Dog Institute, Hospital Road, Randwick, Australia

[3]Fire and Rescue New South Wales, Sydney, Australia

[4]University of New South Wales, School of Psychology, Sydney, Australia

Corresponding Author:
Sadhbh Joyce, MClinNeuroPsych
School of Psychiatry
Faculty of Medicine
University of New South Wales
The Black Dog Institute
Hospital Road
Randwick, 2031
Australia
Phone: 61 401297711
Email: sadhbh.joyce@unsw.edu.au

Abstract

Background: A growing body of research suggests that resilience training can play a pivotal role in creating mentally healthy workplaces, particularly with regard to protecting the long-term well-being of workers. Emerging research describes positive outcomes from various types of resilience training programs (RTPs) among different occupational groups. One specific group of workers that may benefit from this form of proactive resilience training is first responders. Given the nature of their work, first responders are frequently exposed to stressful circumstances and potentially traumatic events, which may impact their overall resilience and well-being over time.

Objective: This study aimed to examine whether a mindfulness-based RTP (the Resilience@Work [RAW] Mindfulness Program) delivered via the internet can effectively enhance resilience among a group of high-risk workers.

Methods: We conducted a cluster randomized controlled trial (RCT) comprising 24 Primary Fire and Rescue and Hazmat stations within New South Wales. Overall, 12 stations were assigned to the 6-session RAW Mindfulness Program and 12 stations were assigned to the control condition. A total of 143 active full-time firefighters enrolled in the study. Questionnaires were administered at baseline, immediately post training, and at 6-month follow-up. Measurements examined change in both adaptive and bounce-back resilience as well as several secondary outcomes examining resilience resources and acceptance and mindfulness skills.

Results: Mixed-model repeated measures analysis found that the overall test of group-by-time interaction was significant (P=.008), with the intervention group increasing in adaptive resilience over time. However, no significant differences were found between the intervention group and the control group in terms of change in bounce-back resilience (P=.09). At 6-month follow-up, the group receiving the RAW intervention had an average increase in their resilience score of 1.3, equating to a moderate-to-large effect size compared with the control group of 0.73 (95% CI 0.38-1.06). Per-protocol analysis found that compared with the control group, the greatest improvements in adaptive resilience were observed among those who completed most of the RAW program, that is, 5 to 6 sessions (P=.002).

Conclusions: The results of this RCT suggest that mindfulness-based resilience training delivered in an internet format can create improvements in adaptive resilience and related resources among high-risk workers, such as first responders. Despite a number of limitations, the results of this study suggest that the RAW Mindfulness Program is an effective, scalable, and practical means of delivering online resilience training in high-risk workplace settings. To the best of our knowledge, this is the first time a mindfulness-based RTP delivered entirely via the internet has been tested in the workplace.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12615000574549; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=368296 (Archived by WebCite at http://www.webcitation.org/75w4xtrpw).

(J Med Internet Res 2019;21(2):e12894) doi:10.2196/12894

KEYWORDS

resilience training; workplace mental health; occupational health; well-being; online intervention; employee resilience; health and safety; psychological health; first responders

Introduction

Background

Poor mental health is the leading cause of absenteeism and long-term disability worldwide [1-3], costing the global economy an estimated US $1 trillion each year in lost productivity [4]. Most of the common mental health difficulties experienced by workers, such as depression and anxiety, are both treatable and, in some cases, preventable [5-7]. Policy makers and researchers are now concentrating their efforts on how to address this public health issue, with a specific focus on what can be done within the workplace. Research continues to highlight the important role the workplace can play in the prevention, treatment, and management of common mental health conditions [8]. Developing mentally healthy workplaces requires a multifaceted approach across all levels of an organization and should include proactive and preventative strategies [9].

One proactive approach receiving increased attention from both public and private sectors is individual psychological resilience training. Resilience can be viewed as a dynamic process reflecting a person's ability to adapt, manage, and recover effectively from stressful experiences and adverse circumstances [10]. Employee resilience has previously been identified as a key component in creating mentally healthy workplaces [9]. An emerging body of research describes positive outcomes from various types of resilience training programs (RTPs) among different groups, including medical specialists, factory workers, nurses, youth workers, and public servants [11-16]. In addition, research among high-risk occupations, such as firefighters, police, paramedics, and military personnel, highlights the benefits of resilience training among those who frequently experience high-stress situations as an inherent aspect of their work [17-19]. Conversely, several larger trials with US army personnel and more recently with London Ambulance Service in the United Kingdom reported more limited improvements following resilience training [20,21].

Determining which RTPs are beneficial to groups such as emergency workers is particularly important for several reasons. First, these workers play a vital role in providing critical services that maintain the health and safety of our communities. Second, given the nature of their work, emergency workers are at higher risk of developing mental health conditions such as depression, anxiety, and alcohol misuse as well as posttraumatic stress disorder (PTSD) [21-24]. Finally, RTPs that are evaluated and found to benefit emergency workers are likely to provide valuable insights in terms of how best to support the mental health of workers in other high-stress occupations (eg, health care, journalism, and military).

Prior Work

RTPs typically share the common aim of enhancing a person's ability to manage stressful situations and adverse circumstances more effectively and with greater emotional insight. Although they may share this common goal, RTPs often differ considerably in terms of content, length, and delivery [25]. More recently, researchers have been drawing on evidence-based psychological therapies such as acceptance and commitment therapy (ACT), cognitive behavioral therapy, and mindfulness-based stress reduction to inform program development [11,12,15,26-30]. These RTPs generally include a combination of psychoeducational material, cognitive strategies, and goal setting, with mindfulness training frequently emerging as a core component.

The growing emphasis on mindfulness training within RTPs is understandable, given the large body of research highlighting the positive benefits of this practice on mental health outcomes [31-38]. Several studies have also described the positive impact of mindfulness on psychological resilience [11,12,26,39]. The idea that mindfulness training may serve to enhance psychological resilience is further supported by findings from a recent meta-analysis that found resilience can indeed be enhanced, particularly by programs that involve both cognitive behavioral strategies and mindfulness training [40].

These specific strategies and skills require time to practice and attain proficiency. It is, therefore, unsurprising that most resilience studies completed thus far describe training programs that involve multiple face-to-face training sessions [25,41]. This is particularly challenging for many employers, as taking workers off the job to attend training can result in considerable disruption to business and services. Face-to-face training is also inherently expensive. As well as direct program costs, employers may face many additional expenses, such as travel, accommodation, venue hire, and the cost of replacement staff. Delivering programs in remote areas can also prove to be particularly difficult and expensive. The scalability of RTPs is thus a critical consideration. The more affordable and accessible RTPs can be made, the more likely they are to be implemented on a large scale.

In response to these challenges, workplace mental health researchers have begun to develop and evaluate novel electronic health (eHealth) programs (online, Web-based training) to help improve accessibility and engagement. Results from a recent meta-analysis found that digital mental health interventions in the workplace can improve psychological well-being [38] and

work effectiveness among employees [42]. Despite the apparent advantages of eHealth, research examining the efficacy of this approach for resilience training remains sparse. A few trials have examined either a blended approach (ie, programs that combine online and face-to-face resilience training) [12,26] or an online approach with an emphasis on stress reduction and enhancing resilience-related factors [43,44]. As with the main resilience literature to date, these studies vary greatly in their approach to measuring program efficacy, and thus, limited conclusions can be drawn regarding the efficacy of online RTPs.

To address these issues, we developed an interactive e-learning program called the Resilience@Work (RAW) Mindfulness Program. This self-paced intervention aims to enhance psychological resilience among workers. It comprises 6 online training sessions, each taking about 20 to 25 min to complete on a tablet or computer. The program involves mindfulness training, psychoeducation, and a range of skills and strategies drawn from evidence-based therapies, including ACT, mindfulness-based cognitive therapy, and compassion-focused therapy. A more detailed overview of this program was published in a recent pilot study, and it was found that the RAW Mindfulness Program is feasible in a workplace setting and that those using the program showed a trend toward increased resilience and psychological flexibility [45].

Hypotheses

The main aim of this study was to build upon the initial pilot research and examine whether the RAW Mindfulness Program can improve resilience among a group of high-risk workers, specifically emergency services personnel. Our primary hypothesis is that first responders receiving the RAW Mindfulness Program will have increased resilience following training, compared with the control group. In addition, we will examine the impact of this training on several secondary outcomes, including acceptance and mindfulness skills, as well as resilience resources such as coping, self-compassion, and optimism. To the best of our knowledge, this is the first time an entirely online mindfulness-based RTP has been evaluated via a randomized controlled trial (RCT) with active emergency services personnel.

Methods

Cluster Randomization

This study was prospectively registered with the Australian New Zealand Clinical Trial Registry (ANZCTR no: 12615000574549). We conducted a cluster RCT comprising 24 NSW Primary Fire and Rescue and Hazmat Stations. Fire and Rescue New South Wales (FRNSW) is the seventh largest urban fire service in the world and responds to firefighting, rescue, and hazardous material emergencies in Sydney, Australia, and surrounding regional areas. Given the nature of their work, employees are known to have an elevated risk of depression, anxiety, and PTSD [24,46]. Ethics approval was obtained via the Human Research Ethics Committee at the University of New South Wales, Australia. Randomization was at the station level using an online random sequence generator. An external researcher completed the randomization process. Overall, 12 stations were assigned to receive the RAW Mindfulness Program, and 12 stations were assigned to receive the attention-matched control intervention.

Participants

Potential participants were full-time firefighters working in the 24 Primary Rescue and Hazmat Stations involved in this trial. Participants were informed about the study during a standard well-being talk facilitated by members of the FRNSW Peer Support Team. These talks are performed several times a year as part of FRNSW's employee well-being initiative. The Peer Support System within NSW Fire and Rescue is a successful and long-running nonhierarchical support service for firefighters. From November 2015 to April 2016, the Peer Support Officers provided a brief presentation on the topic of resilience as well as the aims of the research study and highlighted that participation in the study was voluntary. The presentation slides were provided by the researchers. Following the presentation, firefighters were provided with a participant information sheet and consent form, the study questionnaire, and a stamped address envelop to return their information to the research team if they chose to be a part of this study. Potential participants in the 12 stations randomly assigned to the intervention group were also asked to provide their email address in the consent form to receive information to access the online program. Log-in details were then emailed directly to the participant. When a participant logged into the online RAW program, this was considered as final consent and enrolment in the intervention group. Participants could withdraw at any time from the study. Participants were not aware of the study hypotheses and were blinded to intervention versus control status. Potential participants who were currently engaged in any form of regular psychological therapy with a psychologist and/or psychiatrist were excluded from this study.

Intervention: Resilience@Work (RAW) Mindfulness Program

Firefighters assigned to the intervention group received the RAW Mindfulness Program. FRNSW granted firefighters who were enrolled in the study permission to access the program at work. Tablets (iPads) were made available in the stations for firefighters to complete the online program. The RAW program is a mindfulness-based intervention, which also draws on ACT and has a significant emphasis on self-compassion and acceptance skills. The intervention comprises 6 online training sessions. Each session takes about 20 to 25 min to complete. A combination of interactive exercises, audio, and animation is used to teach resilience skills. An overview of the core strategies and skills taught in the RAW program is outlined in Table 1. Participants were able to download mindfulness tracks to their own device for continued practice. Each session was provided in a sequential order, with completion of the first module unlocking access to the next and so on. There was a 3-day break in between each session to encourage skills practice. Therefore, the minimum amount of time a participant could complete the training was 3.5 weeks and the maximum was 6 weeks. Participants also had the opportunity to sign up for text message and/or email reminders.

Table 1. Overview of skills and topics covered in the Resilience@Work Mindfulness Program.

Session	Resilience topic and skills focus	Mindfulness tracks
1	Introduction to mindfulness, resilience and psychological well-being	Drop anchor; Take 10; Leaves on a stream
2	Mindfulness skills, understanding your reactive mind versus wise mind, recognizing unhelpful mind chatter and managing uncomfortable and unhelpful thoughts (cognitive defusion); Recognizing your values exercise	Mindful Breathing; Defusion Technique; Notice it, Name it, Let it Go (I'm having the thought that…); Defusion technique 2: Thank you Mind
3	Revision of Cognitive Defusion; Introduction to Mindfulness with Emotions, The Reactive Mind and Avoidance, Understanding how values are linked to Emotions; Valued Action check	Creating Space (Mindfulness with emotions); Mindful Body Scan; The Golden Room
4	The problem with Avoidance, Recognizing avoidance strategies versus adaptive strategies	Creating Space; A Mindful Break (mindfulness with words) Surfing Waves
5	Self-care and support, the Compassion myth, barriers to accessing compassion, compassion fatigue, self-compassion actions & resilience, Identifying Mindful Support (compassionate, nonjudgmental and mindful); Valued Action check	A kind and gentle hand (loving-kindness practice); A Safe Place (compassion-focused mindfulness); A bird's eye view
6	Compassion focused Mindfulness; Gratitude practice, optimism and resilience, identify and celebrate the milestones; creating a personalized action plan to practice skills	Breathing in the Present Moment; A Golden Moment exercise; Being Kind to your old wounds

Control Condition: Healthy Living Program

Along with the standard FRNSW well-being talk, firefighters who were assigned to the control group received access to the Healthy Living Program (HLP). FRNSW granted permission to firefighters enrolled in the study to access the program while at work. The HLP comprises 6 modules that provide helpful information on a range of health and well-being topics, for example, healthy skin, healthy home environment, and mobile phone use. The self-paced program was available on tablets (iPads) within the station, with each module taking about 20 min to complete.

Measures

Primary Outcome: Measure of Resilience

The primary outcome of this study was resilience. The means by which resilience is best measured remains a topic of considerable discussion in the literature. In their review of resilience scales, Windle et al [47] concluded that there was "no current 'gold standard.'" They did, however, identify the Connor-Davidson Resilience Scale (CDRISC) and the Brief Resilience Scale (BRS) as 2 of the better measures available to researchers at present. We utilized both measures at baseline and 6-month follow-up in our study to examine 2 inherent constructs of resilience: (1) successful adaptation to stressful life events and circumstances and (2) bounce-back resilience. The 10-item version of the CDRISC was used to examine the ability to successfully adapt and tolerate experiences such as illness, pressure, personal problems, failure, change, and painful feelings [48]. The BRS was employed to specifically examine the concept of *bounce-back* resilience, that is, the ability to recover from stress [49].

Connor-Davidson Resilience Scale

Psychological resilience was measured using the validated short-form 10-item version of the Connor-Davidson Resilience Scale (CDRISC_10) [48]. Participants respond to each item on a 5-point scale, ranging from 0 (not true at all) to 4 (true nearly all of the time). The total score ranges from 0 to 40, with a higher score indicative of higher psychological resilience. Previous studies have found CDRISC_10 to be a reliable and valid measure, with Cronbach alpha ranging from .81 to .88 [50,51] and test-retest reliability of .90 at 6 weeks [51,52].

Brief Resilience Scale

Bounce-back resilience (the ability to recover from stress) was measured using the 6-item BRS [49]. Items are rated on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The total score ranges from 6 to 30, with higher scores signifying greater bounce-back resilience. BRS has demonstrated good internal consistency, with Cronbach alpha ranging from .83 to .90 [49,53].

Secondary Outcomes

A number of secondary outcomes were included to examine the processes by which the RAW Mindfulness Program may enhance resilience. Measures of acceptance and mindfulness skills and several resilience resources were administered.

Acceptance and Mindfulness Skills

Freiburg Mindfulness Inventory

Mindfulness was measured using the short version of the Freiburg Mindfulness Inventory (FMI-14) [54]. Previous studies have found that the FMI-14 is a reliable measure with good internal consistency (Cronbach alpha=.86) [54-56].

Cognitive Fusion Questionnaire

The CFQ is a measure of cognitive fusion and defusion, a core component of the ACT model [57]. Higher scores reflect greater thought entanglement. The CFQ has been found to be a reliable and valid measure, with Cronbach alpha ranging from .89 to .93 [57,58].

Acceptance and Action Questionnaire

The Acceptance and Action Questionnaire (AAQ-II) is a measure of experiential avoidance and psychological inflexibility. Previous research has found the AAQ-II to be a reliable and valid measure, with a Cronbach alpha of .84 and test-retest reliability of .81 at 3-month follow-up [59].

Self-Compassion Scale

The 12-item short form of Self-Compassion Scale (SCS-SF) [60] assesses the level of self-compassion an individual has toward themselves during difficult and challenging times. The measure has high internal consistency (Cronbach alpha=.85) and a nearly perfect correlation with the long form of the SCS [60,61].

Resilience Resources

Optimism: Life Orientation Test-Revised

The Life Orientation Test-Revised (LOT-R) was used to assess levels of optimism [62]. This 10-item scale examines the extent to which individuals anticipate positive outcomes in the future, with higher scores reflecting greater optimism. The LOT-R has been shown to be a reliable and valid measure (Cronbach alpha=.73) [61,63].

Coping: The Brief-Coping Orientation to Problems Experienced

A total of 3 subscales of the Brief-Coping Orientation to Problems Experienced (Brief-COPE) [64] were included *Using Emotional Support* (accepting emotional support, compassion, and sympathy from others), *Using Instrumental Support* (seeking advice, help, or information from others), and *Active Coping* (the process of taking active steps to alter or reduce the impact of a stressor). Higher scores reflect greater use of each coping strategy. These subscales have been found to be reliable measures, with Cronbach alpha values ranging from .64 to .82 [64-66].

Sense of Purpose: Life Engagement Test

The extent to which participants value their daily activities and have a sense of life purpose was examined by the 6-item Life Engagement Test (LET) [67]. The LET has previously been found to be a reliable measure, with Cronbach alpha ranging from .72 to .87 [67,68].

Statistical Analysis

As outlined in the a priori analysis plan recorded in the online trial registry (ANZCTR no: 12615000574549), the primary analysis was undertaken within an intent-to-treat framework utilizing mixed-model repeated measures (MMRM). This approach is recognized as a reliable method of analyzing RCT data [69,70]. MMRM uses all available data and, therefore, does not substitute missing values with estimated values. The dependent variables in these analyses were the absolute values for resilience, as measured by CDRISC_10 and BRS. The relationships between observations at different occasions were modeled using an unstructured variance-covariance matrix. The analysis examined whether the rates of change on resilience scores over time differed significantly between the intervention group and the control group via a group-by-time interaction. In addition, differences between the intervention and control groups at each of the follow-up time points were examined against baseline using planned contrasts. Similar analyses examined the impact of the RAW intervention program compared with the control condition on several secondary outcomes, which included mindfulness/acceptance skills and resilience resources. A priori planned per-protocol analyses were also conducted to assess the effectiveness of the program among participants who completed different numbers of online RAW modules compared with the control group. All tests of treatment effects were conducted using a 2-sided alpha level of .05 and 95% CIs. All analyses were conducted in IBM SPSS version 24. On the basis of our pilot study results of baseline resilience levels [45] and predicted intracluster correlation coefficient of .01, we estimated that a total of 24 fire stations (clusters) including 144 firefighters would need to be recruited to achieve .80 power of detecting a .50 SD improvement in resilience measures with an alpha of .05 (2-sided).

Results

Overview

This study was conducted in accordance with the Consolidated Standards of Reporting Trials (CONSORT) statement and guideline for transparent reporting of RCTs. Figure 1 outlines the flow of recruitment and loss to follow-up. Overall, a total of 143 firefighters were recruited, with 83 firefighters assigned to the control condition and 60 firefighters to the RAW Mindfulness Program. Of those that were recruited, data for the primary outcome were available for 55.2% (79/143) of the sample at 6-week follow-up. At 6 months, there was a decrease in the follow-up rate to 48.3% (69/143). There was a greater loss to follow-up in the control group at post intervention (54% [45/83] compared with 32% [19/60]; P=.01); however, by the 6-month follow-up, this differential loss to follow-up was no longer apparent (43% [36/83] compared with 53% [32/60]; P=.24). Loss to follow-up was unrelated to age, gender, years served as a firefighter, baseline measures of resilience, and prior exposure to traumatic events (P>.05).

Demographic details of firefighters enrolled in the study are shown in Table 2. In line with recommendations from the CONSORT statement [71], statistical tests of baseline differences were not carried out, as any differences were because of the randomization process and, therefore, by definition, because of random chance. However, observation of the data summarized in Table 2 suggested that the participants in the intervention group were slightly older and more experienced than the participants in the control group and, therefore, may have been exposed to more traumatic incidents across their career. These baseline differences were controlled for in later sensitivity analyses.

Figure 1. Consolidated Standards of Reporting Trials flow diagram. RAW: Resilience@Work.

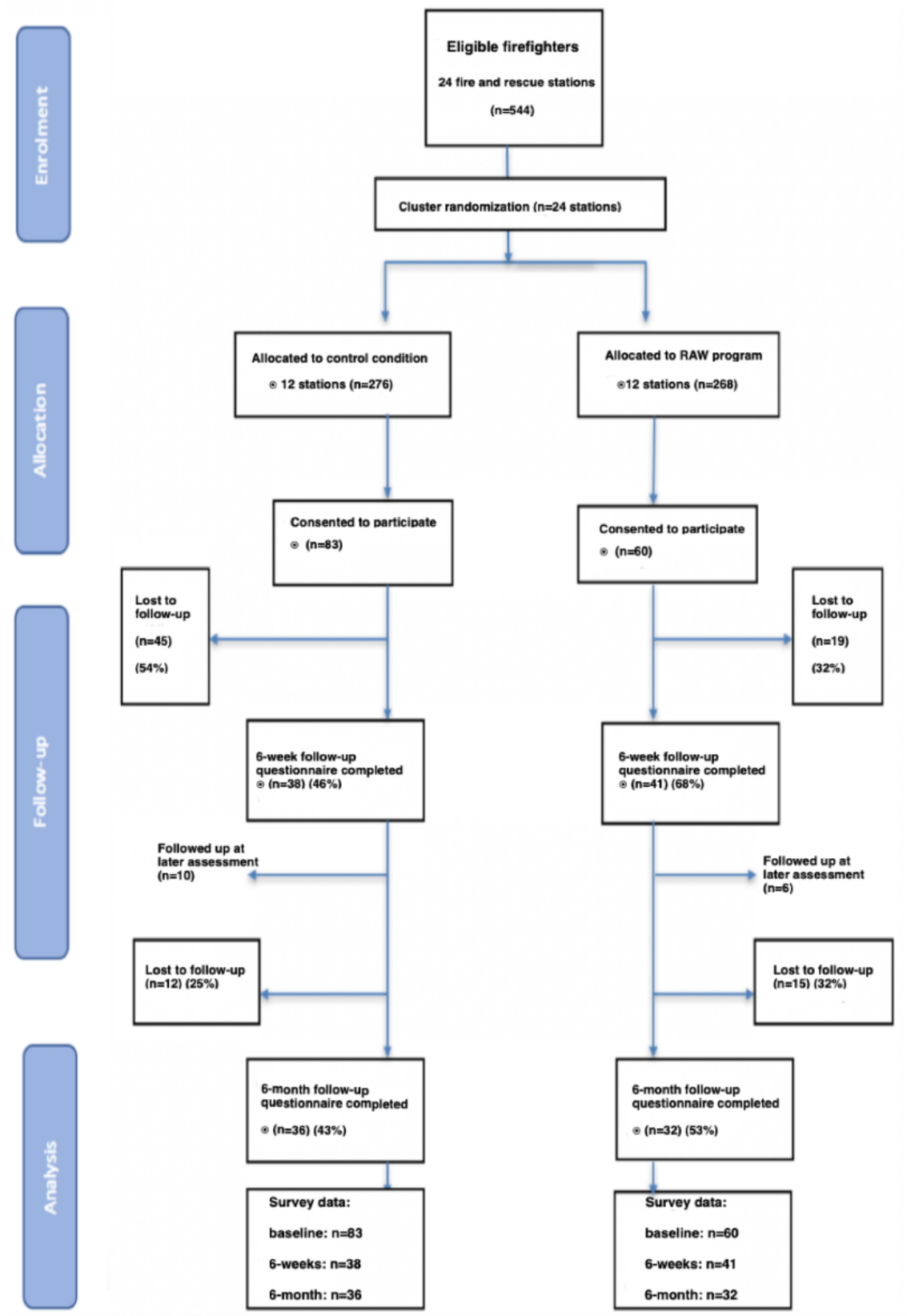


Table 2. Demographics of enrolled participants at baseline.

Demographics	Intervention group (n=60)	Control group (n=83)
Gender, n (%)		
Male	56 (93)	81 (98)
Female	4 (7)	2 (2)
Age (years), mean (SD)	43.9 (7.8)	41.1 (9.2)
Years served as a firefighter, n (%)		
1 to 5 years	5 (8)	9 (11)
6 to 10 years	13 (22)	27 (33)
11 to 15 years	11 (18)	24 (29)
16 to 20 years	6 (10)	6 (7)
More than 20 years	25 (42)	17 (20)
Number of trauma incidents attended over the course of career, n (%)		
1 to 5	3 (5)	14 (17)
6 to 10	7 (12)	17 (21)
11 to 15	5 (9)	10 (12)
16 to 20	6 (10)	5 (6)
More than 20	37 (64)	37 (45)
Measures of resilience, mean (SD)		
CDRISC_10[a]	28.4 (5.3)	1. (5.5)
BRS[b]	22.1 (3.4)	23.0 (3.6)

[a]CDRISC_10: 10-item version of Connor-Davidson Resilience Scale.

[b]BRS: Brief Resilience Scale.

Resilience@Work Program Engagement

The majority of participants (38/60, 63%) completed more than half the RAW program (mean number of sessions completed was 3.5 out of a possible 6; SD 2.0), equating to 60 to 75 min of training. A total of 22 participants (22/60, 37%) went on to complete 5 to 6 sessions (a total of at least 100 to 120 min of training).

Resilience

For the primary outcome, the overall test of group-by-time interaction was significant (*P*=.01), with the intervention group increasing in resilience over time (Figure 2). Although the intervention group increased in resilience at 6 weeks, the difference compared with the control group at this time point fell short of significance (*P*=.09). However, at 6-month follow-up, the intervention group continued to improve in resilience and was significantly different from the control group (*P*=.002). At 6-month follow-up, the group receiving the RAW intervention had an average increase in their resilience score of 1.3, equating to a moderate-to-large effect size compared with the control group of 0.73 (CI: 0.38-1.06). This effect remained when the analysis was repeated with adjustment for baseline age, years of service, and number of traumatic incidents experienced (test of group-by-time interaction, *P*=.008; difference between groups at 6-month follow-up, *P*=.02).

No significant differences were found between the intervention group and the control group in terms of change in bounce-back resilience as measured by BRS (*P*=.09). This finding was replicated when the analysis was repeated to adjust for baseline age, years of service, and number of traumatic incidents experienced (test of group-by-time interaction, *P*=.09).

A per-protocol analysis examined the dose response to the RAW intervention program at 6-month follow-up (Figure 3). Compared with the control group, there was a significant and positive change in CDRISC_10 resilience scores at 6-month follow-up among those who completed most of the RAW program, that is, 5 to 6 sessions (*P*=.002). Within the intervention group, there was no difference between partial and full completers (5-6 sessions) of the RAW program in terms of age (*P*=.20), gender (*P*=.62), and baseline levels of resilience (*P*=.39).

Figure 2. Test of group-by-time interaction (*P*=.01) change on the 10-item version of the Connor-Davidson Resilience Scale. RAW: Resilience@Work.

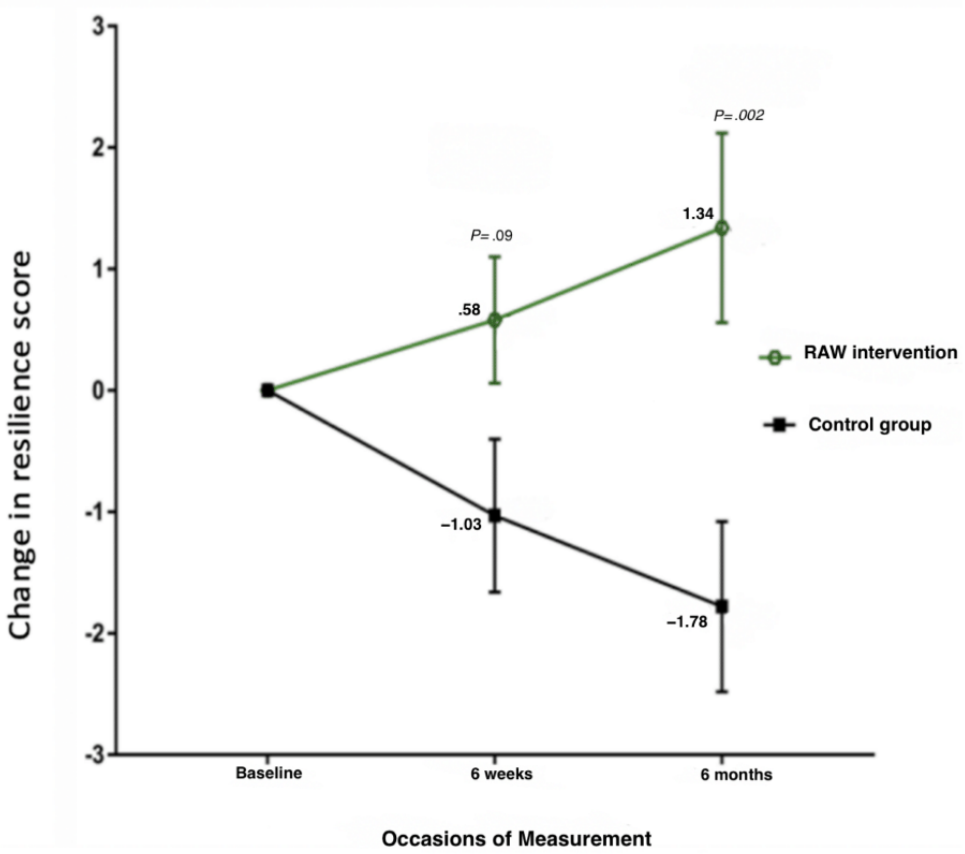


Figure 3. Analysis of primary outcome (change on the 10-item version of the Connor-Davidson Resilience Scale) by number of Resilience@Work sessions. Statistical difference test of significance compared with control group obtained from mixed-methods repeated measures models. RAW: Resilience@Work; CDRISC_10: 10-item version of the Connor-Davidson Resilience Scale.

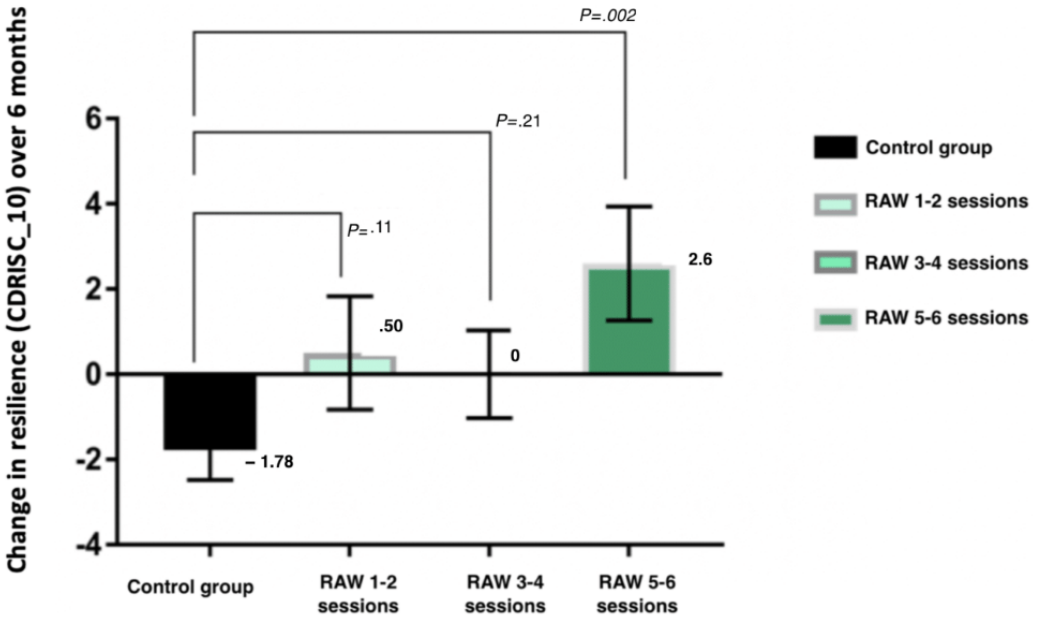


Table 3. Mixed-model repeated measures with secondary outcome variables change at 6 weeks and 6 months for intervention and control groups.

Measure	Description	Change at 6 weeks Mean (SE)		Significance test of difference at 6 weeks[a]	Change at 6 months Mean (SE)		Significance test of difference at 6 months[a]
		Control	RAW		Control	RAW	
Acceptance and mindfulness skills							
Cognitive Fusion (CFQ)	Level of thought entanglement	−0.18 (1.07)	−0.85 (0.74)	.25	0.06 (0.89)	−1.65 (1.18)	.40
Experiential Avoidance(AQQ_II)	Psychological inflexibility/ reactivity, tendency to avoid experiencing internal events	−0.29 (1.38)	−0.83 (0.56)	.31	−0.27 (0.81)	−0.94 (0.81)	.27
Self-Compassion (SCS-SF)	Level of self-compassion during difficult times	0.89 (1.66)	1.31 (0.95)	.42	0.71 (1.24)	1.47 (1.08)	.99
Mindfulness (FMI)	Level of mindfulness and present moment awareness	0.38 (1.06)	1.18 (0.87)	.79	0.37 (0.99)	4.16 (1.05)	.09
Resilience resources							
Life Engagement Test (LET)	Sense of purpose in Life	0.00 (0.64)	0.11(0.57)	.83	−0.69 (0.66)	−0.60 (0.53)	.89
Life Orientation Test-Revised (LOT-R)	Level of optimism	−0.83 (0.55)	1.2 (.43)	.05[a]	−0.86 (0.48)	0.38 (0.68)	.14
Active Coping (AC)	Personal effort and actions to change and improve current situation	0.11 (0.46)	0.40 (0.43)	.09	−0.56 (0.44)	0.21 (0.43)	.046[a]
Use of Emotional Support (ES)	Emotional, empathic and comfort from others	−1.35 (0.48)	0.30 (0.36)	.05[a]	−0.40 (0.34)	0.24 (0.30)	.10
Use of Instrumental Support (IS)	Advice/help from others on what actions to take	−0.47 (0.43)	0.43 (0.36)	.05[a]	−0.29 (0.37)	0.03 (0.38)	.32

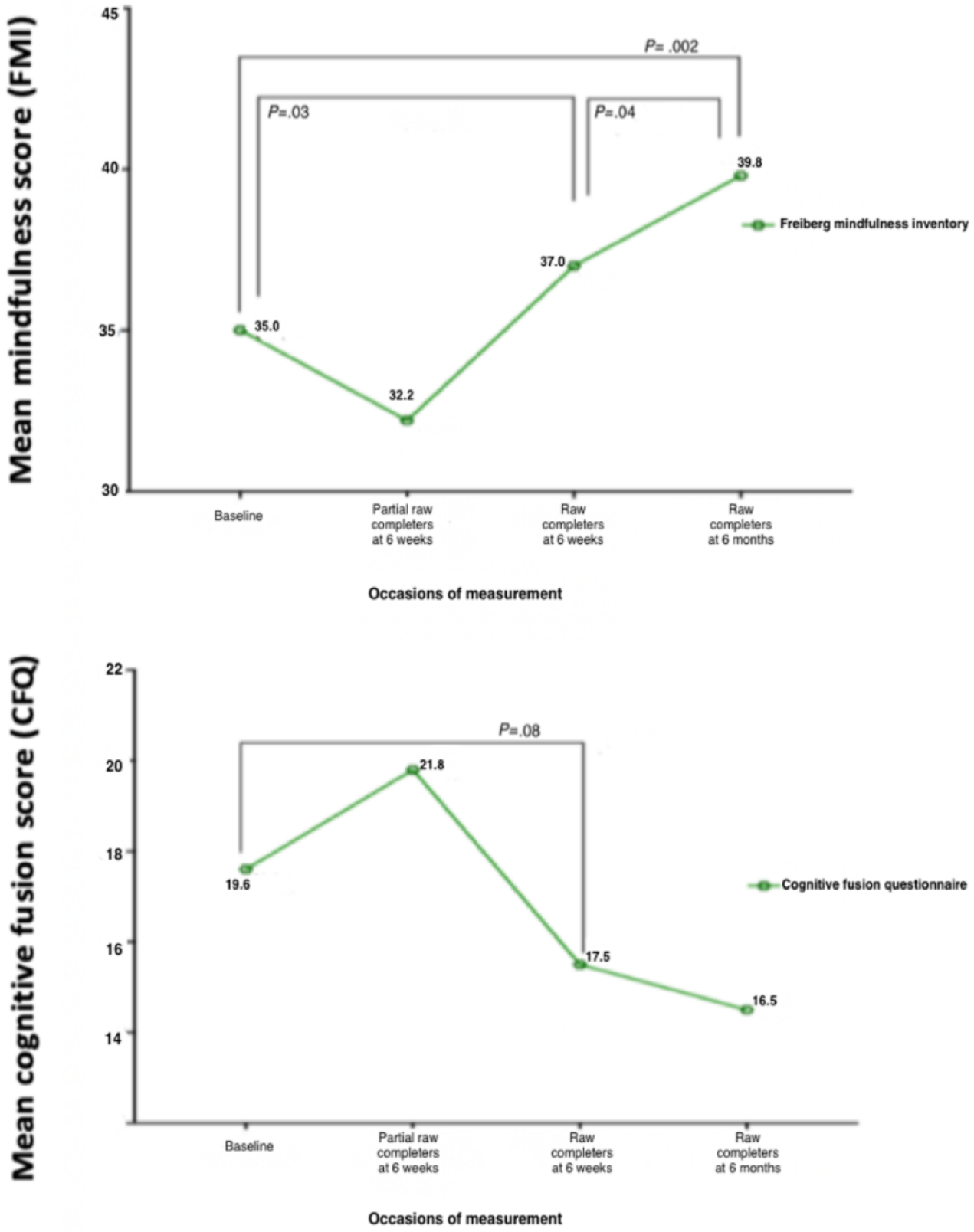
[a]Significant at $P<0.05$. Test of difference between groups at each time point were obtained from planned contrasts utilizing unadjusted mixed-model repeated measures analysis.

Secondary Outcomes

Analyses of change for secondary outcomes are shown in Table 3. Improvements within the intervention group were observed in mindfulness and self-compassion as well as reduced thought entanglement (cognitive fusion) and experiential avoidance; however, these changes were not significantly different from those observed in the control condition. With regard to resilience resources, there were statistically significant improvements noted in optimism (P=.05), use of emotional support (P=.05), and use of instrumental support (P=.05) at 6-week follow-up in the intervention group compared with the control condition; however, this change was not sustained at 6-month follow-up. There was a significant and sustained improvement in active coping among the intervention group when compared with the control group at 6-month follow-up (P=.046).

Finally, we examined how mindfulness (FMI) and cognitive defusion (CFQ) skills changed over time for partial completers (1-4 sessions) and completers (5-6 sessions). Change in scores was examined across 3 time points: baseline, 6-week follow up, and 6-month follow-up. The differences between baseline and different follow-up for partial completers and completers were analyzed using paired t tests (Figure 4). For RAW program completers, significant improvements were observed in mindfulness skills at 6-week follow-up (P=.03) and 6-month follow-up (P=.002) compared with baseline. The improvements in mindfulness from 6-week follow-up to 6-month follow-up were also statistically significant (P=.046). In addition, there was a trend toward reduced cognitive fusion/thought entanglement (CFQ) at 6-week follow-up for program completers (P=.08). However, for partial completers, there was limited change in mindfulness or cognitive defusion.

Figure 4. Analysis of change in mindfulness and cognitive fusion scores over time for partial and full program completers. RAW: Resilience@Work; FMI: Freiburg Mindfulness Inventory; CFQ: cognitive fusion questionnaire.



Discussion

Principal Findings

This study is the first ever RCT to test the ability of an entirely online training program to enhance psychological resilience. Our results demonstrate that the RAW Mindfulness Program enhanced adaptive psychological resilience among active firefighters, increasing their ability to adapt successfully in the face of adverse circumstances and situations. Our per-protocol analysis found that the greatest improvements were seen among

those who completed all 6 RAW sessions—at least 2 hours of training over a minimum of 3.5 weeks. The results support previous findings that mindfulness-based RTPs can create improvements to individual resilience [40,53,72]. The results also further support previous research regarding workplace eHealth initiatives and the positive impact they can have on employee well-being [38,42]. These findings have important implications for other high-risk groups, with workers such as paramedics, police, defense personnel, doctors, nurses, midwives, and journalists, who may also be likely to benefit from similar interventions.

This study utilized 2 of the most commonly used measures of resilience, along with several measures of resilience-related factors. This approach was in response to the absence of a current gold standard resilience measure [47] and the growing consensus among researchers that resilience is a multifaceted construct [48,73]. The RAW program resulted in significant improvements in resilience as measured by the CDRISC_10; however, no changes were observed on the BRS. It is difficult to establish why RAW led to improvements on the CDRISC_10 and not the BRS. One possible explanation is that our baseline sample was already scoring in the average to high range on the BRS and yet low average to average on the CDRISC. This may have limited the opportunity to enhance bounce-back resilience in this group. This observation supports the notion that these scales are measuring 2 different aspects of resilience and that the RAW program may enhance certain aspects of resilience and not others. Endorsement of the CDRISC_10 reflects an individual's ability to tolerate experiences such as change, personal problems, illness, pressure, failure, and painful feelings [48,73,74]. In contrast, Smith et al argue that the BRS is currently the only scale that accurately measures resilience in its most basic form as it focuses solely on a person's ability to *bounce back* or recover from stress rather than the personal qualities that facilitate and promote positive adaptation [49]. Smith et al emphasize that *bounce-back* resilience is particularly important when examining groups that are already unwell or facing health-related stressors, as it reflects the specific ability to recover rather than one's ability to resist illness. In their review of resilience measures, Windle et al note that although items in the BRS correspond with the ability to recover and cope with challenging circumstances, all 6 items reflect a sense of personal agency [47]. Although personal agency can be viewed as an essential part of resilient adjustment to adversity, it is now well documented that resilience is a dynamic multifaceted construct, with no single factor accurately predicting resilience outcomes [47,74]. For groups of relatively healthy workers who do not identify as being unwell physically or psychologically, focusing on factors that can enhance adaptive resilience may serve a greater purpose than examining factors related to personal agency. The enhancement of adaptive resilience and related resources may be central to such groups' abilities to manage potentially stressful and challenging situations in the future. This, in turn, may have a direct impact on their overall psychological well-being. Indeed, a recent study of the relationship between baseline resilience and mental health outcomes [75] found that low adaptive resilience among active first responders (as measured by the CDRISC_10) was an accurate predictor of increased PTSD symptoms at 6-month follow-up. Similarly, researchers in the United Kingdom found that low adaptive resilience at baseline (measured by the CDRISC_25) accurately predicted greater symptoms of depression at 2-year follow-up among active paramedics [76].

In terms of the secondary outcomes examined in this study, the RAW program resulted in significant improvements in overall optimism, use of emotional support (seeking empathy and emotional support from others), and use of instrumental support (actively seeking advice and help from others on what actions to take) post training. At 6-month follow-up, there was a significant improvement in the level of active coping (personal effort and actions to change and improve the current situation). This may further explain why improvements were observed on the CDRISC_10 and not the BRS, as the RAW program focused heavily on the acquisition of skills directly aimed at enhancing resilience resources that are more directly measured by the CDRISC. Surprisingly, although an overall trend of improvement in acceptance and mindfulness skills was observed in the RAW group, when compared with the control group, these changes did not reach statistical significance. As this study's power analysis was based on our primary outcome of resilience, it may have been underpowered to detect a difference on measures of mindfulness and acceptance skills.

Overall, our findings suggest that the RAW program can improve specific resilience resources and aspects of resilience such as distress tolerance, positive adjustment, and perseverance, as measured by the CDRISC_10. It may have less impact on the concept of *bounce back* as measured by the BRS. It is also important to consider why improvements in resilience were more prominent at 6-month follow-up than immediately following training. Time is an essential aspect in any skill's acquisition, allowing the opportunity for ongoing practice, development, and refinement. Given the greater time to practice and develop their skills, it is consistent that enhanced resilience was seen at 6-month follow-up. These findings are in line with results from a recent meta-analysis and systematic review of RTPs, which found that programs delivered over time can improve resilience, whereas one-off training sessions resulted in little to no improvement [40]. This also speaks to the temporal elements of resilience and the importance of including longer-term follow-up in trials examining the efficacy of RTPs [74]. Resilience studies to date have generally limited data collection time points to posttraining and 3-month follow-up, with minimal good-quality studies including 6-month follow-up data [25,40].

Limitations

Limitations to this research include that the workforce was a male-dominated high-risk group, thus limiting the generalizability of the findings to lower-risk and gender-balanced occupational groups. It is also important to acknowledge that during the trial period, the self-reported resilience of the control group reduced. It is unknown whether this may be related to their ongoing trauma exposure or whether it may be the natural trajectory of resilience among firefighters over a 6-month period. The latter may be the case, given that participants enrolled in this study were all active, full-time firefighters based at Primary Rescue and Hazmat Stations. These

stations are the busiest across the state of New South Wales and frequently respond to serious emergencies, critical incidents, and disaster situations, including motor vehicle accidents, suicide, structural fires, hazard material, and body recovery. In light of these environmental factors and the workplace setting, it is perhaps unsurprising that the overall resilience of firefighters in the control group declined over a 6-month period. Importantly, this highlights the significance of the temporal elements of resilience and the value of measuring it over several time points [74]. A further limitation in the study's design was the second step in the consent process for the intervention group, that is, logging into the online program. This is a potential source of bias and may have influenced the different rates of loss to follow-up. Another design limitation was the omission of measurements examining personal motivation and skills implementation. Therefore, we have limited insight into how often the skills were used post training and whether they were implemented during certain work situations. It will be critical to include these measurements in future evaluations of the RAW program. It will also be important to examine whether increased resilience can lead to improvements in job-related factors such as productivity, safety, and reduced on-the-job errors.

It is important to also acknowledge the adherence issues in this trial. Although the majority of firefighters (63%) in our study completed over half the program (3.5 out of 6 modules), only 37% of firefighters went on to complete the entire program. Unfortunately, we did not gather specific follow-up information from participants regarding their reasons to stop training. Some potential reasons include poor internet connection within the station, the program being accessible only on tablets or computers, reduced motivation, or limited available time because of competing work responsibilities. It is also possible that a participant may have found certain skills in the early modules helpful and, therefore, felt they did not require additional training. Alternatively, some participants may have found some aspects of the training repetitive and restrictive (each module had to be completed to unlock the next one) and, therefore, lost interest in the program. In future evaluations of the program, it will be important to examine why participants stop engaging in the RAW program and whether having access to the program on other personal devices such as smartphones may increase program adherence. In addition, as noted by recent research [12], providing greater control and flexibility around how participants access the online intervention may well increase program adherence and long-term engagement.

A final limitation is that the primary outcome of resilience was measured by self-report. Although the measures used are the most validated and widely used measures of resilience available, it is important that future studies are able to assess whether changes in self-reported resilience from interventions such as RAW translate into fewer incidents of mental illness over time. It is important to note that the RAW program was specifically aimed at enhancing personal resilience rather than reducing mental health symptoms. Given the emerging literature on the relationship between low resilience and increased risk of future mental health difficulties [75,76], the next key step in this research will be to examine whether improved resilience results in reduced mental health symptomology and whether this reduction is sustained over time.

Conclusions

Despite these limitations, our study's findings have important implications. First, mindfulness-based resilience training delivered in an online format can create improvements in adaptive resilience and related resources among high-risk workers, such as firefighters. This is particularly significant, given recent findings that low baseline resilience may be a risk factor for increased mental health symptomology in emergency workers [75,76]. As resilience is a potentially modifiable risk factor for mental health conditions in high-risk groups, programs such as RAW may increase an organization's ability to play a proactive role in protecting psychological well-being. The RAW Mindfulness Program and similar interventions may also serve to bolster psychological resilience among other high-risk groups, such as military personnel, journalists, nurses, doctors, and midwives. Further large-scale resilience trials would offer the opportunity to measure outcomes across different workplaces and occupations. Research trials incorporating extended follow-ups beyond 6 months would also be beneficial, to examine long-term impacts and the optimum time for retraining to occur.

In conclusion, the RAW Mindfulness Program is an effective, scalable, and practical means of delivering online resilience training to high-risk groups such as first responders. With the benefit of further research and development, this form of online resilience training may serve to enhance mental health on a broad scale, protecting workers who perform some of our society's most challenging roles.

Acknowledgments

This publication was made possible by an Australian Government Research Training Program Scholarship and the University New South Wales Brain Sciences PhD Grant in aid awarded to SJ for her doctoral studies and funding for SH from NSW Health and the icare foundation (grant number: RM09708). RTP, icare foundation, or NSW Health did not have a role in the study design, collection, analysis, or interpretation of the data; writing the manuscript; or the decision to submit the paper for publication. The researchers would like to express their sincere gratitude to the NSW firefighters who generously gave their time to participate in this study. The authors would also like to thank the FRNSW Peer Support Officers and staff who provided invaluable logistical support throughout the research project, in particular, Tara J Lal, Brendan Mott, Mark Dobson, Brett Farmer, and Bruce Fitzpatrick. Finally, the authors wish to thank and dedicate this paper to the NSW firefighters and first responders globally who choose to be of service in our communities every day and by doing so make our world a safer, kinder, and more compassionate place.

Authors' Contributions

SJ and SBH devised the study. SJ developed the RAW Mindfulness Program; the internet-based format; and collected, scored, and entered the data. TJL assisted with data collection. SJ and SBH analyzed and interpreted the data, and SJ wrote the first draft of the manuscript. All authors read and contributed to subsequent versions and approved the final manuscript.

Conflicts of Interest

SJ and SBH are associated with a company that offers resilience training (RAW Mind Coach). SBH and FS work for the Black Dog Institute, a not-for-profit organization that provides mental health and resilience training to various other organizations.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 753KB - jmir_v21i2e12894_fig.pdf]

References

1. Harvey SB, Henderson M, Lelliott P, Hotopf M. Mental health and employment: much work still to be done. Br J Psychiatry 2009 Mar;194(3):201-203. [doi: 10.1192/bjp.bp.108.055111] [Medline: 19252144]
2. Cattrell A, Harris EC, Palmer KT, Kim M, Aylward M, Coggon D. Regional trends in awards of incapacity benefit by cause. Occup Med (Lond) 2011 May;61(3):148-151 [FREE Full text] [doi: 10.1093/occmed/kqr008] [Medline: 21482620]
3. Murray C, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
4. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424 [FREE Full text] [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
5. Knudsen AK, Harvey SB, Mykletun A, Øverland S. Common mental disorders and long-term sickness absence in a general working population. The Hordaland Health Study. Acta Psychiatr Scand 2013 Apr;127(4):287-297. [doi: 10.1111/j.1600-0447.2012.01902.x] [Medline: 22775341]
6. Knudsen AO, Øverland S, Aakvaag HF, Harvey SB, Hotopf M, Mykletun A. Common mental disorders and disability pension award: seven year follow-up of the HUSK study. J Psychosom Res 2010 Jul;69(1):59-67. [doi: 10.1016/j.jpsychores.2010.03.007] [Medline: 20630264]
7. Mykletun A, Harvey SB. Prevention of mental disorders: a new era for workplace mental health. Occup Environ Med 2012 Dec;69(12):868-869. [doi: 10.1136/oemed-2012-100846] [Medline: 23152590]
8. Joyce S, Modini M, Christensen H, Mykletun A, Bryant R, Mitchell PB, et al. Workplace interventions for common mental disorders: a systematic meta-review. Psychol Med 2016 Mar;46(4):683-697. [doi: 10.1017/S0033291715002408] [Medline: 26620157]
9. Petrie K, Joyce S, Tan L, Henderson M, Johnson A, Nguyen H, et al. A framework to create more mentally healthy workplaces: a viewpoint. Aust N Z J Psychiatry 2018 Jan;52(1):15-23. [doi: 10.1177/0004867417726174] [Medline: 28835112]
10. Comas-Diaz L, Luthar SS, Maddi SR, O' Neill KH, Saakvitne KW, Tedeschi RG. American Psychological Association. 2016. The Road to Resilience URL: https://www.apa.org/helpcenter/road-resilience.aspx [accessed 2019-01-26] [WebCite Cache ID 75jLlFFDR]
11. Victoria CM, Ortiz-Tallo M, Cardenal V, De La Torre-Luque A. Positive psychology group intervention for breast cancer patients: a randomised trial. Psychol Rep 2014 Aug;115(1):44-64. [doi: 10.2466/15.20.PR0.115c17z7] [Medline: 25153949]
12. Kahn J, Collinge W, Soltysik R. Post-9/11 veterans and their partners improve mental health outcomes with a self-directed mobile and web-based wellness training program: a randomized controlled trial. J Med Internet Res 2016 Sep 27;18(9):e255 [FREE Full text] [doi: 10.2196/jmir.5800] [Medline: 27678169]
13. Loprinzi CE, Prasad K, Schroeder DR, Sood A. Stress Management and Resilience Training (SMART) program to decrease stress and enhance resilience among breast cancer survivors: a pilot randomized clinical trial. Clin Breast Cancer 2011 Dec;11(6):364-368. [doi: 10.1016/j.clbc.2011.06.008] [Medline: 21831722]
14. Sood A, Prasad K, Schroeder D, Varkey P. Stress management and resilience training among Department of Medicine faculty: a pilot randomized clinical trial. J Gen Intern Med 2011 Aug;26(8):858-861 [FREE Full text] [doi: 10.1007/s11606-011-1640-x] [Medline: 21279454]
15. Sood A, Sharma V, Schroeder DR, Gorman B. Stress Management and Resiliency Training (SMART) program among department of radiology faculty: a pilot randomized clinical trial. Explore (NY) 2014;10(6):358-363. [doi: 10.1016/j.explore.2014.08.002] [Medline: 25443423]

16. Mealer M, Conrad D, Evans J, Jooste K, Solyntjes J, Rothbaum B, et al. Feasibility and acceptability of a resilience training program for intensive care unit nurses. Am J Crit Care 2014 Nov;23(6):e97-105 [FREE Full text] [doi: 10.4037/ajcc2014747] [Medline: 25362680]
17. Arnetz BB, Nevedal DC, Lumley MA, Backman L, Lublin A. Trauma Resilience Training for Police: Psychophysiological and Performance Effects. J Police Crim Psych 2008 Dec 2;24(1):1-9 [FREE Full text] [doi: 10.1007/s11896-008-9030-y]
18. McCraty R, Atkinson M. Resilience training program reduces physiological and psychological stress in police officers. Glob Adv Health Med 2012 Nov;1(5):44-66 [FREE Full text] [doi: 10.7453/gahmj.2012.1.5.013] [Medline: 27257532]
19. Adler AB, Bliese PD, McGurk D, Hoge CW, Castro CA. Battlemind debriefing and battlemind training as early interventions with soldiers returning from iraq: randomization by platoon. J Consult Clin Psychol 2009 Oct;77(5):928-940. [doi: 10.1037/a0016877] [Medline: 19803572]
20. Adler AB, Williams J, McGurk D, Moss A, Bliese P. Resilience training with soldiers during basic combat training: randomisation by platoon. Appl Psychol Health Well Being 2015 Mar;7(1):85-107. [doi: 10.1111/aphw.12040] [Medline: 25641899]
21. Wild J. University of Oxford. 2016 May 7. An evaluation of Mind's resilence intervention for emergency workers URL: https://www.mind.org.uk/media/4627959/strand-3.pdf [accessed 2019-01-27] [WebCite Cache ID 75jU2vjfZ]
22. McFarlane AC, Bryant RA. Post-traumatic stress disorder in occupational settings: anticipating and managing the risk. Occup Med (Lond) 2007 Sep;57(6):404-410. [doi: 10.1093/occmed/kqm070] [Medline: 17728313]
23. Kaufmann C, Rutkow K, Spira AP, Mojtabai R. Mental health of protective services workers: results from the national epidemiologic survey on alcohol and related conditions. Disaster Med Public Health Prep 2013 Feb;7(1):36-45. [doi: 10.1001/dmp.2012.55] [Medline: 23193220]
24. Harvey SB, Milligan-Saville JS, Paterson HM, Harkness EL, Marsh AM, Dobson M, et al. The mental health of fire-fighters: an examination of the impact of repeated trauma exposure. Aust N Z J Psychiatry 2016 Jul;50(7):649-658. [doi: 10.1177/0004867415615217] [Medline: 26607303]
25. Leppin AL, Gionfriddo MR, Sood A, Montori VM, Erwin PJ, Zeballos-Palacios C, et al. The efficacy of resilience training programs: a systematic review protocol. Syst Rev 2014 Mar 6;3:20 [FREE Full text] [doi: 10.1186/2046-4053-3-20] [Medline: 24602236]
26. Aikens KA, Astin J, Pelletier KR, Levanovich K, Baase CM, Park YY, et al. Mindfulness goes to work: impact of an online workplace intervention. J Occup Environ Med 2014 Jul;56(7):721-731. [doi: 10.1097/JOM.0000000000000209] [Medline: 24988100]
27. Johnson DC, Thom NJ, Stanley EA, Haase L, Simmons AN, Shih PA, et al. Modifying resilience mechanisms in at-risk individuals: a controlled study of mindfulness training in Marines preparing for deployment. Am J Psychiatry 2014 Aug;171(8):844-853 [FREE Full text] [doi: 10.1176/appi.ajp.2014.13040502] [Medline: 24832476]
28. McGonagle AK, Beatty J, Joffe R. Coaching for workers with chronic illness: evaluating an intervention. J Occup Health Psychol 2014 Jul;19(3):385-398. [doi: 10.1037/a0036601] [Medline: 24796227]
29. Yu X, Stewart SM, Chui JP, Ho JL, Li AC, Lam TH. A pilot randomized controlled trial to decrease adaptation difficulties in chinese new immigrants to Hong Kong. Behav Ther 2014 Jan;45(1):137-152. [doi: 10.1016/j.beth.2013.10.003] [Medline: 24411121]
30. Songprakun W, McCann TV. Effectiveness of a self-help manual on the promotion of resilience in individuals with depression in Thailand: a randomised controlled trial. BMC Psychiatry 2012 Feb 16;12:12 [FREE Full text] [doi: 10.1186/1471-244X-12-12] [Medline: 22339984]
31. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
32. Keng SL, Smoski MJ, Robins CJ. Effects of mindfulness on psychological health: a review of empirical studies. Clin Psychol Rev 2011 Aug;31(6):1041-1056 [FREE Full text] [doi: 10.1016/j.cpr.2011.04.006] [Medline: 21802619]
33. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010 Apr;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
34. Kuyken W, Hayes R, Barrett B, Byng R, Dalgleish T, Kessler D, et al. Effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse or recurrence (PREVENT): a randomised controlled trial. Lancet 2015 Jul 4;386(9988):63-73 [FREE Full text] [doi: 10.1016/S0140-6736(14)62222-4] [Medline: 25907157]
35. Kuyken W, Weare K, Ukoumunne OC, Vicary R, Motton N, Burnett R, et al. Effectiveness of the Mindfulness in Schools Programme: non-randomised controlled feasibility study. Br J Psychiatry 2013 Aug;203(2):126-131. [doi: 10.1192/bjp.bp.113.126649] [Medline: 23787061]
36. Khoury B, Lecomte T, Fortin G, Masse M, Therien P, Bouchard V, et al. Mindfulness-based therapy: a comprehensive meta-analysis. Clin Psychol Rev 2013 Aug;33(6):763-771. [doi: 10.1016/j.cpr.2013.05.005] [Medline: 23796855]
37. Krusche A, Cyhlarova E, Williams JM. Mindfulness online: an evaluation of the feasibility of a web-based mindfulness course for stress, anxiety and depression. BMJ Open 2013 Nov 29;3(11):e003498 [FREE Full text] [doi: 10.1136/bmjopen-2013-003498] [Medline: 24293203]

38. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
39. Sood A, Prasad K, Schroeder D, Varkey P. Stress management and resilience training among department of medicine faculty: a pilot randomized clinical trial. J Gen Intern Med 2011 Aug;26(8):858-861 [FREE Full text] [doi: 10.1007/s11606-011-1640-x] [Medline: 21279454]
40. Joyce S, Shand F, Tighe J, Laurent S, Bryant RA, Harvey SB. Road to resilience: a systematic review and meta-analysis of resilience training programmes and interventions. BMJ Open 2018 Jun 14;8(6):e017858 [FREE Full text] [doi: 10.1136/bmjopen-2017-017858] [Medline: 29903782]
41. Robertson IT, Cooper CL, Sarkar M, Curran T. Resilience training in the workplace from 2003 to 2014: a systematic review. J Occup Organ Psychol 2015 Apr 25;88(3):533-562. [doi: 10.1111/joop.12120]
42. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Dec 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
43. Rose RD, Buckey Jr JC, Zbozinek TD, Motivala SJ, Glenn DE, Cartreine JA, et al. A randomized controlled trial of a self-guided, multimedia, stress management and resilience training program. Behav Res Ther 2013 Feb;51(2):106-112. [doi: 10.1016/j.brat.2012.11.003] [Medline: 23262118]
44. Abbott JA, Klein B, Hamilton C, Rosenthal AJ. The impact of online resilience training for sales managers on wellbeing and performance. E-J Appl Psychol 2009 Jun 15;5(1):89-95. [doi: 10.7790/ejap.v5i1.145]
45. Joyce S, Shand F, Bryant RA, Lal TJ, Harvey SB. Mindfulness-based resilience training in the workplace: pilot study of the internet-based Resilience@Work (RAW) mindfulness program. J Med Internet Res 2018 Sep 11;20(9):e10326 [FREE Full text] [doi: 10.2196/10326] [Medline: 30206055]
46. Harvey SB, Milligan-Saville JS, Paterson HM, Harkness EL, Marsh AM, Dobson M, et al. The mental health of fire-fighters: an examination of the impact of repeated trauma exposure. Aust N Z J Psychiatry 2016 Jul;50(7):649-658. [doi: 10.1177/0004867415615217] [Medline: 26607303]
47. Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales. Health Qual Life Outcomes 2011 Feb 4;9:8 [FREE Full text] [doi: 10.1186/1477-7525-9-8] [Medline: 21294858]
48. Campbell-Sills L, Cohan SL, Stein MB. Relationship of resilience to personality, coping, and psychiatric symptoms in young adults. Behav Res Ther 2006 Apr;44(4):585-599. [doi: 10.1016/j.brat.2005.05.001] [Medline: 15998508]
49. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
50. Aloba O, Olabisi O, Aloba T. The 10-Item Connor-Davidson Resilience Scale: factorial structure, reliability, validity, and correlates among student nurses in Southwestern Nigeria. J Am Psychiatr Nurses Assoc 2016;22(1):43-51. [doi: 10.1177/1078390316629971] [Medline: 26929231]
51. Notario-Pacheco B, Martínez-Vizcaíno V, Trillo-Calvo E, Pérez-Yus MC, Serrano-Parra D, García-Campayo J. Validity and reliability of the Spanish version of the 10-item CD-RISC in patients with fibromyalgia. Health Qual Life Outcomes 2014 Feb 1;12:14 [FREE Full text] [doi: 10.1186/1477-7525-12-14] [Medline: 24484847]
52. Wang L, Shi Z, Zhang Y, Zhang Z. Psychometric properties of the 10-item Connor-Davidson Resilience Scale in Chinese earthquake victims. Psychiatry Clin Neurosci 2010 Oct;64(5):499-504 [FREE Full text] [doi: 10.1111/j.1440-1819.2010.02130.x] [Medline: 20923429]
53. Christopher MS, Goerling RJ, Rogers BS, Hunsinger M, Baron G, Bergman AL, et al. A pilot study evaluating the effectiveness of a mindfulness-based intervention on cortisol awakening response and health outcomes among law enforcement officers. J Police Crim Psych 2015 Jan 23;31(1):15-28. [doi: 10.1007/s11896-015-9161-x]
54. Walach H, Buchheld N, Buttenmüller V, Kleinknecht N, Schmidt S. Measuring mindfulness—the Freiburg Mindfulness Inventory (FMI). Pers Individ Dif 2006 Jun;40(8):1543-1555. [doi: 10.1016/j.paid.2005.11.025]
55. Trousselard M, Steiler D, Raphel C, Cian C, Duymedjian R, Claverie D, et al. Validation of a French version of the Freiburg Mindfulness Inventory - short version: relationships between mindfulness and stress in an adult population. Biopsychosoc Med 2010 Aug 12;4:8 [FREE Full text] [doi: 10.1186/1751-0759-4-8] [Medline: 20704696]
56. Chen S, Zhou R. Validation of a Chinese version of the Freiburg Mindfulness Inventory–short version. Mindfulness 2013 Apr 4;5(5):529-535. [doi: 10.1007/s12671-013-0208-8]
57. Gillanders DT, Bolderston H, Bond FW, Dempster M, Flaxman PE, Campbell L, et al. The development and initial validation of the cognitive fusion questionnaire. Behav Ther 2014 Jan;45(1):83-101. [doi: 10.1016/j.beth.2013.09.001] [Medline: 24411117]
58. Ruiz FJ, Suárez-Falcón JC, Riaño-Hernández D, Gillanders D. Psychometric properties of the Cognitive Fusion Questionnaire in Colombia. Rev Latinoam Psicol 2017 Jan;49(1):80-87. [doi: 10.1016/j.rlp.2016.09.006]
59. Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, et al. Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. Behav Ther 2011 Dec;42(4):676-688. [doi: 10.1016/j.beth.2011.03.007] [Medline: 22035996]

60. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the Self-Compassion Scale. Clin Psychol Psychother 2011;18(3):250-255. [doi: 10.1002/cpp.702] [Medline: 21584907]
61. Smeets E, Neff K, Alberts H, Peters M. Meeting suffering with kindness: effects of a brief self-compassion intervention for female college students. J Clin Psychol 2014 Sep;70(9):794-807. [doi: 10.1002/jclp.22076] [Medline: 24691680]
62. Scheier MF, Carver C, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a re-evaluation of the Life Orientation Test. J Pers Soc Psychol 1994 Dec;67(6):1063-1078. [Medline: 7815302]
63. Glaesmer H, Rief W, Martin A, Mewes R, Brähler E, Zenger M, et al. Psychometric properties and population-based norms of the Life Orientation Test Revised (LOT-R). Br J Health Psychol 2012 May;17(2):432-445. [doi: 10.1111/j.2044-8287.2011.02046.x] [Medline: 22106985]
64. Carver C. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
65. Yusoff N, Low WY, Yip CH. Reliability and validity of the Brief COPE Scale (English version) among women with breast cancer undergoing treatment of adjuvant chemotherapy: a Malaysian study. Med J Malaysia 2010 Mar;65(1):41-44 [FREE Full text] [Medline: 21265247]
66. Carver C, Scheier M, Weintraub JK. Assessing coping strategies: a theoretically based approach. J Pers Soc Psychol 1989 Feb;56(2):267-283. [Medline: 2926629]
67. Scheier MF, Wrosch C, Baum A, Cohen S, Martire LM, Matthews KA, et al. The Life Engagement Test: assessing purpose in life. J Behav Med 2006 Jun;29(3):291-298. [doi: 10.1007/s10865-005-9044-1] [Medline: 16565785]
68. Platania S, Gruttadauria S, Citelli G, Giambrone L, Di Nuovo S. Associations of thalassemia major and satisfaction with quality of life: the mediating effect of social support. Health Psychol Open 2017;4(2):2055102917742054 [FREE Full text] [doi: 10.1177/2055102917742054] [Medline: 29379628]
69. Mackinnon A, Griffiths KM, Christensen H. Comparative randomised trial of online cognitive-behavioural therapy and an information website for depression: 12-month outcomes. Br J Psychiatry 2008 Feb;192(2):130-134. [doi: 10.1192/bjp.bp.106.032078] [Medline: 18245031]
70. 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]
71. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux P, Consolidated Standards of Reporting Trials Group. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010 Aug;63(8):e1-37 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.03.004] [Medline: 20346624]
72. Kaplan JB, Bergman AL, Christopher M, Bowen S, Hunsinger M. Role of resilience in mindfulness training for first responders. Mindfulness 2017 Apr 19;8(5):1373-1380. [doi: 10.1007/s12671-017-0713-2]
73. Davidson JR, Connor KM. Connor-Davidson Resilience Scale (CD-RISC).: Unpublished; 2016. Connor-Davidson Resilience Scale (CD-RISC) Manual URL: http://www.cd-risc.com [accessed 2019-01-28] [WebCite Cache ID 75jajgywg]
74. Bonanno GA, Romero SA, Klein SI. The temporal elements of psychological resilience: an integrative framework for the study of individuals, families, and communities. Psychol Inq 2015 Jun 11;26(2):139-169. [doi: 10.1080/1047840X.2015.992677]
75. Joyce S, Tan L, Shand F, Bryant RA, Harvey SB. Can resilience be measured and used to predict mental health symptomology among first responders exposed to repeated trauma? J Occup Environ Med 2018 Dec 18:-. [doi: 10.1097/JOM.0000000000001526] [Medline: 30575696]
76. Wild JS, Smith KV, Thompson E, Béar F, Lommen MJ, Ehlers A. A prospective study of pre-trauma risk factors for post-traumatic stress disorder and depression. Psychol Med 2016 Dec;46(12):2571-2582 [FREE Full text] [doi: 10.1017/S0033291716000532] [Medline: 27348599]

Abbreviations

AAQ-II: Acceptance and Action Questionnaire version 2
ACT: acceptance and commitment therapy
Brief-COPE: Brief-Coping Orientation to Problems Experienced
BRS: Brief Resilience Scale
CDRISC_10: 10-item version of the Connor-Davidson Resilience Scale
CFQ: cognitive fusion questionnaire
eHealth: electronic health
FMI: Freiburg Mindfulness Inventory
FRNSW: Fire and Rescue New South Wales
HLP: Healthy Living Program
LET: Life Engagement Test
LOT-R: Life Orientation Test-Revised
MMRM: mixed-methods repeated measures

NSW: New South Wales
PTSD: posttraumatic stress disorder
RAW: Resilience@Work
RCT: randomized controlled trial
RTP: resilience training program
SCS: Self-Compassion Scale

Edited by G Eysenbach; submitted 27.11.18; peer-reviewed by J Kahn, S Pezaro; comments to author 02.01.19; revised version received 13.01.19; accepted 20.01.19; published 19.02.19.

Please cite as:
Joyce S, Shand F, Lal TJ, Mott B, Bryant RA, Harvey SB
Resilience@Work Mindfulness Program: Results From a Cluster Randomized Controlled Trial With First Responders
J Med Internet Res 2019;21(2):e12894
URL: http://www.jmir.org/2019/2/e12894/
doi:10.2196/12894
PMID:30777846

Original Paper

A Smoking Prevention Program Delivered by Medical Students to Secondary Schools in Brazil Called "Education Against Tobacco": Randomized Controlled Trial

Oscar Campos Lisboa[1]; Breno Bernardes-Souza[1], MD; Luiz Eduardo De Freitas Xavier[1]; Matheus Rocha Almeida[1]; Paulo César Rodrigues Pinto Corrêa[1], MD, MPH; Titus Josef Brinker[2,3,4], MD

[1]School of Medicine, Federal University of Ouro Preto, Ouro Preto, Brazil

[2]National Center for Tumor Diseases, German Cancer Research Center, University of Heidelberg, Heidelberg, Germany

[3]Department of Dermatology, University Hospital Heidelberg, Heidelberg, Germany

[4]German Cancer Consortium, University of Heidelberg, Heidelberg, Germany

Corresponding Author:
Titus Josef Brinker, MD
National Center for Tumor Diseases
German Cancer Research Center
University of Heidelberg
Im Neuenheimer Feld 460
Heidelberg, 69120
Germany
Phone: 49 62213219304
Email: titus.brinker@nct-heidelberg.de

Abstract

Background: Smoking is the largest preventable cause of mortality in Brazil. Education Against Tobacco (EAT) is a network of more than 3500 medical students and physicians across 14 countries who volunteer for school-based smoking prevention programs. EAT educates 50,000 adolescents per year in the classroom setting. A recent quasi-experimental study conducted in Germany showed that EAT had significant short-term smoking cessation effects among adolescents aged 11 to 15 years.

Objective: The aim is to measure the long-term effectiveness of the most recent version of the EAT curriculum in Brazil.

Methods: A randomized controlled trial was conducted among 2348 adolescents aged 12 to 21 years (grades 7-11) at public secondary schools in Brazil. The prospective experimental design included measurements at baseline and at 6 and 12 months postintervention. The study groups comprised randomized classes receiving the standardized EAT intervention (90 minutes of mentoring in a classroom setting) and control classes in the same schools (no intervention). Data were collected on smoking status, gender, social aspects, and predictors of smoking. The primary endpoint was the difference in the change in smoking prevalence between the intervention group and the control group at 12-month follow-up.

Results: From baseline to 12 months, the smoking prevalence increased from 11.0% to 20.9% in the control group and from 14.1% to 15.6% in the intervention group. This difference was statistically significant (P<.01). The effects were smaller for females (control 12.4% to 18.8% vs intervention 13.1% to 14.6%) than for males (control 9.1% to 23.6% vs intervention 15.3% to 16.8%). Increased quitting rates and prevented onset were responsible for the intervention effects. The differences in change in smoking prevalence from baseline to 12 months between the intervention and control groups were increased in students with low school performance.

Conclusions: To our knowledge, this is the first randomized trial on school-based tobacco prevention in Brazil that shows significant long-term favorable effects. The EAT program encourages quitting and prevents smoking onset, especially among males and students with low educational background.

Trial Registration: ClinicalTrials.gov NCT02725021; https://clinicaltrials.gov/ct2/show/NCT02725021

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.7134

(J Med Internet Res 2019;21(2):e12854) doi:10.2196/12854

KEYWORDS

smoking; tobacco; prevention; medical students; schools

Introduction

In 2015, smoking accounted for 156,216 deaths and 3.72 million disability-adjusted life years in Brazil, representing a direct cost for the health system of R$39.4 billion [1]. More than 30.0% of Brazilian boys and 27% of girls aged 13 to 15 years had tried smoking before the age of 12 [2]. Given the time adolescents spend in the school setting, schools represent an excellent opportunity to deliver smoking prevention programs.

Current Knowledge on School-Based Tobacco Prevention

Most school-based tobacco control programs are ineffective, but data from Brazil remain scarce [3-5]. Recent trials on tobacco prevention in the school setting have focused on including school teachers in the intervention [6-8], with others involving families [9,10]. However, these studies concluded that the students' environment (ie, peer group as well as parental behavior and school policies) plays a role in smoking initiation in adolescence.

A randomized controlled trial involving different school-based interventions to reduce the use of various psychotropic substances among 1316 students in Brazil showed mixed effects for different drugs; however, the study design had limitations that precluded interpretation [11]. Another study from Brazil analyzed the effectiveness of an educational intervention by the Brazilian Cancer Institute (INCA) on smoking among school adolescents. Those researchers randomized 32 schools to either control (no intervention) or intervention arms, with a total sample of 2200 students in grades 7 and 8 (aged 13-14 years). INCA members lectured teachers from schools in the intervention arm about tobacco control, with the expectation that those teachers would discuss tobacco-related topics with their students. No change in smoking prevalence was found at the study endpoint, but knowledge about passive smoking had improved [12].

Education Against Tobacco

Education Against Tobacco (EAT) is a network of volunteer medical students and physicians from more than 80 medical schools in 14 countries worldwide that was founded in Germany in 2012 [13]. The network has its roots in school-based interventions delivered by medical students. These interventions cost approximately US $20 per participating class and can reach up to 50,000 students per year worldwide. EAT is also involved in medical education research on smoking cessation counseling, science-based multilanguage apps, and public awareness and advocacy for tobacco control [14-16].

A German quasi-experimental study showed the school-based intervention resulted in a significant reduction in smoking prevalence among secondary school students at 6-month follow-up [17,18]. A randomized follow-up study in Germany indicated effectiveness at the 12-month follow-up; however, the results were not significant because of a large loss-to-follow-up effect [19]. Recent studies indicate that physicians substantially undertreat tobacco addiction compared with other chronic conditions, such as diabetes or hypertension [20-22]. The school-based smoking prevention provided by EAT is thought to sensitize medical students for tobacco control in general, as well as having a direct effect on adolescents [13].

A facial-aging app was implemented as part of the school-based EAT interventions to increase effectiveness. Facial-aging interventions, in which a selfie is altered to predict future appearance, provide motivation for healthier behavioral choices in adiposity prevention, skin cancer prevention, and smoking cessation [14,16,19,23-39]. An explanation for these preliminary results may be the high importance of appearance for a persons' self-concept, especially in adolescence [40]. However, to our knowledge, the only completed randomized trial that investigated the prospective effectiveness of a facial-aging intervention on actual behavior (smoking) was conducted by Burford et al [41]. In that study, five of 80 control group participants (6.3%) suggested they had quit smoking at the 6-month follow-up, and 22 of 80 intervention group participants (27.5%) reported quitting (P<.05).

This study aimed to determine the long-term effectiveness of the school-based EAT intervention in reducing smoking prevalence among secondary school students in Brazil, as per the study protocol [42].

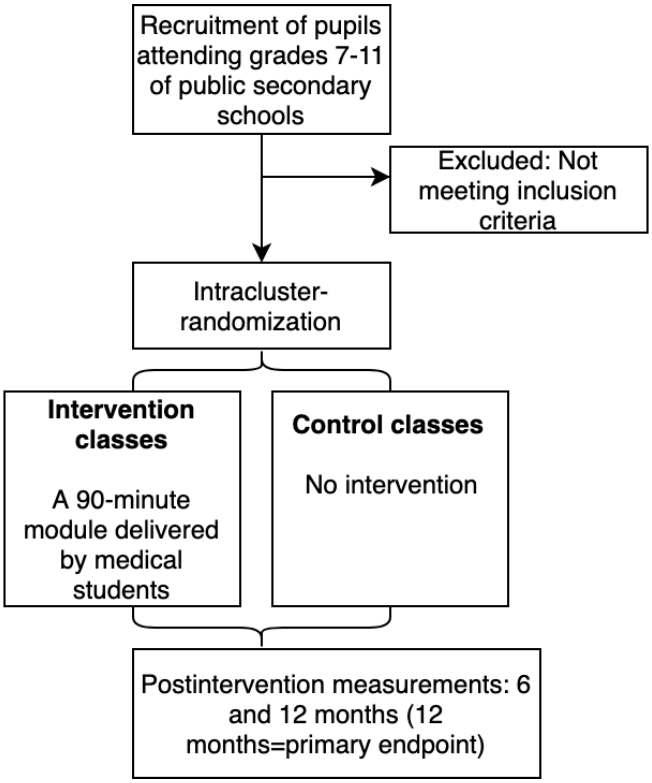
Methods

A randomized controlled trial was conducted among 2384 adolescents in grades 7 to 11 from secondary schools in Brazil from February 2017 to June 2018 (Figure 1). All predefined time points were met. Details of the study design and the development of the questionnaire are outlined in our previously published study protocol [42].

Participants

Students in grades 7 to 11 at secondary public schools were eligible to participate in this study. In total, 2348 secondary school students from 110 classes (from 14 eligible schools), who fulfilled the inclusion criteria, entered baseline data. Baseline data (t1) were collected from February 2017 to May 2017. Follow-up data (t2 and t3) were collected 6 and 12 months after that (from August 2017 to June 2018). Overall, 1353 participants provided data at both t1 and t3, which were used for primary endpoint analyses. The loss to follow-up effect was 42.38% (995/2348).

Figure 1. Study design and flow of participants.



Attrition Analysis

To evaluate attrition bias, participants who dropped out at follow-up (t3) were analyzed using logistic regression analysis (1=dropout and 0=analysis sample; t1 and t3 participation as the dependent variable). There was no systematic bias regarding the main effect of study group (intervention vs control; *P*=.24) and the interactions between study group and gender (*P*=.61), study group and age (*P*=.56), or study group and grade (grades 7-11; *P*=.10). However, there was systematic bias regarding the interaction of study group (0=control and 1=intervention) and smoking status (0=no tobacco consumption during the last 30 days and 1=at least one regular or straw cigarette, water pipe/hookah, or e-cigarette; odds ratio [OR] 0.63, 95% CI 0.42-0.94, *P*=.02). Among smokers, there were fewer dropouts in the intervention group than in the control group. For nonsmokers, the dropout rates were about the same in both study groups (intervention group/smoker: 52.1%, 114/219 dropout; control group/smoker: 63.4%, 116/183 dropout; intervention group/nonsmoker: 39.0%, 409/1084 dropout; control group/nonsmoker: 39.6%, 356/898 dropout).

The dropout probability differed for the main effects of four characteristics: smoking status (OR 2.67, 95% CI 1.95-3.66, *P*<.001), gender (OR 1.57, 95% CI 1.24-2.00, *P*<.001), age (OR 1.14, 95% CI 1.09-1.19, *P*<.001), and academic performance (OR 1.39, 95% CI 1.27-1.53, *P*<.001). Age and academic performance were used as the metric. The ORs represented a change in the characteristic by one level.

Dropouts distorted the remaining sample for the analysis in four directions: younger age, more girls, students with better performance, and fewer smokers. Reasons for loss to follow-up included identifier code not assignable, change of school, unauthorized absence from class (truancy), illness, or grade retention.

Intervention: Education Against Tobacco

The EAT school-based intervention comprises a 90-minute module in the classroom setting (about 25 students per class) delivered by two medical students per classroom. The intervention was implemented by 36 volunteering medical students from the Federal University of Ouro Preto EAT group who received standardized training in advance, which was monitored via a performance questionnaire by the instructor. The program covered features of smoking that students can relate to in their everyday life in a gain-framed and interactive manner, which also involved a three-dimensional facial-aging app, “Smokerface,” developed by EAT. The intervention is described in detail in our study protocol [42].

Outcomes

The primary outcome was the difference in smoking prevalence from baseline (t1) to 12 months of follow-up (t3) in the control group versus the difference from t1 to t3 in the intervention group. The differences in smoking behavior (smoking onset, quit attempts) between the two groups were studied as secondary outcomes, along with gender-specific effects.

Data Entry

Data entry was performed manually at the Federal University of Ouro Preto in Brazil, using Microsoft Excel (Microsoft Corp, Redmond, WA, USA) and SPSS (IBM Corp, Armonk, NY, USA).

Statistical Analyses

We used chi-square tests, *t* tests, and Fisher exact tests to examine baseline differences. The effects of predictors (gender

and social characteristics) on smoking behavior at 12 months were calculated using robust panel logistic regression analysis. The significance level was set at 5% for *t* tests (two-sided) with 95% CIs (two-sided). Statistical analyses were performed using SPSS version 23 (IBM Corp, Armonk, NY, USA). The group allocation of the study sample was based on class level. Statistically robust panel logistic regression was used (SPSS GENLINMIXED procedure) to account for clustering. This procedure was also used to calculate the difference in the smoking prevalence from baseline to 12-month follow-up in the control group versus that in the intervention group (primary outcome). The number needed to treat (NNT) was calculated for the total effect (preventing smoking onset and initiating quit attempts).

Ethics Approval and Consent to Participate

In accordance with Guidelines for Good Epidemiologic Practice [43], the study protocol was submitted for approval to the responsible ethics committee (Federal University of Ouro Preto, Brazil) and consent was obtained. All legal and data protection issues were discussed with the responsible authorities, and all participants were required to provide informed consent.

Results

Baseline Data

The mean age of the 2348 participants at baseline (Multimedia Appendix 1) was 14.8 years (range 12-21 years), and 50.72% (1191/2348) were female. At baseline, the survey identified 7.79% (183/2348) of participants as regular cigarette smokers, 12.05% (283/2348) as straw cigarette smokers, 4.64% (109/2348) as water pipe/hookah smokers, and 2.21% (52/2348) as e-cigarette smokers.

New derived variables were calculated for the analyses. In total, 14.78% (347/2348) of participants had smoked at least one regular or straw cigarette during the past 30 days, 5.88% (138/2348) had used a new tobacco product during the past 30 days (new tobacco product defined as water pipe/hookah or e-cigarettes), and 17.12% (402/2348) had used at least one of these during the past 30 days (regular/straw cigarettes, water pipe/hookah, or e-cigarette). The last characteristic was used as the criterion for the primary outcome and did not differ between the randomly assigned treatment groups at baseline (P=.83, Multimedia Appendix 1). However, there were strong influences of age, grade, and school performance on smoking behavior (Multimedia Appendix 2).

Follow-Up Data

Data analyses were based on the originally assigned groups (Table 1). There were 744 students in the intervention group and 609 in the control group who participated in the survey at both baseline and at 12-month follow-up that could be identified (baseline sample N=2348; prospective sample: n=1353; lost to follow-up: n=995).

From baseline to 12-month follow-up, the smoking prevalence increased from 11.0% (67/609) to 20.9% (127/609) in the control group and from 14.1% (105/744) to 15.6% (116/744) in the intervention group (NNT=19), with an effect for female gender (control 12.4%, 43/346 to 18.8%, 65/346 vs intervention 13.1%, 54/411 to 14.6%, 60/411) and an even greater effect for males (control 9.1%, 24/263 to 23.6%, 62/263 vs intervention 15.3%, 51/333 to 16.8%, 56/333) (Table 1).

Primary Outcome

The difference in the change in smoking prevalence between the control and intervention groups was statistically significant at 8.1% (95% CI 3.5%-12.7%, P<.001, *t* test calculated from estimated parameters). Smoking prevalence increased over the 12 months in both groups. However, in the intervention group, smoking only increased by 1.5% (95% CI −1.5% to 4.5%) compared to a 9.6% (95% CI 6.1%-13.1%) increase in the control group (Table 2).

Table 2 was calculated using the GENLINMIXED procedure (SPSS version 23) for a binomially distributed dependent variable with logit-link and an unstructured covariance matrix. Degrees of freedom were adjusted using the Satterthwaite approximation, and robust estimates were used. The class affiliation was used as random factor, meaning the clustered structure of the experimental design was taken into account. Age and academic performance at baseline were used for model adjustment. Academic performance was dichotomized as 0=good and very good; 1=reasonable, poor, and very poor. The (significant) influence of age and academic performance remained constant for the estimation of prevalence.

Secondary Outcomes

The intervention effect could be explained by the EAT program preventing adolescents from starting smoking as well as encouraging quitting attempts, as can be seen in Table 3.

If the intervention and control group were additionally differentiated according to gender, there were four groups (Table 4). Smoking prevalence increased in all four groups. In the intervention group (both males and females), smoking increased by 1.5%. However, in the control group, the increase in smoking prevalence differed for males and females: 13.6% for males and 6.4% for females.

The difference in the change in smoking prevalence between the control and intervention groups was statistically significant among males (difference 12.1%, 95% CI 5.0%-19.1%, P<.001; *t* test calculated from estimated parameters). However, the difference in the change in smoking prevalence between the control and intervention groups was not statistically significant for females (difference 4.9%, 95% CI −0.7% to 10.4%, P=.09; *t* test calculated from estimated parameters). If only the control group was considered, the difference in the increase in smoking prevalence between males and females was significant (P=.049, *t* test). Figure 2 illustrates the gender-specific findings.

Table 1. Smoking prevalence at baseline (t1) and at the 6- (t2) and 12-month (t3) follow-ups.

Time point[a]	Total sample, n (%)	Intervention group, n (%)	Control group, n (%)	Number needed to treat
Total				
t1	172/1353 (12.7)	105/744 (14.1)	67/609 (11.0)	
t2	150/899 (16.7)	64/472 (13.6)	86/427 (20.1)	
t3	243/1353 (18.0)	116/744 (15.6)	127/609 (20.9)	19/12[b]
Gender				
Female				
t1	97/757 (12.8)	54/411 (13.1)	43/346 (12.4)	
t2	86/518 (16.6)	39/266 (14.7)	47/252 (18.7)	
t3	125/757 (16.5)	60/411 (14.6)	65/346 (18.8)	24
Male				
t1	75/596 (12.6)	51/333 (15.3)	24/263 (9.1)	
t2	64/381 (16.8)	25/206 (12.1)	39/175 (22.3)	
t3	118/596 (19.8)	56/333 (16.8)	62/263 (23.6)	15
Grade at baseline				
7				
t1	25/341 (7.3)	19/213 (8.9)	6/128 (4.7)	
t2	30/208 (14.4)	12/133 (9.0)	18/75 (24.0)	
t3	56/341 (16.4)	28/213 (13.1)	28/128 (21.9)	11
8				
t1	40/296 (13.5)	34/192 (17.7)	6/104 (5.8)	
t2	38/185 (20.5)	24/114 (21.1)	14/71 (19.7)	
t3	52/296 (17.6)	27/192 (14.1)	25/104 (24.0)	10
9				
t1	14/86 (16.3)	6/34 (17.6)	8/52 (15.4)	
t2	14/56 (25.0)	3/10 (30.0)	11/46 (23.9)	
t3	19/86 (22.1)	9/34 (26.5)	10/52 (19.2)	−14
10				
t1	48/317 (15.1)	18/116 (15.5)	30/201 (14.9)	
t2	45/245 (18.4)	14/96 (14.6)	31/149 (20.8)	
t3	61/317 (19.2)	23/116 (19.8)	38/201 (18.9)	−108
11				
t1	45/313 (14.4)	28/189 (14.8)	17/124 (13.7)	
t2	23/205 (11.2)	11/119 (9.2)	12/86 (14.0)	
t3	55/313 (17.6)	29/189 (15.3)	26/124 (21.0)	18
Academic performance[c] (at baseline)				
Very good				
t1	28/335 (8.4)	21/197 (10.7)	7/138 (5.1)	
t2	24/216 (11.1)	12/122 (9.8)	12/94 (12.8)	
t3	43/335 (12.8)	25/197 (12.7)	18/138 (13.0)	283
Good				
t1	68/577 (11.8)	40/299 (13.4)	28/278 (10.1)	

Time point[a]	Total sample, n (%)	Intervention group, n (%)	Control group, n (%)	Number needed to treat
t2	54/411 (13.1)	23/212 (10.8)	31/199 (15.6)	
t3	87/577 (15.1)	37/299 (12.4)	50/278 (18.0)	18
Reasonable				
t1	64/363 (17.6)	34/201 (16.9)	30/162 (18.5)	
t2	53/214 (24.8)	19/105 (18.1)	34/109 (31.2)	
t3	91/363 (25.1)	42/201 (20.9)	49/162 (30.2)	11
Poor				
t1	9/57 (15.8)	7/34 (20.6)	2/23 (8.7)	
t2	14/43 (32.6)	8/25 (32.0)	6/18 (33.3)	
t3	16/57 (28.1)	9/34 (26.5)	7/23 (30.4)	25
Very poor				
t1	3/21 (14.3)	3/13 (23.1)	0/8 (0.0)	
t2	5/15 (33.3)	2/8 (25.0)	3/7 (42.9)	
t3	6/21 (28.6)	3/13 (23.1)	3/8 (37.5)	7

[a]The case number at t1 and t3 corresponds to all participants at t3 (N=1353, the basis for the analysis of the primary outcome). For better comparability, only cases are used for t2 that were also present at t3 (N=899).

[b]The NNT is 12 if the baseline differences are taken into account.

[c]Academic performance was assessed by the students themselves through one single 5-point Likert scale item in the study questionnaire.

Table 2. Change in smoking prevalence in the intervention and control groups from baseline to 12-month follow-up.

Pairwise contrasts	Survey wave pairwise contrasts	Contrast estimate	SE	*t* (*df*)	*P* value[a]	95% CI[b]
Intervention	12-month follow-up from baseline	0.015	0.015	0.975 (883)	.33	−0.015, 0.045
Control	12-month follow-up from baseline	0.096	0.018	5.413 (714)	<.001	0.061, 0.131
Difference in change	Control and intervention groups	0.081	0.023	3.456 (1595)	<.001	0.035, 0.127

[a]The sequential Sidak adjusted significance level is .05.

[b]Confidence interval bounds are approximate.

Table 3. Nominal and percentage effects of the intervention on the smoking status (secondary outcomes) from baseline to 12-month follow-up (N=1353; *P*=.001 [Fisher test]).

Sample	Prospective smoking status (t1-t3)			
	Remained a nonsmoker (n=1041)	Started smoking (n=140)	Quit smoking (n=69)	Remained a smoker (n=103)
Control group, n (%)	460 (75.5)	82 (13.5)	22 (3.6)	45 (7.4)
Intervention group, n (%)	581 (78.1)	58 (7.8)	47 (6.3)	58 (7.8)

The change in smoking behavior was also examined separately for the intervention and control groups. In the control group, there was a 13.5% change from nonsmokers to smokers, but only a 3.6% change from smokers to nonsmokers. The proportion of smokers in the control group increased from 11.0% to 20.9%. This difference was statistically significant (*P*<.001, McNemar test). However, in the intervention group, the changes from smokers to nonsmokers and from nonsmokers to smokers were roughly balanced. The proportion of smokers increased moderately from 14.1% (105/744) to 15.6% (116/744), but the difference was not significant (*P*=.34, McNemar test). This suggests the intervention prevented a further increase in the proportion of smokers.

The differences in change in smoking prevalence from baseline to 12-month follow-up between the intervention and control groups were increased in participants with low school performance (Table 5). Dichotomized academic performance at baseline was used for the calculation (0=good or very good; 1= reasonable, poor, or very poor).

Table 4. Nominal and percentage effects of the intervention on smoking status by gender.

Sample	Prospective smoking status (t1-t3)			
	Remained a nonsmoker (n=1041)	Started smoking (n=140)	Quit smoking (n=69)	Remained a smoker (n=103)
Control group, n (%)				
Total	460 (75.5)	82 (13.5)	22 (3.6)	45 (7.4)
Female	267 (77.2)	36 (10.4)	14 (4.0)	29 (8.4)
Male	193 (73.4)	46 (17.5)	8 (3.0)	16 (6.1)
Intervention group, n (%)				
Total	581 (78.1)	58 (7.8)	47 (6.3)	58 (7.8)
Female	324 (78.8)	33 (8.0)	27 (6.6)	27 (6.6)
Male	257 (77.2)	25 (7.5)	20 (6.0)	31 (9.3)
Total (n)				
Female	591	69	41	56
Male	450	71	28	47

Figure 2. Effects of the intervention on smoking status by gender.

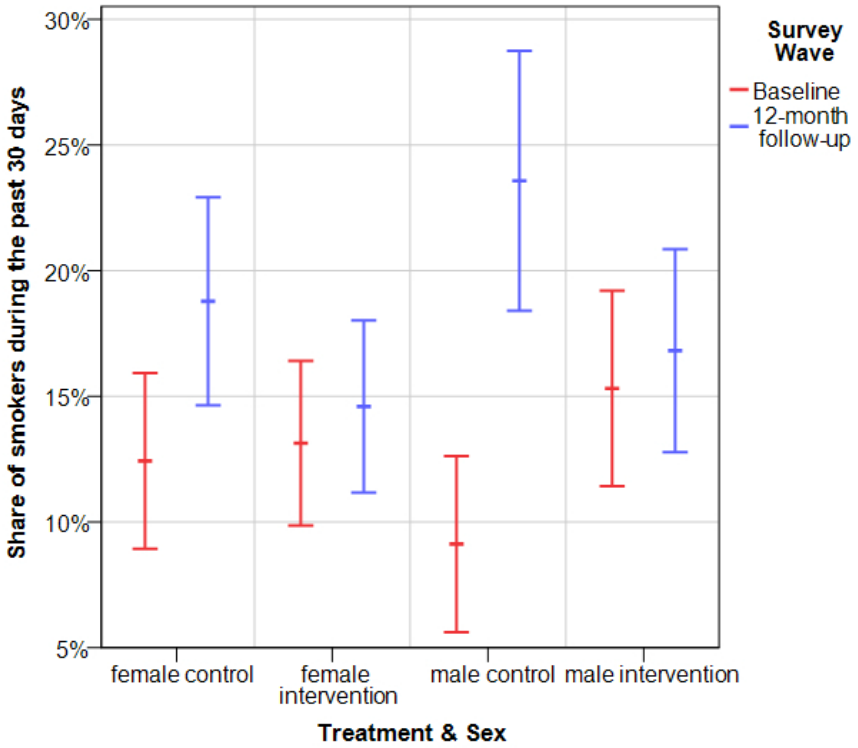


Table 5. Effect of school performance on the change in the prevalence of smoking, calculated with GENLINMIXED.

Academic performance and treatment	Survey wave pairwise contrasts	Contrast estimate	SE	*t* (*df*)	Adjusted significance[a]	95% CI[b]
Not good						
Intervention	12-month follow-up from baseline	0.041	0.038	1.094 (356)	.28	−0.033, 0.115
Control	12-month follow-up from baseline	0.138	0.033	4.141 (1155)	<.001	0.073, 0.203
Good						
Intervention	12-month follow-up from baseline	0.002	0.014	0.145 (2077)	.89	−0.026, 0.031
Control	12-month follow-up from baseline	0.077	0.020	3.936 (754)	<.001	0.039, 0.116

[a]The sequential Sidak adjusted significance level is .05.

[b]Confidence interval bounds are approximate.

Discussion

Principal Findings

To our knowledge, this is the first randomized trial on school-based tobacco prevention in Brazil that showed significant (P<.01) results in favor of the intervention. From baseline to the 12-month follow-up, the smoking prevalence increased from 11.0% to 20.9% in the control group and from 14.1% to 15.6% in the intervention group. This effect was increased for students with low educational background (ie, with low academic performance), which suggests that the intervention may contribute to reducing social inequalities among Brazilian adolescents, which are enhanced by tobacco addiction [44-48]. In addition, this study represents the first time the current (2015) EAT school curriculum was prospectively evaluated globally. Therefore, our findings have high relevance for the global EAT network (educationtobacco.org). In Brazil, EAT is currently established at 15 Brazilian medical schools, with more than 200 medical students volunteering to educate more than 8000 adolescents per year. Therefore, our findings also have high local relevance. Considering the NNT we found in this study (NNT=12 with baseline differences taken into account) and the current structure of EAT in Brazil (8000 adolescents covered annually), we can extrapolate that 667 Brazilian adolescents per year (8000/12) either quit smoking or do not start smoking due to EAT.

Local and international observations indicate that school-based preventions strongly motivated medical students within EAT for further tobacco control activities, such as engaging with politicians or improving medical education in regard to smoking cessation by organizing elective courses for their peers [13]. In Brazil, a national tobacco control award was organized by local medical students involved with EAT to recognize the best idea for enhancing tobacco control [49].

Limitations

Lack of Biochemical Validation

Originally, we planned to biochemically validate our findings via carbon monoxide measurements with a portable CO analyzer "Smokerlyzer piCO+" (Bedfont Scientific, Maidstone, United Kingdom) that was purchased for the study [42]. However, random measurements were not possible as some schools and teachers refused to participate. This made systematic collection impossible, and we were not able to collect a sample for analysis.

Generalizability

As our research was not conducted multinationally, we cannot generalize our results to different countries and cultural backgrounds. However, the similarity between the results found here and the ones found in our German studies [17-19] increases the international validity of our research. Also, as this study was performed in the setting of Brazilian public schools, the results might not be generalizable to private schools. The fact that there were some students aged 18 to 21 years in our study (the normal expected age range for grades 7-11 would be 12-17 years) may reflect higher rates of grade retention in public schools. Nevertheless, after excluding those participants aged 18 to 21 years (117/2348), we repeated all statistical analyses (data not shown) and noted that the significant results found for the total number of students remained the same.

Comparison With Prior Work in Brazil

This is the first randomized trial on school-based tobacco prevention in Brazil that showed significant (P<.01) results in favor of the intervention. Prior work failed to show such an effect mainly due to lack of sample size, funding, and incompletely implemented interventions [11,12]. By using motivated volunteer medical students, we were not dependent on the participation and training of teachers but had a preselected enthusiastic team that achieved complete implementation in the intervention classes. However, although the use of medical students has many benefits [13], it limits the number of adolescents that may be reached by the intervention.

Conclusions

The EAT intervention prevents smoking by encouraging quitting and preventing smoking onset, especially among males and students with low educational background.

Acknowledgments

We would like to thank all participating schools, pupils, teachers, and medical students who helped to organize the classroom visits in Ouro Preto. We also would like to thank Federal University of Ouro Preto, Ouro Preto Government Department of Education, Ouro Preto Government Department of Health, and the Office of Ouro Preto Regional Superintendent of Schools for their contribution in determining the project logistics.

Authors' Contributions

TJB invented, designed, and organized the intervention, provided the original classroom materials translated into Portuguese, wrote the manuscript, provided all figures, conceived the study design, provided the questionnaire, and supported conducting the study, data entry, and performed the statistical analysis. OCL, BBS, LEFX, and PCRPC contributed to the design of the study, assisted with the translation of classroom materials, wrote parts of the manuscript, and reviewed the final version of the manuscript. OCL, BBS, LEFX, and MRA coordinated the logistics of the study in Brazil, data collection, and data entry. All authors declare responsibility for the data and findings presented and have full access to the dataset.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Descriptive characteristics at baseline: all cases including dropouts.

[PDF File (Adobe PDF File), 46KB - jmir_v21i2e12854_app1.pdf]

Multimedia Appendix 2

Some bivariate relationships at baseline: all cases including dropouts.

[PDF File (Adobe PDF File), 32KB - jmir_v21i2e12854_app2.pdf]

Multimedia Appendix 3

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 97KB - jmir_v21i2e12854_app3.pdf]

References

1. Pinto M, Bardach A, Palacios A, Biz A, Alcaraz A, Rodríguez B, et al. Carga de doença atribuível ao uso do tabaco no Brasil e potencial impacto do aumento de preços por meio de impostos. Documento técnico IECS N° 21. Buenos Aires, Argentina: Instituto de Efectividad Clínica y Sanitaria; 2017.
2. Barreto S, Giatti L, Oliveira-Campos M, Andreazzi M, Malta DC. Experimentation and use of cigarette and other tobacco products among adolescents in the Brazilian state capitals (PeNSE 2012). Rev Bras Epidemiol 2014;17 Suppl 1:62-76 [FREE Full text] [Medline: 25054254]
3. Thomas R, McLellan J, Perera R. Effectiveness of school-based smoking prevention curricula: systematic review and meta-analysis. BMJ Open 2015 Mar 10;5(3):e006976 [FREE Full text] [doi: 10.1136/bmjopen-2014-006976] [Medline: 25757946]
4. Peirson L, Ali M, Kenny M, Raina P, Sherifali D. Interventions for prevention and treatment of tobacco smoking in school-aged children and adolescents: a systematic review and meta-analysis. Prev Med 2016 Apr;85:20-31. [doi: 10.1016/j.ypmed.2015.12.004] [Medline: 26743631]
5. Hefler M, Liberato S, Thomas DP. Incentives for preventing smoking in children and adolescents. Cochrane Database Syst Rev 2017 Dec 06;6:CD008645. [doi: 10.1002/14651858.CD008645.pub3] [Medline: 28585288]
6. Bast L, Due P, Ersbøll AK, Damsgaard M, Andersen A. Association of school characteristics and implementation in the X:IT Study-a school-randomized smoking prevention program. J Sch Health 2017 Dec;87(5):329-337. [doi: 10.1111/josh.12500] [Medline: 28382673]
7. Midford R, Cahill H, Lester L, Foxcroft D, Ramsden R, Venning L. Smoking prevention for students: findings from a three-year program of integrated harm minimization school drug education. Subst Use Misuse 2016;51(3):395-407. [doi: 10.3109/10826084.2015.1110173] [Medline: 26886503]
8. Garnham-Lee K, Trigwell J, McGee CE, Knowles Z, Foweather L. Impact and acceptability of the coach and teacher training within a school-based sport-for-health smoking prevention intervention: SmokeFree Sports. J Child Adoles Subst 2016 Jun 17;25(6):606-612. [doi: 10.1080/1067828X.2016.1175982]

9. Brown N, Luckett T, Davidson P, DiGiacomo M. Family-focussed interventions to reduce harm from smoking in primary school-aged children: a systematic review of evaluative studies. Prev Med 2017 Aug;101:117-125. [doi: 10.1016/j.ypmed.2017.06.011] [Medline: 28601619]
10. Chan S, Cheung Y, Fong D, Emmons K, Leung A, Leung D, et al. Family-based smoking cessation intervention for smoking fathers and nonsmoking mothers with a child: a randomized controlled trial. J Pediatr 2017 Dec;182:260.e4-266.e4. [doi: 10.1016/j.jpeds.2016.11.021] [Medline: 27989407]
11. do Nascimento MO, De Micheli D. Evaluation of different school-based preventive interventions for reducing the use of psychotropic substances among students: a randomized study. Cien Saude Colet 2015 Aug;20(8):2499-2510 [FREE Full text] [doi: 10.1590/1413-81232015208.15152014] [Medline: 26221815]
12. Malcon M, Menezes A, Assunção M, Neutzling M, Challal P. Efetividade de uma intervenção educacional em tabagismo entre adolescentes escolares. Rev Bras Epidemiol 2011 Mar;14(1):63-72. [doi: 10.1590/S1415-790X2011000100006]
13. Brinker T, Buslaff F, Haney C, Gaim B, Haney A, Schmidt S, Netzwerk Aufklärung gegen Tabak, et al. [The global medical network Education Against Tobacco-voluntary tobacco prevention made in Germany]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Nov;61(11):1453-1461. [doi: 10.1007/s00103-018-2826-8] [Medline: 30284623]
14. Brinker TJ, Seeger W. Photoaging mobile apps: a novel opportunity for smoking cessation? J Med Internet Res 2015 Jul 27;17(7):e186 [FREE Full text] [doi: 10.2196/jmir.4792] [Medline: 26215210]
15. Education Against Tobacco. URL: http://educationtobacco.org
16. Brinker T, Seeger W, Buslaff F. Photoaging mobile apps in school-based tobacco prevention: the mirroring approach. J Med Internet Res 2016 Dec 28;18(6):e183 [FREE Full text] [doi: 10.2196/jmir.6016] [Medline: 27352819]
17. Brinker T, Stamm-Balderjahn S, Seeger W, Klingelhöfer D, Groneberg DA. Education Against Tobacco (EAT): a quasi-experimental prospective evaluation of a multinational medical-student-delivered smoking prevention programme for secondary schools in Germany. BMJ Open 2015 Sep 18;5(9):e008093 [FREE Full text] [doi: 10.1136/bmjopen-2015-008093] [Medline: 26384722]
18. Brinker T, Stamm-Balderjahn S, Seeger W, Groneberg DA. Education Against Tobacco (EAT): a quasi-experimental prospective evaluation of a programme for preventing smoking in secondary schools delivered by medical students: a study protocol. BMJ Open 2014 Jul 24;4(7):e004909 [FREE Full text] [doi: 10.1136/bmjopen-2014-004909] [Medline: 25059969]
19. Brinker T, Owczarek A, Seeger W, Groneberg D, Brieske C, Jansen P, et al. A medical student-delivered smoking prevention program, Education Against Tobacco, for secondary schools in Germany: randomized controlled trial. J Med Internet Res 2017 Dec 06;19(6):e199 [FREE Full text] [doi: 10.2196/jmir.7906] [Medline: 28588007]
20. Bernstein S, Yu S, Post L, Dziura J, Rigotti NA. Undertreatment of tobacco use relative to other chronic conditions. Am J Public Health 2013 Aug;103(8):e59-e65. [doi: 10.2105/AJPH.2012.301112] [Medline: 23763395]
21. Raupach T, Falk J, Vangeli E, Schiekirka S, Rustler C, Grassi M, et al. Structured smoking cessation training for health professionals on cardiology wards: a prospective study. Eur J Prev Cardiol 2014 Jul;21(7):915-922. [doi: 10.1177/2047487312462803] [Medline: 23008136]
22. Anders S, Strobel L, Krampe H, Raupach T. [Do final-year medical students know enough about the treatment of alcohol use disorders and smoking?]. Dtsch Med Wochenschr 2013 Jan;138(1-2):23-27. [doi: 10.1055/s-0032-1327367] [Medline: 23250691]
23. Flett K, Grogan S, Clark-Carter D, Gough B, Conner M. Male smokers' experiences of an appearance-focused facial-ageing intervention. J Health Psychol 2017 Dec;22(4):422-433. [doi: 10.1177/1359105315603477] [Medline: 26338489]
24. Jiwa M, Burford O, Parsons R. Preliminary findings of how visual demonstrations of changes to physical appearance may enhance weight loss attempts. Eur J Public Health 2015 Apr;25(2):283-285. [doi: 10.1093/eurpub/cku249] [Medline: 25644139]
25. Mahler H, Kulik J, Butler H, Gerrard M, Gibbons FX. Social norms information enhances the efficacy of an appearance-based sun protection intervention. Soc Sci Med 2008 Jul;67(2):321-329 [FREE Full text] [doi: 10.1016/j.socscimed.2008.03.037] [Medline: 18448221]
26. Olson A, Gaffney C, Starr P, Dietrich AJ. The impact of an appearance-based educational intervention on adolescent intention to use sunscreen. Health Educ Res 2008 Oct;23(5):763-769 [FREE Full text] [doi: 10.1093/her/cym005] [Medline: 18039727]
27. Owen A, Grogan S, Clark-Carter D. Effects of an appearance-focussed versus a health-focussed intervention on men's attitudes towards UV exposure. Int J Men Health 2016;15(1):12-26. [doi: 10.3149/jmh.1501.34]
28. Stapleton J, Turrisi R, Hillhouse J, Robinson J, Abar B. A comparison of the efficacy of an appearance-focused skin cancer intervention within indoor tanner subgroups identified by latent profile analysis. J Behav Med 2010 Jun;33(3):181-190. [doi: 10.1007/s10865-009-9246-z] [Medline: 20058183]
29. Tuong W, Armstrong AW. Effect of appearance-based education compared with health-based education on sunscreen use and knowledge: a randomized controlled trial. J Am Acad Dermatol 2014 Apr;70(4):665-669. [doi: 10.1016/j.jaad.2013.12.007] [Medline: 24508292]
30. Brinker T, Holzapfel J, Baudson T, Sies K, Jakob L, Baumert H, et al. Photoaging smartphone app promoting poster campaign to reduce smoking prevalence in secondary schools: the Smokerface Randomized Trial: design and baseline

characteristics. BMJ Open 2016 Dec 07;6(11):e014288 [FREE Full text] [doi: 10.1136/bmjopen-2016-014288] [Medline: 27821601]
31. Brinker T, Schadendorf D, Klode J, Cosgarea I, Rösch A, Jansen P, et al. Photoaging mobile apps as a novel opportunity for melanoma prevention: pilot study. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e101 [FREE Full text] [doi: 10.2196/mhealth.8231] [Medline: 28747297]
32. Burford O, Smith M, Jiwa M, Carter O. Photoageing Intervention (PAINT): a proposal for a randomised controlled trial in Australian primary care. AMJ 2008 Nov 31;1(7):8-12. [doi: 10.4066/AMJ.2009.108]
33. Eastabrook S, Chang P, Taylor MF. Melanoma risk: adolescent females' perspectives on skin protection pre/post-viewing a ultraviolet photoaged photograph of their own facial sun damage. Glob Health Promot 2018 Mar;25(1):23-32. [doi: 10.1177/1757975916639871] [Medline: 27333916]
34. Faria B, Brieske C, Cosgarea I, Omlor A, Fries F, de Faria CO, et al. A smoking prevention photoageing intervention for secondary schools in Brazil delivered by medical students: protocol for a randomised trial. BMJ Open 2017 Dec 10;7(12):e018589 [FREE Full text] [doi: 10.1136/bmjopen-2017-018589] [Medline: 29229659]
35. Mahler H, Kulik J, Gerrard M, Gibbons FX. Effects of photoaging information and UV photo on sun protection intentions and behaviours: a cross-regional comparison. Psychol Health 2013;28(9):1009-1031. [doi: 10.1080/08870446.2013.777966] [Medline: 23537173]
36. Lo Presti L, Chang P, Taylor MF. Young Australian adults' reactions to viewing personalised UV photoaged photographs. Australas Med J 2014;7(11):454-461 [FREE Full text] [doi: 10.4066/AMJ.2014.2253] [Medline: 25550717]
37. Brinker T, Heckl M, Gatzka M, Heppt M, Resende Rodrigues H, Schneider S, et al. A skin cancer prevention facial-aging mobile app for secondary schools in Brazil: appearance-focused interventional study. JMIR Mhealth Uhealth 2018 Mar 09;6(3):e60 [FREE Full text] [doi: 10.2196/mhealth.9794] [Medline: 29523502]
38. Brinker T, Brieske C, Esser S, Klode J, Mons U, Batra A, et al. A face-aging app for smoking cessation in a waiting room setting: pilot study in an HIV outpatient clinic. J Med Internet Res 2018 Aug 15;20(8):e10976 [FREE Full text] [doi: 10.2196/10976] [Medline: 30111525]
39. Brinker T, Klode J, Esser S, Schadendorf D. Facial-aging app availability in waiting rooms as a potential opportunity for skin cancer prevention. JAMA Dermatol 2018 Sep 01;154(9):1085-1086. [doi: 10.1001/jamadermatol.2018.1907] [Medline: 30046805]
40. Baudson T, Weber K, Freund PA. More than only skin deep: appearance self-concept predicts most of secondary school students' self-esteem. Front Psychol 2016;7:1568 [FREE Full text] [doi: 10.3389/fpsyg.2016.01568] [Medline: 27803681]
41. Burford O, Jiwa M, Carter O, Parsons R, Hendrie D. Internet-based photoaging within Australian pharmacies to promote smoking cessation: randomized controlled trial. J Med Internet Res 2013 Mar 26;15(3):e64 [FREE Full text] [doi: 10.2196/jmir.2337] [Medline: 23531984]
42. Xavier L, Bernardes-Souza B, Lisboa O, Seeger W, Groneberg D, Tran T, et al. A medical student-delivered smoking prevention program, Education Against Tobacco, for secondary schools in Brazil: study protocol for a randomized trial. JMIR Res Protoc 2017 Jan 30;6(1):e16 [FREE Full text] [doi: 10.2196/resprot.7134] [Medline: 28137703]
43. BMBF Gesundheitsforschung. Deutsche Arbeitsgemeinschaft fuer Epidemiologie. 2004. Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP) URL: http://wwwgesundheitsforschung-bmbfde/_media/Empfehlungen_GEPpdf [WebCite Cache ID 75kAdw5Qg]
44. Kuntz B, Waldhauer J, Moor I, Rathmann K, Richter M, Orth B, et al. [Trends in educational inequalities in smoking among adolescents in Germany: evidence from four population-based studies]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Jan;61(1):7-19. [doi: 10.1007/s00103-017-2636-4] [Medline: 28980028]
45. Hoebel J, Kuntz B, Kroll L, Finger J, Zeiher J, Lange C, et al. Trends in absolute and relative educational inequalities in adult smoking since the early 2000s: the case of Germany. Nicotine Tob Res 2018 Feb 07;20(3):295-302. [doi: 10.1093/ntr/ntx087] [Medline: 28431153]
46. Kuntz B, Zeiher J, Hoebel J, Lampert T. Soziale Ungleichheit, Rauchen und Gesundheit. Suchttherapie 2016 Aug 16;17(03):115-123 [FREE Full text] [doi: 10.1055/s-0042-109372] [Medline: 10]
47. Kuntz B, Lampert T. Smoking and passive smoke exposure among adolescents in Germany. Dtsch Arztebl Int 2016 Jan 22;113(3):23-30 [FREE Full text] [doi: 10.3238/arztebl.2016.0023] [Medline: 26857509]
48. Kuntz B, Hoebel J, Lampert T. [Educational differences in smoking and smoking cessation among young adults. Results from the "German Health Update" (GEDA) Study 2009 and 2010]. Gesundheitswesen 2014 Oct;76(10):647-654. [doi: 10.1055/s-0033-1364017] [Medline: 24639388]
49. Corrêa PCRP, Pereira RMOS, Temporao JG, Cavalcante TM, Lisboa OC, Azevedo LG, et al. EAT-Brazil Award for Tobacco Control: a brief description of the profile of the proposals submitted in the 1st edition of the award. Revista da Associação Médica Brasileira 2019 (forthcoming).

Abbreviations

EAT: Education Against Tobacco
NNT: number needed to treat

OR: odds ratio

Edited by G Eysenbach; submitted 18.11.18; peer-reviewed by L Post, F Fries; comments to author 21.12.18; revised version received 28.12.18; accepted 05.01.19; published 21.02.19.

Please cite as:
Lisboa OC, Bernardes-Souza B, Xavier LEDF, Almeida MR, Corrêa PCRP, Brinker TJ
A Smoking Prevention Program Delivered by Medical Students to Secondary Schools in Brazil Called "Education Against Tobacco": Randomized Controlled Trial
J Med Internet Res 2019;21(2):e12854
URL: http://www.jmir.org/2019/2/e12854/
doi:10.2196/12854
PMID:30789347

Original Paper

Mobile Technology for Community Health in Ghana: Is Maternal Messaging and Provider Use of Technology Cost-Effective in Improving Maternal and Child Health Outcomes at Scale?

Michelle Willcox[1], MSPH; Anitha Moorthy[2], PhD, MHS; Diwakar Mohan[1], PhD, MBBS; Karen Romano[2], MPH; David Hutchful[2]; Garrett Mehl[3], PhD; Alain Labrique[1], PhD, MSc, MHS; Amnesty LeFevre[1,4], PhD, MHS

[1]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Grameen Foundation Ghana, Accra, Ghana

[3]World Health Organization, Geneva, Switzerland

[4]Health Intelligence Initiative, Division of Epidemiology and Biostatistics, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa

Corresponding Author:
Amnesty LeFevre, PhD, MHS
Health Intelligence Initiative
Division of Epidemiology and Biostatistics, School of Public Health and Family Medicine
University of Cape Town
Falmouth Rd, Observatory
Cape Town, 7925
South Africa
Phone: 27 0767828704
Fax: 27 0767828704
Email: aelefevre@gmail.com

Abstract

Background: Mobile technologies are emerging as tools to enhance health service delivery systems and empower clients to improve maternal, newborn, and child health. Limited evidence exists on the value for money of mobile health (mHealth) programs in low- and middle-income countries.

Objective: This study aims to forecast the incremental cost-effectiveness of the Mobile Technology for Community Health (MOTECH) initiative at scale across 170 districts in Ghana.

Methods: MOTECH's "Client Data Application" allows frontline health workers to digitize service delivery information and track the care of patients. MOTECH's other main component, the "Mobile Midwife," sends automated educational voice messages to mobile phones of pregnant and postpartum women. We measured program costs and consequences of scaling up MOTECH over a 10-year analytic time horizon. Economic costs were estimated from informant interviews and financial records. Health effects were modeled using the Lives Saved Tool with data from an independent evaluation of changes in key services coverage observed in Gomoa West District. Incremental cost-effectiveness ratios were presented overall and for each year of implementation. Uncertainty analyses assessed the robustness of results to changes in key parameters.

Results: MOTECH was scaled in clusters over a 3-year period to reach 78.7% (170/216) of Ghana's districts. Sustaining the program would cost US $17,618 on average annually per district. Over 10 years, MOTECH could potentially save an estimated 59,906 lives at a total cost of US $32 million. The incremental cost per disability-adjusted life year averted ranged from US $174 in the first year to US $6.54 in the tenth year of implementation and US $20.94 (95% CI US $20.34-$21.55) over 10 years. Uncertainty analyses suggested that the incremental cost-effectiveness ratio was most sensitive to changes in health effects, followed by personnel time. Probabilistic sensitivity analyses suggested that MOTECH had a 100% probability of being cost-effective above a willingness-to-pay threshold of US $50.

Conclusions: This is the first study to estimate the value for money of the supply- and demand-side of an mHealth initiative. The adoption of MOTECH to improve MNCH service delivery and uptake represents good value for money in Ghana and should be considered for expansion. Integration with other mHealth solutions, including e-Tracker, may provide opportunities to continue or combine beneficial components of MOTECH to achieve a greater impact on health.

(J Med Internet Res 2019;21(2):e11268) doi:10.2196/11268

KEYWORDS

child health; frontline health workers; Ghana; health information systems; maternal health; mHealth; mobile phone; newborn health

Introduction

Globally, an estimated 273,500 maternal and 6.3 million child deaths occur each year; nearly half are in sub-Saharan Africa [1,2]. Ghana, home to 24.5 million people, is estimated to have 3,100 maternal and 56,000 child deaths annually [3]. While maternal and child mortality rates have both declined by 30% between 2000 and 2012, the rate of reduction annually over the last decade has not been enough for Ghana to achieve the Millennium Development Goals' targets for maternal and child health [3].

Public health programs that aim to capitalize on mobile health (mHealth) technologies are emerging as potentially novel and effective strategies to address critical health system constraints [4-7]. Over the last decade, more than 600 documented mHealth strategies and comprehensive programs have been introduced globally, and innumerable small-scale pilots are in development or underway [8]. Ghana has seen a similar proliferation of mHealth programs alongside increasing rates of phone usage in the population: at least 22 different projects have been piloted since 2004 [5,9]. While evidence exists on the effectiveness of mHealth programs in improving health outcomes, fewer studies have investigated the cost-effectiveness of these new technology-supported solutions [7,10-14]. While there is broad consensus on the potential for mHealth solutions to improve the management of health systems, provider performance, and empower clients [15], several technical complexities, including network coverage, technology use, and data flow, when coupled with limited information on the effectiveness and cost-effectiveness, have impeded efforts to sustain and scale programs in areas with limited resources.

The Mobile Technology for Community Health (MOTECH) program aims to empower pregnant women and new mothers to improve knowledge and utilization of services through mobile phone-based voice messages on maternal, newborn, and child health (MNCH) and system-managed appointment reminders and alerts for important care visits. Simultaneously, MOTECH supports frontline health workers in collecting patient-level clinical information, improving data-reporting processes, and improving the timeliness and continuity of MNCH service delivery for clients in their catchment area [16,17]. The program was first initiated in 2009 in community-level health facilities in the Upper East Region of Ghana and later expanded to the Central, Greater Accra, and Volta Regions. To date, implementation has occurred in all public health centers and rural frontline health facilities in 7 districts: 24 health centers and 83 Community-based Health Planning and Services (CHPS) facilities.

To complement efforts to evaluate the effectiveness of MOTECH [18,19], we sought to determine the incremental cost-effectiveness of MOTECH as compared to existing health services in 170 districts over a 10-year analytic time horizon (2015-2024). Study findings aim to support evidence-based decision making on the value for money of MOTECH. This is the first study of its kind on the cost-effectiveness of an mHealth solution in Ghana and one of few globally on mHealth strategies for enhancing community-based MNCH services through the support of frontline health workers [20].

Methods

Study Setting

Ghana's estimated population of 24.5 million is spread across 216 districts. Health services in the public sector are delivered through community CHPS facilities, health centers, and district and regional hospitals with varying rates of public sector service utilization. MNCH indicators have improved with the help of policies such as the Free Maternal Care Policy and National Health Insurance; however, maternal mortality (380 per 100,000 live births) and infant mortality (49 per 1000 live births) rates did not decrease sufficiently to reach 2015 targets, and socioeconomic and geographic differences continue to reflect inequalities in health service utilization and outcomes [1,3,21]. Utilization of at least 1 antenatal care (ANC) session is nearly universal, and 84% of women attend 4 or more ANC visits, nearly 30% of deliveries are not attended by a skilled birth attendant, and almost 20% do not receive postnatal care [22].

In an effort to catalyze improvements in the quality and continuity of care during pregnancy and postpartum, the MOTECH program launched in August 2010 in Kassena-Nankana West District in the Upper East Region of Ghana and in 2011 in Awutu Senya District in the Central Region. In 2012, with added funding from United States Agency for International Development and the Bill and Melinda Gates Foundation (BMGF), MOTECH was expanded into new districts: Dangme East in Greater Accra, South Tongu in the Volta region, and Gomoa West in Central region. Gomoa West was selected for an independent evaluation led by Healthcare Innovation Technology Lab in partnership with the University of Ghana School of Public Health. Because of the availability of data on health effects, data from Gomoa West were used in this analysis to forecast the costs and consequences of program expansion to 170 districts across Ghana.

Gomoa West is a semirural coastal district of the Central Region with a population of 135,139 [23]. Health services in Gomoa West are delivered through 17 CHPS facilities (1 per 7949 population), 4 health centers (1 per 33,785 population), and 1 district hospital [16]. Population and health systems characteristics for the remaining 169 districts were drawn from the 2010 Population and Housing Census, including district-division and population statistics, data on the number of health facilities, and human resources [23,24].

Ethical Review

This study received ethical approval from the Ethical Review Committee of Ghana Health Service (GHS) as well as the Johns Hopkins School of Public Health Institutional Review Board. Requirement of informed consent was waived.

Program Description

MOTECH, a partnership between Grameen Foundation and GHS, aims to improve service delivery and access to MNCH services through 2 components: Mobile Midwife and Client Data Application (CDA). Mobile Midwife is a demand-side intervention, which aims to improve client knowledge and awareness of key health information during pregnancy and postpartum, with the goal of stimulating best practices and encouraging timely and appropriate service utilization. Under Mobile Midwife, registered pregnant women and mothers of infants are provided with a MOTECH identification number and receive weekly local-language interactive voice response messages on pregnancy and infant care and appointment reminders for routine clinical visits on their mobile phones (Multimedia Appendix 1). To complement demand generation through Mobile Midwife, the CDA of MOTECH allows frontline health workers affiliated with CHPS facilities and health centers to use mobile phones to digitize clinical care information to better track and deliver care to women of reproductive age, including pregnant women and children <5 years of age. CDA consists of simplified digital registers consolidating information previously collected in over a dozen paper-based registers in health centers and CHPS facilities. Data elements recorded onto mobile phones are used to generate monthly reports as well as care reminders and alerts sent to both the clients and nurses along with a weekly list of care defaulters in need of follow-up in their catchment area [16]. The input of digitized health data activates the automated system to cross-reference national care protocols, identify appropriate reminders or alerts to clients and nurses, and send voice messages. The MOTECH program also facilitates data reporting by summarizing facility statistics for a report to the district, although the reporting system is not directly interoperable with the government health information system.

As part of this analysis, we forecasted the costs and consequences of scaling up MOTECH activities to 170 districts across Ghana, mirroring the implementation activities conducted in Gomoa West. Program activities were provided in 3 phases: development (6 months), start-up (6 months), and ongoing support to implementation (Table 1). Development encompasses all central activities, such as national stakeholder meetings, which are necessary to prepare for the initiation of the program. Start-up includes all central- and local-level activities required to initiate the program, including district profiling, content localization, orientations, and training. Implementation refers to ongoing activities required to sustain the program and includes salary support for personnel at all levels for ongoing monitoring and evaluation as well as recurrent telecommunications and data usage costs, which assume no fluctuation for agreements with mobile network operators from what was established in the MOTECH program in Gomoa West. Development and start-up phase activities occur at the program's inception and are considered capital investments that are annualized across the life of the program. Some activities were modified for national scale from the original Grameen Foundation MOTECH project in Gomoa West: instead of mass registration campaigns for enrollment, the national program would assume point-of-care enrollment at health facilities. Recurring costs such as personnel, utilities, and vehicle maintenance are included in all phases.

To initiate MOTECH activities at scale, a 3-year national rollout plan was designed to begin in 3 regions, with 5 districts in each region conducting start-up activities simultaneously supported by key central and regional support staff for training, program administration, and technical support (Table 2). Additional districts could be brought on board within those regions once the first districts completed all development activities and had moved on to start-up or implementation phases of the program. GHS successfully upgraded from District Health Information System (DHIS) version 1 to the current DHIS 2 in only 6 months, and although this is promising for the capacity and experience with digital health solutions, the roll out of MOTECH activities at scale would require significant oversight capacity for additional health worker training, equipment replacement, and data monitoring [25].

Table 1. Program costs by phase and activity for the gradual rollout of Mobile Technology for Community Health (MOTECH) to 170 districts from 2015-2024.

Program activities	National program activity description and input	Total cost (US $)
Development[a]		
Program design	National leadership meetings among central program leadership, regional, and district health management teams	36,800
Telecommunications	Voice message program national negotiations and ongoing partnership with telecommunications companies, infrastructure establishment to manage call system	23,480
Technology	Platform support and server hosting	16,907
Personnel	Central, regional, and district staff time allocated to develop the program for each district	596,955
Subtotal development		698,647
Subtotal development costs per district per year		4,110
Start-up[b]		
District profiling	Profile data on health system facilities and telecommunication infrastructure compiled by district health management teams	150,919
Content localization	Voice health messaging content standardized, translated, recorded, and tested for 9 recognized national languages	30,690
Equipment	Phone purchases for facilities to use for MOTECH mHealth application platform, additional staff equipment	2,471,799
Customer support	Customer service referral system for technical or programmatic issues	249,746
Training	Orientation for leadership, nurse training on simplified registers and data entry application, and subdistrict program orientation	2,406,578
Community mobilization	District launch events, durbars, or other marketing	627,638
Partnership building	Regional steering committee meeting for district program planning and technical steering committees in the district meet monthly for 4 months to direct start-up	10,449
Vehicle maintenance	Cost to maintain and use existing vehicles without new capital cost purchase	172,082
Office maintenance	Central office space	86,277
Telecommunications	Airtime for voice messages, nurse data upload cost	77,166
Technology	Platform support, server hosting, and system scaling by region for increased call capacity	26,318
Personnel & benefits	Central, regional, and district staff time allocated to initiating the program for each district	1,193,847
Subtotal start-up		7,503,508
Subtotal start-up costs per district per year		44,138
Implementation[c]		
Technical groups	District health management teams will include program tasks in current workflow, no added cost	17,810
Monitoring and evaluation	Routine data entry application use monitoring and data validation	1,700,567
Continued training	Refresher training and training of new hires	2,140,605
Equipment & materials	Phone replacement and simplified registers annual set per facility	4,928,480
Vehicle Maintenance	No vehicle cost in national program	1,078,131
Field office maintenance	District office space	209,521
Office maintenance	Central office space	257,036
Telecommunications	Airtime for voice messages and nurse data upload cost	2,343,275
Technology maintenance	Data platform and information technology technical assistance	241,582
Personnel & benefits	Central, regional, and district staff time allocated to maintaining the program	11,218,454
Subtotal implementation		24,135,461

Program activities	National program activity description and input	Total cost (US $)
Subtotal implementation costs per district per year		17,918
Total annualized cost per district per year		66,166

[a]6 months in duration per district.

[b]6 months in duration per district.

[c]9 years of implementation annual costs for 45 districts in which MOTECH program activities were initiated in 2015, 8 years for 67 districts in which MOTECH program activities were initiated in 2016, 7 years for 58 final districts in which MOTECH program activities were initiated in 2017.

Table 2. Rollout plan for the Mobile Technology for Community Health program: Reaching 170 districts over 3 years.

Ghana regions	Number of districts						
	Year 1 January-June	Year 1 July-December	Year 2 January-June	Year 2 July-December	Year 3 January-June	Year 3 July-December	Total
Western	5	5	7	—[a]	—	—	17
Central	5	5	7	—	—	—	17
Greater Accra	5	5	—	—	—	—	10
Volta	—	5	5	8	—	—	18
Eastern	—	5	5	5	6	—	21
Ashanti	—	5	5	5	5	7	27
Brong-Ahafo	—	—	5	5	5	7	22
Northern	—	—	5	5	5	5	20
Upper East	—	—	—	—	5	4	9
Upper West	—	—	—	—	5	4	9
New districts starting up	15	30	39	28	31	27	—
All districts with program costs	15	45	84	112	143	170	170

[a]No start-up costs were being incurred in this district.

Costs

Economic costs were estimated from a program perspective for a 10-year analytic time horizon (2015-2024). Program costs were defined as the costs required to develop, start up, and support ongoing implementation. Costs associated with MOTECH activities from 2012-2014 in the district of Gomoa West were used to generate estimates of the resources required to scale up and sustain implementation into 170 districts. Drawing from Grameen Foundation project financial records in Accra, which detailed capital and recurrent expenditures over time for MOTECH in Gomoa West, as well as informant interviews with program staff, we used an ingredients approach to forecast estimates of costs by activity and level of the health system (central, regional, or district) over time for each district. District population estimates were drawn from the 2010 census in Ghana and the model adapted to incorporate differences across districts in numbers of CHPS facilities and health centers [23]. Once collected, costs were converted into US $, using month-appropriate market exchange rates, and then adjusted into 2014 US $ using the 2014 Consumer Price Index for Ghana [26]. Capital costs were annualized over the lifetime of the project or life span of the item as appropriate and discounted at 3% [27]. Development and start-up phase costs were viewed as one-time activities and similarly annualized over the lifetime of the project. Variable costs were anticipated to change with the number of facilities or nurses (mobile phones purchased, refresher training, etc) and adjusted based on population rates in Gomoa West.

Health Effects

Health effects were modeled based on the average intervention effects observed as part of the independent evaluation of MOTECH carried out in Gomoa West and 1 comparison district. Secondary analyses of exit interview data conducted by our team sought to generate an estimate of the average treatment effect of the MOTECH program based on propensity score values. Findings from this analysis suggested a significant increase in the coverage of skilled birth attendants (11%), facility delivery (10%), and measles immunization (6%) following program implementation. No significant changes were observed in pregnant women reporting 4 or more ANC visits, exclusive breastfeeding, the use of modern methods postpartum family planning, and other indicators. Indicators with significant increases in health effects were inputted in to the Lives Saved Tool (LiST) in Spectrum version 4.7. Increases in coverage were assumed to increase linearly but not anticipated to increase at the same rate over time. When intervention coverage surpassed 75%, the rate of coverage increase was adjusted by half; above 90% overall intervention coverage, the

rate of increase is again halved; and for any intervention, cannot increase beyond 99%. These breakpoints help to adjust expectations of increases in coverage with the model's time horizon. The disability-adjusted life years (DALYs) averted were calculated with a 3% discount rate, no age weighting, and life expectancy of 65.5 years as reported in the 2010 Census [23,28]. In the absence of data on disabilities, DALYs are based on years of life lost only.

Analyses

Analyses of incremental costs and effects were conducted in Microsoft Excel. To test for uncertainty, one-way and probabilistic sensitivity analyses were conducted. The latter was carried out in Microsoft Excel using Monte Carlo simulation with 1000 iterations per analysis. The resulting mean point estimate was obtained by dividing mean costs by mean effects and presented along with the 95% CI. A cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) were used to calculate the probability that the intervention would be cost-effective for each of a number of standard thresholds of cost-effectiveness. Cost-effectiveness was ultimately determined according to thresholds set forth by the Commission for Macroeconomics and Health and World Health Organization in 2002, which stipulate that an intervention is "highly cost-effective" and "cost-effective" at 1 and 3 times the value of per capita gross domestic product per DALY averted, respectively. To facilitate comparison with alternative resource uses, we additionally compared findings against those available in the literature, including the Disease Control Priorities Project, third edition that highlights low-cost, high-priority interventions for key regions for the world, including sub-Saharan Africa.

Results

Costs

Table 1 presents program costs by phase and activity for the 10-year time horizon taking into account the rollout plan shown (Table 2). The mean annualized cost per district over the 10 years of the MOTECH-modeled program is US $237,745: US $4,110 per district for central development and program planning activities over 6 months, US $44,138 to initiate the program over 6 months, and US $17,918 to sustain the program annually. Development phase costs, which include program design, telecommunications negotiation, and system development from existing technology, are all incurred at a central level. Personnel costs accounted for 85% (US $596,955/US $698,647) of total costs in that phase. Start-up costs include some central- and some district-level activities, such as district profiling, content localization for health messaging, customer support, training for all health workers and district health teams, community mobilization through durbars or community meetings, and more. The main cost drivers in this phase are mobile phone device purchase and district-wide training, which together account for over 65% (US $4,878,377/US $7,503,508) of total start-up phase costs. During the years of program implementation, costs at the district level are focused on technical working group meetings, monitoring and evaluation, refresher and new hire training, and technical support and maintenance of the system. Costs in this phase are also driven primarily by personnel (47%, US $11,218,454/US $24,135,461), equipment (20%, US $4,928,480/ US$24,135,461) including phone replacement and annual registry replacement costs as well as airtime and data costs (10%, US $2,343,275/US$24,135,461) associated with health messaging and appointment reminders.

Health Effects

Multimedia Appendix 2 present data on health effects adjusted for district rollout and duration of program implementation. Based on incremental changes in facility delivery, skilled birth attendance, and measles coverage, a total of 6298 maternal deaths, 33,797 child deaths, and 19,811 stillbirths would be averted over the 10-year analytic time horizon. This would correlate to a yearly average decrease in the <5 mortality rate by 4 per 1000 births. While the program is not anticipated to save lives in the first year of implementation, a total of 483 lives would be saved in Year 2; 59.6% (288/483) of these are child deaths averted, 31.5% (152/483) stillbirths averted, and 8.9% (43/483) deaths averted. By Year 10, when implementation would have occurred in all 170 districts for at least 7 years, a total of 11,938 lives would be saved. Cumulative lives saved over the 10-year analytic time horizon are anticipated to correspond to 1,550,028 DALYs averted.

Incremental Cost-Effectiveness

Multimedia Appendix 3 presents data on the incremental cost-effectiveness of MOTECH based on deterministic values draw from 10 year estimates of costs and consequences implementation across 170 districts. The deterministic costs per DALY averted decreased over time with increasing scale and coverage from US $173.62 in Year 1 to US $6.54 in Year 10. Probabilistic estimates of the total cost per DALY averted and total cost per death averted after 10 years are US $20.92 (95% CI US $20.34-$21.55) and US $586.72 (95% CI US $569.36-$604.09), respectively.

Uncertainty Analyses

Figure 1 presents findings from a 1-way sensitivity analysis. Estimates of the cost per DALY averted are most sensitive to changes in health effects parameters, including estimates of lives saved among children, stillbirths, and mothers as well as personnel costs during implementation. Probabilistic sensitivity analyses sought to explore the effects of varying multiple parameters simultaneously (Multimedia Appendix 3). The cost-effectiveness plane presented in Figure 2 depicts the incremental costs and consequences of 1000 simulations, all of which fall in the northeast quadrant. Figure 3 presents incremental CEACs of MOTECH implementation from 2015-2024 versus status quo in 170 districts of Ghana for varying denominators of DALYs based on cumulative lives saved versus DALYs based on child, maternal, or stillbirths averted. With a gross national income per capita for 2015 of US $1,480 as the threshold, program activities have a 100% probability of being cost-effective. At lower willingness-to-pay thresholds, MOTECH has a 90% probability of being cost-effective above a willingness-to-pay threshold of US $400 when maternal lives saved are used to generate DALYs and a 100% probability of being cost-effective above a

willingness-to-pay threshold of US $50 when cumulative lives saved are used to generate DALYs.

Figure 1. Tornado diagram showing one-way sensitivity analyses on key drivers of cost-effectiveness.DALYs: disability-adjusted life years.

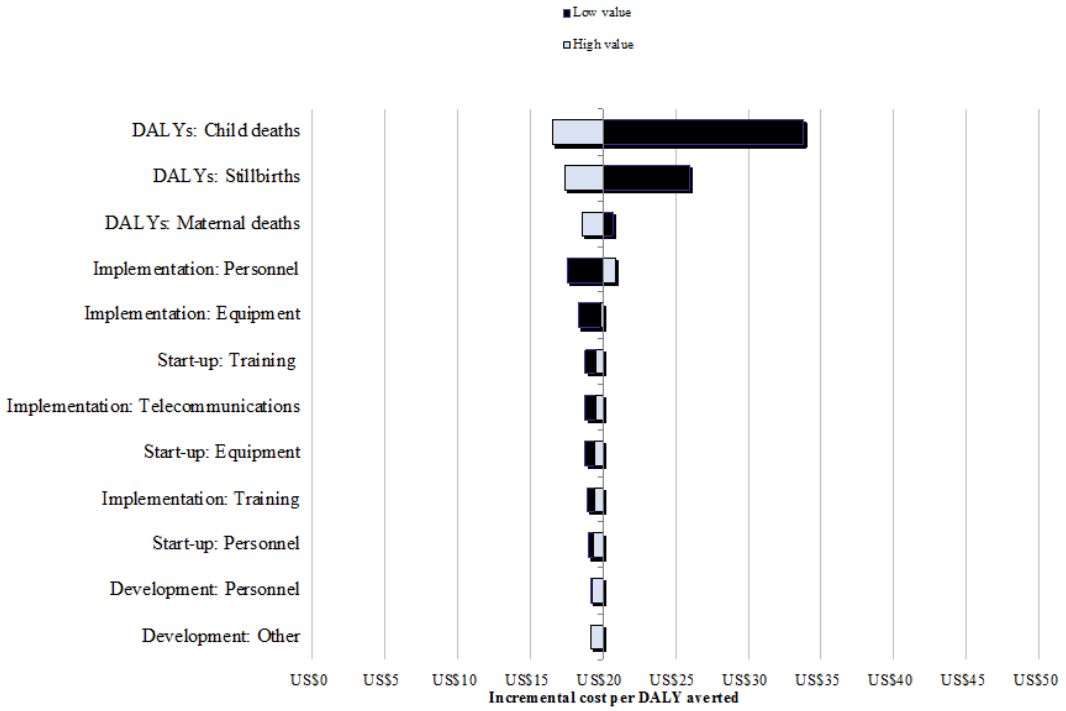
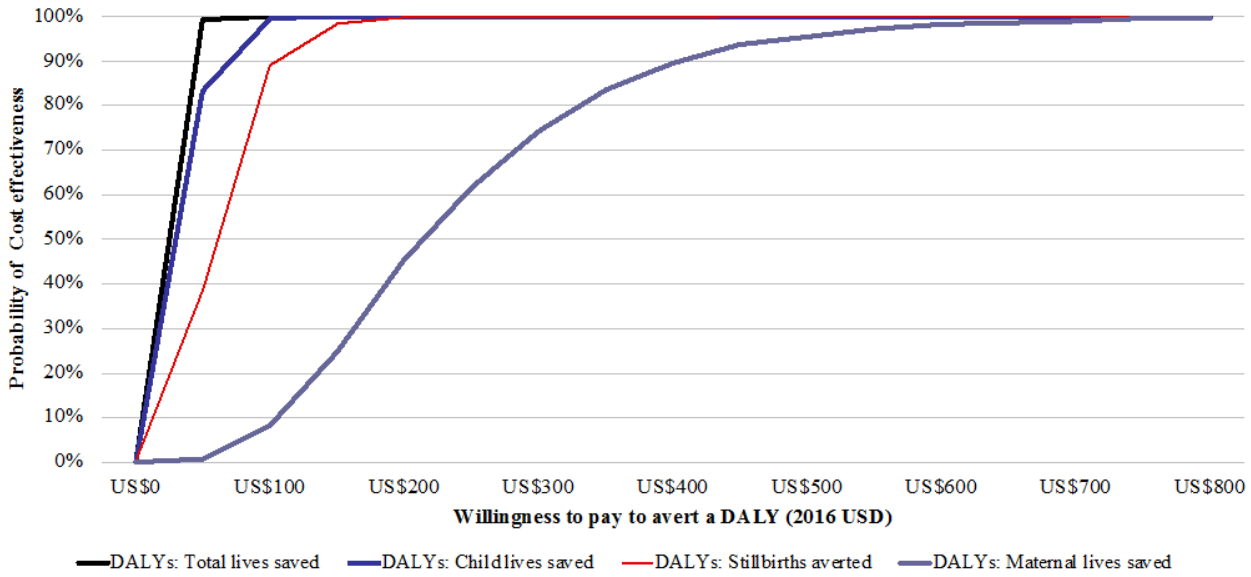


Figure 2. Cost-effectiveness plane of incremental program costs and disability-adjusted life years (DALYs) averted from Mobile Technology for Community Health implementation in 170 districts from 2015-2024.

Figure 3. Incremental cost-effectiveness acceptability curves of Mobile Technology for Community Health implementation from 2015-2024 versus status quo in 170 districts of Ghana. DALYs: disability-adjusted life years.



Discussion

Principal Findings

This study assessed the costs and consequences of implementing an mHealth program–dually providing an interactive voice message system to pregnant and new mothers as well as a health record and reminder system for use by frontline health workers–as compared to the pre-existing health services over 10 years in nearly 80% of districts across Ghana. If implemented across Ghana through 170 districts over 10 years, MOTECH could potentially save an estimated 59,906 lives at a total 10-year cost of US $32 million. Estimates of the total annualized cost per district of US $66,166 were driven by start-up costs including provider training and equipment. The cost to sustain the program once implemented would be US $17,618 on average annually per district. In Ghana's 2013 government budget, US $264.5 million was earmarked for the health sector of which 72% was specifically directed for primary care [29]. Assuming this directive funding is continued in the modeled MOTECH timeframe with equal allocation to 170 districts, then developing and starting the MOTECH program would equate to 4% of the district budget, and sustaining the program would be less than 1.6%.

Efforts to determine the value for money of MOTECH suggest that program activities are highly cost-effective at a total cost per DALY averted of US $20.94 (95% CI US $20.34-$21.55) over 10 years–an estimate that falls well below the gross national income for Ghana of US $1,480 and compares favorably with alternative resource uses [30,31]. As implementation at scale continues and integration with existing or planned digital health solutions occurs it is probable that further declines in costs could be attained, corresponding to even greater improvements in cost-effectiveness. Alternatively, it is also possible that 1 of MOTECH's components (CDA or Mobile Midwife) could be scaled in isolation, which may also affect value for money estimates. Our analyses did not differentiate the costs and consequences of the subcomponents of MOTECH. However, messaging has been shown to be an effective strategy for client engagement across a number of geographies [7,32,33], and evidence is emerging on its probable cost-effectiveness. In South Africa, a small-scale deployment of MAMA maternal messaging forecasted to scale across 1 province over 5 years reported a US $200 cost per DALY averted [34]—a finding nearly 10 times higher than that estimated for Ghana when maternal messaging is provided in conjunction with the supply-side CDA.

Estimates of the value for money of initiatives similar to CDA that digitize patient health records and send alerts and reminders are not yet available in the literature. While such programs may have some effects on improved efficiencies in reporting, in practice, these can be difficult to translate into measurable health outcomes. Mobile e-Tracker, a new feature of the national population health monitoring DHIS2 system in Ghana, is an example of 1 initiative underway that may ultimately take the place of CDA at scale should GHS not opt to expand and sustain MOTECH as this analysis considers. E-Tracker aims to digitize individual client data for aggregation at district, regional, and national levels, a step that may ease reporting and improve the national health information system. To date, implementation has occurred in 200 hospitals around the country, which process, on average, 1000 new admissions per day using ICD-10 codes [25]. Whether e-Tracker will be expanded to the community level and taken up by frontline health workers remains to be seen. In the event that it is, our analyses highlight the potential synergistic effect that concurrent implementation with a demand-generating initiative like Mobile Midwife may have. Early findings on the cost-effectiveness from other programs, such as the Smart Registries Project (OpenSRP), led by the WHO, or the mCARE program in Bangladesh, which have drawn inspiration from the MOTECH strategy of combined client and health worker engagement, echo these results,

suggesting value for money particularly over time and with sufficient scale [35].

Comparison With Alternative Resource Uses

Data on the cost-effectiveness of interventions to improve MNCH coverage are emerging [20]. Incremental cost-effectiveness ratios from these studies range from US $79 per DALY averted, for a 10-year forecasted program that supports delivery care in Zambia and US $214 per DALY averted, for a home-care service package that includes community mobilization and health system strengthening aspects in Bangladesh, to US $302 per DALY averted, for a women's group quality improvement collaborative to improve birth outcomes in rural Nepal [20,36-38]. When compared against studies that report findings on the cost per life saved, MOTECH similarly falls well below program alternatives with a cost per death averted of US $586.72 (95% CI US $569.36-$604.09). The quality improvement collaborative in Nepal reported US $8,670 per life saved, the Zambian program supporting delivery care reported US $1,988 per life saved, and a maternal health mobile unit providing outreach services in The Gambia reported a cost per life saved range of US $1,380-$6,414 [20,37-39]. Collectively, these studies suggest that MOTECH, even in its earliest years of implementation, where the cost per DALY averted is US $173.62, is likely to provide good value for money.

In the context of other digital health programs, only a dozen peer-reviewed articles comprise the body of evidence on the value for money of mHealth solutions, including cost-effectiveness analyses (5 studies [40-44]), cost-utility analyses (2 studies [45,46]), and cost-benefit analyses (4 studies [47-50]). Ours is the first effort to forecast costs and consequences of scaling up an mHealth solution and the only study to date of an integrated behavior change communication and service delivery mHealth solution. The effectiveness results may be further improved if the program implementation targeted districts or regions with a worse health status first, in order to have longer implementation and achieve greater effects in those areas.

Limitations

This study encompasses several important limitations. Efforts to determine the economic costs of MOTECH implementation were conducted from a program perspective, drawing upon data obtained from a single district. While these costs have been modified to best model potential program expenditure, they do not represent planned government resource use or cost structures. In particular, personnel costs are based on Grameen Foundation personnel in Accra and in Gomoa West rather than GHS personnel or salary costs, as this level of granularity in government expenditure was not available at the time of the analysis. To account for these limitations, we achieved a granular level of cost and activity data that are related to the number of nurses, CHPS facilities, and health centers, which we used to vary mHealth program costs across different districts to best estimate national-scale costs in Ghana. Further sensitivity analyses were used to characterize the effects of changes to these parameters on emerging cost-effectiveness estimates. To forecast the costs of scaling up program activities to 170 districts, we drew from the most recent census data available, the 2010 Ghana Population and Housing Report, and generated a rollout plan to estimate associated costs for these 170 districts [23]. In 2012, 46 new districts were created; however, in the absence of updated population or health facility data, the model could not adequately estimate the program scaling costs for these presently defined districts [24]. Finally, it is noteworthy that our analyses did not include costs incurred by pregnant and postpartum women to seek care or to the health system to collect data using CDA or accommodated forecasted increases in utilization stemming from increased demand. Future evaluations of digital health programs should plan economic evaluations from the outset that can prospectively track costs from a societal perspective.

To model health effects, we drew from data collected as part of a randomized controlled trial exploring the effects of MOTECH versus status quo activities on key MNCH outcomes, including skilled birth attendance, facility delivery, and immunizations, conducted in Gomoa West led by the Healthcare Innovation Technology Lab in partnership with the University of Ghana School of Public Health. In brief, methodological limitations were seen in the study design, methods for data collection (sampling, survey tool), and the analytic approach. Among these factors, limitations in the survey tool and wording of indicators as well as the decision to recruit the sample population from community outreach activities organized by GHS community health nurses versus sampling at a household population level has rendered efforts to estimate the effects of MOTECH activities on population-level changes in service utilization challenging. In an effort to work with the data available to us, we generated propensity scores based on values of education, employment, marital status, age, and income at endline. We then estimated average treatment (MOTECH intervention) effects based on nearest neighbor matching based on propensity score values. The end changes in coverage were then inputted into the LiST and used to generate lives saved and, ultimately, a years of life lost only-based DALY. For the health effects included, we further attenuated coverage gains anticipated over time across the time horizon in order to account for declines in the magnitude of gains anticipated as coverage reached thresholds of 75% and above for key interventions. To account for uncertainty in our primary effects data as well as forecasted changes over time, we conducted 1-way and probabilistic sensitivity analyses. We further presented CEACs that included cumulative lives saved as well as those based on maternal, child, and stillbirths averted.

Conclusion

Study findings suggest that supply- and demand-side digital health initiatives like MOTECH may offer value for money when implemented over time and at scale. Further research is needed to validate underpinning assumptions and generate more robust evidence on the costs and consequence of expanding and sustaining the implementation MOTECH in Ghana.

Acknowledgments

We are grateful to Eddie Adimazoya for assistance with defining the overall program; Antoinette Akanlise and Shuaib Tetteh for perspectives and insights on support team activities in Gomoa West; Billey Adamu for his information regarding the monitoring and evaluation for project data; Lisa Wei for her inputs on financial costs and effort on valuing airtime and data; and the rest of the MOTECH team who participated in interviews and provided important support for this study. MOTECH in Ghana was a partnership of the GHS, Grameen Foundation, and the Columbia University Mailman School of Public Health. The initiative was cofinanced by the BMGF; Saving Lives at Birth: A Grand Challenge for Development, a multinational partnership of the United States Agency for International Development, the Norwegian Ministry of Foreign Affairs, Grand Challenges Canada, the UK Department for International Development, the Korea International Cooperation Agency, and the BMGF; and the United Nations Foundation. Support for Johns Hopkins University student and faculty time on this analysis was provided by the World Health Organization through the Innovation Working Group grant mechanism. Publication costs were supported by the BMGF.

Authors' Contributions

ALF and AM conceived the idea for this analysis and, along with KR, oversaw all elements of data collection, analysis, and write up. DM analyzed data on health effects, which were inputted and analyzed in LiST by MW. ALF conducted the probabilistic sensitivity analyses. MW collected the data, conducted the analysis, and wrote the first draft of the manuscript with ALF and inputs from all authors. AL provided inputs to the study design and analyses. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Subsample of mobile midwife postpartum messages occurring in the first 1 month.

[PDF File (Adobe PDF File), 35KB - jmir_v21i2e11268_app1.pdf]

Multimedia Appendix 2

Health effects adjusted for district rollout and duration of MOTECH program implementation.

[PDF File (Adobe PDF File), 32KB - jmir_v21i2e11268_app2.pdf]

Multimedia Appendix 3

Parameters for probabilistic sensitivity analyses.

[PDF File (Adobe PDF File), 71KB - jmir_v21i2e11268_app3.pdf]

References

1. Lozano R, Wang H, Foreman K, Rajaratnam J, Naghavi M, Marcus J, et al. Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: an updated systematic analysis. The Lancet 2011 Sep;378(9797):1139-1165. [doi: 10.1016/S0140-6736(11)61337-8]
2. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn J, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. The Lancet 2015 Jan;385(9966):430-440. [doi: 10.1016/S0140-6736(14)61698-6]
3. UNICEF, World Health Organization. Healthy Newborn Network. Baltimore, MD USA: Countdown 2015; 2014 Jun 30. Fulfilling the Health Agenda for Women and Children: 2014 Conference Report Draft URL: https://www.healthynewbornnetwork.org/resource/fulfilling-the-health-agenda-for-women-and-children-the-2014-report/ [accessed 2018-11-28] [WebCite Cache ID 74GGQ378i]
4. Agarwal S, Labrique A. Newborn health on the line: the potential mHealth applications. JAMA 2014 Jul 16;312(3):229-230. [doi: 10.1001/jama.2014.6371] [Medline: 24953141]
5. World Health Organization. 2011. mHealth: New horizons for health through mobile technologies: second global survey on eHealth URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf
6. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
7. 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]
8. Cargo M, June G. South Africa mHealth Landscape. London: GSMA; 2013.

9. Vest JR, Issel LM, Lee S. Experience of using information systems in public health practice: findings from a qualitative study. Online J Public Health Inform 2014;5(3):227 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4847] [Medline: 24678380]
10. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
11. Källander K, Tibenderana J, Akpogheneta O, Strachan D, Hill Z, ten Asbroek AHA, et al. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review. J Med Internet Res 2013 Jan 25;15(1):e17 [FREE Full text] [doi: 10.2196/jmir.2130] [Medline: 23353680]
12. Whitten P, Mair F, Haycox A, May C, Williams T, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. BMJ 2002 Jun 15;324(7351):1434-1437 [FREE Full text] [Medline: 12065269]
13. Guerriero C, Cairns J, Roberts I, Rodgers A, Whittaker R, Free C. The cost-effectiveness of smoking cessation support delivered by mobile phone text messaging: Txt2stop. Eur J Health Econ 2013 Oct;14(5):789-797 [FREE Full text] [doi: 10.1007/s10198-012-0424-5] [Medline: 22961230]
14. Zurovac D, Larson BA, Sudoi RK, Snow RW. Costs and cost-effectiveness of a mobile phone text-message reminder programmes to improve health workers' adherence to malaria guidelines in Kenya. PLoS One 2012;7(12):e52045 [FREE Full text] [doi: 10.1371/journal.pone.0052045] [Medline: 23272206]
15. Mehl G, Vasudevan L, Gonsalves L, Berg M, Seimon T, Temmerman M. Harnessing mHealth in low-resource settings to overcome health system constraints and achieve universal access to health. In: Marsch LA, Lord SE, Dallery J, editors. Behavioral Health Care and Technology: Using Science-Based Innovations to Transform Practice. New York, NY: Oxford University Press; 2014.
16. Wood T, Osborn J, Hutchful D, Larsen JM, Adimazoya E, Kwarah W. Mobile Technology for Community Health in Ghana. Accra, Ghana 2012:1-50.
17. Nyonator F, Awoonor-Williams J, Phillips J, Jones T, Miller RA. The Ghana community-based health planning and services initiative for scaling up service delivery innovation. Health Policy Plan 2005 Jan;20(1):25-34. [doi: 10.1093/heapol/czi003] [Medline: 15689427]
18. Rothstein J, Jennings L, Moorthy A, Yang F, Gee L, Romano K, et al. Qualitative Assessment of the Feasibility, Usability, and Acceptability of a Mobile Client Data App for Community-Based Maternal, Neonatal, and Child Care in Rural Ghana. Int J Telemed Appl 2016;2016:2515420 [FREE Full text] [doi: 10.1155/2016/2515420] [Medline: 28070186]
19. LeFevre A, Mohan D, Hutchful D, Jennings L, Mehl G, Labrique A, et al. Mobile Technology for Community Health in Ghana: what happens when technical functionality threatens the effectiveness of digital health programs? BMC Med Inform Decis Mak 2017 Dec 14;17(1):27 [FREE Full text] [doi: 10.1186/s12911-017-0421-9] [Medline: 28292288]
20. Mangham-Jefferies L, Pitt C, Cousens S, Mills A, Schellenberg J. Cost-effectiveness of strategies to improve the utilization and provision of maternal and newborn health care in low-income and lower-middle-income countries: a systematic review. BMC Pregnancy Childbirth 2014 Jul 22;14:243 [FREE Full text] [doi: 10.1186/1471-2393-14-243] [Medline: 25052536]
21. Opoku E. Utilization of Maternal Care services in Ghana by Region after the Implementation of the Free Maternal Care Policy. Fort Worth: University of North Texas Health Science Center at Fort Worth; 2009.
22. Victora CG, Requejo JH, Barros AJD, Berman P, Bhutta Z, Boerma T, et al. Countdown to 2015: a decade of tracking progress for maternal, newborn, and child survival. Lancet 2016 May 14;387(10032):2049-2059. [doi: 10.1016/S0140-6736(15)00519-X] [Medline: 26477328]
23. Ghana Statistical Service. 2012 May 31. 2010 Population and Housing Census: Final Results URL: http://www.statsghana.gov.gh/docfiles/2010phc/2010_POPULATION_AND_HOUSING_CENSUS_FINAL_RESULTS.pdf [accessed 2018-11-28] [WebCite Cache ID 74GIV6zPu]
24. Government of Ghana. New districts inaugurated, Metropolitan, Municipal and Districts now 216. 2012. URL: http://www.ghanadistricts.gov.gh/Admin/Account/Login?ReturnUrl=%2Fnews%2F%3Fread%3D46228 [accessed 2018-11-28] [WebCite Cache ID 74GIjSq1m]
25. Health Information Systems Programme.: University of Oslo; 2018. District Health Information Software 2 In Action URL: https://www.dhis2.org/inaction [accessed 2018-11-28] [WebCite Cache ID 74GIw2NRL]
26. International Monetary Fund. 2017. World Economic and Financial Surveys: World Economic Outlook Database URL: https://www.imf.org/external/pubs/ft/weo/2014/02/weodata/index.aspx [accessed 2018-11-28] [WebCite Cache ID 74GJ2jiVm]
27. Tan-Torres ET, Baltussen R, Adam T, Hutubessy R, Acharya A, Evans D. Making Choices in Health: WHO Guide to Cost-effectiveness Analysis. Tan-Torres Edejer T, Baltussen R, ADAM T, Hutubessy R, Acharya A, Evans DB. editors. Geneva: World Health Organization; 2003.
28. Fox-Rushby JA, Hanson K. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. Health Policy Plan 2001 Sep;16(3):326-331. [Medline: 11527874]
29. Ministry of Finance. My Ghana Budget. Accra, Ghana: Ministry of Finance, Government of Ghana; 2013. The 2013 Budget Statement and Economic Policy URL: http://myghanabudget.org/assets/Uploads/2013-Budget-Statement.pdf

30. Laxminarayan R, Chow J, Shahid-Salles SA. Chapter 2 Intervention Cost-Effectiveness: Overview of Main Messages. In: Jamison DT, Breman JG, Measham AR, editors. Disease Control Priorities in Developing Countries. 2nd edition. New York: Oxford University Press; 2006.
31. World Health Organization. Choosing interventions that are cost effective (WHO-CHOICE). Geneva: World Health Organization; 2005. URL: http://www.who.int/choice/costs/CER_thresholds/en/ [accessed 2018-11-28] [WebCite Cache ID 74GJloz1p]
32. Thirumurthy H, Lester R. M-health for health behaviour change in resource-limited settings: applications to HIV care and beyond. Bull World Health Organ 2012 May 01;90(5):390-392 [FREE Full text] [doi: 10.2471/BLT.11.099317] [Medline: 22589574]
33. Tomori C, Risher K, Limaye R, Van Lith LM, Gibbs S, Smelyanskaya M, et al. A role for health communication in the continuum of HIV care, treatment, and prevention. J Acquir Immune Defic Syndr 2014 Aug 15;66 Suppl 3:S306-S310 [FREE Full text] [doi: 10.1097/QAI.0000000000000239] [Medline: 25007201]
34. LeFevre AE, Cabrera-Escobar MA, Mohan D, Eriksen J, Rogers D, Neo Parsons A, et al. Forecasting the Value for Money of Mobile Maternal Health Information Messages on Improving Utilization of Maternal and Child Health Services in Gauteng, South Africa: Cost-Effectiveness Analysis. JMIR Mhealth Uhealth 2018 Jul 27;6(7):e153 [FREE Full text] [doi: 10.2196/mhealth.8185] [Medline: 30054263]
35. Jo Y, Labrique A, Lefevre A, Mehl G, Pfaff T, Walker N, et al. Using the lives saved tool (LiST) to model mHealth impact on neonatal survival in resource-limited settings. PLoS One 2014;9(7):e102224 [FREE Full text] [doi: 10.1371/journal.pone.0102224] [Medline: 25014008]
36. LeFevre A, Shillcutt S, Waters H, Haider S, El Arifeen S, Mannan I, Projahnmo Study Group. Economic evaluation of neonatal care packages in a cluster-randomized controlled trial in Sylhet, Bangladesh. Bull World Health Organ 2013 Oct 01;91(10):736-745 [FREE Full text] [doi: 10.2471/BLT.12.117127] [Medline: 24115797]
37. Sabin L, Knapp A, MacLeod W, Phiri-Mazala G, Kasimba J, Hamer D, et al. Costs and cost-effectiveness of training traditional birth attendants to reduce neonatal mortality in the Lufwanyama Neonatal Survival study (LUNESP). PLoS One 2012;7(4):e35560 [FREE Full text] [doi: 10.1371/journal.pone.0035560] [Medline: 22545117]
38. Borghi J, Thapa B, Osrin D, Jan S, Morrison J, Tamang S, et al. Economic assessment of a women's group intervention to improve birth outcomes in rural Nepal. The Lancet 2005 Nov;366(9500):1882-1884. [doi: 10.1016/S0140-6736(05)67758-6]
39. Fox-Rushby JA, Foord F. Costs, effects and cost-effectiveness analysis of a mobile maternal health care service in West Kiang, The Gambia. Health Policy 1996 Feb;35(2):123-143. [Medline: 10156650]
40. Dodani, Songer T, Ahmed Z, Laporte RE. Building research capacity in developing countries: cost-effectiveness of an epidemiology course taught by traditional and video-teleconferencing methods in Pakistan. Telemed J E Health 2012 Oct;18(8):621-628. [doi: 10.1089/tmj.2011.0262] [Medline: 23061643]
41. White J, Dow W, Rungruanghiranya S. Commitment contracts and team incentives: a randomized controlled trial for smoking cessation in Thailand. Am J Prev Med 2013 Nov;45(5):533-542 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.020] [Medline: 24139765]
42. Chen Z, Fang L, Chen L, Dai H. Comparison of an SMS text messaging and phone reminder to improve attendance at a health promotion center: a randomized controlled trial. J Zhejiang Univ Sci B 2008 Jan;9(1):34-38 [FREE Full text] [doi: 10.1631/jzus.B071464] [Medline: 18196610]
43. Anchala R, Kaptoge S, Pant H, Di Angelantonio E, Franco OH, Prabhakaran D. Evaluation of effectiveness and cost-effectiveness of a clinical decision support system in managing hypertension in resource constrained primary health care settings: results from a cluster randomized trial. J Am Heart Assoc 2015 Jan 05;4(1):e001213 [FREE Full text] [doi: 10.1161/JAHA.114.001213] [Medline: 25559011]
44. Rashid RMA, Ramli S, John J, Dahlui M. Cost effective analysis of recall methods for cervical cancer screening in Selangor--results from a prospective randomized controlled trial. Asian Pac J Cancer Prev 2014;15(13):5143-5147 [FREE Full text] [Medline: 25040965]
45. Hunchangsith P, Barendregt JJ, Vos T, Bertram M. Cost-effectiveness of various tuberculosis control strategies in Thailand. Value Health 2012;15(1 Suppl):S50-S55 [FREE Full text] [doi: 10.1016/j.jval.2011.11.006] [Medline: 22265067]
46. Rachapelle S, Legood R, Alavi Y, Lindfield R, Sharma T, Kuper H, et al. The cost-utility of telemedicine to screen for diabetic retinopathy in India. Ophthalmology 2013 Mar;120(3):566-573. [doi: 10.1016/j.ophtha.2012.09.002] [Medline: 23211635]
47. Latifi K, Lecaj I, Bekteshi F, Dasho E, Doarn C, Merrell R, et al. Cost-benefit analysis on the use of telemedicine program of Kosova for continuous medical education: a sustainable and efficient model to rebuild medical systems in developing countries. Telemed J E Health 2011 Dec;17(10):757-762. [doi: 10.1089/tmj.2011.0102] [Medline: 22011053]
48. Broomhead S, Mars M. Retrospective return on investment analysis of an electronic treatment adherence device piloted in the Northern Cape Province. Telemed J E Health 2012;18(1):24-31. [doi: 10.1089/tmj.2011.0143] [Medline: 22150713]
49. Andrade M, Maia A, Cardoso CS, Alkmim MB, Ribeiro ALP. Cost-benefit of the telecardiology service in the state of Minas Gerais: Minas Telecardio Project. Arq Bras Cardiol 2011 Oct;97(4):307-316 [FREE Full text] [Medline: 21808852]
50. Bernal-Sánchez G, de la Torre-Rodriguez J, Robles-Bonilla C, Campos-Romero A, Otero-Meza C. Análisis costo-beneficio. Telemedicina Anáhuac 2007;75:35 [FREE Full text]

Abbreviations

ANC: antenatal care
BMGF: Bill and Melinda Gates Foundation
CDA: Client Data Application
CEAC: cost-effectiveness acceptability curve
CHPS: Community-Based Health Planning and Services
DALY: disability-adjusted life year
DHIS: District Health Information System
GHS: Ghana Health Service
LiST: Lives Saved Tool
mHealth: mobile health
MNCH: maternal, newborn, and child health
MOTECH: Mobile Technology for Community Health

Edited by G Eysenbach; submitted 10.06.18; peer-reviewed by T Jones, T Cruvinel; comments to author 01.08.18; revised version received 12.09.18; accepted 12.09.18; published 13.02.19.

Please cite as:
Willcox M, Moorthy A, Mohan D, Romano K, Hutchful D, Mehl G, Labrique A, LeFevre A
Mobile Technology for Community Health in Ghana: Is Maternal Messaging and Provider Use of Technology Cost-Effective in Improving Maternal and Child Health Outcomes at Scale?
J Med Internet Res 2019;21(2):e11268
URL: http://www.jmir.org/2019/2/e11268/
doi:10.2196/11268
PMID:30758296

Original Paper

Digital Marketing to Promote Healthy Weight Gain Among Pregnant Women in Alberta: An Implementation Study

Jocelyn E Graham[1*], MSc; Jana L Moore[1*], MSc; Rhonda C Bell[2*], PhD; Terri Miller[1*], RD, MSA

[1]Reproductive Health, Population, Public and Indigenous Health, Alberta Health Services, Calgary, AB, Canada

[2]Department of Agricultural, Food & Nutritional Science, Faculty of Agricultural, Life & Environmental Sciences, University of Alberta, Edmonton, AB, Canada

[*]all authors contributed equally

Corresponding Author:
Jocelyn E Graham, MSc
Reproductive Health
Population, Public and Indigenous Health
Alberta Health Services
10101 Southport Road SW
Calgary, AB, T2W 3N2
Canada
Phone: 1 403 943 1878
Fax: 1 403 943 2211
Email: Jocelyn.Graham@ahs.ca

Abstract

Background: As the use of digital media for health promotion has become increasingly common, descriptive studies exploring current and innovative marketing strategies can enhance the understanding of effective strategies and best practices.

Objective: This study aims to describe the implementation of a provincial digital media campaign using complementary advertising platforms to promote healthy pregnancy weight gain messages and direct a Web audience to a credible website.

Methods: The digital media campaign occurred in 3 phases, each for 8 weeks, and consisted of search engine marketing using Google AdWords and social media advertising through Facebook. All advertising materials directed users to evidence-based pregnancy-related weight gain content on the Healthy Parents, Healthy Children website.

Results: Google Ads received a total of 43,449 impressions, 2522 clicks, and an average click-through rate (CTR) of 5.80%. Of people who clicked on a Google ad, 78.9% (1989/2522) completed an action on the website. Across all Facebook advertisements, there were 772,263 impressions, 14,482 clicks, and an average CTR of 1.88%. The highest-performing advertisement was an image of a group of diverse pregnant women with the headline “Pregnancy weight is not the same for every woman.”

Conclusions: This study supports the use of digital marketing as an important avenue for delivering health messages and directing Web users to credible sources of information. The opportunity to reach large, yet targeted audiences, along with the ability to monitor and evaluate metrics to optimize activities throughout a campaign is a powerful advantage over traditional marketing tactics. Health organizations can use the results and insights of this study to help inform the design and implementation of similar Web-based activities.

(J Med Internet Res 2019;21(2):e11534) doi:10.2196/11534

KEYWORDS

internet; maternal health; mobile phone; pregnant women; search engine; social media

Introduction

Studies have consistently found that the majority of Canadian women do not meet the Health Canada [1] gestational weight gain (GWG) recommendations [2,3]. Exceeding these recommendations is associated with adverse outcomes, such as pre-eclampsia, cesarean delivery, and infants born large for gestational age [4], with an increased risk of childhood obesity [5]. Weight gain below the recommendations can increase risks of preterm birth and infants who are small for their gestational age [6]. Reported barriers to meeting GWG recommendations include misperceptions, lack of advice, and poor knowledge with regard to personal body mass index, GWG

recommendations, and the impact of GWG on health outcomes [7-10].

While the prenatal visit offers an opportunity to provide pregnancy information and support, many women feel that discussions with health professionals are insufficient and use Web-based resources to compensate [11,12]. Women commonly access the internet during pregnancy [13,14] and perceive Web-based health information as reliable and useful, while appreciating features such as anonymity, simplicity, and unrestricted access at any time [15,16]. In a study by Larsson [17], 84% of women reported using the internet to obtain pregnancy-related information, particularly in early pregnancy, and 97% of participants in a study by Lagan et al [12] reported using search engines to find pregnancy webpages. In addition, women use social media, including Facebook, during pregnancy to share their experiences and seek advice, with the majority accessing these platforms, at least, once a day [11,15].

The internet, therefore, provides an opportunity to increase women's awareness and knowledge about healthy GWG and connect women with credible resources, as other Web-based information may not be consistent and evidence-based. For health organizations, advantages of using digital platforms for these purposes include low cost, accessible information, interactivity with users, and the ability to deliver tailored messages with a wide reach to specific audiences [18,19]. Because of these advantages, it has become increasingly common for health organizations to create and maintain social media accounts. Organizations can share health information through these accounts in 2 main ways—unpaid (organic) and paid (advertising). As social media platforms, like Facebook, continue to decrease the amount and frequency of content from organizations that users see organically [20], paid advertising is becoming increasingly important. Descriptive studies about the feasibility and steps for implementing successful Web-based campaigns are needed to support a better understanding of innovative strategies that engage the public in health topics like GWG [18,19,21].

Informed by recent Alberta-based maternal health research [3,10,22], a provincial marketing campaign using digital platforms was developed to promote awareness of healthy GWG among women. This paper aims to describe the development and implementation of the digital campaign, which used search engine marketing using Google AdWords and social media advertising through Facebook. This project was a collaboration between Alberta Health Services (AHS), which is the health authority for the Canadian province of Alberta, and the University of Alberta's ENRICH Research Program.

Methods

Campaign Objectives

The GWG campaign was embedded within *Healthy Parents, Healthy Children* (HPHC), a program created by AHS. This program provides evidence-based information and best practice advice to expectant parents and parents of children up to 6 years of age through printed books, a website [23], and social media activity. Pregnancy information on the website covers topics such as nutrition, exercise, distribution of weight, risks of gaining too much or too little weight, and weight gain guidelines. A Web-based calculator is also available for pregnant women to find their prepregnancy body mass index and recommended weight gain range as well as to track their current weight in relation to their recommendation [24]. All digital advertising campaign materials directed users to this GWG content on the HPHC website.

The campaign objectives were to increase awareness about the importance of healthy GWG among women in Alberta and direct women to the HPHC website. Google AdWords and Facebook Ads were chosen to achieve these objectives as Google is the most popular search engine [25] and Facebook is the most used social media network, with the greatest proportion of users being women aged 18-29 years [26]. In addition, the combination of these platforms reaches 2 audiences—those searching for information (Google AdWords) and those who may not be actively searching (Facebook). A health marketing framework that "involves creating, communicating, and delivering health information and interventions using customer-centered and science-based strategies to protect and promote the health of diverse populations" [27] was used to guide the campaign.

Google AdWords

Google AdWords is a paid service that displays ad text and a website link above, beside, or below a list of Google search results when a user performs a search using keywords that match those selected by the advertiser. In the list of search results, these paid ads are labeled "Ad" to differentiate them from organic search results.

Facebook Ads

On Facebook, ads are paid messages from organizations or businesses displayed to members of a predefined audience. Paid ads look similar to unpaid Facebook content and can appear throughout the social network. Advertisers can set budget, articulate advertising objectives, and define a target audience, all of which determine where, how often, and to whom the ad is shown. For this campaign, the project team developed the ads using Facebook Ads Manager, Facebook's advertising management platform.

Campaign Design

The campaign occurred in 3 distinct phases of 8 weeks each between January 2 and December 3, 2017, with a total budget of Can $7034.44. This design was used to decrease the risk of message fatigue and overexposing the audience, which would have been more likely had this been carried out as a single 24-week campaign. To ensure equitable distribution across the entire province, the budget was allocated to 3 markets—65% (Can $4572.39/Can $7034.44) to 2 major cities in Alberta, 15% (Can $1055.17/Can $7034.44) to smaller cities throughout the province, and 20% (Can $1406.88/Can $7034.44) to rural Alberta. A local vendor was contracted to establish the advertising platforms and manage the first phase of the campaign. The remaining phases of the campaign were maintained by internal staff who attended local and Web-based digital marketing training sessions to gain knowledge and

expertise. At the end of each campaign phase, staff reviewed the campaign results and adapted the strategy for the next phase.

Overall, 3 Google AdWords ads were developed to be used across all phases (Figure 1) and set to appear to users in Alberta searching for pregnancy weight information. One of the ads was duplicated and served only to mobile devices. When a user clicked on the ad, they were brought to the GWG calculator on the HPHC website. Google AdWords uses a bidding system to determine which ads are displayed and the cost to the advertiser. For this digital campaign, the budget was Can $10 per day, and we used automatic cost per click (CPC) bidding, meaning we only paid when the ad was clicked.

For Facebook ads, the audience was defined as female, aged 18-44 years, with interests in pregnancy. Ads were selected to show on all devices (mobile, tablet, and desktop) and in Facebook newsfeeds, instant papers, and in the right column of Facebook's desktop platform. A flexible daily budget of Can $26 was set to have ads run continuously. Like Google AdWords, Facebook uses a bidding system to determine which ads to display. For the digital campaign, Facebook performed automatic bidding based on campaign objectives and audience to achieve the highest number of clicks on the ad at the lowest cost. Allowing for adjustments in daily spending through the automatic bidding process helped to take advantage of opportunities in the ad auction marketplace to optimize results.

Unlike Google AdWords, Facebook ads offer the opportunity for longer messaging, images, and branding. Two new HPHC-branded Facebook ads were developed by staff prior to each phase of the campaign using key messages about healthy pregnancy weight, stock photos, and a link to a relevant HPHC webpage (Figure 2). Ads adhered to Facebook guidelines; however, adjustments to the wording of the ad copy were required when Facebook did not approve ads because of an unexpected application of their prohibited content guidelines that restricted the use of the phrase "weight gain" and the option to ask personal questions (eg, How much weight gain is healthy for you during pregnancy?).

During the first phase, ads incorporated color blocks using HPHC branding colors and text onto the image. For the second phase, women in the images were not visibly pregnant and the color blocks and text were removed. Ads for the third phase incorporated design and content elements of the top performing ads from the previous phases; these elements included an image of a woman talking with a health care provider, an image of diverse women to promote the individuality of weight gain, displaying the AHS logo, and the use of HPHC branding colors.

Facebook users can react and leave comments on paid ads; hence, throughout the campaign, staff monitored the ads for comments and provided timely responses, redirecting users to additional HPHC webpages as applicable. While privacy and profanity issues were not a challenge during the campaign, staff had prepared to hide comments that were inappropriate or revealed personal medical information, and connect with users through private messaging as needed.

Figure 1. Sample design of Google AdWords advertisements promoted during the campaign.

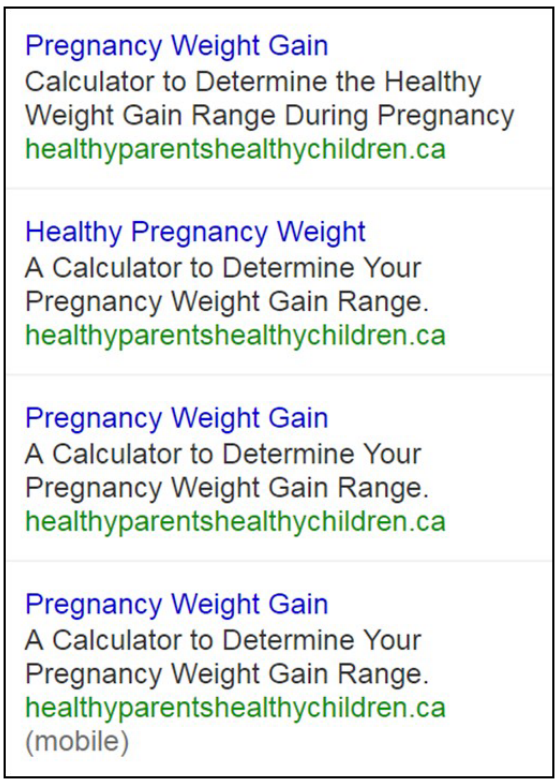


Figure 2. A screenshot of the Facebook advertisements for each campaign phase. (Source: photos purchased from iStock.com and Shutterstock.com).

Data Analysis

Both Google and Facebook provide extensive data for paid ads delivered using their platforms. Ad performance data were collected from the HPHC Google AdWords account and Facebook Ads Manager. Definitions of key metrics can differ across digital platforms. For the results and discussion that follows, we will use the following terms and definitions:

- Impressions: the number of times an ad was displayed to a user
- Engagement: interaction with an ad, including Facebook user reactions, sharing, and comments, as well as ad clicks
- CPC: price an advertiser pays for each click on the ad
- Click-through rate (CTR): the number of clicks on the ad per the number of impressions
- Conversions: actions that a user completes after clicking on an ad (completing the Web-based weight gain calculator, clicking to download the results of the Web-based weight calculator, clicking on a social media share button). These data could be collected because of the existing codes that had been added to the HPHC website by the website developer.

It was predicted that Facebook ad engagement would be low because weight can be a personal and sensitive topic, and many women in early pregnancy delay public announcement. Any reaction, comment, or share of the ads could be seen by the women's Web-based social network, which may deter users from engaging in these ways. For this campaign, the CTR was a better indicator of overall success and user interest in the ad message.

Results

Google AdWords

Ads displayed on Google received a total of 43,449 impressions throughout the campaign. The average position of a campaign ad in the search results list was 1.3, meaning the ad was often displayed first or second when an associated keyword was searched. A total of Can $1913.72 was spent on ads with an average CPC of Can $0.76, which is the average amount charged for a click on an ad.

The Google ads were clicked 2522 times, with an average CTR of 5.80%. The most popular search term that led to an ad click was "calculate weight gain during pregnancy" with 137 clicks (Table 1). The search term "pregnancy weight gain" had the highest CTR at 24.07% (Table 2). There were 1989 conversions as a direct result of the ads, which represents a 78.9% (1989/2522) conversion rate by users who clicked on an ad.

Facebook Ads

During the campaign, Facebook ads received 772,263 impressions, and the average number of times an ad was displayed to the same individual ranged from 2.53 to 3.28. A total of Can $5067 was spent on Facebook ads with an average CPC of Can $0.35. A greater amount was spent on Facebook ads because of the platform's required minimum daily budget for an ad set. Table 3 shows results for Facebook ads during each phase. The overall average CTR was 1.88% (range 0.32%-2.56%), and the most clicks on any one ad were 4741. In comparison, the highest number of clicks for organic, unpaid Facebook posts on the HPHC account during 2017 that included an image and a link to a healthy pregnancy weight gain content page was 95. The highest-performing ad, as determined by the CTR, occurred in the third phase and included the AHS logo, an image of diverse women, and the headline "Pregnancy weight is not the same for every woman" (Figure 2).

Facebook ads were clicked a total of 14,482 times, for an overall CTR of 1.88%. Ads received 43 comments, 28 shares, and 247 reactions. While a rigorous content analysis was not performed, the majority of ad comments were neutral, with many users sharing about their personal weight gain experience. Several comments provided an opportunity to enhance knowledge. For example, when a woman commented on weight gain for carrying twins, our reply encouraged them to talk with her health care provider, and we shared a link to the HPHC webpage containing information about healthy weight gain for a twin pregnancy.

Table 1. Top 5 performing keywords on Google AdWords by ad clicks.

Keyword	Clicks, n (%)	Click-through rate, %	Cost per click, Can $	Cost, Can $
Calculate weight gain during pregnancy	137 (5.43)	10.65	0.65	88.91
Pregnancy weight gain chart calculator	109 (4.32)	17.90	0.54	58.39
Weight gain during pregnancy chart	103 (4.08)	12.06	0.58	59.96
Pregnant weight gain chart calculator	90 (3.57)	13.72	0.55	49.72
Calculate weight gain during pregnancy	89 (3.53)	8.52	0.64	56.65

Table 2. Top 5 performing keywords on Google AdWords by click-through rate.

Keyword	Clicks, n (%)	Click-through rate, %	Cost per click, Can $	Cost, Can $
Pregnancy weight gain	13 (0.52)	24.07	0.39	5.10
Healthy pregnancy weight gain chart	4 (0.16)	20.00	0.65	2.61
Weight gain during pregnancy month by month	6 (0.24)	20.00	0.75	4.50
Pregnancy weight gain chart calculator	109 (4.32)	17.90	0.54	58.39
Healthy pregnancy weight gain chart	6 (0.24)	17.65	0.50	03.01

Table 3. Facebook Ads results during each phase.

Phase and ad headline	Impressions, n (%)	Clicks, n	Click-through rate, %	Cost per click, Can $	Comments, n	Shares, n	Reactions, n
Phase 1							
Find out where the weight goes and why it's important	27,747 (3.59)	485	1.75	0.41	0	0	34
Gaining a healthy amount of weight is important for you and your baby	197,921 (25.63)	4273	2.16	0.37	22	8	89
Phase 2							
Pregnancy weight is not the same for every woman	295,762 (38.30)	4741	1.60	0.34	11	6	47
Pregnancy weight on target?	60,522 (7.84)	196	0.32	0.17	0	2	4
Phase 3							
Pregnancy weight is not the same for every woman	176,288 (22.83)	4521	2.56	0.33	10	12	62
Gaining a healthy amount of weight is important for you and your baby	14,023 (1.82)	266	1.90	0.43	0	0	11

Discussion

Principal Findings

This observational study explored the use of a digital media campaign applying complementary tactics to increase awareness about healthy GWG among women in Alberta. Google AdWords targeted users who were actively searching for pregnancy-related information, while Facebook ads reached an audience who were not actively seeking this information. Both tactics directed users to an established website that had credible, evidence-based information, as well as helpful interactive tools to support women achieve a healthy GWG. Observational studies of digital campaigns and the sharing of lessons learned are important as health organizations continue to increase their presence on the Web and as Facebook increasingly limits the reach of unpaid messaging.

Although determining definitive benchmarks and comparing digital marketing tactics can be difficult because of the many variables involved (eg, budget, target audience, industry, competition, imagery, etc), the GWG digital campaign did exceed recently published standards. A digital marketing company in the United States investigated ad campaign metrics and determined that the average CTR across all industries in Google AdWords was 1.91% and average CPC was US $2.32 (Can $2.92) for paid search ads [28]. The average CTR for Facebook ads was 0.9% and average CPC was US $1.72 (Can $2.19) [29]. Using these overall benchmarks, our campaign had approximately 3 times the CTR for a third of the cost in Google AdWords and had approximately double the CTR for one-fifth of the cost in Facebook. The results of this campaign serve as the baseline for future campaigns and provide insights into target audience preferences and the components that contribute to successful advertising.

Previous health campaigns have used paid digital display ads (ads placed on various websites that are not affiliated with the advertiser) and search engine ads or have focused solely on social media platforms to reach target audiences. Cooper et al [30], for example, used Google ads to direct users searching for gynecologic cancer information to content on the Centers for Disease Control and Prevention website. Visits to the advertised webpages were 26 times higher during the 3-month campaign compared with the period before. Another campaign conducted in the state of Michigan over 11 weeks with a US $15,000 budget, used Facebook ads to increase awareness of a newborn screening and biobanking program [31]; the campaign reached 1.88 million users while achieving 15,958 website clicks, 452 shares, and 542 comments. Facebook ads have also been used to reach women for participation in prenatal research studies [22,32,33]. Compared with traditional tactics, Web-based ads were found to be more efficient and cost-effective for recruitment because of the ability to target women based on specific demographic and geographic features.

In this study, almost 80% of users who clicked on a Google ad completed a conversion on the website, which highlights the ability of these ads to attract highly interested users. Using Facebook ads instead of digital display ads as a complementary tactic also has several advantages, such as more effective targeting capabilities, higher-quality data on ad performance

(which enabled us to refine the ad content between phases), ability to see user comments, the opportunity for 2-way communication, and extended organic reach when users interact with ads (because the engaged users' social networks are seeing their engagement and thereby exposed to the ad message).

Online ads have been shown to impact users and increase awareness of the message even without a click [34]. A study involving a tobacco prevention campaign using Web-based ads to direct users to a campaign website found that exposure to digital display and search ads influenced visits to the website with added visits occurring up to 4 weeks after a user was exposed to an ad [35]. The ads were also found to influence other information-seeking behaviors, with users visiting additional cessation sites following ad exposure; this further emphasizes the benefit of complementary ad tactics as a user exposed to a social media ad may later use a search engine to research the topic, providing another opportunity to advertise and direct the user to a trusted, designated website.

A well-defined strategy including target audience research, aligning metrics with campaign goals, and a monitoring and evaluation plan was essential to the campaign. Lee et al [36] described a similar experience in using digital tactics as part of a campaign to increase awareness among health care professionals in California about smoking cessation resources for patients; the authors also comment on the benefits of having a clear purpose for each campaign tactic, continuous data monitoring to guide activities, and working with a marketing consultant. Hiring a vendor is an additional expense; however, organizations with limited experience may benefit from the technical expertise and the opportunity for staff to learn about campaign design and management. The GWG digital campaign began with the support of an external expert but advertising functions transitioned to internal staff as they grew more comfortable with the technology and system established by the external expert.

The strategy allowed for each phase of the campaign to build upon the data collected and lessons learned throughout the project. The ability to measure, monitor, and adapt the strategy continuously allowed for gradual optimization of campaign elements toward target audience preferences, lower costs, and a wider reach. Using readily available metrics, campaign activities were adjusted on the basis of components that indicated effectiveness. The highest-performing social media ad, for example, occurred in the last phase of the campaign and represented a combination of the most successful elements (eg, text, branding, and imagery) from previous ads. Furthermore, ad fatigue is a common challenge with digital advertising, where ad performance tends to decline over time. Running 2 Facebook ads at a time—and testing the 2 against one another—helped to reduce ad fatigue by keeping the content new and engaging while providing a means to test the performance of various campaign elements among the target audience.

Limitations

This study has a number of limitations. Conversions could only be reported for Google AdWords because of an oversight in placing a tracking code in the HPHC website. Future digital campaigns will include this coding consideration early in the planning stage to ensure that additional data can be collected. There are also limitations in comparing results between marketing tactics as they reach different audiences and have different aims. Determining which tactic is a better value or which had a larger impact on the target audience is dependent on the individual interpretation that considers audiences and objectives—it would not be accurate in this case to directly compare the CTR for Google AdWords with the CTR for Facebook ads, for example.

This study did not assess the impact of the campaign on women's behaviors, knowledge, or attitude; however, 83% of women in a study by Lagan et al [12] reported that the internet was used to influence decision making during their pregnancy. Women felt that their confidence levels to make decisions during pregnancy and their ability to engage in discussions with their health care provider significantly increased after using the internet. Thus, in the case of the GWG campaign, paid digital marketing may offer a means to meet women's informational needs and ensure they receive credible, evidence-based, and personalized recommendations about GWG. Future research exploring the impact of digital ads on health-related attitudes and behavioral outcomes is warranted.

Conclusions

This study adds to the literature by employing search engine marketing and social media advertising as complementary tactics for GWG health promotion goals. While the quality of information on the Web is highly variable, Google AdWords can be used by health organizations to promote credible websites as women search for specific information to fill knowledge gaps and inform decisions about their pregnancy. An advantage of incorporating Facebook ads into an overall digital marketing campaign is the ability to deliver information to women without their intervention, reaching those who may not be aware of health information related to GWG and women who may not have information-seeking skills.

Harnessing the power of search engine marketing and social media advertising together is a promising strategy to reach women who may not receive adequate information about GWG from other sources. Advantages of digital marketing include the opportunity to target a large audience and the ability to monitor and apply user data to optimize digital marketing activities. While the impact of digital marketing on behavior change is unknown, it may play an important role in contributing to increased awareness and knowledge of health topics that inspire action.

Acknowledgments

The authors thank Sheila Tyminski and Dr Kara Nerenberg for their contribution during the design of the Facebook ads. The authors also thank Dr Dolly Bondarianzadeh and Vincent Terstappen for their critical review of the manuscript. The campaign was funded by a Collaborative Research and Innovation Opportunities grant provided by Alberta Innovates.

Conflicts of Interest

None declared.

References

1. Health Canada. Prenatal Nutrition Guidelines for Health Professionals: Gestational Weight Gain. Ottawa, Canada: Government of Canada; 2010.
2. Kowal C, Kuk J, Tamim H. Characteristics of weight gain in pregnancy among Canadian women. Matern Child Health J 2012 Apr;16(3):668-676. [doi: 10.1007/s10995-011-0771-3] [Medline: 21431862]
3. Jarman M, Yuan Y, Pakseresht M, Shi Q, Robson PJ, Bell RC, Alberta Pregnancy Outcomes Nutrition study team, ENRICH team. Patterns and trajectories of gestational weight gain: a prospective cohort study. CMAJ Open 2016;4(2):E338-E345 [FREE Full text] [doi: 10.9778/cmajo.20150132] [Medline: 27525254]
4. Johnson J, Clifton RG, Roberts JM, Myatt L, Hauth JC, Spong CY, Eunice Kennedy Shriver National Institute of Child Health, Human Development (NICHD) Maternal-Fetal Medicine Units (MFMU) Network. Pregnancy outcomes with weight gain above or below the 2009 Institute of Medicine guidelines. Obstet Gynecol 2013 May;121(5):969-975 [FREE Full text] [doi: 10.1097/AOG.0b013e31828aea03] [Medline: 23635732]
5. Lau EY, Liu J, Archer E, McDonald SM, Liu J. Maternal weight gain in pregnancy and risk of obesity among offspring: a systematic review. J Obes 2014;2014:524939 [FREE Full text] [doi: 10.1155/2014/524939] [Medline: 25371815]
6. Han Z, Mulla S, Beyene J, Liao G, McDonald SD, Knowledge Synthesis Group. Maternal underweight and the risk of preterm birth and low birth weight: a systematic review and meta-analyses. Int J Epidemiol 2011 Feb;40(1):65-101. [doi: 10.1093/ije/dyq195] [Medline: 21097954]
7. Herring SJ, Oken E, Haines J, Rich-Edwards JW, Rifas-Shiman SL, Kleinman Sc DP, et al. Misperceived pre-pregnancy body weight status predicts excessive gestational weight gain: findings from a US cohort study. BMC Pregnancy Childbirth 2008 Dec 22;8:54 [FREE Full text] [doi: 10.1186/1471-2393-8-54] [Medline: 19102729]
8. Shub A, Huning EY, Campbell KJ, McCarthy EA. Pregnant women's knowledge of weight, weight gain, complications of obesity and weight management strategies in pregnancy. BMC Res Notes 2013 Jul 18;6:278 [FREE Full text] [doi: 10.1186/1756-0500-6-278] [Medline: 23866845]
9. Gaudet LM, Gruslin A, Magee LA. Weight in pregnancy and its implications: what women report. J Obstet Gynaecol Can 2011 Mar;33(3):227-234. [Medline: 21453562]
10. Nikolopoulos H, Mayan M, MacIsaac J, Miller T, Bell RC. Women's perceptions of discussions about gestational weight gain with health care providers during pregnancy and postpartum: a qualitative study. BMC Pregnancy Childbirth 2017 Dec 24;17(1):97 [FREE Full text] [doi: 10.1186/s12884-017-1257-0] [Medline: 28335749]
11. Kraschnewski JL, Chuang CH, Poole ES, Peyton T, Blubaugh I, Pauli J, et al. Paging “Dr. Google”: does technology fill the gap created by the prenatal care visit structure? Qualitative focus group study with pregnant women. J Med Internet Res 2014 Jun 03;16(6):e147 [FREE Full text] [doi: 10.2196/jmir.3385] [Medline: 24892583]
12. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]
13. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
14. Szwajcer EM, Hiddink GJ, Koelen MA, van Woerkum CMJ. Nutrition-related information-seeking behaviours before and throughout the course of pregnancy: consequences for nutrition communication. Eur J Clin Nutr 2005 Aug;59 Suppl 1:S57-S65. [doi: 10.1038/sj.ejcn.1602175] [Medline: 16052197]
15. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Dec 19;16(1):171 [FREE Full text] [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
16. Bert F, Gualano MR, Brusaferro S, De Vito E, de Waure C, La Torre G, et al. Pregnancy e-health: a multicenter Italian cross-sectional study on Internet use and decision-making among pregnant women. J Epidemiol Community Health 2013 Dec 01;67(12):1013-1018. [doi: 10.1136/jech-2013-202584] [Medline: 24072743]
17. Larsson M. A descriptive study of the use of the Internet by women seeking pregnancy-related information. Midwifery 2009 Feb;25(1):14-20. [doi: 10.1016/j.midw.2007.01.010] [Medline: 17408822]
18. Chou WS, Prestin A, Lyons C, Wen K. 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]
19. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]

20. Mosseri A. Facebook. 2018 Jan 11. Bringing People Closer Together URL: https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together [accessed 2018-05-03] [WebCite Cache ID 6zGRe9qmb]
21. 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]
22. Adam LM, Manca DP, Bell RC. Can Facebook Be Used for Research? Experiences Using Facebook to Recruit Pregnant Women for a Randomized Controlled Trial. J Med Internet Res 2016 Sep 21;18(9):e250 [FREE Full text] [doi: 10.2196/jmir.6404] [Medline: 27655184]
23. Alberta Health Services. Healthy Parents, Healthy Children. 2018. URL: http://www.healthyparentshealthychildren.ca/ [accessed 2018-05-08] [WebCite Cache ID 6zGRVt2Gw]
24. Alberta Health Services. Healthy Parents, Healthy Children: Track Your Weight During Pregnancy. 2018. URL: https://www.healthyparentshealthychildren.ca/resources/tools/weight-gain-calculator/ [accessed 2018-05-08] [WebCite Cache ID 6zGhIWj1S]
25. Purcell K, Brenner J, Rainie L. Pew Research Center. 2012. Search Engine Use 2012 URL: http://www.pewinternet.org/2012/03/09/search-engine-use-2012 [accessed 2018-05-08] [WebCite Cache ID 6zGRohTGY]
26. Greenwood S, Perrin A, Duggan M. Pew Research Center. 2016 Nov 11. Social Media Update 2016 URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016 [accessed 2018-05-08] [WebCite Cache ID 6zGS4oxyx]
27. Centers for Disease Control Prevention. 2011. What is Health Marketing? URL: https://www.cdc.gov/healthcommunication/toolstemplates/WhatIsHM.html [accessed 2018-05-06] [WebCite Cache ID 6zGSCC3s9]
28. Irvine M. WordStream. 2018 Apr 09. Google AdWords Benchmarks for YOUR Industry [Updated!] URL: https://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks [accessed 2018-05-08] [WebCite Cache ID 6zGSGp41A]
29. Irvine M. WordStream. 2018 Apr 09. Facebook Ad Benchmarks for YOUR Industry [New Data] URL: https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks [accessed 2018-05-08] [WebCite Cache ID 6zGSLZIHY]
30. Cooper CP, Gelb CA, Vaughn AN, Smuland J, Hughes AG, Hawkins NA. Directing the public to evidence-based online content. J Am Med Inform Assoc 2015 Apr;22(e1):e39-e41 [FREE Full text] [doi: 10.1136/amiajnl-2014-002701] [Medline: 25053580]
31. Platt T, Platt J, Thiel DB, Kardia SLR. 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]
32. Admon L, Haefner JK, Kolenic GE, Chang T, Davis MM, Moniz MH. Recruiting Pregnant Patients for Survey Research: A Head to Head Comparison of Social Media-Based Versus Clinic-Based Approaches. J Med Internet Res 2016 Dec 21;18(12):e326 [FREE Full text] [doi: 10.2196/jmir.6593] [Medline: 28003174]
33. Arcia A. Facebook Advertisements for Inexpensive Participant Recruitment Among Women in Early Pregnancy. Health Educ Behav 2014 Jun;41(3):237-241 [FREE Full text] [doi: 10.1177/1090198113504414] [Medline: 24082026]
34. Yoo CY. Unconscious processing of Web advertising: Effects on implicit memory, attitude toward the brand, and consideration set. Journal of Interactive Marketing 2008 Jan;22(2):2-18. [doi: 10.1002/dir.20110]
35. Kim A, Hansen H, Duke J, Davis K, Alexander R, Rowland A, et al. Does Digital Ad Exposure Influence Information-Seeking Behavior Online? Evidence From the 2012 Tips From Former Smokers National Tobacco Prevention Campaign. J Med Internet Res 2016 Mar 16;18(3):e64 [FREE Full text] [doi: 10.2196/jmir.4299] [Medline: 26983849]
36. Lee YO, Momin B, Hansen H, Duke J, Harms K, McCartney A, et al. Maximizing the Impact of Digital Media Campaigns to Promote Smoking Cessation: A Case Study of the California Tobacco Control Program and the California Smokers' Helpline. Calif J Health Promot 2014;12(3):35-45 [FREE Full text] [Medline: 28239304]

Abbreviations

AHS: Alberta Health Services
CPC: cost per click
CTR: click-through rate
GWG: gestational weight gain
HPHC: Healthy Parents, Healthy Children

Edited by G Eysenbach; submitted 10.07.18; peer-reviewed by J Most; comments to author 14.10.18; revised version received 27.10.18; accepted 29.10.18; published 01.02.19.

Please cite as:
Graham JE, Moore JL, Bell RC, Miller T
Digital Marketing to Promote Healthy Weight Gain Among Pregnant Women in Alberta: An Implementation Study
J Med Internet Res 2019;21(2):e11534
URL: https://www.jmir.org/2019/2/e11534/
doi:10.2196/11534
PMID:30707100

Original Paper

Mitigation of Participant Loss to Follow-Up Using Facebook: All Our Families Longitudinal Pregnancy Cohort

Nikki Lee Stephenson[1,2], MSc; Erin Hetherington[1], MSc; Shawn Dodd[2], MSc; Alexander Mathews[3], BSc; Suzanne Tough[1,2], PhD

[1]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[2]Department of Paediatrics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[3]O'Brien Centre for the Bachelor of Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Suzanne Tough, PhD
Department of Paediatrics
Cumming School of Medicine
University of Calgary
Owerko Centre
2500 University Drive NW
Calgary, AB, T2N 1N4
Canada
Phone: 1 403 441 8450
Email: stough@ucalgary.ca

Abstract

Background: Facebook, a popular social media site, allows users to communicate and exchange information. Social media sites can also be used as databases to search for individuals, including cohort participants. Retaining and tracking cohort participants are essential for the validity and generalizability of data in longitudinal research. Despite numerous strategies to minimize loss to follow-up, maintaining contact with participants is time-consuming and resource-intensive. Social media may provide alternative methods of contacting participants who consented to follow-up but could not be reached, and thus are potentially "lost to follow-up."

Objective: The aim of this study was to determine if Facebook was a feasible method for identifying and contacting participants of a longitudinal pregnancy cohort who were lost to follow-up and re-engaging them without selection bias.

Methods: This study used data from the All Our Families cohort. Of the 2827 mother-child dyads within the cohort, 237 participants were lost to follow-up. Participants were considered lost to follow-up if they had agreed to participate in additional research, completed at least one of the perinatal questionnaires, did not complete the 5-year postpartum questionnaire, and could not be contacted after numerous attempts via phone, email, or mail. Participants were considered to be matched to a Facebook profile if 2 or more characteristics matched information previously collected. Participants were sent both a friend request and a personal message through the study's Facebook page and were invited to verify their enrollment in the study. The authors deemed a friend request was necessary because of the reduced functionality of nonfriend direct messaging at the time. If the participant accepted the study's friend request, then a personalized message was sent. Participants were considered reconnected if they accepted the friend request or responded to any messages. Participants were considered re-engaged if they provided up-to-date contact information.

Results: Compared with the overall cohort, participants who were lost to follow-up (n=237) were younger (P=.003), nonmarried (P=.02), had lower household income (P<.001), less education (P<.001), and self-identified as being part of an ethnic minority (P=.02). Of the 237 participants considered lost to follow-up, 47.7% (113/237) participants were identified using Facebook. Among the 113 identified participants, 77.0% (87/113) were contacted, 32.7% (37/113) were reconnected, and 17.7% (20/113) were re-engaged. No significant differences were found between those identified on Facebook (n=113) and those who were not able to be identified (n=124).

Conclusions: Facebook identified 47.6% (113/237) of participants who were considered lost to follow-up, and the social media site may be a practical tool for reconnecting with participants. The results from this study demonstrate that social networking

sites, such as Facebook, could be included in the development of retention practices and can be implemented at any point in cohort follow-up.

(J Med Internet Res 2019;21(2):e10441) doi:10.2196/10441

KEYWORDS

social media; social networking; longitudinal studies; patient dropouts; cohort studies; follow-up studies; lost to follow-up

Introduction

Background

Prospective cohort study designs are methodologically valuable as they follow participants over time, which allows for the identification of risk factors that occur before outcomes and allow for the examination of trajectories of health and development [1]. However, this study design is particularly vulnerable to attrition over time as people move or lose interest in participating [1]. Birth and pregnancy cohorts may be particularly susceptible to participant attrition due to high mobility during this life stage and busy schedules of families with small children. In addition to introducing selection bias [2], loss of cohort participants reduces data quality, interpretability, and potential generalizability. Attrition also reduces statistical power and can threaten the accuracy of measures of association (such as odds ratios and risk ratios). Consequently, maximum follow-up rates are attempted [1,3].

The now widespread use of social networking sites, such as Facebook, can provide new opportunities for locating participants of research studies, who are difficult to track. Previous studies have used Facebook to recruit participants—specifically to identify and engage participants with rare conditions—those who are hard to reach or vulnerable [4-6]. Recruitment through Facebook can be more cost-effective than traditional methods, which is appealing to researchers [7-10]. Several studies have also successfully used Facebook to relocate participants after the original research concluded. These include a longitudinal follow-up of an intervention program for at-risk families [11], a longitudinal study of adults who used methamphetamine [12], and members of a graduating class [13]. These studies demonstrate the potential for identifying participants for whom follow-up was unplanned, who were high risk, or highly mobile. Yet, use of these methods to reduce attrition within a low-risk cohort has not been examined.

The All Our Families cohort (formerly the All Our Babies study) is a contemporary ongoing prospective community-based pregnancy cohort situated in Calgary, Alberta, Canada. A detailed overview of the study design, recruitment, eligibility, and data collection is described elsewhere [14-16]. Since recruitment, between 2008 and 2011, participants have completed 7 questionnaires; 3 questionnaires in the perinatal period (22 to 24 and 32 to 36 weeks gestation and 4 months postpartum) and 4 questionnaires in the early childhood period (1, 2, 3, and 5 years postpartum). The All Our Families cohort has utilized many methods to improve response rates and minimize loss to follow-up. These include collection of detailed and appropriate recruitment information, implementing standardized participant-tracking procedures, contacting alternative contacts, identifying 3 or more alternate contacts, increased frequency of participant contact, and offering monetary or other incentives for study participation [1,12,17-19]. However, even when all recommended strategies are implemented, a participant may not be retained in a mobile population [1]. Addresses, telephone numbers, and even emails are becoming less predictable ways of tracking participants in cohorts [20]. Recommendations to obtain personal identifying information to track participants on major databases (ie, driver's license numbers for use with the department of motor vehicles) [20] are not feasible in every country or region as privacy laws may prohibit researchers from accessing this information without explicit prior participant consent. Therefore, in an effort to minimize selection bias and re-engage lost-to-follow-up participants, the All Our Families cohort required new methods of participant tracking.

Objectives

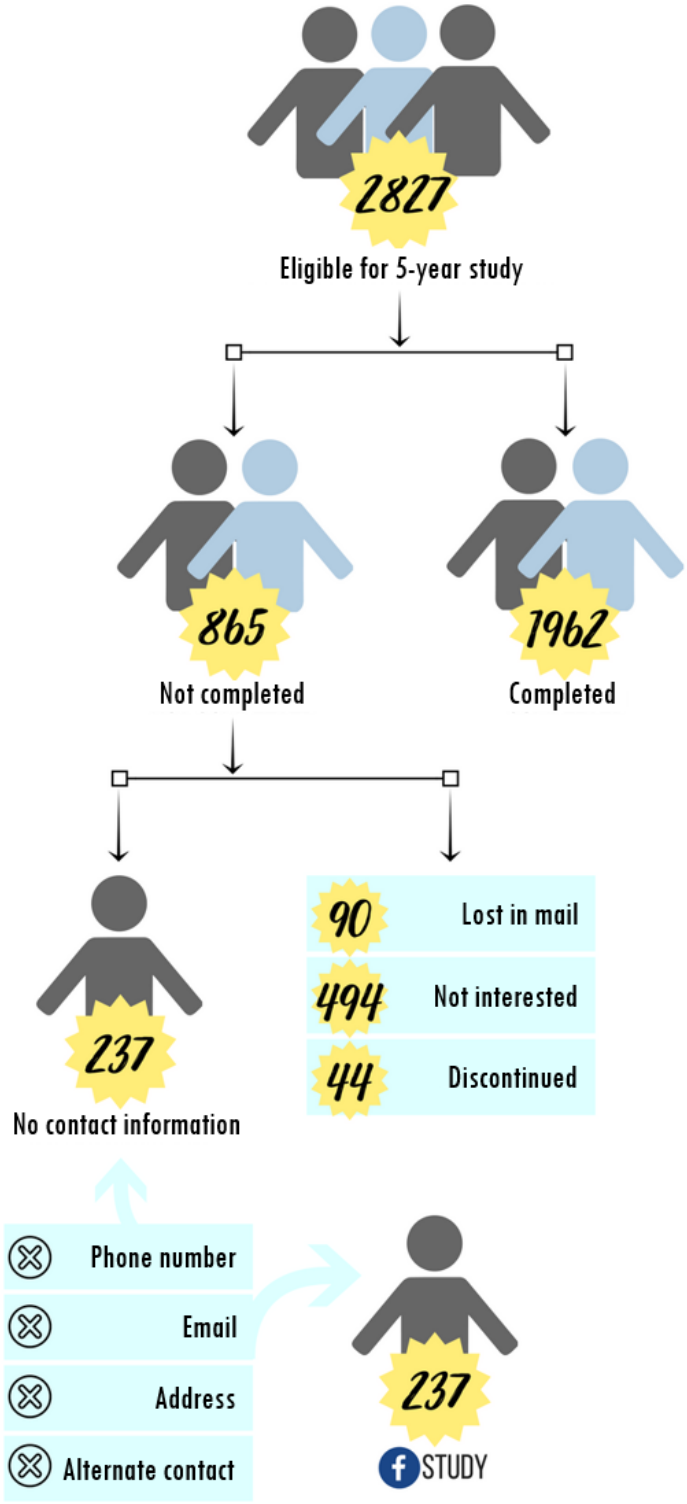
The effectiveness of using social media to recontact participants in a longitudinal pregnancy cohort has not previously been examined. Among social networking sites, Facebook has emerged as one of the dominant platforms reporting 2.23 billion monthly active users as of June 30, 2018 [21]. Unlike traditional communication platforms, Facebook URLs are associated with longer periods of use compared with email addresses [12], which could potentially make retroactive tracking more accessible. Among active Facebook users, people aged between 18 and 24 years (30.9%), 25 and 34 years (22.6%), and 35 and 54 years (27.0%) comprise the largest member groups and continue to grow [22]. For this study, Facebook was chosen as a target social media platform to use as the median age of the All Our Families cohort and the typical age of Facebook users overlapped [23]. This study investigated the feasibility of 1 social networking site, Facebook, for both identifying and recontacting participants without demographic bias as part of a contemporary longitudinal pregnancy cohort situated in Calgary, Alberta, Canada.

Methods

Study Eligibility

Of the 2827 mother-baby dyads who were eligible to complete the 5-year questionnaire, 237 participants were considered lost to follow-up (Figure 1). Participants were considered lost to follow-up if they had agreed to participate in additional research and completed at least one of the perinatal questionnaires, did not complete the 5-year postpartum questionnaire, and could not be contacted after numerous attempts (n=237). Participants were excluded from the re-engagement study if they had previously withdrawn or discontinued from the study or had indicated a lack of interest in the data collection wave.

Figure 1. Eligibility criteria of lost to follow-up participants.



Facebook Re-Engagement Protocol

To contact participants, an All Our Families study Facebook profile was created. Throughout the course of the re-engagement study, the All Our Families Facebook profile included the study logo, a brief paragraph describing the study, the study's contact information, and frequent updates regarding the study.

To re-engage participants, those lost to follow-up were first identified by searching the first and last name on Facebook. To verify the participants' identity, profiles were browsed for known identifiers including birth date, home address, email address, child's name, spouse's name, phone number, and if the alternate contact on file was included in the participant's friends list. If a Facebook profile contained at least two identifiers that matched data previously collected by the study, the participant was considered found on Facebook. Profiles with matching identifiers were then sent friend requests and a personalized message via Facebook's direct messaging service,

asking the profile owner to confirm their participation in the All Our Families study. The friend request was deemed necessary for direct messaging at the time of the study (May-August 2016), as messages from nonfriends were relegated to a separate section within the messages tab on Facebook's website, where messages would not prompt notifications to the user. If participants responded to the initial message, they were encouraged to contact the study team via email or telephone to update their contact information. If participants did not respond to the initial Facebook message after 2 weeks, a follow-up message was sent. A third and final follow-up message was sent to participants if they did not respond within 1 month following the initial message.

Ethical Considerations

Social media platforms such as Facebook provide an innovative means for recruitment and retention; however, the use of social media may increase risks to participant privacy and confidentiality. This study recognized that ethical principles for ensuring privacy and confidentiality of study participant's personal information may be affected by the use of Facebook. To protect participants' privacy on the study's Facebook page, the study's Facebook privacy settings were set to "Disable posts by other people on the Page."

Confidentiality cannot be guaranteed if a participant "shares" or "likes" the study's page; however, by performing these actions, the participant, who has agreed to the terms and conditions of the social media site, has the ability to disclose this information as they wish according to their own privacy settings. In addition, the study's Facebook page accepts friends who are not participants of the study, and as such, being a friend of the study does not imply that a person is a participant—only that the person may have an interest in our research on maternal child health. Nonparticipants have equal opportunity to utilize these social media functions, which further protects confidentiality as the communication is not exclusive to participants. In addition to the above safeguards, Facebook's privacy settings were monitored on a weekly basis for potential updates. No updates or changes occurred during the data collection period. Although the study undertook precautions to protect the identity of participants, we recognized the potential risk of disclosing a participant's identity when using social media as a public database.

The All Our Families study is a population-based cohort, with the eligibility criteria of having a child and being over the age of 18 years. The potential risk associated with the proposed re-engagement strategy was the potential harm that a person may have experienced from the disclosure, collection, and use of personal and sensitive information triggered by accepting the study's friend request and replying to the message through the personal messaging system. This was considered to be minimal within our study population as accepting the friend request was at the discretion of the participant, who had already agreed to be contacted and had previously provided personal information to the study. If the friend request was accepted, the All Our Families study information that would have been visible was information already in the public domain about the study (eligibility criteria). The study viewed nonacceptance of the friend request or nonresponse to the direct message as the participants practicing their right to control their information while participating in social media within a public space. The study's privacy settings were set so that only the study team could view the friend list, only mutual friends (persons who were friends with the study as well as with the newly added participant) would be visible to both the study team and to each individual friend. This was meant to maximize confidentiality for the participant as their privacy settings would determine who is able to view their association with the study. Studies that may be dealing with specific diseases or have identifiable characteristics (ie, use of the same program) may require different privacy safeguards [7]. Furthermore, if the association of the participant and study was revealed through this communication, the personal and integrated nature of our previous contact with the participant and our study's purpose were considered in weighing research utility with participant confidentiality. Messages to participants included sufficient information to potentially confirm their participation (asking for confirmation of the details we found through Facebook) but not lead to involuntary disclosure of additional personal information through social media. In addition, conversations were moved off of social media once a participant had been engaged.

The All Our Families study was approved by the Child Health Research Office, Alberta Health Services, and the Conjoint Health Research Ethics Board of the University of Calgary. Written informed consent for the initial study and follow-up was obtained from the study participants at the time of recruitment, who were also provided copies for their records. Although this consent did include permission to contact for future studies, it did not specify the use of social media as a mode of contact. Additional ethical approval for this participant re-engagement was obtained, considering the specific privacy and confidentiality concerns regarding the use of social media. Data used for identifying participants were stored and analyzed on the 256-bit encrypted server at the University of Calgary, and only researchers named in the ethics file and those who sign confidentiality agreements have access to these data.

Data Analysis

For the purpose of this analysis, participants were considered to be identified on Facebook if their identity could be confirmed on their profile. Participants were considered contacted if their privacy settings allowed for a friend request and a personal message to be sent. If a participant accepted the study's friend request or replied to any of the messages, they were considered as being reconnected with the study. Finally, if a participant responded to the messages and provided their up-to-date contact information, they were considered as being re-engaged with the study.

Bivariate analyses were used to compare demographic variables of active study participants and participants considered lost to follow-up, using chi-square or Fisher exact tests. Bivariate analyses were also used to compare those whose Facebook profiles were identified compared with those whose could not be identified. A *P* value of less than .05 was considered

statistically significant. Statistical analyses were performed using SPSS version 23 (SPSS Inc).

Results

Engagement

Data collection was initiated in June 2016 and ended in August 2016. Of the 237 All Our Families participants considered lost to follow-up, 47.7% (113/237) of participant profiles were identified using Facebook. Among the 113 participant profiles that were identified, 77.0% (87/113) of participants (those who accepted the friend request) were sent messages through the Facebook messenger app. Moreover, 32.7% (37/113) of participants responded to the sent messages, and 17.7% (20/113) of these participants were re-engaged in the study's follow-up.

Participant Characteristics

To understand the sociodemographic differences between active participants (n=2590) and participants considered lost to follow-up (n=237), baseline data were compared on the following variables: maternal age, marital status, born in Canada, household income, education status, ethnicity, and language spoken at home (Table 1). Participants considered lost to follow-up when compared with continuing participants were younger (P=.003), single (P=.023), had less household income (P<.001), lower education (P<.001), and were more likely to identify as ethnic (P=.015).

No significant differences were found between those whose Facebook profiles were identified (n=113) and those whose could not be identified (n=124; Table 2).

Table 1. Characteristics of continuing and lost to follow-up participants.

Demographic characteristics (at recruitment)	Was this participant identified as lost to follow-up?		*P* value
	No (n=2590[a]), n (%)	Yes (n=237[a]), n (%)	
Maternal age (years)			.003[b]
Less than 35	2012 (77.68)	201 (88.2)	
35 or older	503 (19.42)	27 (11.8)	
Marital status			.02[b]
Married or common law	2449 (95.22)	213 (91.8)	
Other	123 (4.78)	19 (8.2)	
Born in Canada			.10
Yes	2052 (79.72)	175 (75.1)	
No	522 (20.28)	58 (24.9)	
Household income			<.001[b]
<Can $60,000	376 (15.09)	62 (27.4)	
Can $60,000 or greater	2115 (84.91)	164 (72.6)	
Education			<.001[b]
High school or less	228 (8.88)	59 (25.2)	
Some or completed postsecondary	2341 (91.12)	175 (74.8)	
Self-identified ethnicity			.02[b]
Minority	501 (19.49)	61 (26.2)	
White	2069 (80.51)	172 (73.8)	
Language spoken at home			.42
English	2298 (89.28)	204 (87.6)	
Other	276 (10.72)	29 (12.4)	

[a]Small variations in totals may exist because of missing data (less than 3%).

[b]Indicates statistical significance (P<.05).

Table 2. Characteristics of participants identified on Facebook.

Demographic characteristics (at recruitment)	Was this participant identified on Facebook?		*P* value
	No (n=124[a]), n (%)	Yes (n=113[a]), n (%)	
Maternal age (years)			>.99
Less than 35	104 (88.1)	97 (88.2)	
35 or older	14 (11.9)	13(11.8)	
Marital status			.69
Married or common law	111 (92.5)	102 (91.1)	
Other	9 (7.5)	10 (8.9)	
Born in Canada			.57
Yes	89 (73.6)	86 (76.8)	
No	32 (26.4)	26 (23.2)	
Household income			.74
<Can $60,000	31 (26.5)	31 (29.4)	
Can $60,000 or greater	86 (73.5)	78 (71.6)	
Education			.5
High school or less	33 (27.0)	26 (23.2)	
Some or completed postsecondary	89 (73.0)	86 (76.8)	
Self-identified ethnicity			.49
Minority	34 (28.1)	27 (24.1)	
White	87 (71.9)	85 (75.9)	
Language			.24
English	103 (85.1)	101 (90.2)	
Other	18 (14.9)	11 (9.8)	

[a]Small variations in totals may exist because of missing data (less than 3%).

[b]Not applicable.

Discussion

Principal Findings

This study has 2 key findings. First, Facebook is a feasible way to identify participants who are lost to follow-up even without prior collection of Facebook identifiers. Using the social networking site, 48% (n=113) of lost to follow-up participants were identified and 17% (n=20) of those identified were re-engaged in the study's follow-up. Finding and contacting participants who are lost to follow-up often depends upon collection of detailed information at the time of recruitment [18]. By using only the participant information collected at recruitment (ie, first and last name, email, and alternative contacts), almost half of the participants who were lost to follow-up were identified. The authors recognized that the privacy norms surrounding communication through social media differ from traditional means of communication by email or telephone. Although this study did not encounter negative perceptions of our use of social media from participants, the concept of privacy within social media norms is situationally dependent, and researchers must understand the context surrounding perceived privacy violations when using social media within their research [24].

Second, Facebook can assist with identifying and re-engaging participants across sociodemographic strata and may be particularly valuable for certain populations, such as those who are younger. In pregnancy cohorts, participants who are lost to follow-up are often those who are younger, unmarried, have a lower household income, less education, and more likely to self-identify as a self-identified ethnic minority compared with active participants [2]. These sociodemographic characteristics are consistent with other studies, where more vulnerable groups are more susceptible to be lost to follow-up [25,26]. Although the definition of vulnerability varies depending on locality, comparative to the relatively affluent cohort population, vulnerability in this study was defined as being younger (aged less than 35 years), lower income (<Can $60,000), or self-identified ethnic minority or not born in Canada. Of note, among those lost to follow-up, there were no sociodemographic differences between those who were identified through Facebook and those who were not, implying the feasibility of using Facebook to connect with participants without bias.

Incorporation of Facebook in the study protocols was time-efficient, and therefore cost-efficient, in comparison with other re-engagement protocols. Moreover, 1 team member actively searched, identified, and contacted each of the 237

participants in 1 week (35 hours or 2100 min), with minimal weekly follow-up messages. Approximately 10 min were spent per participant, and almost half of the participants were identified. Comparatively, the cohort staff spend 20 min per participant on phone calls to participants and alternate contacts, sending emails, and sending request letters via mail. In the case of lost to follow-up participants, *traditional* methods were ineffective as the required information (phone numbers, mailing address, and email) was out-of-date.

Strengths and Limitations

Although this study was able to use Facebook to identify participants without a priori data collection, collection of Facebook-specific identifiers at study outset is recommended, if possible. Facebook's privacy settings are continually changing to keep up with the public demand for privacy and security of data, and it may become more difficult to use Facebook to reconnect with participants. Unlike phone or email, Facebook URLs are associated with longer periods of use [12]. Collection of Facebook ID at the time of recruitment would expedite the process of connecting with participants and would increase confidence that the profile corresponds to the study participant. Researchers using any type of social media to engage participants should remain vigilant to the dynamic nature of privacy settings on social media platforms and would be encouraged to utilize social media early on in the study timeline as part of their knowledge dissemination strategy. This method should be adapted depending on the demographics and subject matter of the particular study. The All Our Families study is a population-based cohort, where eligibility does not imply specific personal health information, and future studies should alter the privacy measures to address the sensitivity of information being exchanged. For this study, Facebook was the best social media platform to use as the median age of the All Our Families cohort and the typical age of Facebook users overlapped [23]. However, when looking at a younger cohort, other social networking sites should be considered as they may be of greater benefit. When considering which social media site to use, the risks and benefits of each were weighed. The advantage of an open platform such as Twitter is that anyone can follow and search for users, whereas a closed platform such as Facebook allows users more control over who can view their profile. Both platforms allow direct messaging, and the accessibility of this messaging is determined by the profile owner [27]. The user interface of Twitter is more accessible to the public, and its intended use is for online social networking and microblogging. Conversely, Facebook's functionality is more diverse and provides stronger privacy and security measures at the control of the user [28]. The decision to use a closed platform such as Facebook, rather than an open platform such as Twitter, was to ensure the privacy and confidentiality of the participants was protected, even if the risk of potential disclosure of personal information was low [29].

Facebook's real name policy was used in this study to identify participants. Although some marginalized communities may prefer to use pseudonyms or other ways of remaining anonymous, this likely represents a small proportion of the study's sample based off a previous cohort analysis [16]. Under this assumption, our method would lead to minimal bias in recontacting the participants from this particular study. Moreover, those identified through Facebook were compared with those not identified through Facebook based on demographic characteristics to assess selection bias. However, in utilizing the real name policy, our results may not be as generalizable to studies who work with vulnerable or marginalized populations, and the implications of this policy should be considered dependent on the specific demographics of the study.

The use of social media to re-engage participants brings in new ethical considerations, and early incorporation of possible social media use in the informed consent process is recommended. Social media technology is innovative for recruitment and retention as it encourages the free sharing of information through an interactive and expanding platform. Although all of the participants agreed to participate in the follow-up studies and signed consent forms allowing us to contact them in the future, explicit informed consent from participants at study outset for use of social networking sites for both knowledge dissemination as well as participant communication would reduce the likelihood of participants finding this form of contact intrusive and would provide support for ethical approval. This study had some limitations, including a small sample size. Only 237 participants were identified as lost to follow-up after the 7th wave of data collection in the All Our Families pregnancy cohort. This sample size limits the statistical power when determining differences between those found and those not found. However, when considering the literature surrounding finding, contacting, and re-engaging lost to follow-up participants, this study's sample size is consistent with other studies [11,26].

An additional limitation is that only 1 social media platform was used to recontact participants, and these efforts were limited to a 3-month period. This proposed method of contacting lost to follow-up participants using Facebook was initiated in June 2016 and completed in August 2016, for a total data collection period of 3 months. The short time frame of this study's social media presence may have limited the effectiveness in contacting and re-engaging lost to follow-up participants, especially considering the summer season. Study participants are mothers of school-aged children, and this same protocol at another time of year may increase the likelihood of contacting and re-engaging participants. In addition, the study's Facebook profile was created specifically for this study in May 2016. Had Facebook or other social networking sites been integrated earlier into the study protocol, it may have been more successful at re-engaging lost-to-follow-up participants or reducing attrition over the 7 waves of follow-up.

The dynamic nature of privacy settings within social media sites was a further limitation to this study. At the time of the study, a friend request was necessary preceding a personalized message, as messages before friend requests were relegated to an archived message folder. Messages in the archive folder do not prompt a notification and were less accessible; however, once a friend request was accepted, a personalized message would be sent to the messenger's standard inbox.

The results from this study demonstrated that social networking sites, such as Facebook, should be included in the development of retention practices and can be implemented at any point in cohort follow-up.

Acknowledgments

Initial funding of the All Our Families cohort to investigate maternal and infant outcomes during the perinatal period was provided by Three Cheers for the Early Years, Alberta Health Services. Expansion of the All Our Families cohort was funded by Alberta Innovates Health Solutions (Interdisciplinary Team Grant No. 200700595). Subsequent collaborations focusing on the developmental trajectories of All Our Families children have been supported by numerous grants, including the Social Sciences and Humanities Research Council, Alberta Health Services, the Alberta Children's Hospital Foundation, the Alberta Centre for Child, Family and Community Research, Upstart the United Way of Calgary, Women and Children's Health Research Institute, the Canadian Foundation for Fetal Alcohol Research, and the Max Bell Foundation. EH receives scholarship support from the Canadian Institutes for Health Research Vanier, Alberta Innovates Health Solutions, and the University of Calgary. AM received stipend support through O'Brien Centre Summer Studentships .

The authors acknowledge the contribution and support of All Our Families team members and their participants. The authors are extremely grateful to all the families who took part in this study and the whole All Our Families team. They are extremely grateful to the investigators, coordinators, research assistants, graduate and undergraduate students, volunteers, clerical staff, and managers.

Conflicts of Interest

None declared.

References

1. Golding J, Birmingham K. Enrollment and response rates in a longitudinal birth cohort. Paediatr Perinat Epidemiol 2009 Jul;23(Suppl 1):73-85. [doi: 10.1111/j.1365-3016.2008.01001.x] [Medline: 19490447]
2. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ. Selection bias due to loss to follow up in cohort studies. Epidemiology 2016 Jan;27(1):91-97 [FREE Full text] [doi: 10.1097/EDE.0000000000000409] [Medline: 26484424]
3. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol 2004;19(8):751-760. [doi: 10.1023/B:EJEP.0000036568.02655.f8] [Medline: 15469032]
4. Burton-Chase AM, Parker WM, Hennig K, Sisson F, Bruzzone LL. The use of social media to recruit participants with rare conditions: Lynch syndrome as an example. JMIR Res Protoc 2017 Jan 23;6(1):e12 [FREE Full text] [doi: 10.2196/resprot.6066] [Medline: 28115298]
5. Adam LM, Manca DP, Bell RC. Can Facebook be used for research? Experiences using Facebook to recruit pregnant women for a randomized controlled trial. J Med Internet Res 2016 Sep 21;18(9):e250 [FREE Full text] [doi: 10.2196/jmir.6404] [Medline: 27655184]
6. Martinez O, Wu E, Shultz AZ, Capote J, López Rios J, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014 Apr 24;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]
7. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
8. 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]
9. 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]
10. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and efficiency of online and offline recruitment methods: a web-based cohort study. J Med Internet Res 2017 Mar 01;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]
11. Mychasiuk R, Benzies K. Facebook: an effective tool for participant retention in longitudinal research. Child Care Health Dev 2012 Sep;38(5):753-756. [doi: 10.1111/j.1365-2214.2011.01326.x] [Medline: 21985571]
12. Bolanos F, Herbeck D, Christou D, Lovinger K, Pham A, Raihan A, et al. Using Facebook to maximize follow-up response rates in a longitudinal study of adults who use methamphetamine. Subst Abuse 2012;6:1-11 [FREE Full text] [doi: 10.4137/SART.S8485] [Medline: 22879750]
13. Amerson R. Facebook: a tool for nursing education research. J Nurs Educ 2011 Jul;50(7):414-416. [doi: 10.3928/01484834-20110331-01] [Medline: 21449530]

14. Gracie SK, Lyon AW, Kehler HL, Pennell CE, Dolan SM, McNeil DA, et al. All Our Babies cohort study: recruitment of a cohort to predict women at risk of preterm birth through the examination of gene expression profiles and the environment. BMC Pregnancy Childbirth 2010 Dec 30;10:87 [FREE Full text] [doi: 10.1186/1471-2393-10-87] [Medline: 21192811]
15. McDonald SW, Lyon AW, Benzies KM, McNeil DA, Lye SJ, Dolan SM, et al. The All Our Babies pregnancy cohort: design, methods, and participant characteristics. BMC Pregnancy Childbirth 2013;13(Suppl 1):S2 [FREE Full text] [doi: 10.1186/1471-2393-13-S1-S2] [Medline: 23445747]
16. Tough SC, McDonald SW, Collisson BA, Graham SA, Kehler H, Kingston D, et al. Cohort profile: The All Our Babies pregnancy cohort (AOB). Int J Epidemiol 2017 Oct 01;46(5):1389-190k [FREE Full text] [doi: 10.1093/ije/dyw363] [Medline: 28180262]
17. Kleschinsky JH, Bosworth LB, Nelson SE, Walsh EK, Shaffer HJ. Persistence pays off: follow-up methods for difficult-to-track longitudinal samples. J Stud Alcohol Drugs 2009 Sep;70(5):751-761 [FREE Full text] [doi: 10.15288/jsad.2009.70.751] [Medline: 19737500]
18. Hunt JR, White E. Retaining and tracking cohort study members. Epidemiol Rev 1998;20(1):57-70. [doi: 10.1093/oxfordjournals.epirev.a017972] [Medline: 9762509]
19. Cotter RB, Burke JD, Stouthamer-Loeber M, Loeber R. Contacting participants for follow-up: how much effort is required to retain participants in longitudinal studies? Eval Program Plann 2005 Feb;28(1):15-21. [doi: 10.1016/j.evalprogplan.2004.10.002]
20. Faden VB, Day NL, Windle M, Windle R, Grube JW, Molina BS, et al. Collecting longitudinal data through childhood, adolescence, and young adulthood: methodological challenges. Alcohol Clin Exp Res 2004 Feb;28(2):330-340 [FREE Full text] [doi: 10.1097/01.ALC.0000113411.33088.FE] [Medline: 15112941]
21. Facebook. Newsroom URL: https://newsroom.fb.com/company-info/ [accessed 2018-02-13] [WebCite Cache ID 71kF8Jb4G]
22. Corbett P. ISL. 2011 Jan 03. 2011 Facebook Demographics and Statistics-Including Federal Employees and Gays in the Military URL: https://isl.co/2011/01/2011-facebook-demographics-and-statistics-including-federal-employees-and-gays-in-the-military/ [accessed 2018-08-17] [WebCite Cache ID 71kHhdSNB]
23. Fletcher D. Time. 2010 May. How Facebook Is Redefining Privacy URL: http://content.time.com/time/magazine/article/0,9171,1990798,00.html [accessed 2018-12-18] [WebCite Cache ID 74lX1Y4Zd]
24. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May 1;110(2-3):298-301 [FREE Full text] [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
25. Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. Epidemiology 2013 Jan;24(1):1-9 [FREE Full text] [doi: 10.1097/EDE.0b013e31827623b1] [Medline: 23211345]
26. Davis LL, Broome ME, Cox RP. Maximizing retention in community-based clinical trials. J Nurs Scholarsh 2002;34(1):47-53. [doi: 10.1111/j.1547-5069.2002.00047.x] [Medline: 11901967]
27. Twitter. About Direct Messages URL: https://help.twitter.com/en/using-twitter/direct-messages [accessed 2018-08-16] [WebCite Cache ID 71iZ9tB1W]
28. Debatin B, Lovejoy JP, Horn AK, Hughes BN. Facebook and online privacy: attitudes, behaviors, and unintended consequences. J Comput Mediat Commun 2009 Nov 17;15(1):83-108 [FREE Full text] [doi: 10.1111/j.1083-6101.2009.01494.x]
29. Kwon SJ, Park E, Kim KJ. What drives successful social networking services? A comparative analysis of user acceptance of Facebook and Twitter. Soc Sci J 2014 Dec 1;51(4):534-544 [FREE Full text] [doi: 10.1016/j.soscij.2014.04.005]

Edited by G Eysenbach; submitted 18.03.18; peer-reviewed by A Cyr, K Reuter; comments to author 19.04.18; revised version received 21.08.18; accepted 25.11.18; published 15.02.19.

Please cite as:
Stephenson NL, Hetherington E, Dodd S, Mathews A, Tough S
Mitigation of Participant Loss to Follow-Up Using Facebook: All Our Families Longitudinal Pregnancy Cohort
J Med Internet Res 2019;21(2):e10441
URL: https://www.jmir.org/2019/2/e10441/
doi:10.2196/10441
PMID:30767901

Original Paper

Older Adults' Perspectives on Using Digital Technology to Maintain Good Mental Health: Interactive Group Study

Jacob A Andrews[1], BA (Hons), MA, PhD; Laura JE Brown[2], BSc (Hons), PhD; Mark S Hawley[1], BSc, PhD, MIPEM, Hon FRCSLT, CSci; Arlene J Astell[3,4,5,6], BSc, PhD, CClin Psychol

[1]Centre for Assistive Technology and Connected Healthcare, School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom
[2]Manchester Academic Health Science Centre, Division of Psychology and Mental Health, University of Manchester, Manchester, United Kingdom
[3]Ontario Shores Centre for Mental Health Sciences, Whitby, ON, Canada
[4]School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom
[5]Department of Occupational Sciences & Occupational Therapy, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[6]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Mark S Hawley, BSc, PhD, MIPEM, Hon FRCSLT, CSci
Centre for Assistive Technology and Connected Healthcare
School of Health and Related Research
University of Sheffield
30 Regent Street
Sheffield, S1 4DA
United Kingdom
Phone: 44 1142220682
Email: mark.hawley@sheffield.ac.uk

Abstract

Background: A growing number of apps to support good mental health and well-being are available on digital platforms. However, very few studies have examined older adults' attitudes toward the use of these apps, despite increasing uptake of digital technologies by this demographic.

Objective: This study sought to explore older adults' perspectives on technology to support good mental health.

Methods: A total of 15 older adults aged 50 years or older, in two groups, participated in sessions to explore the use of digital technologies to support mental health. Interactive activities were designed to capture participants' immediate reactions to apps and websites designed to support mental health and to explore their experiences of using technology for these purposes in their own lives. Template analysis was used to analyze transcripts of the group discussions.

Results: Older adults were motivated to turn to technology to improve mood through mechanisms of distraction, normalization, and facilitated expression of mental states, while aiming to reduce burden on others. Perceived barriers to use included fear of consequences and the impact of low mood on readiness to engage with technology, as well as a lack of prior knowledge applicable to digital technologies. Participants were aware of websites available to support mental health, but awareness alone did not motivate use.

Conclusions: Older adults are motivated to use digital technologies to improve their mental health, but barriers remain that developers need to address for this population to access them.

(J Med Internet Res 2019;21(2):e11694) doi:10.2196/11694

KEYWORDS

mental health; older adults; technology; digital technology; Internet; apps

Introduction

Late-life mental health is often overshadowed by physical health. However, mental illness is common in older adult populations, with estimates suggesting that depression is more common than dementia in later life [1]. Prevalence of depression among individuals over the age of 65 has been suggested to range from 7.2% to as high as 49%, depending on the living situation from which the sample is taken [2]. Depression is characterized by persistent feelings of low mood, lack of motivation, reduced

enjoyment of daily activities, tiredness, and suicidality [3]. Over half of all cases of depression in older adults are first onset [4], highlighting the need for effective preventative strategies for this age group. Generalized anxiety disorder is also common, affecting up to 15% of older adults, with even higher rates of symptom incidence [5]. Anxiety has several symptoms in common with depression and can cause persistent and exaggerated worry, trembling, and panic attacks [3]. Both anxiety and depression are risk factors for suicide [6,7] and can negatively impact physical health and well-being. This is evident in hospital care, where older adults with mental health problems spend longer waiting for emergency room treatment, have longer hospital stays, and have higher rates of readmission than those without them [8]. In addition, the presence of mental health problems in older adults is correlated with increased health care costs, even after accounting for mental health-specific treatment costs [9].

In contrast, there is growing interest in efforts to promote positive mental health, well-being, and flourishing, including among the older adult population. Keyes defines mental health as "a syndrome of symptoms of positive feelings and positive functioning in life," and uses the term *flourishing* to describe states of positive mental health [10]. The New Economics Foundation developed the *Five Ways to Wellbeing*, a series of tools to help individuals to improve well-being and increase flourishing. These tools comprise five types of behavior, each of which is likely to increase positive affect, according to evidence reviewed by the foundation [11]. The report specifically mentions the benefits these behaviors can have for older adults. Huppert reviewed literature on the causes and consequences of psychological well-being and concluded that actions and attitudes were likely to have a greater effect on mental state than external circumstances [12]. These accounts suggest that individuals can have a strong influence over their own mental states by engaging in behaviors known to support good mental health. However, for individuals to take responsibility for their own mental states, awareness of their present mental state and knowledge of the behaviors that may improve it are required.

Across all age groups, the potential of digital technologies for monitoring and supporting good mental health is increasingly being recognized [13]. For example, there is increasing use of digital tools for ecological momentary assessment or experience sampling of mood [14,15], as well as recording of GPS data to monitor mental health using digital technology [16]. There is also a growing range of mobile apps to suggest and facilitate activities to maintain positive mood, for example, MoodMission [17], Happier [18], and HeadSpace [19]. Older adults are also making increased use of digital technologies [20,21], particularly for recreational activities [22] and keeping in contact with people they care about [23]. However, there has been less investigation of their use of technologies to support mental health. Those studies that have explored this area have limitations. For example, while Sauve et al [24] found a significant increase in participants' self-reported mood after playing an online educational game on the topic of well-being, they only examined one game, developed by the authors themselves, limiting generalizability to other digital tools. Similarly limited are the findings from a single case study examining online, therapist-assisted cognitive behavioral therapy, which revealed user satisfaction and positive responses to the use of technology for this purpose [25]. Given these limited results, more research is needed involving larger groups of older adults to increase the range of views expressed.

There is also a need to understand potential barriers to older adults using digital technologies for supporting mental health. For instance, many older adults perceive technology to be expensive [26,27] and often prefer the "old-fashioned" ways of doing things [28]. Prior research has also highlighted that older people can have trouble remembering how to use new technologies [27,28]. However, little is known about older adults' attitudes toward the use of digital technologies specifically to support mental health.

To address these gaps in knowledge, this study sought to use an innovative and interactive methodology to understand what motivates and prevents the use of digital technologies to support mental health among older adults. Therefore, this study was based on the Challenging Obstacles and Barriers to Assisted Living Technologies (COBALT) [29] principle of *user as expert* [30]. According to this principle, older adult users of technology are considered the most knowledgeable on the reasons for their adoption and continued use, or abandonment, of particular technologies. This approach invites older adults to share momentary experiences of using technologies presented within the session and addresses the limitations of traditional focus groups that rely on participants recalling anecdotes from memory [31], perhaps from years ago, where subtle details of particular situations may be lost.

The COBALT approach was cocreated with older adults as a means of ensuring their equal participation in technology development and evaluation. Most older adults have no experience interacting with technology developers, but their input into developing technologies and services aimed at them is essential. The COBALT project created user-centered activities specifically to hear the voices of older adults at any and all stages of the technology development process [32]. These arose in response to older adults, technology developers, and service providers who might commission or provide technologies to older adults, identifying lack of meaningful communication as a major barrier to technology uptake. The principle of *user as expert* and some of the specific activities have been further developed in the TUNGSTEN project [33]. All parts of the interactive sessions are audio- and videotaped, meaning that views participants may only express to one or two of their peers are also captured, reducing the influence of large-group dynamics on results. In this study, each facilitated activity used a set of materials or technologies, developed to create a comfortable environment where older adults could speak freely as experts about the topic of interest, which in this case was technology and late-life mental health.

Methods

Design

This study used a qualitative approach to elicit older adults' views and experiences. It sought to explore why older adults might or might not be motivated to engage with digital technologies to support their mental health. The aim was thus to develop new insights and understanding, which are key aims of a qualitative approach. The COBALT tool kit uses techniques that encourage interaction with technology within sessions, so that immediate reactions and experiences can be captured, and that respect the knowledge held by older adults regarding their motivations and experiences of using technologies.

Participants

In total, 15 participants were recruited for the study: one group of 7 participants and one group of 8, which is in line with COBALT methodology guidelines [34]. Eligibility criteria were as follows: being 50 years of age or over, able to read (with or without glasses), able to hear (with or without a hearing aid), and having no diagnosis of cognitive impairment (self-certified). As the aim of the study was to understand views on apps and websites to maintain good mental health rather than treat mental illness, people experiencing mental illness at the time of recruitment were ineligible for the study. While several organizations describe older adulthood as beginning at age 65, we recruited adults aged 50 and older, in line with the Mental Health Foundation's definition of later life as the period beginning at age 50 [35]. While this meant the age range was large, it ensured that views of people likely to have experienced using digital technology during their working lives were included in the study. This was to ensure the research would remain relevant to those reaching retirement age over the next 15-20 years. No upper age limit was set. Participants were recruited from two volunteer groups aimed at people aged 50 years and older and participants were advised that they did not need to have extensive knowledge of technology to take part.

Ethics and Payment

The study was approved by the Research Ethics Committee of the School of Health and Related Research at the University of Sheffield (approval number 003140). All participants provided written informed consent before taking part. No payment was provided to the participants, although return taxi fares to the session venue were offered and refreshments were provided in all sessions.

Materials

Materials associated with each of the interactive activities are summarized in Table 1. Materials included printed and laminated cards as well as apps presented on two iPad Air devices with 9.7-inch touch screens, one Samsung Galaxy Tab 2 tablet computer with 10.1-inch touch screen, and one Asus EeeTop PC with 15-inch touch screen. The latter was used to demonstrate the Novel Assessment of Nutrition and Ageing (NANA) method, a research tool, which is pictured in Figure 1. In addition, at the beginning of their first session, participants completed a demographics questionnaire, requesting information on age, education, experience using different types of everyday technology, and self-perceived health status. Sessions were videotaped using a Sony HD Handycam on a tripod and were audiotaped using three Olympus digital audio recorders.

Procedure

Four interactive group sessions were conducted with two groups of older adults. Each group attended two sessions of two hours in length. The sessions were conducted in the Home Laboratory in the Centre for Assistive Technology and Connected Healthcare at the University of Sheffield, United Kingdom. The Home Laboratory is a multipurpose space laid out as a furnished, one-bedroom apartment. This setting allows research participants to try out new technologies in a home-like environment. The facilitated sessions were comprised of seven activities in total—four in Session 1 and three in Session 2. In addition to activities where participants worked in pairs or small groups, whole-group discussions allowed participants to reflect on their experiences in the sessions.

Data Analysis

All sessions were audio- and videotaped and subsequently transcribed verbatim by the first author (JAA). All personal identifiable data were removed at the stage of transcription. Transcripts were analyzed using template analysis [36]. Applying this form of analysis involved a number of steps, which led to the development of a final template of themes. First, a small number of a priori themes were developed, which formed the basis of the template. Initial themes for this study related to attitudes toward technology and mental health in general, as well as attitudes toward use of technology for the specific purpose of supporting mental health (see Textbox 1).

After familiarization, the data were coded using NVivo 11 PC software (QSR International), in line with the principles of template analysis. This was an iterative process that involved reading each transcript and identifying meaningful chunks of text that demonstrated participants' views. Once identified, chunks were labeled in one of three ways: (1) they were assigned directly to one of the a priori themes; (2) they were assigned to new codes that were related to one of the themes; or (3) where the view represented was not related to any pre-existing theme, they were assigned a *floating* code. These were combined into new themes where appropriate. Themes were iteratively renamed, reorganized, and removed. This process was conducted by JAA. In order to improve the reliability of the analysis, LJEB and AJA also reviewed some of the data and took part in discussions throughout the process.

Table 1. Activities undertaken in the study.

Activity breakdown	Description	Materials	Purpose
Session 1			
1. Icebreaker (10 minutes)	Participants introduced themselves and said something that cheers them up when they are feeling down.	N/A[a]	This activity aimed to make participants feel comfortable speaking in a group and to begin thinking about ideas associated with mood.
2. Activity sorting (25 minutes)	Participants worked in pairs to decide whether or not they would use technology for a range of tasks, including some related to mental health and some not. They also considered what type of technology they might use to achieve these tasks.	Four packs of 10 laminated cards were used. Each card featured an activity that could be done with or without technology (eg, send a photo or research a health problem). These were written in size 55 text.	This activity aimed to begin to elicit attitudes toward use of technology to support mental health.
Break (15 minutes)	N/A	N/A	N/A
3. Vignettes (25 minutes)	Participants worked in small groups to read and discuss four vignettes describing the experience of anxious feelings and low mood.	Four packs of four vignette cards were used. These cards were written in size 14 text.	This activity aimed to engage participants in considering real-world situations in which technology might be used to support mental health.
4. App interaction (25 minutes)	In small groups, participants tried multiple apps on a mobile phone (Samsung Galaxy S3) and on tablet computers (iPad Air and Samsung Galaxy Tab 2). Participants discussed how useful these were and decided if they (the participants) would use these in their own homes.	Apps selected were WellMind (Dudley and Walsall NHS[b] Mental Health Partnership), Five Ways to Wellbeing (Somerset Public Health), and MindShift (AnxietyBC).	This activity aimed to explore motivators and barriers to the use of apps and websites for the purpose of supporting mental health.
Session 2			
5. Show and tell (20 minutes)	In an activity inspired by the COBALT[c] study [29], participants presented a piece of technology they loved and a piece of technology they had abandoned, along with reasons.	Participants' own self-bought and self-chosen technologies were used.	This was a warm-up activity, aimed to give participants confidence and allow them to feel like experts on the topics at hand [29].
6. App interface evaluation (70 minutes, including a 15-minute break)	In small groups, participants evaluated four different ways of self-reporting mood using different digital technologies in turn. Then, following a 15-minute break, the apps were discussed in the whole group.	Mr Mood and Pacifica were presented on Apple iPad Air devices; Five Ways to Wellbeing was presented on a Samsung Galaxy Tab 2; and NANA[d] Mood was presented on a 15-inch touch screen Asus EeeTop PC.	This activity aimed to explore usability of apps for the purpose of supporting mental health.
7. Imagining a future app (30 minutes)	Participants were asked to consider different ways an app might respond to low mood scores. After discussing different ideas in pairs or threes, ideas were discussed as a group. All discussions were recorded.	A flip chart and pens were used.	This activity aimed to allow participants to consider how data might be used and to understand how this may affect their motivation to use mood-reporting technology.

[a]N/A: not applicable.

[b]NHS: National Health Service.

[c]COBALT: Challenging Obstacles and Barriers to Assisted Living Technologies.

[d]NANA: Novel Assessment of Nutrition and Ageing.

Figure 1. Older adult using the Novel Assessment of Nutrition and Ageing (NANA) home system on an Asus EeeTop PC.

Textbox 1. A priori themes chosen before beginning the coding process.

A priori themes:

- Attitudes toward technology in general
- Attitudes toward mental health conditions
- Attitudes toward technology for supporting mental health
- Motivators to use
- Barriers to use
- Usability

Results

Participant Characteristics

The participants' ages ranged from 52 to 88 years and the mean age was 66 years (SD 8.6). The majority of the participants were female (12/15, 80%). Most had attended further or higher education (11/15, 73%). The majority of the participants (11/15, 73%) reported that they regularly used the Internet. Almost half of the participants reported that they used technology to manage a health condition (7/15, 47%), although no data were available on what type of technology they used or what condition they managed. All participants rated their health as fair (8/15, 53%) or good (7/15, 47%).

Results of the Template Analysis

The final template comprised three themes, each with a number of subthemes (see Table 2). Each of these themes is described in detail in the following sections.

Table 2. Final template including themes and subthemes after coding.

Themes and subthemes	Quotes[a] (n)
1. Motivators to the use of technology to support mental health	
Using technology is preferential to "bothering" people when experiencing low mood	6
Keeping a record of mood using apps and websites can facilitate self-awareness	3
Playing music using technology can be beneficial to mood	11
Games within apps can be used as a distraction from problems	6
2. Barriers to the use of technology to support mental health	
Low mood may affect readiness to engage with digital technologies	4
Fear of consequences may affect readiness to engage with technology	4
Self-diagnosis using websites is problematic	9
Technology is inferior to humans	11
Older adults have some difficulties with usability of apps and websites irrespective of the mental health content	9
3. Awareness of technology to support mental health	
Participants are aware of National Health Service websites as a source of information and resources	4
Participants are aware of online meditation and mindfulness resources	5
Awareness alone is not enough for successful use	6

[a]Number of quotes coded to each theme and subtheme. Coding was completed using NVivo 11 (QSR International).

Theme 1: Motivators to the Use of Technology to Support Mental Health

Participants described several reasons they may be motivated to use technology to support their mental health. First, participants described feeling reticent to discuss if they were experiencing low mood with friends or family members for fear of bringing them down, too. Using technology in place of speaking to a person was seen as a potential way to prevent "bothering" others and thus to be self-reliant.

> *I'd be worried about passing misery on, you know what I mean.* [Participant 11, female, age 66]

> *If your computer was set up that way, if you spoke to it and said, "I'm in a low mood today, what should I do?" ...because the computer's programmed to s...well like one of them Android things [sic], just to say, "well I suggest you go for a walk" or whatever it churns out, really. I think that's good. Cos [because] you don't always want to bother somebody, like these ladies have said, you don't.* [Participant 15, female, age 55]

Loneliness was a particular source of concern for participants. Many recognized that loneliness was detrimental to health and mood. Participants discussed how they used technology to alleviate loneliness.

> *I will switch the radio on sometimes, but the visual contact of the TV makes me feel that I'm not on my own... It's company, you see people.* [Participant 15, female, age 55]

Participants mentioned that they were motivated to use technology to play music because it helped to improve their mood. Participants used a wide variety of different technologies to source music, including YouTube [37].

> *If I want a lift, an uplift, I go and put rock and roll on, or summat [something] a bit lively.* [Participant 14, male, age 75]

> *I put music on. Put YouTube on and type in what I fancy listening to.* [Participant 10, female, age 68]

> *Yeah, I like the YouTube [sic] for all sorts. Er, but music is, does relax me.* [Participant 9, female, age 69]

Apps designed to support mental health were considered helpful in normalizing the experience of poor mental health. After testing apps that described commonly experienced symptoms of depression, anxiety, and stress, participants highlighted how reading these accounts provided them with a sense of normality.

> *It makes you feel normal, because you're seeing it written down.* [Participant 12, female, age 56]

Video games were also seen as helpful, as a way to distance oneself from one's problems and improve current mental state through distraction.

> *It's if like, summat's [something has] happened or you know, so it will take your mind if you're concentrating on one thing. Just gives you that bit of respite.* [Participant 14, male, age 65]

Participants recognized these benefits in a game within an app entitled WellMind that was explored during the study session.

> *Oh there's playing a game with a kind of snake and an apple, you have to move a bar and hit the apple and things...I suppose it's a kind of concentration of the mind.* [Participant 12, female, age 56]

Theme 2: Barriers to the Use of Technology to Support Mental Health

Participants also recognized a number of barriers to the use of digital technology to support mental health. First, participants suggested that low mood was likely to affect readiness to engage with digital technologies.

> *I do wonder if you could be so low, that you couldn't be bothered to go on a computer.* [Participant 11, female, age 66]

Participants also commented that the fear of the consequences of using technology to support mental health may also prevent them from engaging with mood-reporting tools, such as apps, to support mental health.

> *Where the fear is of something unpleasant happening to you, like you're being taken, taken into an institution, then you're not going to admit any symptoms that you might think could be interpreted that way.* [Participant 8, female, age 68]

Some participants mentioned that using technology to self-diagnose could be problematic, since it could result in patients mistakenly believing they had a rare or extreme medical condition.

> *To research online a specific problem, you know, say diabetes, is ok, but to research and put in what problems you've got that come up with some outlandish thing is wrong. I don't agree with that. The doctor tells you that.* [Participant 6, male, age 69]

As well as these concerns, there was a sense that human contact was superior to technology when it comes to mental health and that dealing with people was preferable to using technology.

> *If you really feel down and you really need some help, can a computer...don't you think physical contact or speaking to someone is far more important?* [Participant 13, female, age 60]

Some technology usability issues were discussed, which, while not specific to apps to support good mental health, nonetheless created barriers to successful engagement. Many of the apps evaluated by participants assumed a certain level of prior knowledge on the part of their users. For example, some apps use a cross (X) symbol on buttons to close a text-entry box, while others used a swipe action to move between options. For some participants, these assumptions caused difficulties, as they did not always have the required prior knowledge, skills, or experience.

> *If you push that cross does that mean it cancels it out? Yes. Right. You do it.* [Participant 12, female, age 56]
>
> *No I think it, does it, oh.* [Participant 13, female, age 60]

Dexterity was also an issue raised by several participants. One of the apps tested by participants was Pacifica. This app featured a circular slider for users to report on how they were feeling. Participants suggested this action would be particularly difficult for someone who had had a stroke.

> *I think, if they've come out of hospital and had a stroke, that turning one [ie, Pacifica] might be quite difficult for them.* [Participant 15, female, age 55]

They also suggested that poor dexterity may be a barrier to the use of games within apps to support mental health.

> *I think that you've...that snake game, your hands have to be quite good.* [Participant 12, female, age 56]
>
> *Yeah, dexterous.* [Participant 11, female, age 66]
>
> *Yeah, your dexterity has to be pretty good for that one.* [Participant 12, female, age 56]

Further, usability-related barriers found in this study included difficulties reading small fonts; for example, Participant 15 stated, "The print's too small, can you make it bigger?" There was also a sense that using technology is effortful and time-consuming; for example, Participant 2 stated, "When you go on a computer, you've got to log on and all that blah, blah, blah." Some participants also mentioned that they experienced fear when completing certain activities on digital platforms; for example, Participant 13 stated, "I don't know why I'm afraid of downloading apps."

Theme 3: Awareness of Technology to Support Mental Health

Participants demonstrated an awareness of multiple uses of digital technologies to support mental health. These included online doctor's appointment booking systems, National Health Service (NHS) and other health websites for researching health conditions, meditation apps, and YouTube videos for relaxation. One participant expressed confidence using the NHS website and saw the potential benefits of using it. Here, she reacts to a vignette in which the person described is imagined to be experiencing symptoms of depression.

> *I suppose for this one, this red one, you could like, look up a...NHS site, Moodzone, they do audio tapes that, they do tapes that actually give you some inspiration towards feeling better.* [Participant 13, female, age 60]

Many participants referred to CDs, tapes, and online resources for mindfulness, though one participant was unsure about digital delivery.

> *How does mindfulness work online?* [Participant 1, female, age 66]

Another participant found it difficult to use such resources alone.

> *I've got those tapes that they've given me to teach me how to relax. I find that, I can't do it on me [sic] own. But I can do it say if you was [sic] with me, and we just, meditated together, I could do that. But on my own, I'm finding, "oh I haven't washed the pots, I haven't done this, I haven't done that," you know.* [Participant 15, female, age 55]

Thus, although participants showed an awareness of various ways in which digital technologies can be used to support mental health, awareness alone may not be enough for participants to start using such tools successfully. Indeed, not everyone was convinced there is a role for technology in certain situations.

I don't know how technology would help. [Participant 9, female, age 69]

Discussion

Principal Findings

This study used the COBALT principle of user as expert to understand the motivators and barriers to older adults' use of technology to support mental health. New findings presented in this work include that self-reliance, averting loneliness, normalization, facilitated expression, and improving mood can be motivators for older adults to use technology to support their mental health. Some apps and websites included in the study were found to offer functions that assisted users in understanding their mental states and in facilitating behaviors that would promote well-being, flourishing, and mental health. The study has also illuminated possible reasons older adults might avoid using technology to support mental health, including low mood itself, fear of consequences, and a preference for human contact.

The participants expressed an interest in using technology to alleviate low mood, for example, by listening to music on YouTube or playing games within apps. They expressed how playing games, such as that within the app WellMind, could be a helpful distraction from unwanted thoughts and may help to concentrate the mind. Prior research has highlighted that technology can provide older adults with opportunities for enjoyment and fun [22]. This study extends this to the added benefit of alleviating low mood and providing distraction. Older adult participants also highlighted how the framework provided by mood-reporting tools within apps could be helpful for exploring their own feelings and being able to express these. Furthermore, reading descriptions of the experience of symptoms of poor mental health was felt to be normalizing. Our findings suggest the inclusion of elements such as the ability to play music, play games, read about symptoms, and report on mood within an app to support mental health is likely to be motivating and beneficial for older adult users and could promote flourishing, as described by Keyes [10]. However, further research would be required to ascertain the true level of benefit each of these elements could provide.

Findings here reflect how older adults may turn to technology to manage symptoms of low mood and alleviate feelings of loneliness by themselves, without reaching out to friends or family. Participants described how using apps to record their mood and receive activity suggestions may be preferable to speaking with others when experiencing low mood, to avoid being seen as a burden. Peek et al [28] also found that older adults were keen to avoid being a burden and used technology to this end. The determination of older adults to be self-reliant can thus be seen as a motivator for them to engage with technology to support their mental health. However, some participants felt that, if affected by a mental health condition, they would prefer to speak with a human about their experience rather than use technology. Views on the superiority of humans over technology thus contrasted with views around the benefits of technology when an individual did not feel confident speaking with a person. These differing views on the best way to support mental health demonstrate there is no "one-size-fits-all" in the use of digital solutions to support mental health and that there are situations in which technology can be seen as having advantages over human contact. Prior work has also highlighted both the preference of older adults for human management of health care [38] and, contrastingly, the benefits of using technology to support health care in order to avoid being a burden on others [28]. This tension could be resolved in future applications of digital technology for the management of older adults' health and well-being through a decision aid to help decide whether contact should be made with a health professional and/or a formal or informal carer.

The participants identified a number of barriers to using technology to support mental health. Discussions revealed how symptoms of poor mental health may affect readiness to engage with technology. Comments included the fact that feeling depressed was likely to magnify difficulties with usability of digital technologies. The Diagnostic and Statistical Manual of Mental Disorders lists lack of motivation and reduced pleasure in daily activities as symptoms of major depression [3] and participants' comments in this regard are thus in line with these symptoms of depression. Further research would be required to understand this phenomenon more fully, for example, exploring what severity of depression or anxiety causes disengagement with technology and in what populations this is most common. The use of prompts, either originating from the devices themselves or from family members or health care practitioners, could also be explored as a potential method to mitigate this disengagement.

Some usability issues with the apps and websites were identified. For example, the use of interactive COBALT activities revealed that some participants struggled to understand the meaning of relatively standard symbols and dialogues within apps and websites. This suggests that apps and websites must address accessibility and universal design features to maximize their usability and reach to all potential beneficiaries, rather than assuming prior knowledge. Similar findings were reported by Eisma et al [39], who demonstrated that scrollbars were not familiar to older adults, and Grindrod et al [40] who discussed how "Cancel" buttons, scrolling functionality, autocorrect, sample text in grey fonts, and peripheral buttons can confuse older adult users with less experience using technology. Vaportzis et al [41] reported on older adults' struggle to use digital technologies where physical health complaints, such as poor eyesight or limited dexterity, affected interaction with devices. Our findings support these points and further demonstrate the need for lower dexterity to be kept in mind when developers are designing apps for older adults. Thus, to ensure older adults feel comfortable and confident using digital technologies to support their mental health, apps and websites should use intuitive layouts with easy-to-use function buttons and clear explanations, while avoiding the use of technical jargon.

Another perceived barrier to using technology to support mental health was the perception that using digital technology might result in automated decisions about treatment being taken. When discussing the use of digital technology to report on mood, participants suggested that fear of being institutionalized could affect honesty in self-reports. Prior work has indicated the

benefits and psychological attachments older adults have to ageing in their own homes [42,43]; the findings in our study indicate that the possibility of losing this independence could provoke anxiety for older adults using digital technology to monitor mental health. This finding could inform the development of documents and guidance provided alongside apps and websites to address users' fears about the implications of using the technology. For example, these materials might explain that these technologies are designed to be informative rather than diagnostic and that their use would not lead to any automatic decisions about their future care or treatment being made.

Limitations

Although this study has produced novel insights into older adults' attitudes toward the use of technology to support mental health, it does have some limitations. First, there were few participants with very low levels of education or very limited experience of technology, meaning results may reflect a greater awareness of, or ability with, technology among participants than is representative of the population at large. Second, the majority of participants in this study were female, meaning views of older males may be underrepresented. Third, while the inclusion of participants aged 50 and over enabled us to gain a more diverse range of perspectives, it also means that not all points may be relevant to all age groups. Future studies that examine the relevance of each theme to more homogenous subgroups of older people would therefore be useful for learning more about the different types of potential users of mental health technologies.

Conclusions

This study has revealed that older adults are motivated to use digital technology to support their mental health, in particular to promote self-reliance, avert loneliness, and improve mood. However, low mood, fear of consequences, and a preference for human contact may prevent successful engagement with these technologies. Further work is needed to understand how the use of music and games within apps and websites might help to address symptoms of mental illness in older adults and to understand how different states of mind, such as low mood, might affect older adults' readiness to engage with technology. Developers of apps and websites designed to support mental health that could be used by older adults should be mindful that technical jargon and commonly used symbols may not be understandable to all older adults. They should also be aware that guidance provided with such tools could be helpful in reassuring users that using these tools will not reduce human contact with health professionals or result in adverse consequences.

Acknowledgments

This work was funded by a PhD scholarship as part of the Technology for Healthy Ageing and Wellbeing network and was awarded to JAA by the University of Sheffield.

Conflicts of Interest

None declared.

References

1. Allan CE, Valkanova V, Ebmeier KP. Depression in older people is underdiagnosed. Practitioner 2014 May;258(1771):19-22, 2-3. [Medline: 25065018]
2. Djernes JK. Prevalence and predictors of depression in populations of elderly: A review. Acta Psychiatr Scand 2006 May;113(5):372-387. [doi: 10.1111/j.1600-0447.2006.00770.x] [Medline: 16603029]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th edition (DSM-5). Washington, DC: American Psychiatric Publishing; 2013.
4. Fiske A, Wetherell JL, Gatz M. Depression in older adults. Annu Rev Clin Psychol 2009;5:363-389 [FREE Full text] [doi: 10.1146/annurev.clinpsy.032408.153621] [Medline: 19327033]
5. Bryant C, Jackson H, Ames D. The prevalence of anxiety in older adults: Methodological issues and a review of the literature. J Affect Disord 2008 Aug;109(3):233-250. [doi: 10.1016/j.jad.2007.11.008] [Medline: 18155775]
6. Conwell Y, Duberstein PR, Caine ED. Risk factors for suicide in later life. Biol Psychiatry 2002 Aug 01;52(3):193-204. [Medline: 12182926]
7. O'Connell H, Chin A, Cunningham C, Lawlor B. Recent developments: Suicide in older people. BMJ 2004 Oct 16;329(7471):895-899 [FREE Full text] [doi: 10.1136/bmj.329.7471.895] [Medline: 15485967]
8. Adams LY, Koop P, Quan H, Norris C. A population-based comparison of the use of acute healthcare services by older adults with and without mental illness diagnoses. J Psychiatr Ment Health Nurs 2015 Feb;22(1):39-46. [doi: 10.1111/jpm.12169] [Medline: 25430792]
9. Unützer J, Schoenbaum M, Katon WJ, Fan M, Pincus HA, Hogan D, et al. Healthcare costs associated with depression in medically ill fee-for-service medicare participants. J Am Geriatr Soc 2009 Mar;57(3):506-510. [doi: 10.1111/j.1532-5415.2008.02134.x] [Medline: 19175438]
10. Keyes C. The mental health continuum: From languishing to flourishing in life. J Health Soc Behav 2002 Jun;43(2):207-222. [Medline: 12096700]

11. Aked J, Marks N, Cordon C, Thompson S. Five Ways to Wellbeing. London, UK: The New Economics Foundation URL: https://neweconomics.org/uploads/files/8984c5089d5c2285ee_t4m6bhqq5.pdf [accessed 2018-10-30] [WebCite Cache ID 73YJ07BFP]
12. Huppert F. Psychological Well-being: Evidence regarding its causes and consequences. Appl Psychol Health Well Being 2009;1(2):137-164. [doi: 10.1111/j.1758-0854.2009.01008.x]
13. Cotton R, Irwin J, Wilkins A, Young C. The Future's Digital: Mental Health and Technology. London, UK: Mental Health Network, NHS Confederation; 2014. URL: http://www.nhsconfed.org/-/media/Confederation/Files/Publications/Documents/the-futures-digital.pdf [accessed 2018-07-23] [WebCite Cache ID 71827Cc9Z]
14. Brown LJE, Astell AJ. Assessing mood in older adults: A conceptual review of methods and approaches. Int Psychogeriatr 2012 Aug;24(8):1197-1206. [doi: 10.1017/S1041610212000075] [Medline: 22340813]
15. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
16. Mohr DC, Zhang M, Schueller SM. Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 Dec 08;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
17. MoodMission. URL: http://moodmission.com [accessed 2018-12-18] [WebCite Cache ID 74lSvvCHF]
18. Happier. URL: https://www.happier.com/app/ [accessed 2018-12-18] [WebCite Cache ID 74lTj3Qpi]
19. Headspace. URL: https://www.headspace.com/ [accessed 2018-12-18] [WebCite Cache ID 74lTyhHNr]
20. Adults' Media Use and Attitudes: Report 2015. London, UK: Ofcom; 2015 May. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0014/82112/2015_adults_media_use_and_attitudes_report.pdf [accessed 2018-07-15] [WebCite Cache ID 7183t9rJe]
21. Adults' Media Use and Attitudes: Report 2017. London, UK: Ofcom; 2017 Jun. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/102755/adults-media-use-attitudes-2017.pdf [accessed 2018-07-15] [WebCite Cache ID 7183ot9fS]
22. Astell A. Technology and fun for a happy old age. In: Sixsmith A, Gutman G, editors. Technologies for Active Aging. New York, NY: Springer Science+Business Media; 2013:169-187.
23. Cotten SR, Anderson WA, McCullough BM. Impact of Internet use on loneliness and contact with others among older adults: Cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
24. Sauvé L, Renaud L, Kaufman D, Duplàa E. Can digital games help seniors improve their quality of life? In: Proceedings of the 8th International Conference on Computer Supported Education, CSEDU 2016. Cham, Switzerland: Springer Nature; 2017 Presented at: 8th International Conference on Computer Supported Education, CSEDU 2016; April 21-23, 2016; Rome, Italy p. 179-192.
25. Pugh N, Hadjistavropoulos H, Klein B, Austin D. A case study illustrating therapist-assisted Internet cognitive behavior therapy for depression. Cogn Behav Pract 2014 Feb;21(1):64-77 [FREE Full text] [doi: 10.1016/j.cbpra.2013.08.002]
26. Greenhalgh T, Wherton J, Sugarhood P, Hinder S, Procter R, Stones R. What matters to older people with assisted living needs? A phenomenological analysis of the use and non-use of telehealth and telecare. Soc Sci Med 2013 Sep;93:86-94 [FREE Full text] [doi: 10.1016/j.socscimed.2013.05.036] [Medline: 23906125]
27. Chen K, Chan AH. Use or non-use of gerontechnology: A qualitative study. Int J Environ Res Public Health 2013 Sep 30;10(10):4645-4666 [FREE Full text] [doi: 10.3390/ijerph10104645] [Medline: 24084674]
28. Peek STM, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016;62(2):226-237 [FREE Full text] [doi: 10.1159/000430949] [Medline: 26044243]
29. Astell AJ, Williams EA, Hwang F, Norton M, Adlam T, Lethbridge K. COBALT: Challenging obstacles and barriers to assistive living technologies. Gerontechnology 2012;11(2):151. [doi: 10.4017/gt.2012.11.02.426.00]
30. Astell AJ, Andrews J, Bennion M, Clayton D, Dove E, Ellis M, et al. COBALT, TUNGSTEN, and THAW: New tools for engaging older adults as technology experts. Gerontechnology 2018 Apr 24;17(suppl):76s. [doi: 10.4017/gt.2018.17.s.076.00]
31. Kitzinger J. Qualitative research: Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302 [FREE Full text] [Medline: 7633241]
32. COBALT Project. URL: https://cobaltproject.net/about/ [accessed 2018-12-18] [WebCite Cache ID 74lUdWUTO]
33. TUNGSTEN Training. URL: http://tungsten-training.com/ [accessed 2018-12-18] [WebCite Cache ID 74lUq2UGu]
34. Astell AJ, Hwang F, Williams EA, Archer L, Harney-Levine S, Wright D, et al. Well-being and HCI in later life: What matters? In: Proceedings of the 18th International Conference on Human Interface and the Management of Information. Cham, Switzerland: Springer International Publishing; 2016 Presented at: 18th International Conference on Human Interface and the Management of Information; July 17-22, 2016; Toronto, Ontario p. 445-453 URL: https://doi.org/10.1007/978-3-319-40397-7_42 [doi: 10.1007/978-3-319-40397-7_42]
35. Mental Health Foundation. 2018. Mental health statistics: Older people URL: https://www.mentalhealth.org.uk/statistics/mental-health-statistics-older-people [accessed 2018-10-30] [WebCite Cache ID 73YJndF54]

36. King N. Doing template analysis. In: Symon G, Cassell C, editors. Qualitative Organizational Research: Core Methods and Current Challenges. London, UK: SAGE Publications; 2012:426-450.
37. YouTube. URL: https://www.youtube.com/ [accessed 2018-12-19] [WebCite Cache ID 74mhfyZmt]
38. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
39. Eisma R, Dickinson A, Goodman J, Syme A, Tiwari L, Newell AF. Early user involvement in the development of information technology-related products for older people. Univers Access Inf Soc 2004 Jun;3(2):131-140 [FREE Full text] [doi: 10.1007/s10209-004-0092-z]
40. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: A usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
41. Vaportzis E, Clausen M, Gow A. Older adults perceptions of technology and barriers to interacting with tablet computers: A focus group study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
42. Wiles JL, Leibing A, Guberman N, Reeve J, Allen RES. The meaning of "aging in place" to older people. Gerontologist 2012 Jun;52(3):357-366. [doi: 10.1093/geront/gnr098] [Medline: 21983126]
43. Gott M, Seymour J, Bellamy G, Clark D, Ahmedzai S. Older people's views about home as a place of care at the end of life. Palliat Med 2004 Jul;18(5):460-467. [doi: 10.1191/0269216304pm889oa] [Medline: 15332424]

Abbreviations

COBALT: Challenging Obstacles and Barriers to Assisted Living Technologies
NANA: Novel Assessment of Nutrition and Ageing
NHS: National Health Service

Edited by G Eysenbach; submitted 30.07.18; peer-reviewed by M Duncan, HY Huang, B Renn; comments to author 11.10.18; revised version received 07.11.18; accepted 26.11.18; published 13.02.19.

Please cite as:
Andrews JA, Brown LJE, Hawley MS, Astell AJ
Older Adults' Perspectives on Using Digital Technology to Maintain Good Mental Health: Interactive Group Study
J Med Internet Res 2019;21(2):e11694
URL: http://www.jmir.org/2019/2/e11694/
doi:10.2196/11694
PMID:30758292

Review

Design and Delivery Features That May Improve the Use of Internet-Based Cognitive Behavioral Therapy for Children and Adolescents With Anxiety: A Realist Literature Synthesis With a Persuasive Systems Design Perspective

Ashley D Radomski[1], MSc; Lori Wozney[2], PhD; Patrick McGrath[3,4,5], PhD; Anna Huguet[6], PhD; Lisa Hartling[1], PhD; Michele P Dyson[1], PhD; Kathryn Bennett[7], PhD; Amanda S Newton[1], PhD

[1]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada
[2]Centre for Research in Family Health, IWK Health Centre, Halifax, NS, Canada
[3]Department of Psychology, Dalhousie University, Halifax, NS, Canada
[4]Department of Pediatrics, Dalhousie University, Halifax, NS, Canada
[5]Deparment of Psychiatry, Dalhousie University, Halifax, NS, Canada
[6]Department of Community of Health and Epidemiology, Dalhousie University, Halifax, NS, Canada
[7]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Amanda S Newton, PhD
Department of Pediatrics
University of Alberta
3-526 Edmonton Clinic Health Academy
11405 - 87 Avenue
Edmonton, AB, T6G 1C9
Canada
Phone: 1 780 248 5581
Email: mandi.newton@ualberta.ca

Abstract

Background: Internet-based cognitive behavioral therapy (iCBT) is a *persuasive system* as its design combines therapeutic content, technological features, and interactions between the user and the program to reduce anxiety for children and adolescents. How iCBT is designed and delivered differs across programs. Although iCBT is considered an effective approach for treating child and adolescent anxiety, rates of program use (eg, module completion) are highly variable for reasons that are not clear. As the extent to which users complete a program can impact anxiety outcomes, understanding what iCBT design and delivery features improve program use is critical for optimizing treatment effects.

Objective: The objectives of this study were to use a realist synthesis approach to explore the design and delivery features of iCBT for children and adolescents with anxiety as described in the literature and to examine their relationship to program use outcomes.

Methods: A search of published and gray literature was conducted up to November 2017. Prespecified inclusion criteria identified research studies, study protocols, and program websites on iCBT for child and adolescent anxiety. Literature was critically appraised for relevance and methodological rigor. The persuasive systems design (PSD) model, a comprehensive framework for designing and evaluating persuasive systems, was used to guide data extraction. iCBT program features were grouped under 4 PSD categories—Primary task support, Dialogue support, System credibility support, and Social support. iCBT design (PSD Mechanisms) and delivery features (Context of use) were linked to program use (Outcomes) using meta-ethnographic methods; these relationships were described as Context-Mechanism-Outcome configurations. For our configurations, we identified key PSD features and delivery contexts that generated moderate-to-high program use based on moderate-to-high quality evidence found across multiple iCBT programs.

Results: A total of 44 documents detailing 10 iCBT programs were included. Seven iCBT programs had at least one document that scored high for relevance; most studies were of moderate-to-high methodological rigor. We developed 5 configurations that highlighted 8 PSD features (Tailoring, Personalization [Primary task supports]; Rewards, Reminders, Social role [Dialogue

supports]; and Trustworthiness, Expertise, Authority [System credibility supports]) associated with moderate-to-high program use. Important features of delivery Context were adjunct support (a face-to-face, Web- or email-based communications component) and whether programs targeted the prevention or treatment of anxiety. Incorporating multiple PSD features may have additive or synergistic effects on program use.

Conclusions: The Context-Mechanism-Outcome configurations we developed suggest that, when delivered with adjunct support, certain PSD features contribute to moderate-to-high use of iCBT prevention and treatment programs for children and adolescents with anxiety. Standardization of the definition and measurement of program use, formal testing of individual and combined PSD features, and use of real-world design and testing methods are important next steps to improving how we develop and deliver increasingly useful treatments to target users.

(J Med Internet Res 2019;21(2):e11128) doi:10.2196/11128

KEYWORDS

internet; cognitive behavioral therapy; computer-assisted therapy; persuasive communication; anxiety; children; adolescents; review; adherence

Introduction

Background

Cognitive behavioral therapy (CBT) is recommended as a first-line treatment for children and adolescents with anxiety [1-4]. Trained mental health professionals have traditionally delivered CBT, but there is an increasing interest in the internet as a delivery platform to circumvent the multiple barriers to receiving in-person treatment. These barriers can include direct and incidental costs to families, lack of trained deliverers, and inconvenient service times and locations [5]. Internet-based CBT (iCBT) is also proposed to preserve adolescent autonomy, appeal to user preferences, reduce health care system costs, and improve the time it takes to receive treatment [6-8]. iCBT is recognized as an important treatment option [9-11] to meet the increasing demands of a population where anxiety is often undetected and untreated, but with whom early access to care can improve long-term outcomes [3,12-14].

iCBT uses technological features (ie, multimedia and email) to deliver treatment content through the Web or a software application to generate interactions between the user and the program [15]. In this way, iCBT aims to reinforce, change, or shape attitudinal or behavioral health outcomes and aligns with the concept of a *persuasive system* [16-18]. Although recent efforts have been made to provide guidance on the design and delivery features of iCBT [19], considerable differences exist across programs both in terms of their features and the health outcomes they produce. Recent systematic reviews and meta-analyses have found a range of iCBT programs to be effective at reducing anxiety in children and adolescents [9,10,20,21]; however, poor and highly variable rates of completion can be found across programs (up to 50% of participants not reaching the end of a program) [5,7,10,21,22]. The term *program use* captures the various, typically objective, outcomes used across studies (ie, adherence, compliance, and number of program activities or homework completed) that describe the extent to which users interact with a program.

Understanding the factors that influence iCBT program use is important as there have been indications that greater program use is associated with better outcomes [23,24]. Studies of iCBT in children and adolescents with anxiety [25,26] have found that certain participant demographics (eg, gender, age, location, and anxiety severity) and delivery features (eg, parental or therapist support) relate to or predict program use. However, the relationship between technological design features of a program, the ability of those features to enhance the *persuasiveness* of a program, and actual iCBT program use by children and adolescents with anxiety, has received minimal attention in the literature.

Objectives

Recognizing iCBT as a persuasive system, we conducted a realist synthesis to examine the technological design and program delivery features of iCBT for children and adolescents with anxiety in order to document their potential relation to persuading program use. The realist synthesis approach provided a framework to answer 2 main questions: (1) what design and delivery components (technological features, treatment content, and interactions) are described for iCBT programs for children and adolescents with anxiety? and (2) what components may explain reported program use outcomes?

Methods

Study Design

This realist synthesis was conducted using steps recommended by Pawson and Tilley [27,28] and is reported in accordance with the Realist and Meta-narrative Evidence Synthesis: Evolving Standards (RAMESES II) [29]. Realist synthesis is theory-driven in that the synthesis searches for and refines explanations of intervention effects by asking: “What works, for whom, and in what circumstances?” [27]. In this synthesis, we examined the relationships that exist between the therapeutic and technological features of iCBT (Mechanisms) and program use (Outcomes), and the program delivery formats and interactions (Contexts) that support them. We expressed these relationships as Context-Mechanism-Outcome configurations.

Context-Mechanism-Outcome Configuration Development

We began the synthesis by developing *candidate* Context-Mechanism-Outcome configurations for how iCBT programs may work. The development process consisted of brainstorming activities, with the research team reviewing

literature on human-technology interaction and studies of iCBT programs for anxiety to identify relevant and pre-existing theories, models, or frameworks to work from. The persuasive systems design (PSD) model [17] emerged as a key framework for understanding how iCBT, as a persuasive system, was intended to work, and we used this model to develop the initial list of Context-Mechanism-Outcome configurations. The model describes 28 PSD (technological) features, subdivided across the following 4 categories, which can be implemented by programs to guide the user toward their health-related goal: (1) Primary task support, (2) Dialogue support, (3) System credibility support, and (4) Social support.

Using the PSD model, we identified which PSD features (Mechanisms) might be associated with iCBT program use (Outcomes) to formulate Mechanism-Outcome dyads. We then hypothesized which program delivery formats, interactions, and conditions for use (Context) might promote the occurrence of the Mechanism-Outcome interactions. Together, these steps led to the generation of 5 candidate Context-Mechanism-Outcome configurations (Multimedia Appendix 1) [30-34]. The configurations were as comprehensive and justifiable as possible, referencing literature that supported their development and inclusion in the list. These configurations would undergo refinement and testing during the analysis stage of the synthesis, whereby we used evidence from the literature to validate their explanatory usefulness and applicability for answering our research questions.

Literature Search

We used 3 main strategies to identify literature for iCBT programs. The first search strategy involved an information specialist conducting a systematic and comprehensive search of 8 electronic databases: MEDLINE, EMBASE, ERIC, PsycINFO, CINAHL, Cochrane Library, ProQuest Dissertations & Theses Global, and PubMed for the period 1990 to 2017. The second strategy involved a manual search using Google, an internet search engine, and gray literature repositories (Association for Computing Machinery Digital Library, Open Grey, Institute of Electrical and Electronics Engineers Digital Library, and Canadian Agency for Drugs and Technologies in Health) to identify conference proceedings, program evaluations, and government or technical reports. For both search strategies, MeSH terms and text words were based on mental health condition (anxiety and phobias), intervention modality (internet-based and mobile app), intervention type (prevention and treatment), and therapeutic approach (CBT; Multimedia Appendix 2). The third strategy involved manually searching the table of contents in the *Journal of Medical Internet Research*, *Internet Interventions*, *Journal of Cyber Therapy & Rehabilitation*, and *Journal of Telemedicine and Telecare*, and a review of reference lists of included documents and reviews (eg, systematic reviews).

We employed the search strategies on an iterative and recurrent basis until November 2017 to ensure the review was up to date and inclusive. Before discontinuing the literature search, a test of saturation was applied to the search strategies, which involved verifying that further searching would not yield any new results [35].

Literature Selection

In this study, 2 independent, trained reviewers (authors ADR and LW) screened the identified documents for eligibility using a 2-stage approach. During this process, reviewer discrepancies were resolved by consensus or third party arbitration (author ASN). At stage 1, all documents were screened for eligibility using the title and abstract. During this stage, we randomly selected 100 citations to assess inter-rater agreement for inclusion or exclusion decisions; Cohen kappa was 0.74 between raters, reflecting substantial agreement [36]. All documents that were screened "yes, include" or "unsure to include" moved to stage 2. At stage 2, the full text of documents was reviewed by 1 reviewer (ADR), in consultation with another (ASN), with a resulting decision to either include or exclude a document from the synthesis.

For an iCBT program to be included in this synthesis, supporting documents needed to be published in English and provide information on treatment Context, program design (PSD) and delivery features, and program use Outcomes. Each document did not need to provide details on all 3, but all 3 needed to be represented in the total documents for an iCBT program. In addition, at least one published study on the iCBT program needed to be available for inclusion so that we could assess the methodological quality of the study providing program use outcome data.

Intervention studies (eg, clinical trials) were eligible for inclusion if they evaluated iCBT anxiety programs with children (aged <14 years) or adolescents (aged 12-19 years). As some iCBT programs were designed for and evaluated with participants from a broader age range (eg, programs also geared toward young adults), only those studies that provided separate data for participants aged ≤19 years were included. We also required that the type of iCBT program under evaluation be designed for an anxiety disorder(s) or anxiety symptoms associated with a disorder as classified according to the Diagnostic and Statistical Manual of Mental Disorders, fifth edition [37] such as social phobia (social anxiety disorder), generalized anxiety disorder, panic disorder, separation anxiety disorder, or specific phobia. Transdiagnostic programs (ie, programs designed for anxiety plus another diagnosis such as depression) were also eligible for inclusion. We also required that the iCBT program consisted of modules designed for use by the child or adolescent (and not solely delivered to or facilitated by a parent or therapist) as child or adolescent program use was our outcome of interest. Theoretical papers, mixed-methods and qualitative studies, and policy or implementation documents were also eligible if they included a focus on how an iCBT program was proposed to work.

Literature Appraisal

Documents were assessed for relevance and rigor based on consensus between 2 reviewers (authors ADR and LW). Relevance was assessed based on the level of contribution a document provided for an iCBT program in 3 domains: (1) underpinning theory and/or the context and sequence for delivery (Context), (2) PSD features (Mechanism), and (3) program use outcomes (Outcomes). The level of contribution for each domain was rated *low* if little or no information was

provided, *medium* if some information was provided, and *high* if information was well described. Exemplar documents with a *high* level of contribution across the 3 domains informed decision rules for the rating of all other documents.

The methodological quality (rigor) of research studies was assessed using the Mixed Methods Appraisal Tool (MMAT) [38,39]. The MMAT is a reliable, efficient, and valid tool that provides different sections for assessing studies of qualitative, randomized, nonrandomized studies, descriptive studies, and mixed-method designs [38-40]. Multiple publications of the same study (thesis + journal publication) received the same MMAT score. MMAT scores could range from 0% to 100%, with a greater score indicating that more quality criteria were met.

Data Extraction and Coding

Documents for each iCBT program were grouped together during data extraction and coding. Two reviewers (ADR and LW) cross-referenced extraction and coding decisions, with a random subset of 10 documents; the remaining documents were coded by 1 reviewer (ADR). In addition to document characteristics (type and year of publication), iCBT program data were extracted for the following:

- Participants: eligibility criteria and participant demographics.
- iCBT program Context: number of modules, module workflow and sequence, delivery setting, adjunct support, and level of prevention according to the Institute of Medicine model [41].
- Theory or proposed Mechanisms: program features, content and components, including PSD features, and information on why the iCBT program was designed a certain way or how the program was proposed to work.
- Program use Outcomes: information related to how many Web-based modules, exercises, or activities were completed by users or how many users completed certain aspects of the program, measured at posttreatment.

Context and Mechanism Data

Adjunct support details were coded when human-derived technological or therapeutic communication was provided to users to complement iCBT program delivery.

Therapeutic content in programs was coded according to the 5 main CBT components found in the American Academy of Child and Adolescent Psychiatry (AACAP) practice parameter [42]: psychoeducation, somatic management skills, cognitive restructuring, exposure methods, and relapse prevention.

As most authors did not use PSD terminology or concepts, program descriptions were coded as PSD features using a codebook and glossary [17] (Multimedia Appendix 3). PSD features were coded when they were executed by the technology (intrinsic to the design and delivery of the internet-based program) and not by human action (eg, congratulatory comments provided in person by a parent or teacher), which is in line with the use of the PSD model by others [15]. When available, suggested mediators and moderators of program use were extracted, as was information on partial or full Context-Mechanism-Outcome configurations, as discussed by the original authors of the included documents.

Outcome Data

We found inconsistent and heterogeneous measurement and reporting of program use Outcomes—such as program adherence, compliance, and completion. These limitations have been noted by others [22].

For each study, program use Outcomes are reported as published by original authors and are collectively referred to in this study using the umbrella term *program use*. As no applicable cut-off scores have yet been established, when possible, Outcomes were converted into percentages (based on quartiles) to assist with interpreting program use. We used the following parameters to summarize program use: *high* use (≥75%), *moderate* use (50%-74%), *low* use (25%-49%), or *very low* use (≤24%). Study dropout or attrition data were not included in the analysis because these data may not directly reflect program use (eg, program completion), but rather rates of study participants who did not fulfill the research protocol (eg, filling out questionnaires) [15]. Corresponding authors were contacted to provide clarity and completeness of unclear or unreported information and to ensure accurate application of the PSD model for coding iCBT program features. An author for each of the included programs responded to our requests (n=10).

Data Analysis and Synthesis Process

We used a multistep approach to data analysis that was structured according to Pawson's techniques [27,28] and meta-ethnography [43,44]. The first step involved determining recurrent patterns or themes (demi-regularities) across documents for each iCBT program for delivery Context, PSD features and program Mechanisms, and Outcomes related to program use. The purpose was to use evidence from the literature to (1) identify PSD Mechanisms in each program most frequently associated with the program use Outcomes to refine the candidate Mechanism-Outcome dyads, and (2) incorporate delivery Context of each iCBT program into the dyads to refine the overall Context-Mechanism-Outcome configurations. Context-Mechanism-Outcome configurations that were supported by evidence from at least two different iCBT programs progressed to the next step of analysis.

The second step in the analysis involved reciprocal translation analysis, a meta-ethnographic technique that involved reviewing the Context-Mechanism-Outcome configurations across iCBT programs [44,45]. Configurations that were found to have mixed (ie, more heterogeneous support with no larger trend) or confounding evidence across programs, or could not be refined by better describing or recombining the Context, Mechanism or Outcome factors, did not progress to the next stage of analysis. What remained were configurations that provided the best support, across multiple programs, to explain the relationship of design and delivery components of iCBT with program use.

In the final step, we used lines-of-argument synthesis, a theorizing technique [44] that involved assessing how well each Context-Mechanism-Outcome configuration could explain why the same PSD Mechanism(s), operating in different iCBT

program Contexts, might result in particular program use Outcomes. We took into consideration the quality and quantity of evidence supporting the configuration and held a meeting to discuss, amend, and finalize configurations with individuals from across Canada with expertise in electronic health interventions. Configuration refinement continued until we felt that it reflected a pattern that would remain consistent despite differences in small- or large-scale details across iCBT programs. At that point in time, we considered the configuration to be adequately developed.

Results

Included Documents

The literature search and selection progress are presented in Figure 1. The search strategy yielded 11,511 unique documents for stage 1 screening, after duplicates were removed. Of these, 801 documents underwent stage 2 screening. In total, 44 documents detailing 10 iCBT programs were included in the realist synthesis. Documents were published studies (n=20), theses (n=5), published or registered protocols for trials (n=12), study or program websites (n=6), and a study flyer (n=1).

Characteristics of Internet-Based Cognitive Behavioral Therapy for Anxiety in Children and Adolescents

Program and Participant Characteristics

Table 1 presents an overview of iCBT program and user characteristics. The majority of child and adolescent users were white, English speakers, of middle-to-high socioeconomic status, who lived in urban centers with both biological parents. Programs designed to treat an anxiety disorder tended to be longer in duration (included more modules) than prevention-based programs. Treatment-based programs were delivered in the community (some included occasional health care clinic visits) and involved weekly Web- or email-based therapist interaction and parent-dedicated modules. Prevention-based programs were often provided in schools, with a teacher facilitating program administration. Most iCBT programs were adaptations of previously developed mental health prevention or treatment resources [46-55].

Figure 1. Flow diagram of the literature search and selection process. iCBT: internet-based cognitive behavioral therapy.

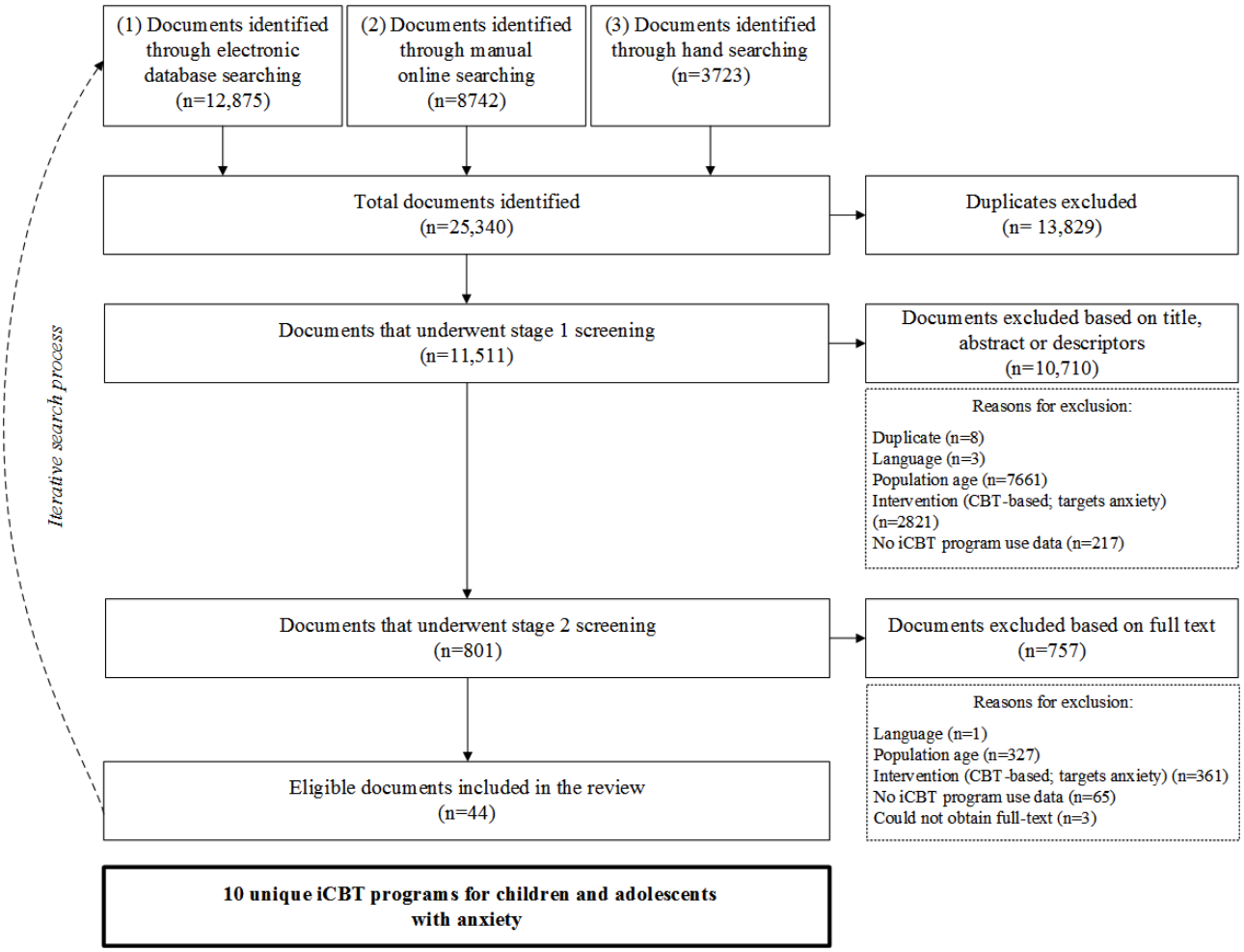


Table 1. Overview of the internet-based cognitive behavioral therapy user, program, and delivery characteristics.

Numbered list of programs[a]	User details	Delivery		Therapist support in program			Adjunct support
		Setting	Number of modules and duration	Web or email	Phone	In-person	
Treatment programs							
1. BRAVE-ONLINE	Children and adolescents with anxiety	Clinic or community	10 weekly modules + 2 booster modules, 60 min each	✓	✓	—[b]	Parent
2. iCBT[c] for children and adolescents with dental anxiety	Children and adolescents with anxiety	Community plus clinic	12 weekly modules	✓	—	—	Parent, Dental professional[d]
3. Internet-delivered CBT for children with anxiety disorders	Children with anxiety	Community	11 modules over a 10-week period[e]	✓	✓	—	Parent
4. Internet-delivered CBT for children with specific phobia	Children with anxiety	Community	11 modules over a 6-week period, 15-45 min each	✓	✓	—	Parent
5. SmartCAT App for children with anxiety disorders	Children with anxiety	Community	Daily app entries completed over 8 in-person modules, 3-4 min each	✓	—	✓	Parent
Indicated prevention programs							
6. Internet cognitive behavioral skills-based program	Children with anxiety	Community	3 modules over a 12-week period[f]	—	✓[g]	—	Parent
7. REACH for success app[h]	Children with anxiety	School	5 activities, 20-30 min for each activity	—	—	✓	Research assistant[i]
8. Individually tailored iCBT for adolescents[h]	Adolescents with anxiety, or anxiety and depression	Clinic	6-9 prescribed modules over a 6-18-week period[j]	✓	✓	✓	Therapist (optional)
Universal prevention programs							
9. The e-couch anxiety and worry program	Adolescents with anxiety	School	6 weekly modules, 30-40 min each	—	—	—	Teacher[k], Mental health service provider[l]
10. MoodGYM	Adolescents with anxiety, or anxiety and depression	School or community	5 weekly modules, 30-60 min each	—	—	—	Teacher[k]

[a]Categorized according to the Level of Prevention Model [41]: universal prevention: target participants have not been identified on the basis of individual risk (ie, no symptoms required); selective prevention: target participants have a higher risk of developing an anxiety disorder than others; indicated prevention: target participants are high risk, and who have anxiety signs or symptoms, but do not currently meet diagnostic levels; and treatment: target participants are diagnosed with an anxiety disorder.

[b]N/A: not applicable.

[c]iCBT: internet-based cognitive behavioral therapy.

[d]A dental professional (a dentist, dental hygienist, or dental assistant) provided exposure at a dental clinic.

[e]Five versions depending on diagnosis.

[f]Two blocks of modules (containing multiple sections) dedicated to mothers and 1 module block (containing multiple sections) dedicated to child + mother.

[g]Therapist completed a brief (15 min), nontherapeutic, check-in telephone call with the mother (not the child).

[h]Program was designed for indicated prevention or treatment (early intervention).

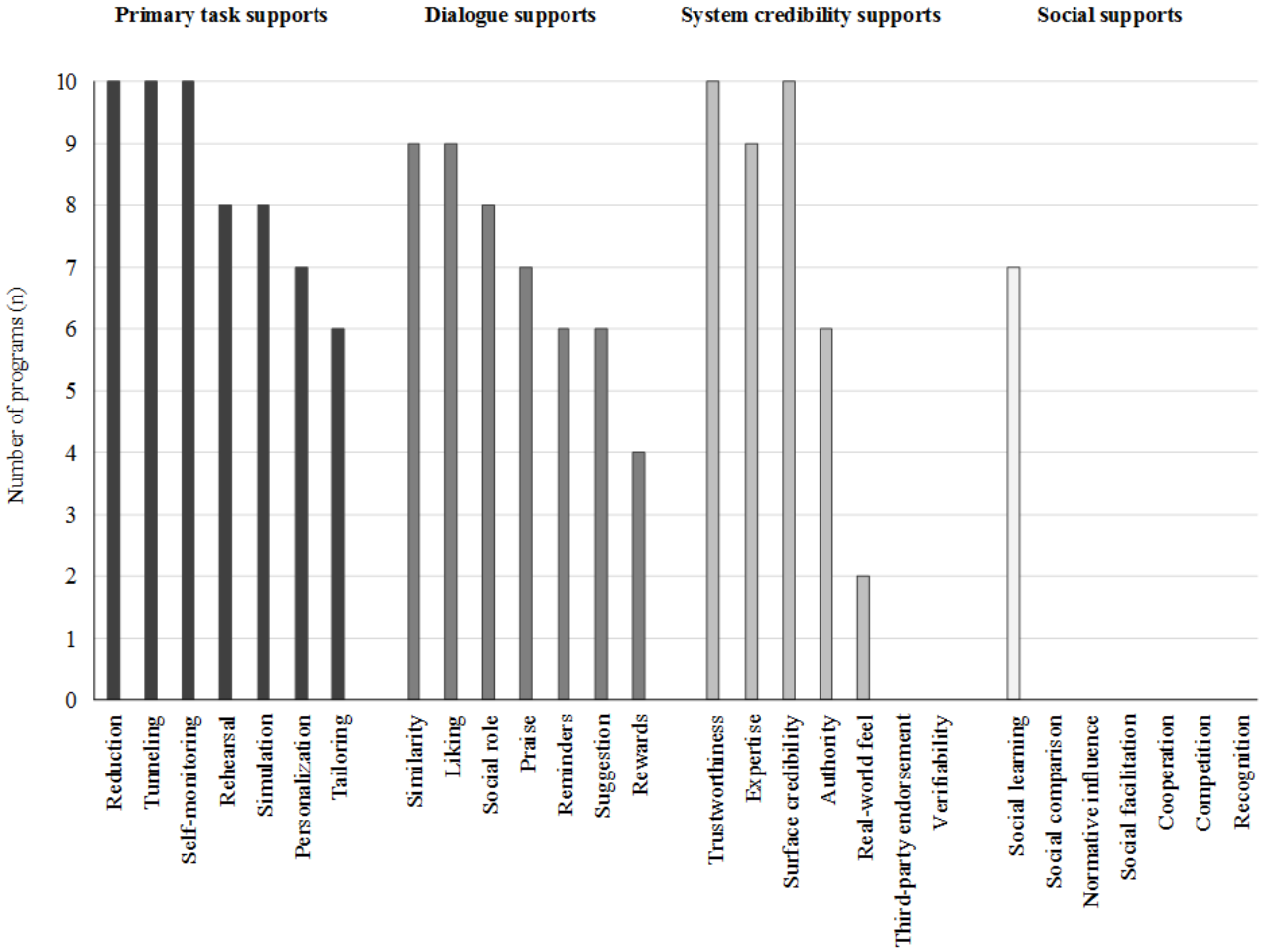
[i]Research assistant or graduate student was present to facilitate aspects of the study such as assessment and troubleshoot technical issues.

[j]Out of a possible 17 modules, based on symptoms.

[k]Teacher facilitated program administration and was available for general guidance or if questions arose but did not provide an active therapeutic role.

[l]Mental health service provider was present in 1 study of the program to facilitate program administration or address student questions [56].

Figure 2. Overview of the persuasive systems design features across the 10 internet-based cognitive behavioral therapy programs included in the synthesis.



Cognitive Behavioral Therapy Components and Persuasive Systems Design Features

All programs described themselves as CBT-based and contained at least 3 of the 5 AACAP-recommended CBT components (most commonly psychoeducation, somatic management skills, and cognitive restructuring). The workflow of the programs presented the more foundational and simpler treatment components (eg, psychoeducation and symptom identification) before more challenging skills (eg, desensitization or exposure). Many programs also integrated interpersonal therapy skills [57], such as assertiveness training and problem solving, to reduce environmental stressors and enhance social support.

The frequency of the PSD features used in the iCBT programs is shown in Figure 2. All programs incorporated Reduction and Tunneling (Primary task supports) to regulate the pace and sequence by which users moved through the program. Self-monitoring, a Primary task support, was also used in all programs to track users' progress or mood over time. A total of 8 programs used Rehearsal by providing recommended or required Web-based homework or practice activities. Moreover, 7 programs provided a Personalized review (eg, progress report and feedback) of homework or module content before the next module was accessible to users. Most programs used Tailoring (n=6) or Personalization (n=7) (Primary task supports) to adapt the content to users' characteristics such as their primary anxiety concern, age, or name. Incorporating a Social role, such as a Web-based or virtual therapist or cartoon guide, was one of the most commonly used Dialogue supports (n=8). In addition, 9 of the 10 programs had program content, technology, and interaction features that were relatable and appealing to target users (Similarity and Liking features). As all programs were part of a research study, Trustworthiness and Surface credibility (System credibility supports) were considered inherent to their design (ie, programs were ad-free, not marketed for commercial use, and accessed through a secure platform), although few documents explicitly reported this. Authority (System credibility support) was incorporated when Web- or email-based therapist support was provided. Social support features were seldom used among programs.

Level of Contribution and Methodological Quality

Details of the quality appraisal are provided in Multimedia Appendix 4 [58-75]. Documents tended to provide the greatest level of contribution to understanding program Context and Outcomes with relatively few details available for understanding program Mechanisms. A total of 7 iCBT programs had at least one document that scored *high* for level of contribution for program Context, Mechanisms, and Outcomes. We evaluated 29 research studies (found in 25 documents) for methodological rigor. A total of 19 studies met all 4 MMAT criteria (100%), 7

met 3 criteria (75%), and 3 met 2 criteria (50%). Lower ratings reflected an unclear description of: processes for recruitment, selection, randomization, or allocation; how group differences (if any) were controlled for; the percentage of outcome data obtained; or withdrawal or dropout rates were not within acceptable limits.

Summary of Key Internet-Based Cognitive Behavioral Therapy Contexts, Mechanisms, and Outcomes

Tables 2 and 3 present an overview of the delivery Context and PSD Mechanisms that were most frequently or consistently associated with moderate-to-high program use Outcomes across iCBT programs. Contexts most indicative of program use were the adjunct support person and the communication approach (eg, Web [the program platform or internal messaging system], email, or in-person) provided to the user—both of which typically varied based on the level of prevention the program was designed for. The type of adjunct support also depended on the age of program users. Children generally received the most extensive adjunct support (ie, therapist and parent), and program use was often greater among this age group than among adolescents. PSD features identified as having a high value to encouraging program use were Tailoring and Personalization (Primary task supports); Social role, Reminders, and Rewards (Dialogue supports); and Authority, Expertise, and Trustworthiness (System credibility supports). Program use Outcomes most commonly reported either the total proportion or average proportion of program modules, homework, or activities completed by users. More than half of the Outcomes indicated *high* or *moderate* program use.

When finalizing our proposed configurations, we considered both what program Contexts and Mechanisms were combined (using Tables 2 and 3) and in what way and for what purpose they may have been combined (using excerpts from individual documents; see Table 4) to explain how moderate-to-high program use Outcomes were generated.

Proposed Internet-Based Cognitive Behavioral Therapy Contexts and Mechanisms for Moderate-to-High Program Use

Key Contexts and PSD Mechanisms that may have led to moderate-to-high iCBT program use Outcomes are described in Table 4 alongside examples from contributing programs.

Context-Mechanism-Outcome Configuration 1: Tailoring and Personalization

Evidence from 8 iCBT programs suggested that indicated prevention and treatment programs that provided adjunct support along with Tailoring and Personalization resulted in greater program use. Studies supporting this configuration had a mean MMAT score of 90.2%.

Tailoring and Personalization were used to provide users with more information to increase the individualized feel of the program and portray that the program knew and could meet the user's needs. Programs most commonly tailored content (ie, therapeutic elements and examples) based on the user's age or mental health condition (eg, specific phobic and social anxiety). User information was often collected by the adjunct support person during study enrollment (ie, part of eligibility screening). Personalization was a feature that could be initiated through program automation (ie, user's name appeared on the home screen; his or her pronouns) or via program communications (ie, individualized weekly emails and secure messages). The adjunct support person acted as an extension of the Tailored or Personalized program experience through their contact with users within (eg, by providing personalized feedback on Web-based homework [50,76,83]) or outside of the iCBT program (eg, by tailoring in-person session objectives [55]). Programs with a combination of Tailoring and Personalization reported some of the highest program use Outcomes.

Context-Mechanism-Outcome Configuration 2: Reminders

Evidence from 6 iCBT programs suggested that indicated prevention and treatment programs that provided adjunct therapist support along with Reminders also resulted in greater program use. Studies supporting this configuration had a mean MMAT score of 92.1%.

The programs contributing to this Context-Mechanism-Outcome configuration involved multiple modules or activities (the number of modules ranged from 6-11 or app use ranged from daily-weekly activities); therefore, users were required to log into the program over numerous instances. Reminders were used to encourage the user to take program action, either by promoting skills practice [46] or by "prompt[ing] participants who [were] late in completing a module" [78] to access the newly available content. Programs provided generic and automatized Reminders through email or the Web-based platform. Reminders were also embedded in the regular, electronic, personalized communications (eg, feedback and progress check-ins) sent by the adjunct therapist to the child or adolescent. If users remained absent from the program beyond the recommended treatment schedule (eg, longer than 1 week), the adjunct therapist provided additional in-person or telephone follow-up, encouraging users to access the next available module.

Context-Mechanism-Outcome Configuration 3: Rewards

Evidence from 4 iCBT programs suggested that programs and mobile apps that provided adjunct support along with Rewards resulted in greater program use. Studies supporting this configuration had a mean MMAT score of 85.0%.

Table 2. An overview of the delivery Context and persuasive systems design features that may explain program use Outcomes across internet-based cognitive behavioral therapy treatment programs.

Program and document	Context: Target users and adjunct support	Mechanism: PSD[a] features	Outcome: Posttreatment findings (program use summary[b])
Program 1: BRAVE-ONLINE for children and adolescents with anxiety disorders			
[49]	Child users; Therapist support: in-person, Web, email, phone; Parent support: in-person, module	Primary task support: Tailoring and Personalization; Dialogue support: Social role and Reminders; System credibility support: Authority, Expertise, and Trustworthiness	91% of homework completed (high use)
[76]	Child users; Therapist support: Web, email, phone; Parent support: modules	Same as above	95% of module activities completed (high use)
[77,78]	Same as above	Same as above	Average of 7.5/10 modules completed (high use); 33.3% of users completed all 10 modules (low use)
[79]	Same as above	Same as above	Average of 4.88/10 modules completed (low use)
[76]	Adolescent users; Therapist support: Web, email, phone; Parent support: modules	Same as above	85% of module activities completed (high use)
[80]	Same as above	Same as above	Average of 7.5/10 modules completed (high use); 39% of users completed all 10 modules (low use)
[80]	Child and adolescent users; Therapist support: Web, email, phone; Parent support: modules	Same as above	Average of 7.9/10 modules completed (high use); 42.6% of users completed all 10 modules (low use); 73.5% of module tasks completed (moderate use); Treatment expectancy predicted compliance (N/A[c])
[81]	Same as above	Same as above	Average of 85% module tasks completed (high use); Average of 8.9/10 modules completed (high use)
[82][d]	Same as above	Same as above	Average of 6.7 /10 modules completed (moderate use); 19% of users completed all 10 modules (very low use)
[23][e]	Same as above	Same as above	Average of 4.8/10 modules completed by children (low use); Average of 4.0/10 modules completed by adolescents (low use)
Program 2: iCBT[f] for children and adolescents with dental anxiety			
[48]	Child and adolescent users; Therapist sup- port: Web; Parent support: in-person; Dental professional support: in-person	Primary task support: Personalization; Dialogue support: Social role; System credibility support: Authority, Expertise, and Trustworthiness	Average of 9.2/12 modules completed (high use)
Program 3: Internet-delivered CBT for children with anxiety disorders			
[54,83]	Child users; Therapist support: Web, email, phone; Parent support: modules	Primary task support: Tailoring and Personalization; Dialogue support: Social role and Reminders; System credibility support: Authority, Expertise, and Trustworthiness	Average of 9.7/11 modules completed (high use)
[54,84]	Same as above	Same as above	83% of users completed ≥9 of 11 modules (high use)
[85]	Same as above	Same as above	Average of 6.0/12 modules completed[g] (moderate use); 53% of users reached at least module 4 (first exposure exercise; moderate use)
Program 4: Internet-delivered CBT for children with specific phobia			
[50]	Child users; Therapist support: Web, email, phone; Parent support: modules	Primary task support: Personalization; Dialogue support: Social role and Reminders; System credibility support: Authority, Expertise, and Trustworthiness	80% of users completed ≥9 of 11 modules (high use)
Program 5: SmartCAT App for children with anxiety disorders			

Program and document	Context: Target users and adjunct support	Mechanism: PSD[a] features	Outcome: Posttreatment findings (program use summary[b])
[46]	Child users; Therapist support: in-person, mobile app; Parent support: in-person	Primary task support: Tailoring and Personalization; Dialogue support: Social role, Reminders, and Rewards; System credibility support: Authority, Expertise, and Trustworthiness	Average of 82.8% of practice entries completed (high use)

[a]PSD: persuasive systems design.

[b]Program use summary was calculated by dividing the reported value by 100 or converting it to a percentage. High use (≥75%), moderate use (50-74%), low use (25-49%), or very low use (≤24%).

[c]Not applicable

[d]All participants were diagnosed with a high functioning autism spectrum disorder and anxiety disorder.

[e]This study compared participants who were randomized to 1 of 2 iCBT conditions: iCBT-generic (iCBT relevant to multiple types of anxiety; ie, social, separation, and generalized anxiety) or iCBT-social anxiety (iCBT specific to social anxiety).

[f]iCBT: internet-based cognitive behavioral therapy.

[g]Data available for 15 out of 17 participants.

A total of 3 iCBT programs regularly incorporated Rewards into modules to encourage ongoing program use and promote the completion of essential treatment exercises [46,47,51]. The iCBT program for dental anxiety opted for a final Reward and presented users with a virtual diploma at the end of their treatment [45]. Rewards were also used as a proxy to track program progress, including completion of exposure activities [46,47,51]. Unlike computer-based programs, the open and flexible design approach to mobile apps gave users the option to select what treatment content and tasks they wanted to access and when. Progressive reward incentives were used to persuade users to complete more of the app's content and critical components. In the REACH for success app, a cartoon character provided regular guidance and feedback to users and entertained them with animations following task completion (Reward) [47]. In the SmartCAT app, a point system tied to external prizes (Reward) was a feature managed by the adjunct therapist [46]. In-person sessions with a therapist or parent also provided positive reinforcement of program use (ie, Praise, Rewards); sessions were also used to instruct users on how to incorporate Rewards into their anxiety management activities outside of the program [48,51].

Context-Mechanism-Outcome Configuration 4: Therapist, Social Role, Authority, Expertise, and Trustworthiness

Evidence from 6 iCBT programs suggested that indicated prevention and treatment programs with adjunct Web- or email-based therapist support that also provided a Social role component, and conveyed Authority, Expertise, Trustworthiness, resulted in greater program use. Studies supporting this configuration had a mean MMAT score of 91.3%.

The Social role component of iCBT programs was often fulfilled by a therapist or coach (who received specialized training with the program but may not have been a licensed psychologist). Therapists engaged in regular, Web- or email-based communication with the user and served 2 roles by (1) facilitating program delivery by providing technical support and answering users' questions and (2) promoting program completion by providing reminders and encouragement, reinforcing program concepts, and ensuring accurate comprehension and application of CBT skills. Together, the Social role feature and therapist emails complemented (had overlap with) other PSD features such as Reminders, Praise, and Suggestion. Therapists had secure access to users' written responses or logged data so that they could send specific communications to users, demonstrating therapists' credibility and competence with both the therapeutic process and individual responsiveness (Authority, Expertise, and Trustworthiness). Moreover, 3 studies measured child-reported therapeutic alliance with their iCBT program therapist and found it to be strong and program use to be high [76,81]. One of these studies correlated therapeutic alliance with program use and found a significant, positive relationship [81].

Context-Mechanism-Outcome Configuration 5: Therapist + Parent, Social Role, Authority, Expertise, and Trustworthiness

Evidence from 5 treatment programs suggested that iCBT programs with adjunct therapist and parent support that also included a Social role component, and conveyed Authority, Expertise, and Trustworthiness, had greater program use. Studies supporting this configuration had a mean MMAT score of 90.8%.

Table 3. An overview of the delivery Context and persuasive systems design features that may explain program use Outcomes across internet-based cognitive behavioral therapy indicated prevention and universal prevention programs.

Program and document	Context: Target users and adjunct support	Mechanism: PSD[a] features	Outcome: Posttreatment findings (program use summary[b])
Indicated prevention programs			
Program 6: Internet cognitive-behavioral skills-based program			
[51]	Child users; Therapist support: phone; Parent support: modules	Primary task support: Tailoring; Dialogue support: Rewards; System credibility support: Trustworthiness	Average of 82.6% modules completed; Users who immediately accessed the program completed more sections (average=17.35) than those who had delayed access (average=8.0); Immediate access users spent more time in the program (average=183.3 min) than those who had delayed access (average=77.6 min); Use time was positively correlated with number of sections completed (high use)
Program 7: REACH for success app			
[47]	Child users; Therapist support: in-person	Primary task support: Tailoring and Personalization; Dialogue support: Social role, Reminders, and Rewards	93.2% of users completed relaxation practice (high use); 91.7% of users completed hypothetical cognitive self-control practice (high use); 15.2% of users completed applied (very low use) cognitive self-control practice (very low use); 45.5% of users completed self-monitoring (low use); The proportion of users who attempted an activity was higher than those who completed an activity (N/A[c])
[47]	Same as above	Same as above	Users completed more activities before an evaluation module (N/A); App use was highest in week 1 and decreased over 6 weeks (N/A); 100% of users completed relaxation practice (high use); 100% of users completed hypothetical cognitive self-control practice (high use); 60.0% of users completed self-monitoring (moderate use); 0% of users completed exposure practice (very low use)
Program 8: Individually tailored iCBT[d] for adolescents			
[55, 86]	Adolescent users; Therapist support: in-person, email, phone	Primary task support: Tailoring and Personalization; Dialogue support: Social role and Reminders; System credibility support: Authority, Expertise, and Trustworthiness	Average of 6.5/9 modules completed (moderate use)
Universal prevention programs			
Program 9: The e-couch anxiety and worry program			
[87]	Adolescent users; Teacher support: in-person	Dialogue support: Social role	45% of users completed all modules (low use)
[24]	Same as above	Same as above	50% of users completed all modules (moderate use)
[24]	Adolescent users; Teacher support: in-person; Mental health provider support: in-person	Same as above	36% of users completed all modules (low use)
Program 10: MoodGYM			
[53]	Adolescent users; Teacher support: in-person	Dialogue support: Social role	Average of 3.2/5 modules completed (moderate use)
[25]	Same as above	Same as above	Average of 9.4/28 exercises completed (low use); >25% of users completed all modules (low use)
[25]	Adolescent users	Same as above	Average of 3.1/28 activities completed (very low use)
[26]	Adolescent users; Teacher support: in-person	Same as above	<1% of users completed all activities (very low use)

[a]PSD: persuasive systems design.

[b]Program use summary was calculated by dividing the reported value by 100 or converting it to a percentage. High use (≥75%), moderate use (50-74%), low use (25-49%), or very low use (≤24%).

[c]Not applicable.

[d]iCBT: internet-based cognitive behavioral therapy.

Table 4. Configuration summaries of the key Contexts and persuasive systems design Mechanisms that may have led to moderate-to-high program use Outcomes.

Context	Mechanism		Program #
	PSD[a] feature(s) and proposed purpose	Example	
Indicated prevention and treatment programs with adjunct support	Configuration 1: Tailoring +/or Personalization to increase relevance of program content	• Through email the therapist provided "written feedback on worksheets" and was available to "answer questions and clarify treatment content, increase motivation and to help solve problems" [83]. • A participant's name was populated in modules and feedback messages [76].	1, 2, 3, 4, 5, 6, 7, and 8
	Configuration 2: Reminders to increase awareness of program availability and progress	• "Participants receive automated, computer-generated, standardized, weekly e-mails both before each module (as a reminder to complete their modules) and after each module (to congratulate them on finishing their module)" [78]. • "Each skills coach entry ends with a customized motivational message from the therapist (entered weekly via the [app] portal) that includes encouragement as well as a reminder to complete any assigned home-based exposure or skills practice" [46].	1, 3, 4, 5, 7, and 8
	Configuration 3: Rewards to recognize and encourage achievement	• Following task completion, the user received a reward in the form of Bob's abilities or tricks, with more complicated tricks being unlocked as users completed more of the treatment protocol [47]. • Program progress was presented and tracked with a virtual sticker chart. A cartoon figure would jump to the next rung of the ladder indicating successful completion of an exposure hierarchy activity [51].	2, 5, 6, and 7
Indicated prevention and treatment programs with adjunct therapist support	Configuration 4: Social role to increase program interaction; Authority+Expertise+Trustworthiness to improve perceived value of information or support	• Participants received "comments and feedback from their therapist on all exercises, and the technical platform also allowed participants to comment on worksheets" [50]). • The therapist portal and secure messaging features in the app allowed the participants and therapist to securely exchange information such as messages, documents, or audio or video files related to treatment [46].	1, 2, 3, 4, 5, and 8
Treatment programs with adjunct therapist support plus parent support	Configuration 5: Social role to increase program interaction; Authority+Expertise+Trustworthiness to improve perceived sense of reliance and cooperation toward program progress	• Parents were provided with their own modules during treatment. "In this way, the parent was empowered to help their child acquire and use the skills presented in the program, and to handle situations in which their child became anxious" [77]. • Check-in telephone calls from the therapist consisted of 4 elements: (1) progress updates, (2) symptom assessments, (3) encouragement to use the program, and (4) troubleshooting [51].	1, 2, 3, 4, and 5

[a]PSD: persuasive systems design.

Treatment-based iCBT programs were designed to be child-parent combined or parent-supported child-based interventions. Combined interventions required parents to complete parent-specific modules (eg, psychoeducation, relaxation training, problem solving, and modeling adaptive behaviors) either before or alongside their child as they completed their own child-directed modules. In parent-supported interventions, parents may have also been responsible for explaining program instructions and assisting their children with their modules [50,54,76], coaching or supporting their child with in vivo exposure or practice activities [46,48,50], and overseeing their child's treatment schedule [54]. The support the adjunct therapist provided to children was also extended to parents. Parents had the opportunity to ask clarifying questions, receive expert advice, and troubleshoot difficulties with their child's iCBT progress with the program therapist. Studies found that both parent and child ratings of therapeutic alliance and program use were high [76,81]. One study correlated therapeutic alliance and program use and found significant, positive relationships for both parents and children [81]. Anderson et al [81] hypothesized that therapist emails may have contributed to fostering a strong therapeutic alliance.

Discussion

Principal Findings

The role of technological, persuasive components in iCBT programs is an understudied aspect of program design and evaluation. The extent of iCBT program use may be a fundamental indication of its *persuasiveness* and its potential to assist the user with their goals of the program [15,88-91]. This realist synthesis identified 5 possible relationships as to

how the use of specific PSD features (technological Mechanisms), supported by some key user and delivery features (Context), may generate moderate-to-high program use (Outcomes) in iCBT for children and adolescents with anxiety.

The 5 Context-Mechanism-Outcome configurations provide support for several persuasive strategies to improve iCBT program use: Tailoring and Personalization as Primary task supports; Rewards, Reminders and Social role features in programs serving as Dialogue supports; and Trustworthiness, Expertise, and Authority features serving as System credibility supports for a program. Traditionally, PSD features that stimulate human-computer communication, such as Dialogue supports, have been the most widely used and studied features for improving program use [6,15,89,92,93]. However, this synthesis demonstrated that having credibility (System credibility supports) and supporting users in completing their target behavior (Primary task supports) may also promote moderate-to-high program use. We hypothesize that using multiple PSD features, both within and across the different support categories, may produce additive or synergistic effects on program use; however, there was insufficient evidence available for our analysis to explain the impact of PSD feature combinations that involved more than a few features (and proposed functions) at a time. This is because the authors of the original studies included in our review typically discussed or formally tested the relationship of only 1 or 2 PSD features and program use at a time. Therefore, our configurations present the fewest possible PSD features that could be linked to higher program use (ie, we uncoupled features as much as possible—an approach that may make testing of their effects more efficient in future studies).

Moreover, we suspect that not all PSD features may have equal influence on program use. Depending on the program, some PSD features may be *necessary* for program use (part of the basic requirements or foundational design framework of iCBT), whereas others may be *complementary* to program use (have an impact by enhancing the design framework of iCBT); although both types of features together may influence program use. In this realist synthesis, all 10 iCBT programs described a purposeful design that incorporated Reduction and Tunneling (Primary task supports) and Similarity and Liking (Dialogue supports) to create a logical, incremental, relevant, and aesthetically pleasing experience for users—these may be the *necessary* PSD features for iCBT for children and adolescents with anxiety. The PSD features described in our configurations are hypothesized to be *complementary*—building on the persuasiveness of necessary PSD features to further improve or optimize iCBT program use.

A meta-analysis of PSD features used in internet-based interventions for mental health demonstrated that determining the amount and type of persuasive principles to include may be a delicate balance, as some principles seem to work together, whereas when other principles co-occur, they may have an unapparent or diminishing effect [90]. As was found for this synthesis, it is not necessarily the number of PSD features used in iCBT, but it is their proposed purpose or implementation that is particularly critical for optimizing program use outcomes [90]. For example, when comparing 2 indicated prevention programs, we observed that the internet cognitive behavioral skills-based program [51] had fewer PSD features than the individually tailored iCBT program [55], although the former reported greater program use. At this time, our understanding of how to best bring together PSD features, such as Personalized Reminders [93], in the design and delivery of iCBT for child and adolescent anxiety is limited. Therefore, further research on the theory, function, quality, and effectiveness of individual PSD features is needed to deliberately use and combine them for idealized treatment outcomes. Moreover, involving target users in the (co)design and testing of treatments is recommended to improve the acceptability, feasibility, and effectiveness of iCBT with children and adolescents [19,94,95]. These participatory research efforts may provide important guidance on the usefulness and functionality of select (PSD) features of iCBT programs in the *real world* and from the user's perspective [96]; therefore, facilitating greater program use [97].

In this synthesis, 3 important potential relationships were identified: (1) adjunct support seemed to improve program use even when input or support was minimal (eg, in-person, classroom-based program administration with no treatment advice given) or when it was provided by a nonexpert (eg, teacher) [24,25]; (2) users of treatment programs demonstrated higher program use than users of universal prevention programs; and (3) a trend for increased program use among programs for children (more so than for adolescents) was identified. Within these relationships, multiple contextual aspects or user characteristics may have also had an additive or synergistic effect on program use. For example, the level of expertise the adjunct support person had (eg, teacher vs therapist) and the degree of their guidance or therapeutic involvement (eg, in-person program administration vs personalized feedback emails) increased from prevention-based to treatment programs. In the literature, little is known about how much, when, and what type of support is necessary for enhancing program use and efficacy [98,99]. Although some evidence suggests that layperson support is as effective as clinician support [100], this synthesis suggests that the person providing support as well as the intensity of their support activities (frequency; inclusion of therapeutic elements) may have a noteworthy effect on program use. The nature of the role adjunct support plays in iCBT program use is also unclear. It has been suggested that adjunct support may leverage the advantages of therapeutic alliance [101,102], which might include principles of persuasion (eg, users feel the need to respond to social cues [18]), it may establish process expectations and social accountability [103], or it may develop a sense of legitimacy or credibility of the program [103] (see Santarossa et al's study [104] for further suggestions). Programs that had both therapist and parent involvement may have (1) reinforced child's understanding of and confidence in treatment content and (2) increased the child's interaction with the program by creating a perceived sense of cooperation (shared goals) and accountability toward treatment progress. In this realist synthesis, adjunct support may have been used to complement or replace the use of some PSD features in iCBT, particularly Dialogue supports. For example, in-person therapist sessions or telephone calls provided opportunities for Reminders, Personalized feedback, or Praise to be conveyed to users [49,50,54,55]. Consideration of how

and when to provide adjunct support is critical when preparing for the implementation and integration of iCBT within routine practice, such as allotting for therapist time, making changes in clinical workflow, and when conducting economic analyses.

Future Directions

This realist synthesis not only provides support for incorporating some of the well-studied and highly used PSD features into iCBT (ie, Reminders [93,105,106]) but also draws attention to underutilized features that can be incorporated in the designs of new treatments. For example, Rewards were only occasionally used by iCBT programs included in this study (4 out of 10 programs) but are more commonly used techniques for increasing program use in internet-based interventions targeting physical activity or dietary behaviors [107]. Recent efforts into improving the gamification of technology-based CBT for pediatric mental health (see SPARX [108]), where incorporating game-design elements such as Praise and Rewards are regularly used to enhance program use and engagement [109], demonstrate the potential benefits that the use of these features may have.

Although this synthesis and other recent reviews have been helpful for identifying PSD features of interest for improving program use of internet-based interventions [15,89,90], the next step is to formally isolate and evaluate the effectiveness of these PSD Mechanisms in producing optimal program use. This synthesis suggests 8 features that may be a priority for further examination. Modeling, factorial designs or the multiphase optimization strategy [110-112] (see also Baker et al's study [113] for other suggestions) can be used to evaluate the best set (individual or combination) of program features to use under different conditions (eg, delivery setting and start or end of treatment). Studies with multiple, active treatment arms would also allow the examination of the comparative effects of select program features [114] or in different delivery contexts (ie, varying the type of adjunct support). From this synthesis, only 3 studies of 3 different programs conducted these comparisons. These studies provided important insights into the impact of delivery medium [49], type of adjunct support [24], and delivery location [25] on iCBT program use. Qualitative studies or self-report data would also provide meaningful information on the factors affecting program use from child and adolescent or health care provider standpoints. Another important line of inquiry relates to defining and measuring program use to ensure its validity and reliability for future studies. Designing studies that incorporate in vivo, objective measurements or automatic data capture of program use [115,116] could improve our awareness of program use predictors beyond user demographics (ie, age), for example, to actual usage behaviors (ie, number of Web pages viewed). This method would allow for iCBT program use to not only be measured at end of the intervention but also throughout the program access period to assess usage patterns over time [15], when certain design or delivery features may be more or less *activated* or present.

Strengths and Challenges of Realist Synthesis

This is the first study to examine PSD features as they relate to program use in iCBT for children and adolescents with anxiety. A strength of this synthesis is the inclusion of diverse and high-quality evidence (ie, MMAT scores>75% [38]) from both the published and gray literature. Approaching our research questions using a single theoretical framework (ie, the PSD model) allowed for systematic and incremental accumulation of knowledge about how iCBT may work from a trackable, technological perspective.

The lack of operationalization of how PSD features and aspects of Context and program use Outcomes were defined, described, and measured by authors affected our data extraction and coding strategies. As adherence to recent recommendations [19] and reporting guidelines, such as Consolidated Standards of Reporting Trials-electronic health (CONSORT-EHEALTH) [117], become mandatory for publication, the opportunity to identify the active ingredients of iCBT will improve. Clarifying PSD features with original authors was an attempt to mitigate the potential bias that lies in coding technological program features and interpreting the papers using the PSD model [92]. However, few details about the time or the quality of communication by the adjunct support person(s) were available, limiting our understanding of the important role this contextual feature played in program use. For reasons of inclusivity, we described the heterogeneous outcomes using the umbrella term *program use*. Adoption of a recent standardized definition and calculation of *adherence* [118] can clarify what is meant by specific program use terms and allow for comparisons of outcomes across programs. iCBT programs' widespread implementation and ability to meet the health goals of users will involve an understanding of the expectations and actuality of program use in the real world (ie, using true effectiveness studies or formative program evaluation), and setting benchmarks for an *effective dose* in different delivery settings. Finally, like others [119-121], we recognized the outcomes of persuasive systems depended on multiple factors, many of which were not examined in this synthesis. However, it was rare to have information on treatment or technology preferences of users (eg, early completers [22,122]), their psychological characteristics or cognitions (eg, motivation, personality, expectations, and treatment perceptions [123]), or personal circumstances (eg, program access [124])—factors that are also considered critical to program use and could be used to construct and validate more intricate Context-Mechanism-Outcome configurations.

Conclusions

The Context-Mechanism-Outcome configurations identified by this realist synthesis provide an initial understanding of how, why, and for whom iCBT programs for children and adolescents with anxiety work from a persuasive systems' perspective. Appreciating that the effectiveness of iCBT programs may hinge on whether and to what extent programs are used, this study is an important step toward successfully implementing and integrating iCBT into routine clinical care. Recognizing that multiple PSD features are incorporated in iCBT program designs and that individual features may affect each other differently, further knowledge and testing of the purpose and function of these features will help determine the number and combination to use in certain delivery contexts (eg, adjunct support included; level of prevention a program is designed for). As PSD features are modifiable, iCBT program designers and developers can

look to create more persuasive programs that promote greater use and improved treatment outcomes.

Acknowledgments

The authors of this study would like to thank the authors who responded to our request for more information about their iCBT program. The authors would also like to thank Robin Featherstone for developing the search strategy and conducting the electronic database search and Dr Purnima Sundar for reviewing the manuscript for clarity of content. This study was supported by a Canadian Institutes of Health Research (Knowledge Synthesis Grant CIHR: KRS2014) awarded to ASN (Principal Investigator) and an Alberta Innovates - Graduate Studentship in Health awarded to ADR (Trainee). During the conduct of this review, ASN and LH held CIHR new investigator awards and PM holds a Tier 1 Canada Research Chair in Child Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Candidate Context-Mechanism-Outcome configurations.

[PDF File (Adobe PDF File), 62KB - jmir_v21i2e11128_app1.pdf]

Multimedia Appendix 2

Document electronic search strategy.

[PDF File (Adobe PDF File), 8KB - jmir_v21i2e11128_app2.pdf]

Multimedia Appendix 3

The persuasive systems design (PSD) model.

[PDF File (Adobe PDF File), 98KB - jmir_v21i2e11128_app3.pdf]

Multimedia Appendix 4

The level of contribution and methodological quality of documents for the included internet-based cognitive behavioral therapy programs.

[PDF File (Adobe PDF File), 64KB - jmir_v21i2e11128_app4.pdf]

References

1. 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 May;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
2. Compton SN, March JS, Brent D, Albano AM, Weersing R, Curry J. Cognitive-behavioral psychotherapy for anxiety and depressive disorders in children and adolescents: an evidence-based medicine review. J Am Acad Child Adolesc Psychiatry 2004 Aug;43(8):930-959. [Medline: 15266189]
3. Albano A, Kendall P. Cognitive behavioural therapy for children and adolescents with anxiety disorders: clinical research advances. Int Rev Psychiatry 2009 Jul 11;14(2):129-134. [doi: 10.1080/09540260220132644]
4. Seligman LD, Ollendick TH. Cognitive-behavioral therapy for anxiety disorders in youth. Child Adolesc Psychiatr Clin N Am 2011 Apr;20(2):217-238 [FREE Full text] [doi: 10.1016/j.chc.2011.01.003] [Medline: 21440852]
5. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
6. Calear AL, Christensen H. Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents. Med J Aust 2010;192(11 Suppl):S12-S14. [Medline: 20528700]
7. 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]
8. Skinner H, Biscope S, Poland B, Goldberg E. How adolescents use technology for health information: implications for health professionals from focus group studies. J Med Internet Res 2003 Dec 18;5(4):e32 [FREE Full text] [doi: 10.2196/jmir.5.4.e32] [Medline: 14713660]

9. Podina IR, Mogoase C, David D, Szentagotai A, Dobrean A. A meta-analysis on the efficacy of technology mediated CBT for anxious children and adolescents. J Rat-Emo Cognitive-Behav Ther 2015 Nov 17;34(1):31-50. [doi: 10.1007/s10942-015-0228-5]
10. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: a systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
11. Spence S, March S, Vigerland S, Serlachius E. Internet-based therapies for child adolescent emotional behavioral problems. Guid internet-based Treat psychiatry Cham: Springer International Publishing; 2016:197-217.
12. Essau CA, Lewinsohn PM, Olaya B, Seeley JR. Anxiety disorders in adolescents and psychosocial outcomes at age 30. J Affect Disord 2014 Jul;163:125-132 [FREE Full text] [doi: 10.1016/j.jad.2013.12.033] [Medline: 24456837]
13. Neil AL, Christensen H. Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety. Clin Psychol Rev 2009 Apr;29(3):208-215. [doi: 10.1016/j.cpr.2009.01.002] [Medline: 19232805]
14. Donovan CL, Spence SH. Prevention of childhood anxiety disorders. Clin Psychol Rev 2000 Jun;20(4):509-531. [Medline: 10832552]
15. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
16. Chatterjee S, Price A. Healthy living with persuasive technologies: framework, issues, and challenges. J Am Med Inform Assoc 2009 Mar;16(2):171-178 [FREE Full text] [doi: 10.1197/jamia.M2859] [Medline: 19074300]
17. Oinas-Kukkonen H, Harjumaa M. A Systematic Framework for Designing and Evaluating Persuasive Systems. In: PERSUASIVE 2008: Persuasive Technology. Berlin, heidelberg: Springer; 2008 Presented at: International Conference on Persuasive Technology; 2008; Oulu, Finland p. 164-176. [doi: 10.1007/978-3-540-68504-3_15]
18. Fogg BJ. Persuasive technology: using computers to change what we think and do. Ubiquity 2002 Dec 01;2002(5):-. [doi: 10.1145/764008.763957]
19. Hill C, Creswell C, Vigerland S, Nauta MH, March S, Donovan C, et al. Navigating the development and dissemination of internet cognitive behavioral therapy (iCBT) for anxiety disorders in children and young people: a consensus statement with recommendations from the #iCBTLorentz Workshop Group. Internet Interv 2018 Jun;12:1-10 [FREE Full text] [doi: 10.1016/j.invent.2018.02.002] [Medline: 30135763]
20. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, et al. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
21. Rooksby M, Elouafkaoui P, Humphris G, Clarkson J, Freeman R. Internet-assisted delivery of cognitive behavioural therapy (CBT) for childhood anxiety: systematic review and meta-analysis. J Anxiety Disord 2015 Jan;29:83-92. [doi: 10.1016/j.janxdis.2014.11.006] [Medline: 25527900]
22. 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]
23. Spence SH, Donovan CL, March S, Kenardy JA, Hearn CS. Generic versus disorder specific cognitive behavior therapy for social anxiety disorder in youth: a randomized controlled trial using internet delivery. Behav Res Ther 2017 Dec;90:41-57. [doi: 10.1016/j.brat.2016.12.003] [Medline: 27988427]
24. Calear AL, Batterham PJ, Poyser CT, Mackinnon AJ, Griffiths KM, Christensen H. Cluster randomised controlled trial of the e-couch Anxiety and Worry program in schools. J Affect Disord 2016 May 15;196:210-217. [doi: 10.1016/j.jad.2016.02.049] [Medline: 26926660]
25. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
26. Calear A, Christensen H, Mackinnon A, Griffiths K. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
27. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005 Jul;10(Suppl 1):21-34. [doi: 10.1258/1355819054308530] [Medline: 16053581]
28. Pawson R, Tilley N. Realistic Evaluation. London: Sage; 1997.
29. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. BMC Med 2016 Dec;14(1):96 [FREE Full text] [doi: 10.1186/s12916-016-0643-1] [Medline: 27342217]
30. Crutzen R, Cyr D, de Vries NK. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012 Mar 09;14(2):e45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]

31. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
32. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
33. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
34. Gasser R, Brodbeck D, Degen M, Luthiger J, Wyss R, Reichlin S. Persuasiveness of a mobile lifestyle coaching application using social facilitation. Berlin, Heidelberg: Springer; 2006 May 18 Presented at: International Conference on Persuasive Technology; May 18, 2006; Eindhoven, The Netherlands p. 27-38. [doi: 10.1007/11755494_5]
35. Walker JL. The use of saturation in qualitative research. Can J Cardiovasc Nurs 2012 Apr;22(2):37-41. [Medline: 22803288]
36. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012 Oct;22(3):276-282 [FREE Full text] [Medline: 23092060]
37. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5. Washington: American Psychiatric Pub; 2013.
38. Pluye P, Gagnon M, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
39. Pluye P, Robert E, Cargo M, Bartlett G. Mixed Methods Appraisal Tool. 2011 Jan 19. Proposal: A mixed methods appraisal tool for systematic mixed studies reviews URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/84371689/MMAT%202011%20criteria%20and%20tutorial%202011-06-29updated2014.08.21.pdf [WebCite Cache ID 5tTRTc9yJ]
40. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. Int J Nurs Stud 2012 Jan;49(1):47-53. [doi: 10.1016/j.ijnurstu.2011.07.002] [Medline: 21835406]
41. Institute of Medicine. In: Mrazek PJ, Haggerty RJ, editors. Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research. Washington: The National Academies Press; 1994.
42. Connolly SD, Bernstein GA, Work Group on Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with anxiety disorders. J Am Acad Child Adolesc Psychiatry 2007 Feb;46(2):267-283. [doi: 10.1097/01.chi.0000246070.23695.06] [Medline: 17242630]
43. Atkins S, Lewin S, Smith H, Engel M, Fretheim A, Volmink J. Conducting a meta-ethnography of qualitative literature: lessons learnt. BMC Med Res Methodol 2008 Apr 16;8(1):21 [FREE Full text] [doi: 10.1186/1471-2288-8-21] [Medline: 18416812]
44. Noblit GW, Hare RD. Meta-Ethnography: Synthesizing qualitative studies. Newbury Park, California: SAGE; 1988.
45. Byng R, Norman I, Redfern S. Using realistic evaluation to evaluate a practice-level intervention to improve primary healthcare for patients with long-term mental illness. Evaluation 2016 Jul 24;11(1):69-93. [doi: 10.1177/1356389005053198]
46. Pramana G, Parmanto B, Kendall PC, Silk JS. The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. Telemed J E Health 2014 May;20(5):419-427 [FREE Full text] [doi: 10.1089/tmj.2013.0214] [Medline: 24579913]
47. Patwardhan M. Assessing the impact of usability design features of an mHealth app on clinical protocol compliance using a mixed methods approach. Arizona State University: ProQuest Dissertations Publishing; 2016. URL: https://repository.asu.edu/attachments/172769/content/Patwardhan_asu_0010N_16210.pdf [accessed 2018-11-28] [WebCite Cache ID 74G7RNGAp]
48. Shahnavaz S. Karolinska Institutet. 2016 Dec 02. Cognitive behavioral therapy for children and adolescents with dental phobia URL: https://openarchive.ki.se/xmlui/bitstream/handle/10616/45319/Thesis_Shervin_Shahnavaz.pdf?sequence=5&isAllowed=y
49. Spence SH, Holmes JM, March S, Lipp OV. The feasibility and outcome of clinic plus internet delivery of cognitive-behavior therapy for childhood anxiety. J Consult Clin Psychol 2006 Jun;74(3):614-621. [doi: 10.1037/0022-006X.74.3.614] [Medline: 16822117]
50. Vigerland S, Thulin U, Ljótsson B, Svirsky L, Ost L, Lindefors N, et al. Internet-delivered CBT for children with specific phobia: a pilot study. Cogn Behav Ther 2013;42(4):303-314. [doi: 10.1080/16506073.2013.844201] [Medline: 24245708]
51. Keller ML. An Internet Cognitive-Behavioral Skills-Based Program for Child Anxiety. Charleston, SC: Proquest, Umi Dissertation Publishing; 2011.
52. Calear AL, Christensen H, Griffiths KM, Mackinnon A. The Y-Worri Project: study protocol for a randomised controlled trial. Trials 2013 Mar 19;14:76 [FREE Full text] [doi: 10.1186/1745-6215-14-76] [Medline: 23506049]
53. Calear AL, Christensen H, Mackinnon A, Griffiths KM, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]

54. Vigerland S. Karolinska Institutet. 2015. Internet-delivered CBT for children with anxiety disorders: effect and prerequisites for implementation within public health care URL: https://openarchive.ki.se/xmlui/bitstream/handle/10616/44877/Thesis_Sarah__Vigerland.pdf?sequence=1&isAllowed=y
55. Silfvernagel K, Gren-Landell M, Emanuelsson M, Carlbring P, Andersson G. Individually tailored internet-based cognitive behavior therapy for adolescents with anxiety disorders: a pilot effectiveness study. Internet Interv 2015 Sep;2(3):297-302. [doi: 10.1016/j.invent.2015.07.002]
56. Calear AL, Batterham PJ, Griffiths KM, Christensen H. Generalized anxiety disorder stigma in adolescents: personal and perceived stigma levels and predictors. Stigma and Health 2017 Aug;2(3):208-215. [doi: 10.1037/sah0000046]
57. Lipsitz JD, Markowitz JC. Mechanisms of change in interpersonal therapy (IPT). Clin Psychol Rev 2013 Dec;33(8):1134-1147 [FREE Full text] [doi: 10.1016/j.cpr.2013.09.002] [Medline: 24100081]
58. Spence S. Australian New Zealand Clinical Trials Registry (ANZCTR). 2011. Factors predicting therapy outcome in online cognitive behaviour therapy (with minimal therapist assistance) in the treatment of anxiety disorders among children and adolescents URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=343375&isReview=true [accessed 2017-11-15]
59. Spence S. Australian New Zealand Clinical Trials Registry (ANZCTR). 2011. A comparison of the effects of online versus clinic-based delivery of CBT in reducing anxiety disorders and anxiety symptoms among clinically anxious adolescents URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=343374&isReview=true [accessed 2017-11-15]
60. Stasiak K. Australian New Zealand Clinical Trials Registry (ANZCTR). 2012. Open trial of the effectiveness of BRAVE - an internet-based cognitive behavioural therapy for children with anxiety following the Canterbury earthquakes URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=347855&isReview=true
61. Spence SH. The University of Queensland. 2017 Jan 01. BRAVE-Online - Helping young people overcome anxiety URL: http://www.brave-online.com/ [accessed 2018-11-28] [WebCite Cache ID 74GnozddA]
62. Shahnavaz S. ClinicalTrials.gov. 2015. Internet-based cognitive behavioral therapy for children with dental anxiety URL: https://clinicaltrials.gov/ct2/show/NCT02588079 [WebCite Cache ID 6xrmqzKrS]
63. Serlachius E. ClinicalTrials.gov. Internet-delivered cognitive behavior therapy (CBT) for children age 8-12 years with anxiety disorders URL: https://clinicaltrials.gov/ct2/show/NCT01533402?term=internet+cbt+anxiety+children&rank=1 [WebCite Cache ID 6xrmmXI3T]
64. Serlachius E. Clinicaltrials.gov. 2014. Internet-delivered treatment for children with anxiety disorders in a rural area; An open trial in a clinical setting URL: https://clinicaltrials.gov/ct2/show/NCT02306356?term=internet+cbt+anxiety+children&rank=5 [WebCite Cache ID 6xrmisrng]
65. Serlachius E. ClinicalTrials.gov. 2016. Internet-based CBT for children with anxiety disorders: implementation in clinical settings URL: https://clinicaltrials.gov/ct2/show/NCT02926365?term=internet+cbt+anxiety+children&rank=12 [WebCite Cache ID 6xrmYyHIX]
66. Vigerland S. ClinicalTrials.gov. 2016. BIP in Jämtland Härjedalen: increased access to CBT within regular health care in northern Sweden URL: https://clinicaltrials.gov/ct2/show/NCT02926365?term=internet+cbt+anxiety+children&rank=12 [WebCite Cache ID 6xrlBvLOc]
67. Stockholm's Child and Adolescent Psychiatry (BUP). 2013. [About BUP] URL: http://www.bup.se/sv/Om-BUP/Forskning/BIP/ [WebCite Cache ID 6xrl0u9AX]
68. Silk J. ClinicalTrials.gov. 2014. Smartphones to enhance the treatment of childhood anxiety (Smart-CAT) URL: https://clinicaltrials.gov/ct2/show/NCT02259036?term=smartcat&rank=1 [WebCite Cache ID 6xrjvDTlG]
69. Parmanto B, Silk J, Pramana G. The Office of the National Coordinator for Health Information Technology (ONC). 2013. SmartCAT: Smartphone-enhanced Child Anxiety Treatment URL: https://behavioralhealth.devpost.com/submissions/17038-smartcat-smartphone-enhanced-child-anxiety-treatment [accessed 2018-03-13] [WebCite Cache ID 6xrrmJjCh]
70. Keller M. Anxiety and Depression Research Center at UCLA. 2009. Internet skills-based program for child anxiety URL: https://anxiety.psych.ucla.edu/childinternetstudy [accessed 2018-11-28] [WebCite Cache ID 74GpY938l]
71. Keller M, Craske M. Anxiety and Depression Research Center at UCLA. 2009. Internet cognitive-behavioral skills-based program study flyer URL: http://anxiety.psych.ucla.edu/files/CBSBP_Flyer.pdf [WebCite Cache ID 6xrkat4hH]
72. Calear A. Australian New Zealand Clinical Trials Registry (ANZCTR). 2010. Effect of the e-coucAnxiety and Worry program on generalised anxiety disorder in an adolescent school-based population URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=335498&isReview=true [WebCite Cache ID 6xrnCcz5N]
73. Griffiths K, Tayler G, Christensen H. Australian National University. 2017. e-couch Self Help URL: https://ecouch.anu.edu.au/new_users/welcome03 [WebCite Cache ID 6xrn27S1F]
74. Christensen H, Griffiths K. ehub Health. 2017. MoodGYM CBT training program URL: https://moodgym.com.au/ [WebCite Cache ID 6xrjOa47y]
75. Spence S. Australian New Zealand Clinical Trials Registry (ANZCTR). 2011. Is disorder-specific CBT more effective than transdiagnostic CBT in reducing social phobia in the online-delivered treatment of socially phobic children and adolescents? URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=343377&isReview=true [accessed 2017-11-15]

76. Spence S, Donovan C, March S, Gamble A, Anderson R, Prosser S, et al. Online CBT in the treatment of child and adolescent anxiety disorders: issues in the development of BRAVE-ONLINE and two case illustrations. Behav Cognit Psychother 2008 Jun 25;36(04):411-430. [doi: 10.1017/S135246580800444X]
77. March S. Investigation of an internet-based treatment for childhood anxiety (BRAVE for Children-ONLINE): An evaluation of efficacy and factors associated with treatment outcome. St. Lucia, Queensland: The University of Queensland; 2008.
78. 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]
79. Stasiak K, Merry SN, Frampton C, Moor S. Delivering solid treatments on shaky ground: feasibility study of an online therapy for child anxiety in the aftermath of a natural disaster. Psychother Res 2018 Jul;28(4):643-653. [doi: 10.1080/10503307.2016.1244617] [Medline: 27781568]
80. Spence SH, Donovan CL, March S, Gamble A, Anderson RE, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
81. Anderson RE, Spence SH, Donovan CL, March S, Prosser S, Kenardy J. Working alliance in online cognitive behavior therapy for anxiety disorders in youth: comparison with clinic delivery and its role in predicting outcome. J Med Internet Res 2012;14(3):e88 [FREE Full text] [doi: 10.2196/jmir.1848] [Medline: 22789657]
82. Conaughton RJ, Donovan CL, March S. Efficacy of an internet-based CBT program for children with comorbid High Functioning Autism Spectrum Disorder and anxiety: a randomised controlled trial. J Affect Disord 2017 Aug;218:260-268. [doi: 10.1016/J.JAD.2017.04.032]
83. Vigerland S, Ljótsson B, Thulin U, Öst L, Andersson G, Serlachius E. Internet-delivered cognitive behavioural therapy for children with anxiety disorders: a randomised controlled trial. Behav Res Ther 2016 Jan;76:47-56 [FREE Full text] [doi: 10.1016/j.brat.2015.11.006] [Medline: 26649465]
84. Vigerland S, Serlachius E, Thulin U, Andersson G, Larsson J, Ljótsson B. Long-term outcomes and predictors of internet-delivered cognitive behavioral therapy for childhood anxiety disorders. Behav Res Ther 2017 Dec;90:67-75. [doi: 10.1016/j.brat.2016.12.008] [Medline: 28012300]
85. Jolstedt M, Ljótsson B, Fredlander S, Tedgård T, Hallberg A, Ekeljung A, et al. Implementation of internet-delivered CBT for children with anxiety disorders in a rural area: a feasibility trial. Internet Interv 2018 Jun 01;12:121-129. [doi: 10.1016/J.INVENT.2017.11.003]
86. Silfvernagel K. Individually tailored internet-based cognitive behavioural therapy for adolescents, young adults and older adults with anxiety. Linköping, Sweden: Linköping University; 2017.
87. Calear AL, Christensen H, Brewer J, Mackinnon A, Griffiths KM. A pilot randomized controlled trial of the e-couch anxiety and worry program in schools. Internet Interv 2016 Nov;6:1-5. [doi: 10.1016/j.invent.2016.08.003]
88. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
89. Wozney L, Huguet A, Bennett K, Radomski AD, Hartling L, Dyson M, et al. How do eHealth programs for adolescents with depression work? A realist review of persuasive system design components in internet-based psychological therapies. J Med Internet Res 2017 Aug 09;19(8):e266 [FREE Full text] [doi: 10.2196/jmir.7573] [Medline: 28793983]
90. Wildeboer G, Kelders SM, van Gemert-Pijnen JE. The relationship between persuasive technology principles, adherence and effect of web-based interventions for mental health: a meta-analysis. Int J Med Inform 2016 Dec;96:71-85. [doi: 10.1016/j.ijmedinf.2016.04.005] [Medline: 27117057]
91. Baumel A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. Transl Behav Med 2018 Sep 08;8(5):793-798. [doi: 10.1093/tbm/ibx037] [Medline: 29471424]
92. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]
93. Langrial SU, Al Araimi FA. Promoting Pediatric Healthcare through Persuasive Information Systems: A Qualitative Study. 2017 Presented at: PACIS 2017 Proceedings; 2017; Langkawi, Malaysia p. 30.
94. Anderson R. New MRC guidance on evaluating complex interventions. Br Med J 2008 Oct 22;337:a1937. [Medline: 18945728]
95. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. JMIR Hum Factors 2015;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
96. Rooksby J, Rouncefield M, Sommerville I. Testing in the wild: the social and organisational dimensions of real world practice. Comput Supported Coop Work 2009 Sep 3;18(5-6):559-580. [doi: 10.1007/s10606-009-9098-7]
97. Howe D, Batchelor S, Coates D, Cashman E. Nine key principles to guide youth mental health: development of service models in New South Wales. Early Interv Psychiatry 2014 May;8(2):190-197. [doi: 10.1111/eip.12096] [Medline: 24251956]

98. Andersson G, Carlbring P, Berger T, Almlöv J, Cuijpers P. What makes Internet therapy work? Cogn Behav Ther 2009 Jan;38(Suppl 1):55-60. [doi: 10.1080/16506070902916400] [Medline: 19675956]
99. 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]
100. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
101. 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]
102. Howgego IM, Yellowlees P, Owen C, Meldrum L, Dark F. The therapeutic alliance: the key to effective patient outcome? A descriptive review of the evidence in community mental health case management. Aust N Z J Psychiatry 2003 Apr;37(2):169-183. [Medline: 12656956]
103. 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]
104. 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 Apr 11;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
105. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
106. O'Leary K, Liu L, Mcclure J, Ralston J, Pratt W. Persuasive reminders for health self-management. 2016 Presented at: AMIA Annu Symp Proc American Medical Informatics Association; 2016; Chicago, Illinois p. 994-1003.
107. Lewis ZH, Swartz MC, Lyons EJ. What's the point?: A review of reward systems implemented in gamification interventions. Games Health J 2016 Apr;5(2):93-99. [doi: 10.1089/g4h.2015.0078] [Medline: 26812253]
108. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. Br Med J 2012;344:e2598 [FREE Full text] [Medline: 22517917]
109. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
110. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: efficient tools for evaluation of intervention components. Am J Prev Med 2014 Oct;47(4):498-504 [FREE Full text] [doi: 10.1016/j.amepre.2014.06.021] [Medline: 25092122]
111. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
112. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, Smith SS, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. Ann Behav Med 2011 Apr;41(2):208-226 [FREE Full text] [doi: 10.1007/s12160-010-9253-x] [Medline: 21132416]
113. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
114. Alfonsson S, Olsson E, Hursti T. The effects of therapist support and treatment presentation on the clinical outcomes of an Internet based applied relaxation program. Internet Interv 2015 Sep 01;2(3):289-296. [doi: 10.1016/j.invent.2015.07.005]
115. Rocha A, Camacho R, Ruwaard J, Riper H. Using multi-relational data mining to discriminate blended therapy efficiency on patients based on log data. Internet Interv 2018 Jun 01;12:176-180 [FREE Full text] [doi: 10.1016/j.invent.2018.03.003] [Medline: 30135781]
116. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb;40(1):85-98. [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]
117. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. Stud Health Technol Inform 2013;192:657-661. [Medline: 23920638]
118. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
119. Halko S, Kientz JA. Personality and persuasive technology: an exploratory study on health-promoting mobile applications. Berlin, Heidelberg: Springer; 2010 Presented at: International conference on persuasive technology; June 7, 2010; Copenhagen, Denmark p. 150-161.
120. Karanam Y, Filko L, Kaser L, Alotaibi H, Makhsoom E, Voida S. Motivational affordances and personality types in personal informatics. 2014 Sep 13 Presented at: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication; 2014; Seattle, WA p. 79-82. [doi: 10.1145/2638728.2638800]

121. Xu Y, Poole E, Miller A, Eiriksdottir E, Kestranek D, Catrambone R, et al. This is not a one-horse race: understanding player types in multiplayer pervasive health games for youth. : ACM; 2012 Feb 11 Presented at: Proceedings of the ACM 2012 conference on computer supported cooperative work; February 11-15, 2012; Seattle, WA p. 843-852. [doi: 10.1145/2145204.2145330]
122. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):e20 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
123. 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]
124. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]

Abbreviations

AACAP: American Academy of Child and Adolescent Psychiatry
CBT: cognitive behavioral therapy
CIHR: Canadian Institutes of Health Research
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials-electronic health
iCBT: internet-based cognitive behavioral therapy
MMAT: mixed method appraisal tool
PSD: persuasive system design

Edited by G Eysenbach; submitted 25.05.18; peer-reviewed by S Kelders, J Povey; comments to author 05.07.18; revised version received 26.09.18; accepted 20.10.18; published 05.02.19.

Please cite as:
Radomski AD, Wozney L, McGrath P, Huguet A, Hartling L, Dyson MP, Bennett K, Newton AS
Design and Delivery Features That May Improve the Use of Internet-Based Cognitive Behavioral Therapy for Children and Adolescents With Anxiety: A Realist Literature Synthesis With a Persuasive Systems Design Perspective
J Med Internet Res 2019;21(2):e11128
URL: https://www.jmir.org/2019/2/e11128/
doi:10.2196/11128
PMID:30720436

Original Paper

How Do Adolescents Use Electronic Diaries? A Mixed-Methods Study Among Adolescents With Depressive Symptoms

Kiki Metsäranta[1*], RN, MNSc; Marjo Kurki[2*], PhD; Maritta Valimaki[2,3*], PhD; Minna Anttila[2*], PhD

[1]Department of Nursing Science, University of Turku, Turun yliopisto, Finland

[2]Department of Nursing Science, University of Turku, Turku, Finland

[3]Hong Kong Polytechnic University, Hong Kong, China

[*]all authors contributed equally

Corresponding Author:
Kiki Metsäranta, RN, MNSc
Department of Nursing Science
University of Turku
Hoitotieteen laitos
Joukahaisenkatu 3-5
Turun yliopisto, 20014
Finland
Phone: 358 456716156
Email: kianme@utu.fi

Abstract

Background: Depression in adolescence is common. Less than half of the adolescents with depression receive mental health care; furthermore, treatment tends to be suspended, and its success rates are low. There is a need for these adolescents to have a safe place to share their thoughts. Studies have shown that writing may be a useful treatment method for people with mental health problems.

Objective: This study aims to describe the use of an electronic diary (e-diary) among adolescents with depressive symptoms.

Methods: This paper describes a substudy of a randomized controlled trial. We used a mixed-methods approach to understand the way in which e-diaries were used by participants in the intervention under the randomized controlled trial. Data were collected during 2008-2010 at 2 university hospitals in Finland. Study participants (N=89) were 15-17-year-old adolescents who had been referred to an adolescent outpatient psychiatric clinic due to depressive symptoms. Participants were instructed to use the e-diary at least once a week to describe their thoughts, feelings, and moods. The content of the e-diary data was analyzed using descriptive statistics and inductive content analysis.

Results: Overall, 53% (47/89) of the adolescents used the e-diary. Most of them (39/47, 83%) logged into the program during the first week, and about one-third (19/47, 40%) logged into the e-diary weekly as suggested. The number of words used in the e-diary per each log ranged between 8 and 1442 words. The 3 topics most often written about in the e-diary were related to mental health problems (mental disorder), social interaction (relationship), and one's own development (identity).

Conclusions: An e-diary may be a usable tool to reflect experiences and thoughts, especially among adolescents who have signs of depression. The results of this study can be used to develop user-centered electronic health applications that allow users to express their own thoughts and experiences in ways other than systematic mood monitoring.

(J Med Internet Res 2019;21(2):e11711) doi:10.2196/11711

KEYWORDS

adolescent; depression; electronic diary; mental health; mobile phone; outpatient care

Introduction

Depression in adolescence is common, and it causes serious impairments of one's social, academic, and emotional lives [1]. Depression is often unrecognized [2] and is an undertreated disorder [1]. It may increase the risk of drug use [3] and suicide among adolescents [2]. More than half of the adolescents with depression do not receive mental health care [4]. Those who have access to care tend to drop out prematurely, or treatment is often unsuccessful [5].

Adolescents need a safe channel for sharing their concerns and feelings [6]. Therefore, new treatment methods, such as Web-based interventions, need more attention. Web-based interventions have already been widely used among adolescents with depression [7]. A commonly used therapeutic approach in these interventions is the cognitive behavioral therapy [7-9]. Besides structured mood tracking with validated instruments, Web-based interventions often include a variety of elements [7], such as email remainders, diaries, discussion forums, or worksheets, focusing on goal setting, problem solving, and thought evaluation [8]. On the other hand, multiple elements in interventions make it difficult to assess the elements that are most influential in increasing the effectiveness of the intervention [7]. In addition, systematic and structured mood tracking does not allow adolescents to express their own emotions, feelings, or current concerns. Therefore, it may be valuable to assess how usable electronic diaries (e-diaries) are to adolescents in supporting them in expressing their burden using written and visual formats.

In this study, an e-diary was used as a part of a Web-based intervention. We assumed that adolescents would be motivated to express their own perceptions of their life situation. Therefore, they had an opportunity to share their thoughts on specific weekly themes [10]. At the same time, they had a chance to share their thoughts freely in the e-diary via expressive writing [11]. Expressive writing is suitable for different therapeutic orientations [11]. It has already been used as a treatment method in mental health [11-13] as well as for those people with depression and anxiety disorders [14]. Writing can support a person as they approach their emotions or modulate the emotional intensity of those who may be less in touch with their emotions [15]. Writing can also decrease stress and generate a desired change [13].

In general, diary writing can help adolescents to cope with numerous developmental challenges [13,15]. Diary writers describes or comments on an event that has happened and explores their feelings that occur almost simultaneously to writing in the diary [16]. Diary writing can create a narrative of events, thoughts, hopes, and emotions. Writing in a diary can, therefore, be seen as the self-reflection of one's own life [17]. However, knowledge is still lacking on how a Web-based e-diary could be used for adolescents with depression. In this study, we aimed is to describe how adolescents with depressive symptoms use e-diary, which was a part of a Web-based program for adolescents with depression.

This study is a substudy of the Depis.Net-study on the effectiveness of a Web-based intervention for adolescents with depression (ISRCTN80379583). The study included 6 adolescent psychiatry outpatient clinics at 2 Finnish University Hospitals, where psychiatric examination and treatment were offered to adolescents between 13 and 17 years of age. Data were gathered using the Depis.Net-program, developed at the University of Turku, Department of Nursing Science [10]. The theoretical framework of the program is based on the self-determination theory [18], and it is tailored to improve adolescents' self-management skills and awareness of their own well-being and mental health [10].

Methods

Design

A mixed-methods study design with quantitative and qualitative descriptive data [19] was used to provide a deep understanding of the studied phenomena and the descriptive approach. In this study, the quantitative data were used to describe how adolescents used the e-diary, while the qualitative data were used to identify the contents adolescents discussed in their e-diaries. We assumed that using quantitative and qualitative approaches together will allow complementary information to arise—the quantitative data augmented the qualitative data, which were analyzed with the method of content analysis [20]. The criteria of Tong et al [21] were followed for reporting the qualitative data.

Participants

Participants (N=89) in this study were adolescents who were between 15 and 17 years of age at the time of obtaining informed consent. This study is a part of a two-arm randomized controlled trial (Depis.Net), and participants of this paper were adolescents who were randomized solely into an intervention group. The inclusion criteria for the randomized controlled trial were as follows: those aged 15-17 (at the time of recruitment) years and able to speak Finnish. Adolescents with serious mental disorders, such as psychotic depression, bipolar disorder, substance abuse, or a primary eating disorder, were excluded; those admitted to psychiatric hospital wards or involved in a brief intervention at an outpatient clinic (≤3 appointments) were also excluded. Participants in the intervention group were given personal identifiers (IDs) for logging into the Depis.Net program, which included an e-diary. They could use the program from any computer with internet access. We included in the analysis all potential participants of the intervention group who had an opportunity to use the e-diary.

Data Collection

Participants' information, including their age, gender, education, previous use of mental health services, previous depression, and medication use during the last 6 months, was collected from medical records by nurses at the outpatient clinics. Outpatient clinic nurses assessed the overall severity of disturbance using the Children's Global Assessment Scale (C-GAS) [22]. Depressive symptoms were identified using the Beck Depression Inventory (BDI-21) [23], which is a self-assessment instrument that has been validated for adolescents [24].

The Depis.Net intervention took 6 weeks—a 1-week introduction and 5 weeks during which themes changed weekly (well-being, home and family, adolescents' rights and responsibilities, adolescent depression, and treatment of adolescents' depression). The intervention also included health information, self-monitoring, and the use of e-diary where adolescents could reflect on their own life situation. The e-diary included technical instructions as well as personal writings that other adolescents could not see. Adolescents were given ideas related to the weekly theme, which aimed to support them in the use of the e-diary. They were asked to describe their current thoughts, feelings, and moods at least once a week, whenever

they had time and the willingness to do so. They were also encouraged to use extra characters in the text, such as emoticons, pictures, song lyrics, or comics. They had access to material from all weeks and the e-diary during the 6-week program. It was possible to use the e-diary at any time within any week because the weeks did not close after they had been activated.

To support adolescents in their use of the e-diary, supportive short message service (SMS) text messages were sent every Monday. They informed participants of the theme of that week. Feedback based on previous week was also offered to participants. In addition, those who had not used the e-diary by Thursday received an additional SMS text message to encourage them to write in it. Those who had used the e-diary during the week received feedback from the trained tutor through the program. Trained tutors read adolescents' e-diary entries daily. If any severe concern was identified in an adolescent's text, such as a threat of suicide, the trained tutors informed it to the responsible nurse or physician.

Data Analysis

Descriptive quantitative statistical methods (frequencies, averages) were used [25], and qualitative data were analyzed using content analysis [20]. First, using descriptive statistics, the content of e-diaries was described by how often and how much adolescents wrote and used pictures (eg, emoticons, combination of signs, photos, or drawings). The number of words was calculated, and the pictures were checked. For example, *I should do lots of schoolwork at some point (I hate school >:/)....i.e. it seems in fact, that I have quite a lot of stress! 8O So typical* >_<"" included 27 words and 3 pictures (*>:/, 8O,* >_<""). After this, an Excel table was generated to gather detailed information from 1 week at a time. The table was used to obtain a general overview of the 6-week period. The e-diary program made one automatic timestamp when an adolescent logged in. Automatic timestamps and the time of logging in were calculated. Background information on adolescents' baseline characteristics and BDI-21 and C-GAS scores were analyzed by frequencies and percentages.

Then, data were analyzed using inductive content analysis made from the manifest content. The sentence was used as a unit of analysis because, with smaller (word) or larger analysis units, the overall vision or some essential information could go unnoticed. For the inductive content analysis, printed e-diary entries were read a number of times to understand their content clearly. After this, the sentences that corresponded to the question "what issues have been addressed in the diaries" were manually highlighted [20]. These sentences were transferred into a separate document and categorized by relevant phrases [20,26], which were coded (a coding phase) [20]. The content of the e-diaries was then exported to the qualitative data analysis software, NVivo, where the coding of the sentences continued [20,27]. The codes and their content were compared to identify differences and similarities; subcategories were further formed. In the next step, these subcategories were abstracted and organized into categories based on the similarity of their content. A total of 7 subcategories were formed, which were further combined into 3 categories and named through reflection and discussions with the cowriters. [20]

To increase the validity of the analysis during the analysis process, the data were coded independently by 2 coders (KM and MK) and categories were formed by 3 authors (KM, MK, and MA) [21]. Comments and thoughts were written in a notebook, which aimed to make the internal thinking visible. The notes also allowed the ideas related to the interpretation of the data to be revisited. Conclusions in all the phases were compared with the original data [26].

Ethical Issues

The ethical approval was received from the ethics committee (#R08075H). Hospital administrators granted permissions to conduct the study. Good scientific practices [28] and the principles for medical research and legislation [29] were adhered to. The research assistants provided verbal and written information about the purpose and process of the study to adolescents. Adolescents' participation in the trial was voluntary, and they gave written informed consent. It was possible to withdraw from the study without giving any reason, and it did not affect treatment. If there were severe concerns, such as suicidal ideation, a research assistant informed the responsible nurse or physician to ensure the safety of the adolescent. Anonymity was ensured by having personal access to the program only with a username and password. Unidentifiable codes (IDs) were used, and only the research assistants had access to the program. Anonymity was guaranteed in all phases of the study and reporting of the results. Printed e-diaries included only the unidentifiable IDs. They were stored in a locked space, and the e-diary texts were protected by passwords.

Results

Characteristics of Electronic Diary Users

Of the 89 participants, 47 used the e-diary. Almost all e-diary users were females. The majority of them were in high school compared to nonusers, many of whom were in a vocational school. More e-diary users had had previous episodes of depression. They had also used mental health services and medication more frequently than nonusers. According to BDI-21, more e-diary users had moderate or severe depression symptoms than nonusers. However, nonusers had slightly more serious problems in functioning (C-GAS) than did e-diary users. Characteristics of the e-diary users and nonusers are described in Table 1.

Table 1. Characteristic of electronic diary users and nonusers.

Characteristics	Users, n (%)	Nonusers, n (%)
Age (Years)		
15	16 (34)	9 (22)
16	17 (36)	21 (51)
17	13 (28)	11 (27)
18[a]	1 (2)	0 (0%)
Gender		
Female	45 (96)	21 (51)
Male	2 (4)	20 (49)
Education		
Comprehensive school	20 (43)	18 (47)
High school	24 (51)	9 (24)
Vocational school	3 (6)	11 (29)
Previous depression		
Yes	32 (71)	23 (61)
No	13 (29)	15 (49)
Previous use of mental health services		
Yes	32 (71)	26 (68)
No	13 (29)	12 (32)
Medication		
Yes	21 (46)	14 (37)
No	25 (54)	24 (63)
Beck Depression Inventory 21 score		
<10	5 (11)	14 (34)
10-16	6 (13)	10 (24)
17-29	19 (41)	10 (24)
30-63	16 (35)	7 (17)
Children's Global Assessment Scale score		
90-81	1 (2)	1 (3)
80-71	3 (6)	6 (16)
70-61	18 (38)	11 (30)
60-51	14 (30)	7 (19)
50-41	10 (21)	8 (22)
40-31	1 (2)	4 (11)

[a]Respondents who turned 18 years of age after the recruitment.

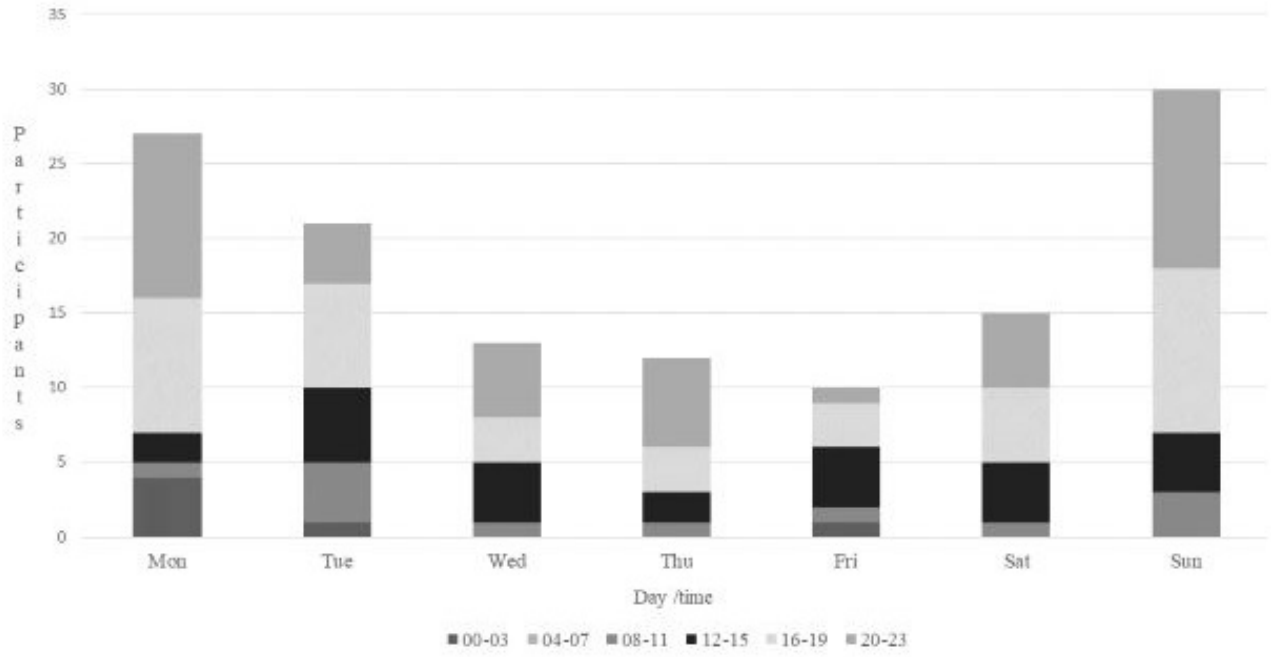
Usage of Electronic Diary

Overall, 53% (47/89) of the participants used the e-diary during data collection. Most of the adolescents (16/47, 34%) logged into the e-diary within 2 weeks, and 21% (10/47) logged into the e-diary at least once during each of the 5 weeks. The adolescents logged into the program a total of 147 times (range 1-13 times; median 2; mean 3.1, SD 2.3). Most adolescents (39/47, 83%) logged into the e-diary during the first week, and over half (27/47, 57%) of the e-diary logs were made during the 2 first weeks. In addition, 21% (10/47) of the adolescents relogged into their e-diary and modified or added their output later. Nearly all (22766/22855, 99.61%) of the e-diary logs were made using text form (words), while pictures were rarely used (89/22855, 0.39%). The most logged themes, based on the number of log-ins, were "well-being" and "adolescents rights and responsibilities" (Table 2). The number of words used ranged from 8 to 1442 per each log (mean 174.2, SD 204.4).

Table 2. Usage of the electronic diary within 1 week and all 5 weeks.

Topic	Words (n=22,766), n (%)	Pictures (n=89), n (%)	Total (N=22,855), n (%)	Logging frequency (N=147), n (%)
Well-being	8081 (35.50)	23 (25.84)	8104 (35.46)	51 (35.70)
Home and family	4286 (18.83)	20 (22.47)	4306 (18.84)	33 (22.45)
Adolescents rights and responsibilities	4877 (21.42)	29 (32.58)	4906 (21.47)	24 (16.33)
Adolescent depression	2742 (12.04)	5 (5.62)	2747 (12.02)	21 (14.29)
Treatment of adolescents' depression	2780 (12.21)	12 (13.48)	2792 (12.22)	18 (12.24)

Figure 1. Logging days and hours.



Sundays to Tuesdays were the most common days for writing in the e-diary, and evenings were the most common time of the day (Figure 1). Adolescents did not use the e-diary at all after 2 am. Logs made in the program at night time (from 23 pm to 2 am) or in the mornings (from 09 am to 12 pm) were also very limited.

Topics of the Electronic Diary

In this study, 3 main categories were formed to represent the topics of the adolescents' e-diaries—mental disorder, relationship, and identity (Table 3).

Mental Disorder

Adolescents discussed and described their symptoms of mental disorder, for example, self-mutilation, substance abuse, and eating problems. They also discussed the destructive thoughts that reflected their moods and maintained their symptoms. They discussed how mental disorder and fatigue decreased their functional and coping capacity. Adolescents described themselves being alone and alienated from others.

They wrote about outpatient care, medication, and how they had tried to help themselves. They highlighted that they feel easier when they have succeeded in adjusting their own thoughts. For example, writing and making e-diary entries in the program were mentioned as helping to clarify feelings and thoughts. Adolescents described different kinds of self-help methods, with which they had tried to find solutions to their mental health problems and improve their well-being. Even though they pointed out what might help them, they wrote that they do not know or have not found ways to obtain relief from mental disorders; there was understanding as to what would help but no strength to accomplish it. They expressed that the things that could help to dissolve bad feeling can also cause them, for example, music or writing.

Adolescents expressed their hope of recovery from mental disorder. They described how they had experienced positive changes that had occurred within their mental health. However, there were descriptions of despondency about never getting better.

Table 3. Topics of the electronic diary.

Main category	Code	Expression[a]
Mental disorder		
Symptoms	Self-harm	I took some scissors and cut long wounds in my legs. (ID[b] 431)
	Suicidal thoughts	I now have fewer thoughts of suicide. I still think about it, but not all the time. (ID 439)
	Tiredness	When I am tired and start thinking about problems, I feel bad. (ID 5)
	Loneliness	I feel lonely when I am with my friends, like I'm in a bubble. (ID 116)
	Distress	I can't find time to relax, or if I have a free moment, I get stressed because I know that there is still school work or something else to be done. This depression never ends. (ID 299)
Treatment	Outpatient care	After the conversation /with the nurse/ my thoughts went from a blob to well-structured body. (ID 79)
	Medication	Compared to the previous weeks I feel better. Probably because the doctor increased my night drugs. (ID 357)
	Self-help	I can write more. This is very relieving. Surprisingly so. I can deal with my thoughts. (ID 290)
		Music always makes me feel both good and bad feelings. (ID 508)
		I've been thinking a lot and have tried many ways to get rid of this bad feeling, but I haven't really found help. (ID 299)
Recovery	Hope	Five years from now, I hope I will be a healthy young woman. (ID 475)
	Despondency	I feel like I'm not getting better and therefore I can't imagine my future. (ID 438)
Relationship		
Supportive	Resource of everyday life	Family, friends, and schoolmates can be a resource. (ID 8)
	Reciprocity	Typically, I am the one who listens to others' problems. Sometimes I'd like for my friends to listen to me, to how I feel. (ID 440)
		I was active at school (I don't usually speak during lessons, just with friends during breaks), and I even spoke with my classmates. (ID 26)
Nonsupportive	Bullying	I have experienced so many different things in my circle of friends, I've been insulted, discriminated against and, bullied in every possible way. (ID 499)
		I'm afraid to bring my own opinions into the classroom because everyone would laugh. (ID 475)
	Conflicts	On Saturday morning was terrible, my dad complained to me about everything and I cried and told him that I did my best and my mum tried to calm him down. (ID 431)
		When my mum is normal it is much easier to be at home. (ID 64)
Identity		
Ego development	Temporal changes	For almost two years my self-esteem has been higher, because I learned not to think too much about others' opinions. (ID 436)
	Strengths	Before, around age 9-12, I had self-esteem problems, but I got over them quite well. (ID 100)
	Weaknesses	I consider myself to be a weak person. Vulnerable, small, pathetic. (ID 449)
	Capacities	I like my way of thinking; my positivity also shows. (ID 116)
Future expectations	Dreams	My plans and dreams cover just the next two years, and during that time I want to go to USA and get a dog and an apartment. And I want to be HAPPY. (ID 147)
	Fears	I think that all my dreams have been crushed, and I don't expect anything from the future. (ID 120)

[a]Translated from Finnish.

[b]ID: Identifier.

Relationship

Adolescents described their involvement in relationships in different situations and events in their communities such as at home, at school, and during leisure-time activities. They discussed their interpersonal relationships with family members, friends, peers, and loved ones. In these writings, adolescents described how they had acted or what had happened, and how relationships have influenced their experience of their environment. In other words, adolescents discussed and described how they feel in different situations with different people.

The nature of relationships was described as being supportive or demanding depending on the behavior and its interpretation. Adolescents expressed the fact that supportive relationships were important resources in everyday life. At the same time, one major fear was losing someone, especially a loved one. They also discussed reciprocity in relationships, how it worked at its best, and what hindered it.

Demanding relationships were discussed in bullying situations and with different kind of difficulties and conflicts with people. Experiences of bullying by friends and schoolmates were described. Conflicts and problems were pointed out, especially with family members. In addition, conflicts with friends and schoolmates were mentioned. Adolescents discussed explanations for these conflicts as being the interpretations of interaction situations, and they discussed how their own and other people's behavior affect how they interact with other people.

Identity

Adolescents discussed their ego development by reflecting on how temporal changes had influenced the way they see themselves and their abilities. They described and discussed their positive and negative self-experiences. Positive expressions were related to adolescents' strengths and things they liked about themselves. Negative expressions were related to weaknesses and things they did not like about themselves; in these writings, adolescents revealed dissatisfaction and difficulty with accepting themselves. They also considered their capacities to reflect on how the experience of oneself was related to the opinions of others and how these opinions affect them. It emerged that thinking of others' opinions undermines self-esteem.

Adolescents wrote about their future expectations, for example, what kind of person they would like to be. Future dreams were related to education, work, having their own home and family. Beside this, there were fears expressed about a future in which dreams and aspirations seemed distant or impossible to achieve.

Discussion

Principal Findings

In this study, we assumed that e-diary writing might be a usable tool for reflecting thoughts for those who have depressive symptoms. Indeed, adolescents included different topics in their e-diaries. They described and discussed negative and positive events in their lives as well as difficulties and good experiences. They also pointed out their hopes and concerns [6] about their future and recovery from mental disorder. Despite the structure and content of the weekly program, adolescents had the chance to bring up concerns and other issues that were important to them, regardless of the week session. We can therefore assume that the instruction of the e-diary supported, but did not restrict, adolescents in dealing with important issues in e-diaries.

On the other hand, 47% (42/89) of the participants did not use the e-diary at all. About one-third (19/47, 40%) logged into the e-diary each week. Välimäki et al [7], in their systematic review and meta-analysis, have identified high dropout rates in Web-based interventions for adolescents. Reasons for low usage in Web-based interventions remain unknown. First, we can assume that adolescents simply may not like to use Web-based interventions because they are out of date or they do not meet their expectations. Second, there is some evidence that young girls in particular prefer to use the internet and smartphone apps for social communication and entertainment but not for purposes related to mental health [9]. Third, adolescents who did not log into the e-diary may have been reluctant to do so if they did not experience any advantage in writing or if they had a fear of disclosing painful or difficult topics [11]. Fourth, as shown in a previous study [12], adolescents with depression may also have difficulties with motivation and concentration, and they may have found Web-based intervention rather demanding. On the other hand, we found in this study that e-diary users more often had a history of depression and previous experience of mental health care services, and their depressive symptoms were more severe compared with those of nonusers. We can therefore assume that, despite burdening depressive symptoms, young people may find an e-diary useful.

We still found that boys were in a clear minority (2/47, 4%) when it came to writing in the e-diary. This is surprising because boys are active users of information technology. On the other hand, diary keeping is a more common habit for girls, especially at the age of 14-15 years [15]. To satisfy boys' needs for technology use, more action and gaming elements might be more usable for young male participants. We assumed at the beginning of the study that adolescents might be reluctant to continue the intervention. Therefore, we sent them reminder SMS text messages and supportive feedback to encourage and motivate their program use [10]. Despite the SMS text message reminder, we found that both participation and e-diary entries decreased after the first week. We can, therefore, assume that SMS text messages did not have much value in supporting adolescents' participation, although a more recent study with adults has showed that participants preferred receiving SMS text message reminders on Mondays and that they preferred slightly humorous SMS text messages [30]. It is also possible that the feedback and SMS text messages were not supportive enough. Gender may also be one explanation since intervention dropouts have previously more often been among women rather than men [31]. While participants' engagement is an important factor for Web-based interventions [23], more studies are still needed to determine whether participation in electronic health applications could be increased [22,24,32]. In future, it would be important to explore whether enhancing real-time support [23] can increase engagement in the Web-based interventions.

Limitations

This study has limitations that should be taken into consideration. First, the data were limited because the number of nonusers increased during the follow-up. Second, the data were collected during 2008-2010, and it could be questioned whether the data are still up to date. However, writing diary entries has a long history, and generally, as in this study, it includes descriptions or comments about events that occur or will occur at about the same time as they are written about [16]. Therefore, we can assume that these data provide us relevant knowledge about topics that concern the minds of adolescents.

On the other hand, technology use has been developed during the last decade, especially mobile technology. If a mobile version of the e-diary was used in this study, the number and time of logs could be different [33]. Third, e-diary entries were written in Finnish, and they represent the quite homogenous nature of Finnish culture. Therefore, the results of this study may not be culturally generalizable. Even so, writing in itself is beneficial for ethnically diverse groups and with a broad spectrum of people, including those who are less socially, academically, or physically advantaged [34]. In general, the results can therefore be valuable in different lingual and cultural regions when engaging Web-based interventions are being designed for young people with mental health problems.

Conclusions

Not all adolescents in this study were willing to use the e-diary to express their emotions or feelings [7]. We also found that adolescents who used the e-diary were more often girls and had current or previous experiences of depression and mental health care services. These adolescents were willing to describe their thoughts about mental health problems and their need for recovery. They also had the ability to reflect on their thoughts in writing, which was supported by specific themes. For future purposes, it is important to explore the reasons for nonuse and suspensions as well as how to support engagement in Web-based interventions.

We are already aware that Web-based interventions with structured monitoring are effective for adolescents with depression [7,9]. Based on this study, we can only assume that an e-diary with an element of expressive writing may have an additional value for some young people, especially for girls, to help them reflect their own thoughts and feelings about their burdening situation. On the other hand, the methods could not reach young boys, who still need more engaging technology where they can share their thoughts, feelings, and moods. In general, participant engagement is an important factor for Web-based interventions [23]. More studies are still needed to identify how participant engagement in Web-based interventions could be supported [22,24,32]. Results of this study can still be used to develop user-centered electronic health applications, which can allow users to express their own thoughts and feelings in ways other than systematic mood monitoring.

Acknowledgments

This study was funded by the Academy of Finland (#8214245), The Hospital District of Southwest Finland (#13893), The Paulo Foundation, the Finnish Cultural Foundation, the Yrjö Jahnsson Foundation (#5920), and the Foundations' Professor Pool.

Conflicts of Interest

None declared.

References

1. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues in clinical neuroscience 2009;11(1):7-20 [FREE Full text] [Medline: 19432384]
2. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. Lancet 2012 Mar 17;379(9820):1056-1067 [FREE Full text] [doi: 10.1016/S0140-6736(11)60871-4] [Medline: 22305766]
3. Gladstone TG, Marko-Holguin M, Rothberg P, Nidetz J, Diehl A, DeFrino DT, et al. An internet-based adolescent depression preventive intervention: study protocol for a randomized control trial. Trials 2015 May 01;16:203 [FREE Full text] [doi: 10.1186/s13063-015-0705-2] [Medline: 25927539]
4. Costello EJ, He J, Sampson NA, Kessler RC, Merikangas KR. Services for adolescents with psychiatric disorders: 12-month data from the National Comorbidity Survey-Adolescent. Psychiatr Serv 2014 Mar 01;65(3):359-366 [FREE Full text] [doi: 10.1176/appi.ps.201100518] [Medline: 24233052]
5. Harpaz-Rotem I, Leslie D, Rosenheck RA. Treatment retention among children entering a new episode of mental health care. Psychiatr Serv 2004 Sep;55(9):1022-1028. [doi: 10.1176/appi.ps.55.9.1022] [Medline: 15345762]
6. Anttila K, Anttila M, Kurki M, Hätönen H, Marttunen M, Välimäki M. Concerns and hopes among adolescents attending adolescent psychiatric outpatient clinics. Child Adolesc Ment Health 2014 Aug 13;20(2):81-88. [doi: 10.1111/camh.12074]
7. 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]

8. Renton T, Tang H, Ennis N, Cusimano MD, Bhalerao S, Schweizer TA, et al. Web-based intervention programs for depression: a scoping review and evaluation. J Med Internet Res 2014 Sep 23;16(9):e209 [FREE Full text] [doi: 10.2196/jmir.3147] [Medline: 25249003]
9. Stasiak K, Fleming T, Lucassen MFG, Shepherd MJ, Whittaker R, Merry SN. Computer-Based and Online Therapy for Depression and Anxiety in Children and Adolescents. J Child Adolesc Psychopharmacol 2016 Apr;26(3):235-245. [doi: 10.1089/cap.2015.0029] [Medline: 26465266]
10. Välimäki M, Kurki M, Hätönen H, Koivunen M, Selander M, Saarijärvi S, et al. Developing an internet-based support system for adolescents with depression. JMIR Res Protoc 2012 Dec 12;1(2):e22 [FREE Full text] [doi: 10.2196/resprot.2263] [Medline: 23612485]
11. Cummings JA, Hayes AM, Saint DS, Park J. Expressive writing in psychotherapy: A tool to promote and track therapeutic change. Professional Psychology: Research and Practice 2014;45(5):378-386. [doi: 10.1037/a0037682]
12. Baikie KA, Geerligs L, Wilhelm K. Expressive writing and positive writing for participants with mood disorders: an online randomized controlled trial. J Affect Disord 2012 Feb;136(3):310-319. [doi: 10.1016/j.jad.2011.11.032] [Medline: 22209127]
13. Boniel-Nissim M, Barak A. The therapeutic value of adolescents' blogging about social-emotional difficulties. Psychol Serv 2013 Aug;10(3):333-341. [doi: 10.1037/a0026664] [Medline: 23937092]
14. Kerner EA, Fitzpatrick MR. Integrating writing into psychotherapy practice: A matrix of change processes and structural dimensions. Psychotherapy (Chic) 2007 Sep;44(3):333-346. [doi: 10.1037/0033-3204.44.3.333] [Medline: 22122259]
15. Seiffge-Krenke I. Imaginary companions in adolescence: sign of a deficient or positive development. J Adolesc 1997 Apr;20(2):137-154. [doi: 10.1006/jado.1996.0072] [Medline: 9104650]
16. Cucu-Oancea O. Using Diaries - A Real Challenge for the Social Scientist. Procedia - Social and Behavioral Sciences 2013 Oct;92:231-238. [doi: 10.1016/J.SBSPRO.2013.08.665]
17. Travers C. Unveiling a reflective diary methodology for exploring the lived experiences of stress and coping. Journal of Vocational Behavior 2011 Aug;79(1):204-216. [doi: 10.1016/j.jvb.2010.11.007]
18. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
19. Kettles AM, Creswell JW, Zhang W. Mixed methods research in mental health nursing. J Psychiatr Ment Health Nurs 2011 Aug;18(6):535-542. [doi: 10.1111/j.1365-2850.2011.01701.x] [Medline: 21749560]
20. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
21. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
22. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? Expert Rev Pharmacoecon Outcomes Res 2007 Jun;7(3):291-297. [doi: 10.1586/14737167.7.3.291] [Medline: 20528315]
23. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014 Sep 16;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
24. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain 2004 Feb;107(3):213-219. [Medline: 14736583]
25. Grove S, Burns N, Gray J. The Practice of Nursing Research. 7th Edition. Appraisal, Synthesis, and Generation of Evidence. Philadelphia, USA: Saunders; Aug 29, 2012.
26. Holloway I, Wheeler S. Qualitative research in nursing and healthcare. 3rd Edition. Hoboken, NJ: Wiley-Blackwell; 2015.
27. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
28. The National Advisory Board on Social Welfare and Health Care Ethics: Ministry of Social Affairs and Mental Health. Ethical grounds for the social and health care field. Helsinki, Finland: Etene Publications; 2012. URL: http://julkaisut.valtioneuvosto.fi/handle/10024/69923 [WebCite Cache ID 70sY6lXZP]
29. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
30. Kauppi K, Kannisto KA, Hätönen H, Anttila M, Löyttyniemi E, Adams CE, et al. Mobile phone text message reminders: Measuring preferences of people with antipsychotic medication. Schizophr Res 2015 Oct;168(1-2):514-522. [doi: 10.1016/j.schres.2015.07.044] [Medline: 26293215]
31. Kannisto KA, Korhonen J, Adams CE, Koivunen MH, Vahlberg T, Välimäki MA. Factors Associated With Dropout During Recruitment and Follow-Up Periods of a mHealth-Based Randomized Controlled Trial for Mobile.Net to Encourage Treatment Adherence for People With Serious Mental Health Problems. J Med Internet Res 2017 Dec 21;19(2):e46 [FREE Full text] [doi: 10.2196/jmir.6417] [Medline: 28223262]

32. Krogh A, Larsson B, Salvesen, Linde M. A comparison between prospective Internet-based and paper diary recordings of headache among adolescents in the general population. Cephalalgia 2016 Apr;36(4):335-345. [doi: 10.1177/0333102415591506] [Medline: 26092285]
33. Lee SW, Kim I, Yoo J, Park S, Jeong B, Cha M. Insights from an expressive writing intervention on Facebook to help alleviate depressive symptoms. Computers in Human Behavior 2016 Sep;62:613-619. [doi: 10.1016/j.chb.2016.04.034]
34. Lumley MA, Provenzano KM. Stress management through written emotional disclosure improves academic performance among college students with physical symptoms. Journal of Educational Psychology 2003;95(3):641-649. [doi: 10.1037/0022-0663.95.3.641]

Abbreviations

BDI-21: Beck Depression Inventory-21
C-GAS: Children's Global Assessment Scale
e-diary: electronic diary
ID: identifier
SMS: short message service

Edited by G Eysenbach; submitted 26.07.18; peer-reviewed by S Hetrick, C Păsărelu; comments to author 21.09.18; revised version received 16.11.18; accepted 23.11.18; published 20.02.19.

Please cite as:
Metsäranta K, Kurki M, Valimaki M, Anttila M
How Do Adolescents Use Electronic Diaries? A Mixed-Methods Study Among Adolescents With Depressive Symptoms
J Med Internet Res 2019;21(2):e11711
URL: http://www.jmir.org/2019/2/e11711/
doi:10.2196/11711
PMID:30785408

Original Paper

Causal Relationships Among Pollen Counts, Tweet Numbers, and Patient Numbers for Seasonal Allergic Rhinitis Surveillance: Retrospective Analysis

Shoko Wakamiya[1,2,3], PhD; Shoji Matsune[4], MD, PhD; Kimihiro Okubo[5], MD, PhD; Eiji Aramaki[1,2,3], PhD

[1]Institute for Research Initiatives, Nara Institute of Science and Technology, Ikoma, Japan

[2]Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Japan

[3]Data Science Center, Nara Institute of Science and Technology, Ikoma, Japan

[4]Musashikosugi Hospital, Nippon Medical School, Kawasaki, Japan

[5]Nippon Medical School Hospital, Nippon Medical School, Bunkyo, Japan

Corresponding Author:
Shoko Wakamiya, PhD
Institute for Research Initiatives
Nara Institute of Science and Technology
8916-5 Takayama-cho
Ikoma, 630-0192
Japan
Phone: 81 743726053
Fax: 81 743726065
Email: wakamiya@is.naist.jp

Abstract

Background: Health-related social media data are increasingly used in disease-surveillance studies, which have demonstrated moderately high correlations between the number of social media posts and the number of patients. However, there is a need to understand the causal relationship between the behavior of social media users and the actual number of patients in order to increase the credibility of disease surveillance based on social media data.

Objective: This study aimed to clarify the causal relationships among pollen count, the posting behavior of social media users, and the number of patients with seasonal allergic rhinitis in the real world.

Methods: This analysis was conducted using datasets of pollen counts, tweet numbers, and numbers of patients with seasonal allergic rhinitis from Kanagawa Prefecture, Japan. We examined daily pollen counts for Japanese cedar (the major cause of seasonal allergic rhinitis in Japan) and hinoki cypress (which commonly complicates seasonal allergic rhinitis) from February 1 to May 31, 2017. The daily numbers of tweets that included the keyword “kafunshō” (or seasonal allergic rhinitis) were calculated between January 1 and May 31, 2017. Daily numbers of patients with seasonal allergic rhinitis from January 1 to May 31, 2017, were obtained from three healthcare institutes that participated in the study. The Granger causality test was used to examine the causal relationships among pollen count, tweet numbers, and the number of patients with seasonal allergic rhinitis from February to May 2017. To determine if time-variant factors affect these causal relationships, we analyzed the main seasonal allergic rhinitis phase (February to April) when Japanese cedar trees actively produce and release pollen.

Results: Increases in pollen count were found to increase the number of tweets during the overall study period (P=.04), but not the main seasonal allergic rhinitis phase (P=.05). In contrast, increases in pollen count were found to increase patient numbers in both the study period (P=.04) and the main seasonal allergic rhinitis phase (P=.01). Increases in the number of tweets increased the patient numbers during the main seasonal allergic rhinitis phase (P=.02), but not the overall study period (P=.89). Patient numbers did not affect the number of tweets in both the overall study period (P=.24) and the main seasonal allergic rhinitis phase (P=.47).

Conclusions: Understanding the causal relationships among pollen counts, tweet numbers, and numbers of patients with seasonal allergic rhinitis is an important step to increasing the credibility of surveillance systems that use social media data. Further in-depth studies are needed to identify the determinants of social media posts described in this exploratory analysis.

(J Med Internet Res 2019;21(2):e10450) doi:10.2196/10450

KEYWORDS

seasonal allergic rhinitis; social media; Twitter; causal relationship; infoveillance; disease surveillance

Introduction

The rapid growth of the internet has been accompanied by an increase in the use of social media data (from sources such as Twitter and Facebook) to explore and understand various phenomena. This form of social media monitoring can facilitate an effective analysis of large quantities of social media data produced in real time.

Large-scale quantitative analyses have been conducted using health-related social media data [1,2], and the use of these data for disease surveillance (referred to as "infoveillance") is gaining interest [3]. In particular, major advances have been made in the use of social media data to track the prevalence and spread of infectious diseases and other conditions [4-6]. These studies have contributed to public health by demonstrating moderately high correlations between fluctuations in the number of relevant social media posts and patients for a specific disease. Public health authorities have also started adopting and applying currently available tools that use social media for influenza surveillance, such as HealthTweets.org [7,8], Sickweather [9], and Now Trending [10].

Despite the reported correlations between actual disease prevalence and social media posts in previous research, the mechanism underlying this relationship is poorly understood. In other words, the causal relationship between disease occurrence and the behavior of social media users remains unclear. For example, some individuals may only begin posting on social media after an existing condition becomes more severe. In contrast, others may start posting when experiencing only mild symptoms and seek medical treatment after their conditions worsen. Due to the presence of these individual-level variations, the relationship between the numbers of social media posts and patient numbers remains inconclusive.

The lack of understanding of this relationship may have reduced the perceived reliability of disease surveillance based on social media data, as exemplified by the failure of Google Flu Trends [11]. This web service consistently overestimated influenza prevalence during the 2012-2013 season by over 50%, which led to a precipitous decline in its credibility as a surveillance system. The overestimations may have been influenced by the unusually early start of the 2012-2013 influenza season, which made it a frequent topic of discussion in many media outlets. As a consequence, this may have increased the number of people searching for influenza-related topics on Google. Although the system algorithm was eventually updated, Google Flu Trends was shut down in 2015. In order to increase the credibility of disease surveillance based on social media data, there is a need to determine if there is a causal relationship between the behavior of social media users and the actual number of patients for a target disease.

Seasonal allergic rhinitis is an allergic disease that is so widely prevalent in Japan that it can be considered a national affliction. In particular, a large number of patients suffer from seasonal allergic rhinitis induced by Japanese cedar pollen between February and April each year [12-14]. Although pollen is the main cause of seasonal allergic rhinitis, symptoms only occur if a person is exposed to a quantity of pollen that exceeds his/her threshold level. As a result, there is no strong association between pollen count and patient numbers. In addition, seasonal allergic rhinitis in Japan can also be triggered by pollen from other plant species (eg, hinoki cypress and common ragweed), which complicates disease surveillance. However, the predicted prevalence of seasonal allergic rhinitis in Japan is currently based solely on Japanese cedar pollen counts.

This study aimed to clarify the causal relationships among the prevalence of seasonal allergic rhinitis, the behavior of social media users, and the actions of the users in the real world. For example, we examined if more patients sought care after increases in pollen count, if Twitter users tweeted more after visiting a health care institute, and if more patients visited hospitals after seasonal allergic rhinitis received increased attention on social media. The results of this analysis may support the use of social media in seasonal allergic rhinitis surveillance and shed light on the previously unknown behavior of patients with seasonal allergic rhinitis. In addition, we discuss the content of some tweet examples.

Methods

Data

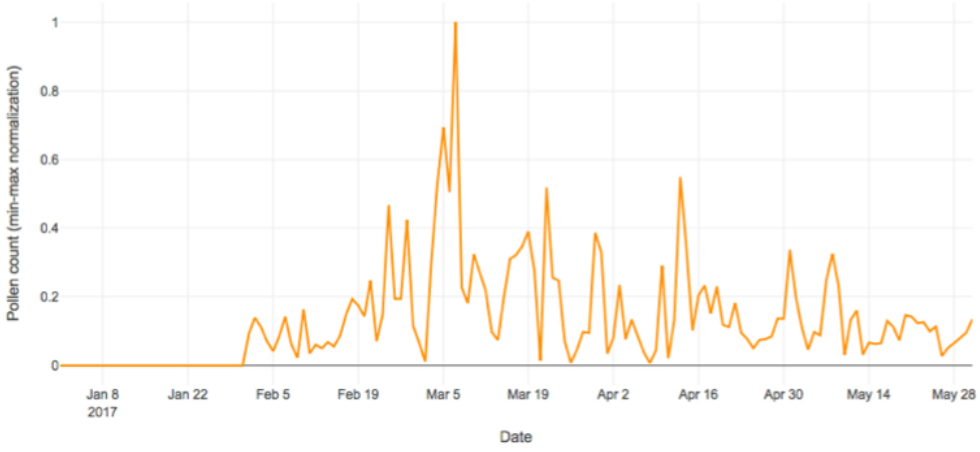
Measures

This analysis was conducted using the following datasets of pollen count, tweets, and numbers of patients with seasonal allergic rhinitis (Figure 1). For this study, we focused on Kanagawa Prefecture in the Kantō region of Japan, which has the second largest population (over 9 million) in Japan, after Tokyo and before Osaka. The population density of Kanagawa Prefecture is estimated to be 3,791.56 people/km^2 [15], which is ranked third in Japan, after Tokyo and Osaka. According to the report on the usage ratio of social networking services of internet users in each prefecture [16], Kanagawa Prefecture has a social networking services usage ratio of 56.5%, with the highest rank. According to the information by Kanagawa Prefectural Government [17], Japanese cedar pollen dispersal in Kanagawa Prefecture occurs from February to April each year, which corresponds with other areas in Kantō region including Tokyo and Saitama.

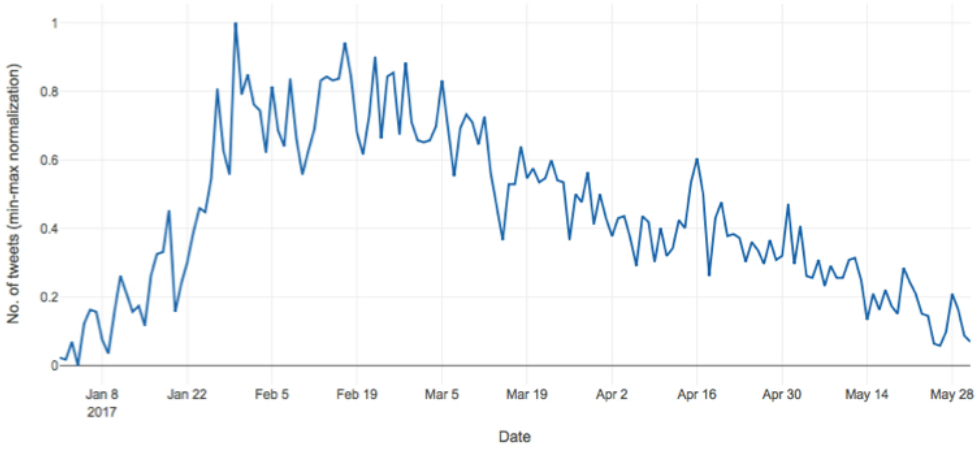
Pollen Count

We examined daily pollen counts for Japanese cedar (the major cause of seasonal allergic rhinitis in Japan) and hinoki cypress (which commonly complicates seasonal allergic rhinitis). Japanese cedar and hinoki cypress pollen counts are measured hourly by automatic pollen counters located nationwide. Each prefecture has two to three of these devices, which are placed in urban areas with high population densities and in mountainous regions that are the major source of pollen production. Each prefecture's mean daily pollen count is calculated based on aggregated hourly counts from multiple observation sites.

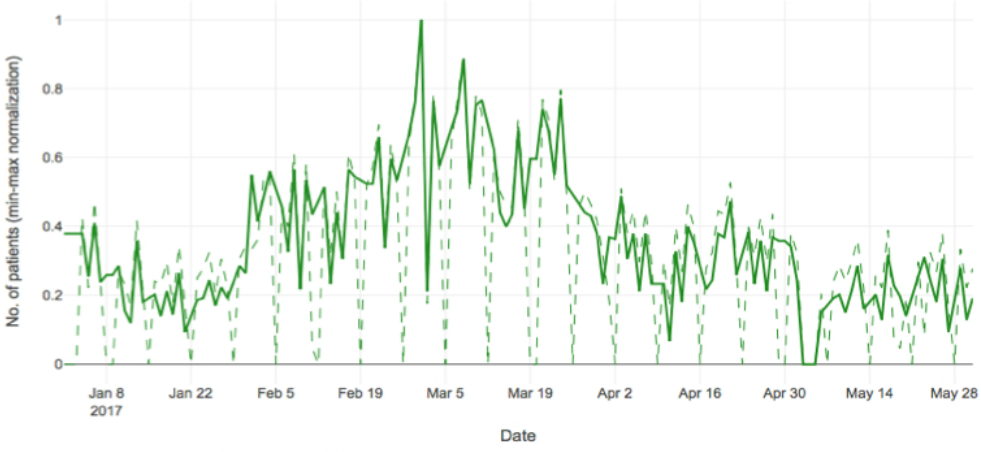
Figure 1. Time-based changes in data in Kanagawa Prefecture, Japan. The X-axes represent the date and the Y-axes represent data counts, to which min-max normalization is applied for the following variables: (a) changes in pollen count (mean daily pollen count from three observation sites within Kanagawa Prefecture), (b) changes in the number of SAR-related tweets, and (c) changes in the number of SAR patients (daily number of patients from three participating health care institutes within Kanagawa Prefecture). The solid line represents the changes in the number of patients on nonconsultation days (ie, days when an institute is closed) that were supplemented by the patient numbers from the preceding and proceeding days; the dashed line represents the changes in the reported number of patients. Our analysis used the supplemented patient numbers. SAR: seasonal allergic rhinitis.



(a) Mean daily pollen count



(b) Daily number of SAR-related tweets



(c) Daily number of SAR patients

Hourly pollen counts from February 1 to May 31, 2017, were obtained from the Japanese Ministry of the Environment's pollen observation system (designated "Hanakosan") for analysis [18,19]. The data were recorded at three observation sites within Kanagawa Prefecture (the Second Annex of the Kanagawa Prefectural Government Building, the Kawasaki Life Science & Environment Research Center, and the Kanagawa Environmental Research Center), and the mean daily pollen counts across these three sites were calculated. Figure 1 shows

the changes in mean daily pollen count in Kanagawa Prefecture from February to May 2017.

Seasonal Allergic Rhinitis–Related Tweets Numbers

In an analysis of Twitter posts, we calculated the daily number of tweets that included the Japanese keyword "kafunshō" (or seasonal allergic rhinitis). These tweets have been crawled using the Twitter Streaming app's programming interface. From among these tweets, we identified Twitter user profiles with location information (such as an area of residence that was freely written by a user in his/her Twitter profile or latitude/longitude data acquired from the satellite-based positioning systems such as global positioning system); the tweets were then classified according to the prefecture. We prioritized the latitude/longitude data, if available. Subsequently, we used retweets. We identified 185,538 tweets from Japan that contained the target keyword between January 1 and May 31, 2017. Figure 1 also shows the changes in the daily number of relevant tweets in Kanagawa Prefecture during this period.

Number of Patients With Seasonal Allergic Rhinitis

We analyzed the daily number of outpatients diagnosed with seasonal allergic rhinitis from January 1 to May 31, 2017. Specifically, electric health records were used to determine whether patients were diagnosed with seasonal allergic rhinitis, causing double counting of patients who saw a doctor twice or more often during the period, although this was a rare occurrence. Although daily patient numbers are not generally publicized, we obtained these data from three health care institutes that agreed to participate in the study. These institutes were Sasaki Hospital (Yokohama City), Kawasaki Saiwai Clinic (Kawasaki City), and Kosugi ENT Clinic (Kawasaki City). Figure 1 shows the changes in the daily number of seasonal allergic rhinitis patients that visited the participating health care institutes from January to May 2017. Single imputation methods such as last observation carried forward [20] were used to deal with missing values. However, missing values on nonconsultation days occurred regularly, and patients who did not want to wait until the following days may have visited a doctor on the preceding days. Thus, we interpolated the missing values from the average values.

Analysis

We aimed to examine the causal relationships among the level of attention gained by seasonal allergic rhinitis on Twitter (number of tweets), pollen count, and the number of patients with seasonal allergic rhinitis. Although pollen counts may directly affect the number of tweets and patients with seasonal allergic rhinitis, the vice-versa is highly unlikely. As a result, we did not analyze the effects of tweet numbers and patient numbers on pollen count. Because this analysis used a data-driven approach, we employed the Granger causality test [21]. This statistical hypothesis test determines if a particular time series is predictive of another time series.

To determine if time-variant factors affect the causal relationships among seasonal allergic rhinitis prevalence, behavior of social media users, and the actions of users in the real world, we conducted additional analyses where the study period (February to May 2017) was divided into two phases. The first was the main seasonal allergic rhinitis phase, which generally occurs from February to April in the Kantō region for Japanese cedar-induced seasonal allergic rhinitis [12]. The second was the concluding phase of the season, which generally occurs in May in the Kantō region for Japanese cedar-induced seasonal allergic rhinitis. The data were analyzed as a differential time series of the differences between each day and the preceding day.

Ethics Statement

This study utilizes only the participants' count information that was nonlinkable, anonymized, and deidentified prior to analysis. As this research did not use personally identifiable information, it was exempt from institutional review board approval in accordance with the Ethical Guidelines for Medical and Health Research Involving Human Subjects stipulated by the Japanese national government.

Results

Overview

In addition to analyzing the overall study period (February to May 2017), the Granger causality test was applied to the main seasonal allergic rhinitis phase (February to April 2017). The results of the overall study period and the main seasonal allergic rhinitis phase are presented in Table 1 and Figure 2.

Effect of Pollen Count on the Number of Seasonal Allergic Rhinitis–Related Tweets

As shown in Table 1 and Figure 2, the Granger causality test rejected the null hypothesis that pollen count has no effect on the number of seasonal allergic rhinitis–related tweets during the overall study period (P=.04); this indicates that pollen count has a causal effect on the number of seasonal allergic rhinitis–related tweets. In contrast, the test did not reject this null hypothesis in the main seasonal allergic rhinitis phase (P=.05; Table 1; Figure 2). We were unable to apply the Granger causality test to the concluding phase (May 2017) for these two variables.

Effect of Pollen Count on Patient Numbers

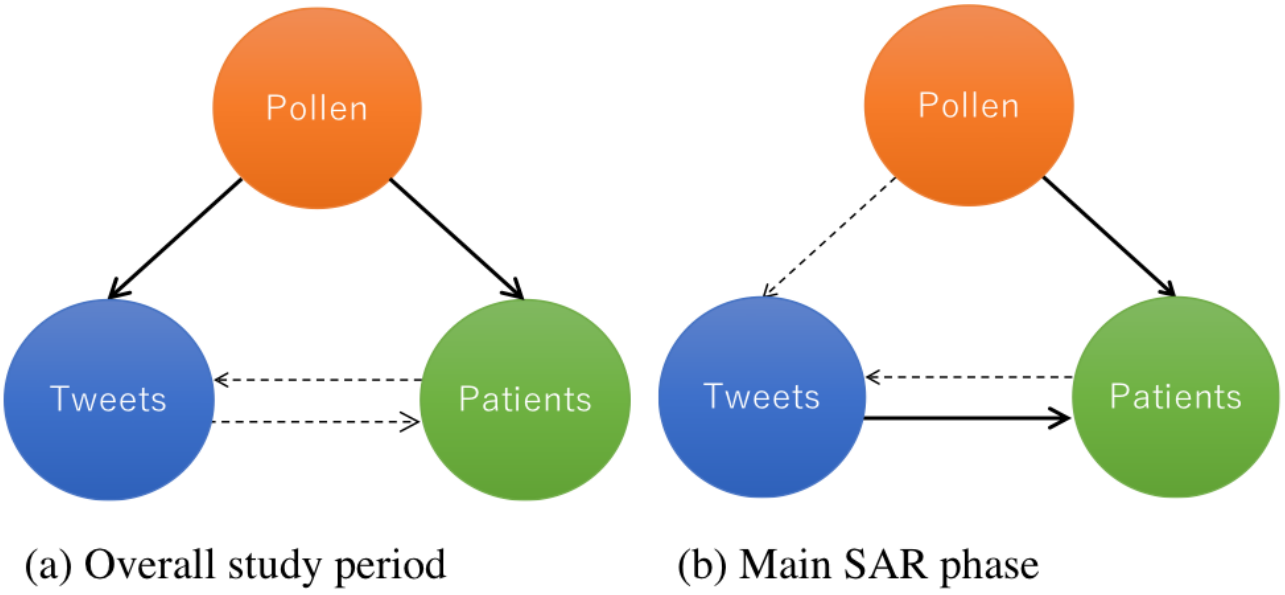
As shown in Table 1 and Figure 2, the Granger causality test rejected the null hypothesis that pollen count has no effect on patient numbers during the overall study period (P=.04); this indicates that pollen count has a causal effect on the number of seasonal allergic rhinitis–related tweets. In addition, the test also rejected the null hypothesis in the main seasonal allergic rhinitis phase (P=.01; Table 1; Figure 2). We were unable to apply the Granger causality test to the concluding phase (May 2017) for these two variables. Consequently, we confirmed that more patients sought care after increases in pollen count occurred during the overall study period.

Table 1. Results of the Granger causality test for pollen count, number of seasonal allergic rhinitis–related tweets, and number of patients with seasonal allergic rhinitis.

Cause/effect	*P* value	
	Number of tweets	Number of patients
Overall study period (February to May 2017)		
Pollen count	.04	.04
Number of tweets	—[a]	.89
Number of patients	.24	—
Main seasonal allergic rhinitis phase (February to April 2017)		
Pollen count	.05	.01
Number of tweets	—	.02
Number of patients	.47	—

[a]Not available.

Figure 2. Causal relationships between pollen count, number of SAR-related tweets, and number of patients with SAR (Granger causality test results) for (a) the overall study period and (b) the main SAR phase. The Granger causality test did not reveal any causal relationships between these variables in the concluding phase of the season. SAR: seasonal allergic rhinitis.



Effect of the Number of Seasonal Allergic Rhinitis–Related Tweets on Patient Numbers

As shown in Table 1 and Figure 2, the Granger causality test did not reject the null hypothesis that tweet numbers have no effect on patient numbers during the overall study period (*P*=.89). In contrast, the test rejected this null hypothesis in the main seasonal allergic rhinitis phase (*P*=.02; Table 1; Figure 2); this indicates that the number of seasonal allergic rhinitis–related tweets have a causal effect on patient numbers during this phase. We were unable to apply the Granger causality test to the concluding phase (May 2017) for these two variables. Therefore, we confirm that more patients visited hospitals after seasonal allergic rhinitis received increased attention on social media during the main seasonal allergic rhinitis phase.

Effect of Patient Numbers on the Number of Seasonal Allergic Rhinitis–Related Tweets

In both the overall study period and the main seasonal allergic rhinitis phase, the Granger causality test did not reject the null hypothesis that patient numbers have no effect on the number of seasonal allergic rhinitis–related tweets (*P*=.24 and *P*=.47, respectively). We were unable to apply the Granger causality test to the concluding phase (May 2017) for these two variables.

Discussion

Content of Seasonal Allergic Rhinitis–Related Tweets

In this analysis of the 2017 Japanese cedar pollen-induced seasonal allergic rhinitis season in Kanagawa Prefecture, our results indicated that the level of attention gained by seasonal allergic rhinitis on Twitter and pollen count may be able to predict the number of patients with seasonal allergic rhinitis. In addition, we examined the content of seasonal allergic

rhinitis–related tweets posted during the study period (January to May 2017) to gain further insight into these relationships.

The tweets in Textbox 1 are examples that were posted on January 30, 2017, which had the highest number of seasonal allergic rhinitis–related tweets before the pollen count, and the number of patients with seasonal allergic rhinitis increased. Pollen count is thought to increase in response to an increase in temperature and a decrease in humidity. In accordance with the expectations, there was a sudden increase in the ambient temperature on January 30, 2017 (Figure 3), which may have caused more sensitive users to identify and report symptoms ascribed to seasonal allergic rhinitis. This, in turn, may have led to a high level of attention to this topic on Twitter on that day.

The tweets in Textbox 2 are examples that were posted on March 1, 2017, which had the highest number of patients with seasonal allergic rhinitis seeking care at the three participating health care institutes during the study period. Coincidentally, there was an extremely low pollen count on this day. Although there appeared to be a slight decrease in tweet numbers (Figure 3), the tweets included those from patients with seasonal allergic rhinitis who were going or had gone to seek treatment.

Finally, the tweets in Textbox 3 are examples that were posted on March 7, 2017, which had the highest pollen count during the 2017 seasonal allergic rhinitis season. As shown in Figure 3, there was a sudden increase in pollen count for several days before the peak on this day. Although there was a reduction in the number of tweets, the tweets included those from people who had seasonal allergic rhinitis symptoms for the first time this season as well as reports of the worst symptoms for this season. There was also an increase in the number of patients with seasonal allergic rhinitis, and we confirmed that there were tweets where patients reported seeking treatment at health care institutes.

Different Characteristics of Variables

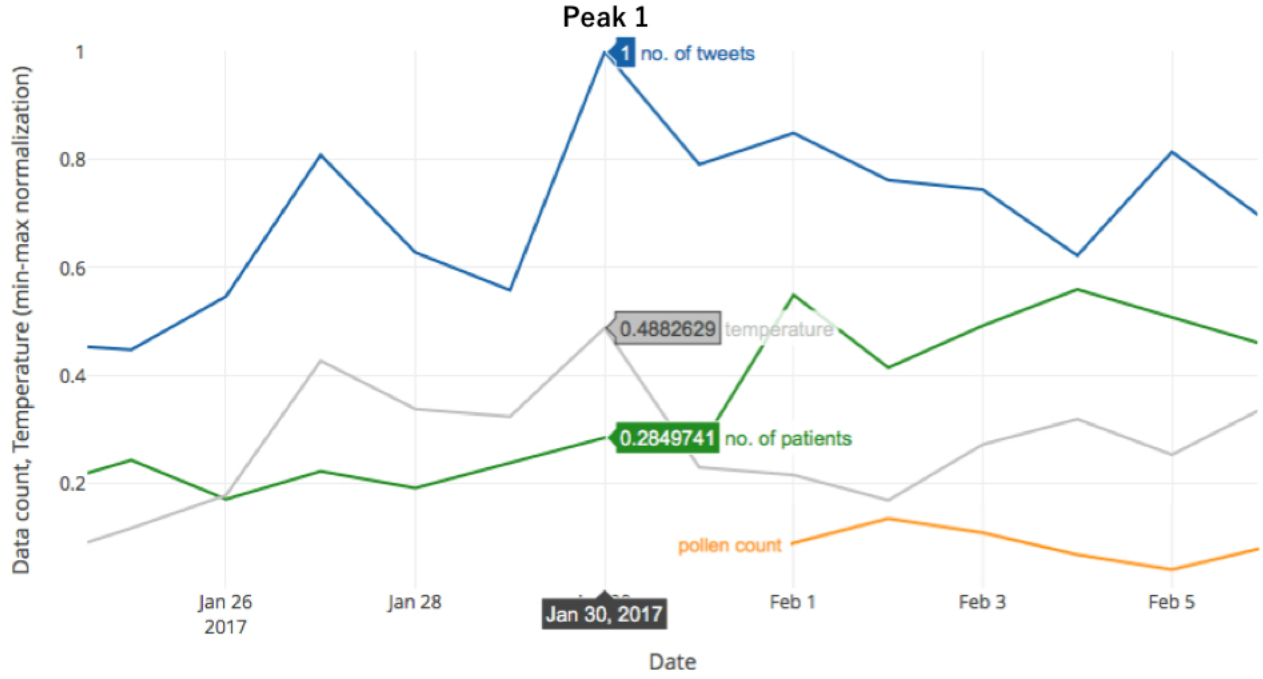
Through our analysis of pollen count and tweet numbers, we were able to observe differences in data characteristics between these variables. The pollen count was affected by external factors (such as temperature, rainfall, wind speed, and wind direction), and there were substantial fluctuations throughout the study period (Figure 1). Accordingly, it was difficult to predict at a glance when the pollen season would end. On the other hand, the number of tweets showed some fluctuations (Figure 1), but there was a general increase from January, a peak in February and March, and a steady downward trend thereafter. Thus, we were able to visualize the trend toward the conclusion of the seasonal allergic rhinitis season. Our analysis showed that the combined use of data with different characteristics not only provided information on the prevalence of seasonal allergic rhinitis but also enabled observation of the seasonal allergic rhinitis season as it progressed.

The development of an analytical model that accounts for the different characteristics of the datasets while providing insight into the causal relationships may enable highly reliable disease surveillance.

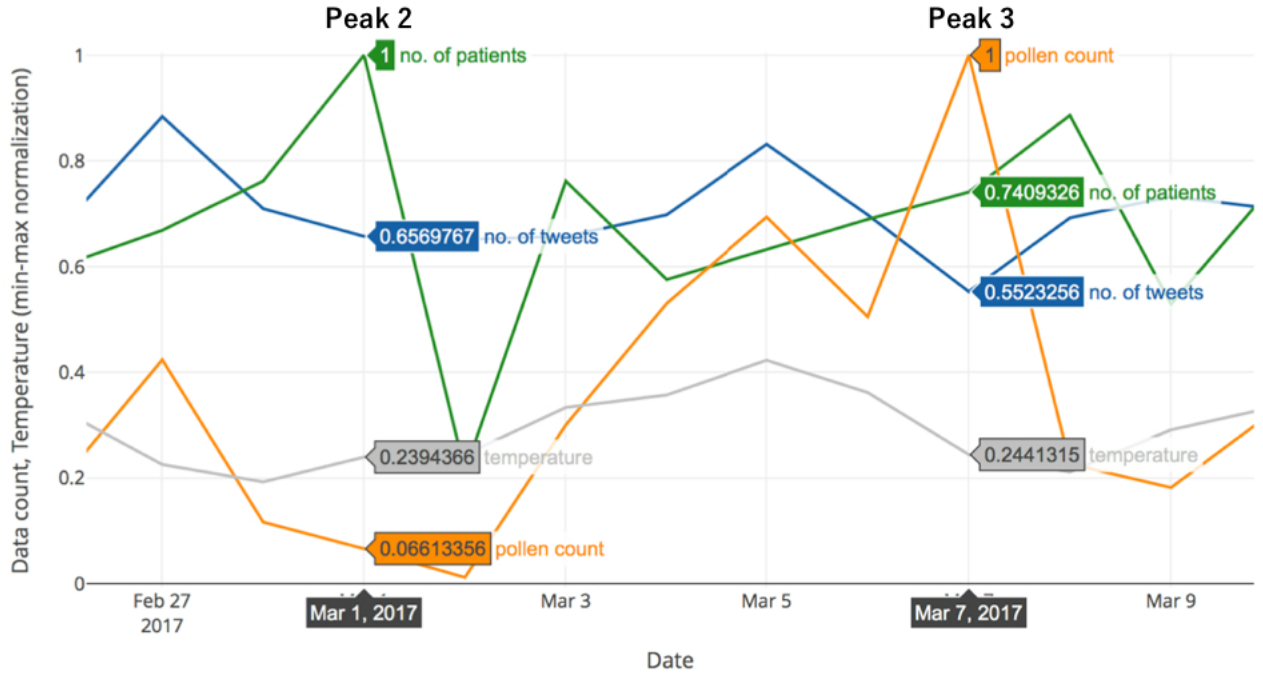
Textbox 1. Examples of tweets posted on January 30, 2017.

- I’ve been down with hay fever since that windy day. I don’t need this lol.
- Pollen is here, isn’t it? My hay fever’s not that bad, but I could feel it’s “arrival” 2/3 days ago. This morning I was hard-core sneezing and my nose can’t stop running...
- Looks like it’s hay fever.
- Good morning. I keep hearing about hay fever these few days…everyone’s most hated season is coming again, eh? You gotta eat lotus roots! Have a great day, everyone!
- Uh oh. I haven’t got any tissues. I’m dying. My runny nose won’t stop. I’m about to have hay fever. This is gonna be rough.
- I’m totally convinced that once it starts to get warm, then hay fever comes along.
- I haven’t got hay fever, but I can kind of feel the pollen flying.
- When it suddenly gets warmer I can feel the hay fever coming
- I don’t know if its hay fever or just the temperature differences in winter…but I’m snuffling.
- Today, it’s warm and windy, and I’m sneezing lots — is the pollen flying? I also heard that this year’s pollen count is 4.4 times more than last year’s. I already had a pretty rough time last year with my severe hay fever…stuffy nose, itchy eyes...I hate this. During this period I’d like to rip out my nose and eyeballs. Seriously.
- I hate this my nose won’t stop running, I don’t even know if this is the remnants of a cold or the effects of hay fever for real lol.
- I kind of feel like this year’s hay fever is already starting. I’ll stock up on OTC meds tomorrow.
- It...it’s finally here...hay fever, my eyes are itching and my nose is running and my head is heavy...it’s usually empty tho.
- My face is suddenly swollen today, and my eyes are bleary. it’s hay fever!!!
- I’m home! maybe it’s the warm weather, but I feel a bit hay fever-ish...what a pain
- When it’s this warm, hay fever’s here---- #nhk11

Figure 3. Peak number of SAR-related tweets, number of SAR patients, and pollen count. The X-axes represent the date and the Y-axes represent the min-max normalized data values for the three variables. The blue line represents the number of tweets, the green line represents the number of patients, the orange line represents the pollen count, and the grey line represents the mean temperature. SAR: seasonal allergic rhinitis.



(a) Peak number of tweets (Peak 1: January 30, 2017)



(b) Peak number of patients (Peak 2: March 1, 2017) and pollen count (Peak 3: March 7, 2017)

Textbox 2. Examples of tweets posted on March 1, 2017.

- I'm fine when I'm outside, but I suddenly get hay fever symptoms around midnight. What's up with this delayed attack.
- I woke up cos I couldn't breathe cos my nose was stuffed cos of hay fever.
- My sleeping time and concentration are dwindling away because of hay fever
- Argh. This hay fever headache is massive…
- OH NO. I forgot to take my hay fever meds.
- My eyes are totally red cos of hay fever, But I still wanna use my colored contacts ~.
- I just took hay fever meds. I'll sleep a bit more. I dunno why my shoulders are so stiff lately.
- This anti-hay fever mask has NO effect LOL
- I forgot my eye drops and nasal spray for hay fever but I haven't got any symptoms. Maybe the non-drowsy oral meds are enough? Or is the real deal still to come?
- I'm using a mask but my nose is running from hay fever...
- Came to see the doctor for hay fever before work but it's really crowded
- I'm here at the ENT. The doctor recommended an actual treatment to cure hay fever and not just suppress the symptoms but it looks like a pain to keep coming here. It seems I have to come here every month for 5 years, even in the off season. I want to do it if I have time next year!!
- So this is hay fever. If I go outside without a mask my nose becomes a waterfall...
- When I woke up today I was sniffing more than usual and my throat hurts, is this hay fever or a cold...if this continues tomorrow, I'll probably come back tomorrow in even worse condition, last time I thought it was a cold and felt pretty confident, but it got worse and I lost my voice LOL

Textbox 3. Examples of tweets posted on March 7, 2017.

- Hay fever sucks
- Good morning. Hay fever has arrived; my eyes and nose are so itchy~ I'll try to be cheerful today as well.
- I've run out of hay fever meds, so if I seem to keep sneezing, I'll probably have to get more meds from the doc.
- Oh man, my face is swollen and painful. Hay fever? Or maybe allergies?
- I've got hinoki hay fever and dust sensitivity, so I've GOT to have a mask for this old house in Yamanashi.
- My hay fever is going completely crazy today. Forgetting the tissues was a fatal mistake.
- Sudden spike in the number of people with hay fever symptoms!
- I thought I was sick, but the hospital told me I had hay fever — unbelievable...
- My hay fever is horrible today. This is the worst for this year.
- I just came back from the hospital, and I've gotta go again...my hay fever debut orz
- I was totally phased out with hay fever and almost forgot to go to school.
- My nose won't stop running. I think it's hay fever.
- My hay fever is crazy so I've got shades and a mask on. Totes feel like a celeb.
- Today's hay fever is really rough. I'm completely done in with this triple play of sneezing, runny nose, and itchy eyes.
- The hay fever eye drops work only for just over 10 minutes. I really got to get some oral meds.
- I haven't even stepped out of the house but the hay fever is so bad I don't feel like doing anything.
- I'm sleepy already, this could also be the hay fever at work
- Hay fever where you don't stop sneezing is a real pain.
- It's not a cold, but my head hurts, and when that's settling I've got a runny nose, am sneezing and coughing all at once. This is the start of legit hay fever.
- Been a while since I've had this full-blown hay fever. I blew my nose and it started bleeding, so I spent almost the whole day with a nosebleed. It won't stop.
- The inside of my mouth itches cos of hay fever

Limitations and Future Direction

This study is subject to several limitations, which could cause statistical and methodological limitations. First, we discuss the limitations caused by characteristics of social media data. This study assumed that there would be no lag effect in reporting symptoms on the same day on twitter. This study also assumed that the location where tweets originated would be the location that twitter users registered in their profile. Although ideally, only geolocation-enabled tweets must be used, the number of these tweets was small (less than 1% in our dataset) due to a recent increase in the number of users who care about their privacy and turn location sharing off. Notably, a lower quality of social media data is regarded as one of the common issues of most research using social media data. In addition, we used retweets as well as the other tweets because we assumed that retweeting tweets was one of the actions of Twitter users over social media. However, the retweet numbers tend to be significantly high when a public health agency or a celebrity tweets about an outbreak. Therefore, we should carefully consider the effect of a sudden increase in retweet numbers. Fortunately, existing research showed that the data-quality issue could be solved by the data quantity and applying appropriate preprocessing techniques to data such as location estimation [22,23], bot detection [24,25], positive/negative classification [26], and retweet handling [27]. In our future work, we plan to apply these techniques to our social media data to improve their quality.

In this analysis, we did not assess the effects of tweet numbers and number of patients with seasonal allergic rhinitis on pollen count because the number of tweets and patients would not directly affect the pollen count. Specifically, although tweet numbers and patient numbers are indicators of the behavior of social media users and their actual actions, pollen count observed outside is a natural phenomenon, which cannot be easily controlled by social media users. However, it will be possible to analyze the effects of tweet numbers and patient numbers on pollen count if the data observed indoor are available, because pollen count indoor can be controlled by people who take measures to remove pollen by, for example, using air cleaners. In our future work, we will try to analyze the unconsidered effects by gathering a variety of pollen count data and rule out any possible mathematical artefacts in an empirical fashion.

Further in-depth studies are needed to identify the determinants of social media posts described in this exploratory analysis. In addition, this study focused on one seasonal allergic rhinitis season, and downstream studies should be conducted using data that are continuously collected over multiple seasons. Furthermore, although the analyses in this paper were performed while focusing on one area due to the difficulty of obtaining the number of patients with seasonal allergic rhinitis, analyses using data that are collected in other areas should be conducted. Practically, it is very difficult to obtain the number of patients with seasonal allergic rhinitis in each area because no reports on seasonal allergic rhinitis patient numbers are provided in Japan and we have to identify health institutes (three or more in each area would be required) that can participate in this study. If we can obtain the patient numbers in other areas, we will conduct further analyses to interpret the results such as population density-based subgroup analysis. Finally, there is a need to conduct intervention-based prospective studies to gain a more accurate understanding of the causal relationships among these variables.

Comparison With Prior Work

Health-related social media data have been used for large-scale quantitative analyses [1,2], referred to as “infoveillance” [3]. In particular, major advances have been made in the use of social media data to track the prevalence and spread of infectious diseases and other conditions. Among the infectious diseases targeted for surveillance, researchers have most actively applied these data to influenza surveillance [26-34]. Surveillance studies have also been conducted on enterohemorrhagic *Escherichia coli* outbreaks [35] and dengue fever [36]. In addition, social media data have been used to improve our understanding of Ebola [37] and Zika virus infections [38-40]. Although disease surveillance efforts tend to focus on acute infectious diseases, studies have also been conducted on chronic diseases such as cancer [41], hypertension [41], asthma [41-43], diabetes [44], and seasonal allergic rhinitis [45-47]. Systematic reviews have also been conducted on disease surveillance based on social media data [4-6].

Although these previous researches reported correlations between actual disease prevalence and social media posts, the mechanism underlying this relationship is still not investigated. In other words, the causal relationship between disease occurrence and the behavior of social media users remains unclear. Thus, this paper clarified the causal relationships among multisource data such as pollen count, the posting behavior of social media users, and the number of patients with seasonal allergic rhinitis in the real world, which is a novel point of this study compared with previous work.

Conclusions

Although social media data are increasingly used in disease surveillance, there is a need to improve the credibility of these surveillance systems in order to promote their implementation and acceptance in society. Understanding the causal relationships between the behavior of social media users and actual patient numbers is an important step to increase the credibility of these surveillance systems. In this study, we analyzed data on pollen count, the number of tweets, and the number of patients during the 2017 seasonal allergic rhinitis season in Japan using the Granger causality test and shed light on the causal relationships among these variables. Increases in pollen count were found to increase the number of tweets and patients. In addition, increases in the number of social media posts (ie, tweets) also increased the patient numbers, suggesting that patients with seasonal allergic rhinitis or nonseasonal allergic rhinitis were motivated by increases in social media posts and went to the hospital. The main seasonal allergic rhinitis phase and the concluding phase of the season appeared to have different characteristics. Accordingly, disease surveillance based on social media data should be adjusted to account for these time-based differences.

Acknowledgments

This work was supported in part by an ACT-I research grant from the Japan Science and Technology Agency and the Research Program on Emerging and Re-Emerging Infectious Diseases from Japan Agency for Medical Research and Development (#16768699). The authors appreciate the advice and cooperation of the Japan Anti-Pollinosis Council and NPO Therapeutic Promotion Society for Pollinosis and Rhinosinusitis for this study. The authors are also grateful to the physicians and staff of the three participating hospitals for providing the patient statistics.

Authors' Contributions

SW and EA conceived and designed the model and method. SW, EA, and SM collected the data. SW and EA analyzed the data. SW, EA, SM, and KO prepared the manuscript.

Conflicts of Interest

None declared.

References

1. Lee JL, DeCamp M, Dredze M, Chisolm MS, Berger ZD. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. J Med Internet Res 2014;16(10):e237 [FREE Full text] [doi: 10.2196/jmir.3765] [Medline: 25591063]
2. Alnemer KA, Alhuzaim WM, Alnemer AA, Alharbi BB, Bawazir AS, Barayyan OR, et al. Are Health-Related Tweets Evidence Based? Review and Analysis of Health-Related Tweets on Twitter. J Med Internet Res 2015 Oct;17(10):e246 [FREE Full text] [doi: 10.2196/jmir.4898] [Medline: 26515535]
3. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
4. Charles-Smith L, Reynolds T, Cameron M, Conway M, Lau E, Olsen J, et al. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PLoS One 2015 Oct 05;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
5. Paul MJ, Sarker A, Brownstein JS, Nikfarjam A, Scotch M, Smith KL, et al. Social Media Mining for Public Health Monitoring and Surveillance. In: Pacific Symposium on Biocomputing 2016, PSB 2016.: World Scientific Publishing Co. Pte Ltd; 2016 Presented at: The 21st Pacific Symposium on Biocomputing (PSB '16); 2016; Big Island, United States p. 468-479.
6. Al-Garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: A systematic review. J Biomed Inform 2016 Dec;62:1-11 [FREE Full text] [doi: 10.1016/j.jbi.2016.05.005] [Medline: 27224846]
7. Dredze M, Cheng R, Paul M, Broniatowski D. HealthTweets.org: A Platform for Public Health Surveillance Using Twitter. 2014 Presented at: The AAAI-14 Workshop on World Wide Web and Public Health Intelligence; 2014; Québec Convention Centre p. 2-3 URL: https://www.aaai.org/ocs/index.php/WS/AAAIW14/paper/view/8723
8. HealthTweets. URL: http://www.healthtweets.org/accounts/login/?next=/ [accessed 2018-02-13] [WebCite Cache ID 6xBuKfrLT]
9. Sickweather. URL: http://www.sickweather.com [accessed 2018-02-13] [WebCite Cache ID 6xBoVXhiw]
10. US Department of Health and Human Services. Now Trending. URL: https://nowtrending.hhs.gov [accessed 2018-02-12] [WebCite Cache ID 6xBoNundL]
11. Google Flu Trends. URL: https://www.google.org/flutrends/about/ [accessed 2018-02-06] [WebCite Cache ID 6x1qfvJ4F]
12. Yamada T, Saito H, Fujieda S. Present state of Japanese cedar pollinosis: the national affliction. J Allergy Clin Immunol 2014 Mar;133(3):632-639 [FREE Full text] [doi: 10.1016/j.jaci.2013.11.002] [Medline: 24361081]
13. Okubo K, Kurono Y, Ichimura K, Enomoto T, Okamoto Y, Kawauchi H, Japanese Society of Allergology. Japanese guidelines for allergic rhinitis 2017. Allergol Int 2017 Apr;66(2):205-219 [FREE Full text] [doi: 10.1016/j.alit.2016.11.001] [Medline: 28214137]
14. Oseroff C, Pham J, Frazier A, Hinz D, Sidney J, Paul S, et al. Immunodominance in allergic T-cell reactivity to Japanese cedar in different geographic cohorts. Ann Allergy Asthma Immunol 2016 Dec;117(6):680-689.e1 [FREE Full text] [doi: 10.1016/j.anai.2016.10.014] [Medline: 27979027]
15. Higashide M. Population, Area, Population Density based Prefecture Ranking in Japan (in Japanese). URL: https://uub.jp/rnk/p_j.html [accessed 2018-11-08] [WebCite Cache ID 73le5kbM0]
16. Ministry of Internal Affairs Communications. Communications Usage Trend Survey, 2016 (in Japanese). URL: https://www.e-stat.go.jp/stat-search/

files?page=1&layout=datalist&toukei=00200356&tstat=000001102495&cycle=0&tclass1=000001102517&stat_infid=000031591554 [accessed 2018-11-07] [WebCite Cache ID 73leLw8kc]
17. About Hay Fever. URL: http://translate.google.co.jp/translate?hl=ja&sl=ja&tl=en&u=http%3A%2F%2Fwww.pref.kanagawa.jp%2Fdocs%2Fcz6%2Fcnt%2Ff848%2F&sandbox=1 [accessed 2018-11-07] [WebCite Cache ID 73ledOx6r]
18. The Ministry of the Environment. Pollen observation system (Hanakosan; in Japanese). URL: http://kafun.taiki.go.jp/index.aspx [accessed 2018-02-06] [WebCite Cache ID 6x1qnURNw]
19. Data Download of Pollen Observation system (in Japanese). URL: http://kafun.taiki.go.jp/DownLoad1.aspx [accessed 2018-02-06] [WebCite Cache ID 6x1qvqIyY]
20. Tanaka S, Fukinbara S, Tsuchiya S, Suganami H, Ito YM. Current Practice in Japan for the Prevention and Treatment of Missing Data in Confirmatory Clinical Trials: A Survey of Japanese and Foreign Pharmaceutical Manufacturers. Ther Innov Regul Sci 2014 Nov 16;48(6):717-723 [FREE Full text] [doi: 10.1177/2168479014530974] [Medline: 30227476]
21. Granger C. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 1969 Aug;37(3):424-438 [FREE Full text] [doi: 10.2307/1912791]
22. Jurgens D, Finethy T, McCorriston J, Xu Y, Ruths D. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice, In Proc. In: Proceedings of the Ninth International AAAI Conference on Web and Social Media (ICWSM '15). 2015 Presented at: The Ninth International AAAI Conference on Web and Social Media (ICWSM '15); 2015; Oxford p. 188-197.
23. Ajao O, Hong J, Liu W. A survey of location inference techniques on Twitter. Journal of Information Science 2015 Nov 20;41(6):855-864 [FREE Full text] [doi: 10.1177/0165551515602847]
24. Subrahmanian V, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, et al. The DARPA Twitter Bot Challenge. Computer 2016 Jun;49(6):38-46 [FREE Full text] [doi: 10.1109/MC.2016.183] [Medline: 27295638]
25. Varol O, Ferrara E, Davis C, Menczer F, Flammini A. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In: Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017). 2017 Presented at: The Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017); 2017; Montréal, Québec, Canada p. 280-289 URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587/14817
26. Aramaki E, Maskawa S, Morita M. Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.: Association for Computational Linguistics; 2011 Presented at: The 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2011; Edinburgh, Scotland, UK p. 1568-1576 URL: https://aclanthology.coli.uni-saarland.de/papers/D11-1145/d11-1145
27. Wakamiya S, Kawai Y, Aramaki E. Twitter-Based Influenza Detection After Flu Peak via Tweets With Indirect Information: Text Mining Study. JMIR Public Health Surveill 2018 Sep 25;4(3):e65 [FREE Full text] [doi: 10.2196/publichealth.8627]
28. Paul MJ, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. 2011 Presented at: International AAAI Conference on Web and Social Media; 2011; Barcelona, Spain p. 265-272 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2880
29. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
30. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014 Oct 28;6:1 [FREE Full text] [doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117] [Medline: 25642377]
31. Iso H, Wakamiya S, Aramaki E. Forecasting Word Model: Twitter-based Influenza Surveillance and Prediction. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics.: The COLING 2016 Organizing Committee; 2016 Presented at: The International Conference on Computational Linguistics (COLING); 2016; Osaka, Japan p. 76-86 URL: http://aclweb.org/anthology/C16-1008
32. Zhang Q, Perra N, Perrotta D, Tizzoni M, Paolotti D, Vespignani A. Forecasting Seasonal Influenza Fusing Digital Indicators and a Mechanistic Disease Model. In: Proceedings of the 26th International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee; 2017 Presented at: The International Conference on World Wide Web (WWW); 2017; Perth, Australia p. 311-319 URL: https://doi.org/10.1145/3038912.3052678
33. Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: A Comparative Analysis. JMIR Public Health Surveill 2016 Oct 20;2(2):e161 [FREE Full text] [doi: 10.2196/publichealth.5901] [Medline: 27765731]
34. Brownstein JS, Chu S, Marathe A, Marathe MV, Nguyen AT, Paolotti D, et al. Combining Participatory Influenza Surveillance with Modeling and Forecasting: Three Alternative Approaches. JMIR Public Health Surveill 2017 Nov 01;3(4):e83 [FREE Full text] [doi: 10.2196/publichealth.7344]
35. Diaz-Aviles E, Stewart A, Velasco E, Denecke K, Nejdl W. Towards personalized learning to rank for epidemic intelligence based on social media streams. In: Proceedings of the 21st International Conference on World Wide Web (WWW '12 Companion).: ACM; 2012 Presented at: The International Conference on World Wide Web (WWW) Companion; 2012; Lyon, France p. 495-496 URL: https://dl.acm.org/citation.cfm?id=2188094

36. Nsoesie EO, Flor L, Hawkins J, Maharana A, Skotnes T, Marinho F, et al. Social Media as a Sentinel for Disease Surveillance: What Does Sociodemographic Status Have to Do with It? PLoS Curr 2016 Dec 07;8:1 [FREE Full text] [doi: 10.1371/currents.outbreaks.cc09a42586e16dc7dd62813b7ee5d6b6] [Medline: 28123858]
37. Fung IC, Duke CH, Finch KC, Snook KR, Tseng P, Hernandez AC, et al. Ebola virus disease and social media: A systematic review. Am J Infect Control 2016 Dec 01;44(12):1660-1671. [doi: 10.1016/j.ajic.2016.05.011] [Medline: 27425009]
38. Fu KW, Liang H, Saroha N, Tse ZTH, Ip P, Fung ICH. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 01;44(12):1700-1702 [FREE Full text] [doi: 10.1016/j.ajic.2016.04.253] [Medline: 27566874]
39. Chandrasekaran N, Gressick K, Singh V, Kwal J, Cap N, Koru-Sengul T, et al. The Utility of Social Media in Providing Information on Zika Virus. Cureus 2017 Oct 23;9(10):e1792 [FREE Full text] [doi: 10.7759/cureus.1792] [Medline: 29282437]
40. Vijaykumar S, Nowak G, Himelboim I, Jin Y. Virtual Zika transmission after the first U.S. case: who said what and how it spread on Twitter. Am J Infect Control 2018 May;46(5):549-557 [FREE Full text] [doi: 10.1016/j.ajic.2017.10.015] [Medline: 29306490]
41. Yin Z, Fabbri D, Rosenbloom ST, Malin B. A Scalable Framework to Detect Personal Health Mentions on Twitter. J Med Internet Res 2015 Jun 05;17(6):e138 [FREE Full text] [doi: 10.2196/jmir.4305] [Medline: 26048075]
42. Ram S, Zhang W, Williams M, Pengetnze Y. Predicting asthma-related emergency department visits using big data. IEEE J Biomed Health Inform 2015 Jul;19(4):1216-1223 [FREE Full text] [doi: 10.1109/JBHI.2015.2404829] [Medline: 25706935]
43. Zhang W, Ram S, Burkart M, Pengetnze Y. Extracting Signals from Social Media for Chronic Disease Surveillance. In: Proceedings of the 6th International Conference on Digital Health Conference. 2016 Presented at: The 6th International Conference on Digital Health Conference (DH '16); 2016; Montréal, Québec, Canada p. 79-83 URL: https://doi.org/10.1145/2896338.2897728
44. Liu Y, Mei Q, Hanauer DA, Zheng K, Lee JM. Use of Social Media in the Diabetes Community: An Exploratory Analysis of Diabetes-Related Tweets. JMIR Diabetes 2016 Nov 07;1(2):e4 [FREE Full text] [doi: 10.2196/diabetes.6256] [Medline: 30291053]
45. Gesualdo F, Stilo G, D'Ambrosio A, Carloni E, Pandolfi E, Velardi P, et al. Can Twitter Be a Source of Information on Allergy? Correlation of Pollen Counts with Tweets Reporting Symptoms of Allergic Rhinoconjunctivitis and Names of Antihistamine Drugs. PLoS One 2015;10(7):e0133706 [FREE Full text] [doi: 10.1371/journal.pone.0133706] [Medline: 26197474]
46. Quincey E, Kyriacou T, Pantin T. #hayfever; A Longitudinal Study into Hay Fever Related Tweets in the UK. In: Proceedings of the 6th International Conference on Digital Health Conference.: ACM; 2016 Presented at: The 6th International Conference on Digital Health Conference (DH '16); 2016; Montréal, Québec, Canada p. 85-89 URL: https://doi.org/10.1145/2896338.2896342
47. Bastl K, Berger U, Kmenta M. Evaluation of Pollen Apps Forecasts: The Need for Quality Control in an eHealth Service. J Med Internet Res 2017 May 08;19(5):e152 [FREE Full text] [doi: 10.2196/jmir.7426] [Medline: 28483740]

Edited by G Eysenbach; submitted 20.03.18; peer-reviewed by J Jonnagaddala, M Thibaudon; comments to author 17.08.18; revised version received 08.11.18; accepted 10.12.18; published 20.02.19.

Please cite as:
Wakamiya S, Matsune S, Okubo K, Aramaki E
Causal Relationships Among Pollen Counts, Tweet Numbers, and Patient Numbers for Seasonal Allergic Rhinitis Surveillance: Retrospective Analysis
J Med Internet Res 2019;21(2):e10450
URL: http://www.jmir.org/2019/2/e10450/
doi:10.2196/10450
PMID:30785411

Original Paper

Tweet Classification Toward Twitter-Based Disease Surveillance: New Data, Methods, and Evaluations

Shoko Wakamiya[1,2,3], PhD; Mizuki Morita[4], PhD; Yoshinobu Kano[5], PhD; Tomoko Ohkuma[6], PhD; Eiji Aramaki[1,2,3], PhD

[1]Institute for Research Initiatives, Nara Institute of Science and Technology, Ikoma, Japan

[2]Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Japan

[3]Data Science Center, Nara Institute of Science and Technology, Ikoma, Japan

[4]Okayama University, Okayama, Japan

[5]Shizuoka University, Hamamatsu, Japan

[6]Fuji Xerox Co., Ltd., Yokohama, Japan

Corresponding Author:
Eiji Aramaki, PhD
Institute for Research Initiatives
Nara Institute of Science and Technology
8916-5 Takayama-cho
Ikoma, 630-0192
Japan
Phone: 81 743 72 6053
Fax: 81 743 72 6065
Email: socialcomputing-office@is.naist.jp

Abstract

Background: The amount of medical and clinical-related information on the Web is increasing. Among the different types of information available, social media–based data obtained directly from people are particularly valuable and are attracting significant attention. To encourage medical natural language processing (NLP) research exploiting social media data, the 13th NII Testbeds and Community for Information access Research (NTCIR-13) Medical natural language processing for Web document (MedWeb) provides pseudo-Twitter messages in a cross-language and multi-label corpus, covering 3 languages (Japanese, English, and Chinese) and annotated with 8 symptom labels (such as cold, fever, and flu). Then, participants classify each tweet into 1 of the 2 categories: those containing a patient's symptom and those that do not.

Objective: This study aimed to present the results of groups participating in a Japanese subtask, English subtask, and Chinese subtask along with discussions, to clarify the issues that need to be resolved in the field of medical NLP.

Methods: In summary, 8 groups (19 systems) participated in the Japanese subtask, 4 groups (12 systems) participated in the English subtask, and 2 groups (6 systems) participated in the Chinese subtask. In total, 2 baseline systems were constructed for each subtask. The performance of the participant and baseline systems was assessed using the exact match accuracy, F-measure based on precision and recall, and Hamming loss.

Results: The best system achieved exactly 0.880 match accuracy, 0.920 F-measure, and 0.019 Hamming loss. The averages of match accuracy, F-measure, and Hamming loss for the Japanese subtask were 0.720, 0.820, and 0.051; those for the English subtask were 0.770, 0.850, and 0.037; and those for the Chinese subtask were 0.810, 0.880, and 0.032, respectively.

Conclusions: This paper presented and discussed the performance of systems participating in the NTCIR-13 MedWeb task. As the MedWeb task settings can be formalized as the factualization of text, the achievement of this task could be directly applied to practical clinical applications.

(J Med Internet Res 2019;21(2):e12783) doi:10.2196/12783

KEYWORDS

text mining; social media; machine learning; natural language processing; artificial intelligence; surveillance; infodemiology; infoveillance

Introduction

Medical reports using electronic media are now replacing those of paper media [1,2]. As a result, the importance of natural language processing (NLP) techniques in various medical fields has increased significantly. Currently, the development of practical tools to assist precise and timely medical decisions has been encouraged.

To contribute to the progress of information retrieval research, a series of *shared tasks* (or contests, competitions, challenge evaluations, and critical assessments) is being used. Thus far, several shared tasks related to medical or health care have already been organized and provided datasets for various NLP tasks. These include the Informatics for Integrating Biology and the Bedside (i2b2) tasks [3], the Text Retrieval Conference (TREC) Medical Records track [4], TREC Clinical Decision Support or Precision Medicine tracks [5-9], the Cross-Language Evaluation Forum for European Languages (CLEF) eHealth [10], and NII Testbeds and Community for Information access Research (NTCIR) Medical tasks and Medical Natural Language Processing (MedNLP) workshops [11-16]. Generally, these shared tasks provide clinical records.

On the other hand, with the widespread use of the internet, considerable material concerning medical or health care has been shared on the Web, and several Web mining techniques for utilizing the material have been developed. One of the most popular medical applications of Web mining is flu surveillance [17-27]. Although most previous studies have relied on shallow textual clues in messages, such as the number of occurrences of specific keywords (eg, *flu* or *influenza*), such simple approaches have difficulty coping with the volume of noisy messages. Typical examples of noisy tweets on Twitter are those that simply express concern or awareness about flu (such as “Starting to get worried about swine flu”). To increase their accuracy, one of the most reasonable approaches employs a binary classifier to filter out noisy messages.

Given this situation, the NTCIR-13 [28] Medical Natural Language Processing for Web Document (MedWeb) task [29,30] is designed for obtaining health-related information by exploiting data on the Web, focusing on social media sites such as Twitter [31] and Facebook [32]. Specifically, we propose a generalized task setting that determines whether a message is written about a patient affected by a specific symptom for public health surveillance, referring to the following 2 characteristics:

1. Multi-label: This task handles not only a single symptom (such as influenza) but also multiple symptoms such as cold, cough or sore throat, diarrhea or stomach ache, fever, hay fever, headache, and runny nose. As a single message can contain multiple symptoms, this is a multi-labeling task.
2. Cross-language: In contrast to the previous shared tasks, this task covers multiple languages, such as Japanese, English, and Chinese. To build parallel corpora, we translated the original Japanese messages to English and Chinese.

In the NTCIR-13 MedWeb, we distributed each corpus to the participants [33-41], of whom 9 groups (7 academia groups, an industry group, and a joint group) submitted results (37 systems). Specifically, 8 groups (19 systems) participated in the Japanese subtask, 4 groups (12 systems) participated in the English subtask, and 2 groups (6 systems) participated in the Chinese subtask (see Multimedia Appendix 1). This report presents the results of these groups, along with discussions, to clarify the issues that need to be resolved in the field of medical NLP.

Methods

Materials

Data

The MedWeb task uses a collection of tweets that include at least one keyword of target diseases or symptoms (for brevity, we refer to these simply as *symptoms* hereafter). We set 8 symptoms, including cold, cough or sore throat (which we refer to as *cough*), diarrhea or stomachache (*diarrhea*), fever, hay fever, headache, influenza (*flu*), and runny nose.

Owing to the Twitter Developer Policy on data redistribution [42], the tweet data crawled using the application programming interface [43] are not publicly available. Therefore, our data consist of pseudotweets created by a crowdsourcing service.

To obtain the pseudotweets, we first collected Japanese tweets related to each symptom from Twitter. Then, we classified these tweets as positive or negative based on the previous study [19]. Next, we extracted keyword sets that appeared frequently in the positive and negative tweets of the symptom by calculating term frequency and inverse document frequency. We call these keywords *seed words*.

We then had a group of people create pseudotweets consisting of 100 to 140 characters, which included a symptom and at least one of the seed words of the symptom. Each person created 32 pseudotweets (2 tweets × 2 keyword sets [positive and negative] × 8 symptoms). As a result, 80 people were able to generate 2560 Japanese pseudotweets.

In the last step, we had the Japanese pseudotweets translated into English and Chinese by relevant first-language practitioners. Therefore, we also had 2560 pseudotweets in both English and Chinese. The corpora are available in a previous paper [44]. Textbox 1 shows samples of each set of pseudotweets, whose ratios of positive labels are presented in Table 1. This table shows the ratio of positive labels out of each symptom’s 320 pseudotweets and the number of positive labels out of all symptoms’ 2560 pseudotweets. Common symptoms such as a runny nose, fever, headache, and cold tend to appear with the other symptoms. Then, the number of tweets of the flu labeled as *p* (positive) is relatively less than the others, indicating that the flu is likely to be a topic even if people did not suffer from flu. On the other hand, tweets concerning several symptoms such as a cough, headache, runny nose, and diarrhea are described in many cases when people suffered from them.

Textbox 1. Samples of pseudotweets of the 8 symptoms. Note that English messages and Chinese messages were translated from Japanese messages.

1. Cold
   - 風邪を引くと全身がだるくなる
   - The cold makes my whole body weak.
   - 一感冒就浑身酸软无力。
2. Cough
   - あかん。咳込みすぎて頭まで痛くなってきた
   - This is not good. I coughed too much and I got a headache from it.
   - 糟了。咳得太厉害，头都疼起来了。
3. Diarrhea
   - 下痢ひどすぎで笑うわ
   - I gotta laugh. My diarrhea is so bad.
   - 腹泻过于严重，很搞笑。
4. Fever
   - 熱が出なくてもリンパが腫れることがよくある。
   - It’s not unusual for lymph nodes to get swollen, even when there’s no fever.
   - 很多时候就算不发热淋巴也肿。
5. Hay fever
   - 花粉症の症状が出てきたのは久し振りだ。
   - It’s been a while since I’ve had allergy symptoms.
   - 好久没有出现花粉症的症状了。
6. Headache
   - 頭痛がやばいから帰宅して寝るー
   - My headache is killing me, so I’m going to go home and sleep.
   - 因为头疼得厉害，我回家睡觉了。
7. Flu
   - インフルエンザのワクチン打ちに行ってきた。
   - I went to get vaccinated for the flu.
   - 去打了流感的疫苗。
8. Runny nose
   - 鼻づまりで今日は休むわー
   - I’m not going today, because my stuffy nose is killing me.
   - 因为鼻塞，今天休息吧！

Table 1. Ratio of positive labels.

Symptom	Ratio of number of positive tweets to the number of each symptom's tweets (N=320 tweets)	Ratio of number of positive tweets to the total number of all symptoms' tweets (N=2560 tweets)
Cold, n (%)	220 (0.6875)	355 (0.1387)
Cough, n (%)	295 (0.9219)	306 (0.1195)
Diarrhea, n (%)	230 (0.7188)	246 (0.0961)
Fever, n (%)	220 (0.6875)	438 (0.1711)
Hay fever, n (%)	208 (0.6500)	209 (0.0816)
Headache, n (%)	260 (0.8125)	328 (0.1281)
Flu, n (%)	128 (0.4000)	130 (0.0508)
Runny nose, n (%)	257 (0.8031)	499 (0.1949)

Table 2. Samples of the training data corpus for the English subtask.

Tweet ID	Message	s_1[a]	s_2	s_3	s_4	s_5	s_6	s_7	s_8
1en[b]	The cold makes my whole body weak.	p[c]	n[d]	n	n	n	n	n	n
2en	It's been a while since I've had allergy symptoms.	n	n	n	n	p	n	n	p
3en	I'm so feverish and out of it because of my allergies. I'm so sleepy.	n	n	n	p	p	n	n	p
4en	I took some medicine for my runny nose, but it won't stop.	n	n	n	n	n	n	n	p
5en	I had a bad case of diarrhea when I traveled to Nepal.	n	n	n	n	n	n	n	n
6en	It takes a millennial wimp to call in sick just because they're coughing. It's always important to go to work, no matter what.	n	p	n	n	n	n	n	n
7en	I'm not going today, because my stuffy nose is killing me.	n	n	n	n	n	n	n	p
8en	I never thought I would have allergies.	n	n	n	n	p	n	n	p
9en	I have a fever but I don't think it's the kind of cold that will make it to my stomach.	p	n	n	p	n	n	n	n
10en	My phlegm has blood in it and it's really gross.	n	p	n	n	n	n	n	n

[a]s_1, s_2, s_3, s_4, s_5, s_6, s_7, and s_8 are IDs of the 8 symptoms (cold, cough, diarrhea, fever, hay fever, headache, flu, and runny nose).

[b]ID corresponds to the corpora of other languages (eg, the tweet of *1en* corresponds to the tweets of *1ja* and *1zh*).

[c]p indicates the positive label.

[d]n indicates the negative label.

Symptom Labeling

This section describes the criteria used for symptom labeling: basic criteria and symptom-specific criteria [45,46]. In this study, 2 annotators attached positive or negative labels of the 8 symptoms to tweets (Table 2).

Basic Criteria

The most basic criterion is that the labeling is examined from a clinical viewpoint, considering the medical importance of the information. Thus, nonclinical information should be disregarded. For example, older information (by several weeks) and nonsevere symptoms (a headache due to overdrinking) should be labeled as *n* (negative).

The following 3 criteria describe the basic principles:

1. Factuality: The Twitter user (or someone close to the user) should be affected by a certain disease or have a symptom of the disease. A tweet that includes only the name of a disease or a symptom as a topic is removed by labeling it as *n* (negative).
2. Tense (time): Older information, which is meaningless from the viewpoint of surveillance, should be discarded. Such information should also be labeled as *n* (negative). Here,

we regard 24 hours as the standard condition. When the precise date and time are ambiguous, the general guideline is that information within 24 hours (eg, information related to the present day or previous day) is labeled as *p* (positive).

3. Location: The location of the disease should be specified as follows. If a Twitter user is affected, the information is labeled as *p* (positive) because the location of the user is the place of onset of the symptom. In cases where the user is not personally affected, the information is labeled as *p* (positive) if it is within the same vicinity (prefecture) as that of the user, and as *n* (negative) otherwise.

Symptom-Specific Criteria

There are several exceptions to the fundamental annotation principles. For example, a remark about a *headache* might not relate to that about a clinical disease (such as headache due to excessive drinking). When conducting disease surveillance, such statements should be regarded as noise. To deal with disease-specific phenomena, we build a guideline that addresses exceptions for each disease. For example, cases such as *excessive drinking*, *medication*, *pungently flavored food (including irritant)*,*spiritual*,*motion sickness*,*morning*, and *menstrual pain* should be excluded for *headache*. The exceptions are summarized in Table 3.

Task Settings

In the MedWeb task, we organized 3 subtasks: a Japanese subtask, an English subtask, and a Chinese subtask. The procedure of the MedWeb task is as follows:

Step 1. Training corpus distribution: The training data corpus and the annotation criteria were sent to the participant groups for development. The training data corpus comprises 1920 messages (75.00% of the whole corpus), with labels. Each message is labeled *p* (positive) or *n* (negative) for each of the 8 symptoms.

Step 2. Formal run result submission: After about a 3-month development period, the test data corpus was sent to each participant group. The test data corpus consists of 640 messages (25.0% of the whole corpus), without labels. Then, the participant groups developed their systems (Table 4) and submitted their annotated results within 2 weeks. Multiple results with up to 3 systems were allowed to be submitted.

Step 3. Evaluation result release: After a 1-month evaluation period, the evaluation results and annotated test data were sent to each participant group.

Systems

Baseline Systems

As a baseline, 2 systems were constructed using a support vector machine (SVM) based on unigram and bigram features. For feature representation, the bag-of-words model was used in each system. A tweet message was segmented using MeCab, created by Kudo et al [47] for Japanese messages, natural language toolkit (NLTK) TweetTokenizer, created by Bird [48,49] for English messages, and jieba, created by Junyi [50] for Chinese messages. The 2 systems had a linear kernel, and the parameter for regularization, *C*, was set to 1. The baseline systems were implemented using scikit-learn (sklearn) [51,52].

Participating Systems

In all, 37 systems (of 9 groups) participated and had their results submitted in the MedWeb. Of these, 19 systems (of 8 groups) submitted results for the Japanese subtask, 12 systems (of 4 groups) for the English subtask, and 6 systems (of 2 groups) for the Chinese subtask. The participating systems for the Japanese, English, and Chinese subtasks are summarized in Table 4.

Table 3. Exceptions for symptom labels.

Symptom	Expressions with suspicion	Just a symptom word	Exceptions	
			Regarded as symptom	Not regarded as symptom
Cold	Accept	Accept	—[a]	—
Cough	Accept	Accept	Alcohol drinking and pungently flavored food	—
Diarrhea	Accept	Accept	Overeating, indigestion, alcohol drinking, medication, and pungently flavored food	—
Fever	Accept	Only *slight fever*	Hay fever and side effect due to any injection	—
Hay fever	Accept	Accept	—	—
Headache	Accept	Accept	—	Due to a sense of sight or smell
Flu	Not accept	Not accept	—	—
Runny nose	Accept	Not accept	Hay fever	Change in temperature

[a]Indicates there are no exceptions.

Table 4. Participating systems in subtasks. A total of 19 participating systems and 2 baseline systems are constructed for the Japanese subtask, 12 participating systems and 2 baseline systems are constructed for the English subtask, and 6 participating systems and 2 baseline systems are constructed for the Chinese subtask.

System ID	Models or methods	Language resources
AITOK-ja [33]	Keyword-based, logistic regression, and SVM[a,b]	—[c]
AKBL-ja and AKBL-en [34]	SVM and Fisher exact test	Patient symptom feature word dictionary and Disease-X feature words dict1 and dict2
DrG-ja [35]	Random forest	—
KIS-ja [36]	Rule-based and SVM	—
NAIST-ja, NAIST-en, and NAIST-zh [37]	Ensembles of hierarchical attention network and deep character-level convolutional neural network with loss functions (negative loss function, hinge, and hinge squared)	—
NIL-ja [38]	Rule-based	—
NTTMU-ja [39]	Principle-based approach	Manually constructed knowledge for capturing tweets that conveyed flu-related information, using common sense and ICD-10[d]
NTTMU-en [39]	SVM and recurrent neural network	Manually constructed knowledge for capturing tweets that conveyed flu-related information, using common sense and ICD-10
TUA1-zh [40]	Logistic regression, SVM, and logistic regression with semantic information	Updated training samples using active learning unlabeled posts downloaded with the symptom names in Chinese
UE-ja [41]	Rule-based and random forest	Custom dictionary consisting of nouns selected from the dry-run dataset and heuristics
UE-en [41]	Rule-based, random forests, and skip-gram neural network for word2vec	Custom dictionary consisting of nouns selected from the dry-run dataset and heuristics
Baseline	SVM (unigram and bigram)	—

[a]SVM: support vector machine.

[b]It indicates that the method was tested after the submission of the formal run, and thus, it was not included in the results.

[c]It indicates that any language resources were not used.

[d]ICD: International Codes for Diseases.

As for the Japanese subtask, most of the groups applied machine learning approaches, such as SVM (as in the baseline systems), random forests, and neural networks. Several groups constructed their own resources to enhance the original training corpus. Similarly, for the English subtask, most of the groups applied machine learning approaches, such as SVM, random forests, and neural networks. The Chinese subtask had 2 participating groups: one applied the same methods as the other subtasks and the other used logistic regression and SVM and updated the training data using active learning.

Evaluation Metrics

The performance in the subtasks was assessed using the exact match accuracy, F-measure (beta=1) (*F1*) based on precision and recall, and Hamming loss [53].

The details of the metrics are as follows.

- Exact match accuracy: If $y'^{(i)}$ indicates the predicted symptom label values of the *i*-th tweet and $y^{(i)}$ is the corresponding true labels, then the fraction of correct predictions over the test data corpus (N=640) is calculated as follows: $accuracy(y, y')=1/N\cdot\sum_{i=1}^{N}1(y'^{(i)}=y^{(i)})$, where *1(.)* is the indicator function, which returns 1 if the entire set of predicted labels for a tweet strictly matches with the true set of labels.
- Precision (micro or macro): It is defined as the number of true positives (T_p) over the sum of the number of true positive and false positives (F_p): $precision=T_p/(T_p+F_p)$.
- Recall (micro or macro): It is defined as the number of true positives (T_p) over the sum of the number of true positive and false negatives (F_n): $recall=T_p/(T_p+F_n)$.
- F1 (micro or macro): The harmonic mean of precision and recall is calculated as follows: $F1=2\cdot precision\cdot recall/(precision+recall)$.
- Hamming loss: It computes the average Hamming loss between two sets of labels. If $y'^{(i)}$ is the predicted value for the *j*-th label of the *i*-th tweet, $y^{(i)}_j$ is the corresponding true value, N is the number of test data (N=640), and L is the number of labels (L=8), then the Hamming loss between the predicted and correct labels is calculated as follows: $L_{Hamming}(y,y')=1/N\cdot 1/L\cdot\sum_{i=1}^{N}\sum_{j=1}^{L}1(y'^{(i)}_j\neq y^{(i)}_j)$, where *1(.)* is the indicator function. Note that lower scores are better.

Note that *micro* is to calculate metrics globally by counting all true positives, false negatives, and false positives: $F1micro=2 \cdot precisionmicro \cdot recallmicro/(precisionmicro+recallmicro)$.

On the other hand, *macro* calculates the metrics for each symptom label and then determines their unweighted mean: $F1macro=1/L \cdot \sum_{j=1}^{L} F1_j$. Therefore, label imbalance is not taken into account.

Ethics Statement

This study did not require the participants to be involved in any physical and/or mental intervention. As this research did not use personally identifiable information, it was exempted from the institutional review board approval in accordance with the Ethical Guidelines for Medical and Health Research Involving Human Subjects stipulated by the Japanese national government.

Results

Symptom Labeling Reliability

To show the reliability of symptom labeling to the corpus, the interannotation agreement ratios of the respective symptoms were measured (Table 5). The total interannotator agreement ratio (n=2) was 0.9851 (ie, 20,174 / [2560 × 8]).

Performance of Baseline Systems

The performance of the baseline was measured using all evaluation metrics. Tables 6-8 show the results for the Japanese, English, and Chinese subtasks, respectively.

For the Japanese and Chinese subtasks, unigram SVM performed better than bigram SVM. On the other hand, bigram SVM outperformed unigram SVM in the English subtask. The highest average of exact match accuracy was 0.791 (English subtask) and the lowest was 0.756 (Japanese subtask).

Performance of Participating Systems

The performance of the participating systems was also measured using all evaluation metrics. Tables 6-8 show the results for the Japanese, English, and Chinese subtasks, respectively, ordered by the exact match accuracy of the systems.

For the Japanese subtask, the best system, NAIST-ja-2, achieved 0.880 in exact match accuracy, 0.920 in F-measure, and 0.019 in Hamming loss, as shown in Table 6. The averages across the participating groups and the baseline systems were 0.720, 0.820, and 0.051, respectively. The rank order of the top 4 systems was the same in all measures. The systems of the AKBL and KIS groups were constructed using an SVM, as in the baseline systems. The AKBL group's results indicated that their system was effective in terms of using additional language resources. The KIS group switched their methods between an SVM and a rule-based method, depending on the confidence factor.

For the English subtask, the best system, NAIST-en-2, achieved 0.880 in exact match accuracy, 0.920 in F-measure, and 0.019 in Hamming loss, as shown in Table 7. The system was constructed using the same method as that used in the Japanese subtask. The averages across the participating groups and the baseline systems were 0.770, 0.850, and 0.037, respectively.

For the Chinese subtask, the best system, NAIST-zh-2, achieved 0.880 in exact match accuracy, 0.920 in F-measure, and 0.019 in Hamming loss, as shown in Table 8. The system was constructed using the same method as that used in the Japanese and English subtasks. The averages across the participating groups and the baseline systems were 0.810, 0.880, and 0.032, respectively.

Table 5. Interannotator agreement ratio.

Symptom	Agreement ratio (number)
Cold	0.9945 (2546/2560)
Cough	0.9934 (2543/2560)
Diarrhea	0.9785 (2505/2560)
Fever	0.9922 (2540/2560)
Hay fever	0.9918 (2539/2560)
Headache	0.9773 (2502/2560)
Flu	0.9734 (2492/2560)
Runny nose	0.9793 (2507/2560)
Total	0.9851 (20,174/20,480)

Table 6. Performance in the Japanese subtask (19 participating systems and 2 baseline systems).

System ID[a]	Exact match[b]	F1		Precision		Recall		Hamming loss
		Micro	Macro	Micro	Macro	Micro	Macro	
NAIST-ja-2	0.880	0.920	0.906	0.899	0.887	0.941	0.925	0.019
NAIST-ja-3	0.878	0.919	0.904	0.899	0.885	0.940	0.924	0.019
NAIST-ja-1	0.877	0.918	0.904	0.899	0.887	0.938	0.921	0.020
AKBL-ja-3	0.805	0.872	0.859	0.896	0.883	0.849	0.839	0.029
UE-ja-1	0.805	0.865	0.855	0.831	0.819	0.903	0.902	0.033
KIS-ja-2	0.802	0.871	0.856	0.831	0.815	0.915	0.904	0.032
AKBL-ja-1	0.800	0.869	0.847	0.889	0.873	0.849	0.825	0.030
UE-ja-3	0.800	0.866	0.855	0.823	0.812	0.913	0.911	0.033
AKBL-ja-2	0.795	0.868	0.849	0.891	0.875	0.846	0.827	0.030
KIS-ja-3	0.784	0.855	0.831	0.840	0.816	0.871	0.850	0.034
SVM-unigram	0.761	0.849	0.835	0.843	0.828	0.854	0.842	0.036
KIS-ja-1	0.758	0.849	0.833	0.798	0.782	0.906	0.899	0.038
SVM-bigram	0.752	0.843	0.830	0.838	0.820	0.848	0.845	0.037
NTTMU-ja-1	0.738	0.835	0.829	0.770	0.761	0.913	0.921	0.042
UE-ja-2	0.706	0.815	0.803	0.696	0.702	0.983	0.984	0.052
NIL-ja-1	0.680	0.749	0.742	0.862	0.845	0.662	0.671	0.052
DrG-ja-1	0.653	0.777	0.774	0.825	0.808	0.734	0.779	0.049
NTTMU-ja-3	0.614	0.775	0.773	0.740	0.720	0.814	0.840	0.055
NTTMU-ja-2	0.597	0.770	0.753	0.741	0.706	0.801	0.813	0.056
AITOK-ja-2	0.503	0.706	0.696	0.726	0.738	0.687	0.767	0.067

[a]The system ID comprises the group ID (see Multimedia Appendix 1), the abbreviation of subtask (ja indicates Japanese subtask), and the system number from 1 to 3 since each group can submit three systems per subtask.

[b]The results are ordered by exact match accuracy.

Table 7. Performance in the English subtask (12 participating systems and 2 baseline systems).

System ID[a]	Exact match[b]	F1		Precision		Recall		Hamming loss
		Micro	Macro	Micro	Macro	Micro	Macro	
NAIST-en-2	0.880	0.920	0.906	0.899	0.887	0.941	0.925	0.019
NAIST-en-3	0.878	0.919	0.904	0.899	0.885	0.940	0.924	0.019
NAIST-en-1	0.877	0.918	0.904	0.899	0.887	0.938	0.921	0.020
SVM-bigram	0.800	0.866	0.856	0.865	0.849	0.868	0.865	0.031
UE-en-1	0.789	0.858	0.848	0.846	0.831	0.871	0.876	0.034
SVM-unigram	0.783	0.858	0.845	0.851	0.830	0.864	0.864	0.033
NTTMU-en-2	0.773	0.856	0.849	0.807	0.796	0.911	0.918	0.036
NTTMU-en-3	0.758	0.845	0.828	0.836	0.818	0.854	0.844	0.037
UE-en-2	0.745	0.821	0.809	0.861	0.838	0.786	0.800	0.040
UE-en-3	0.739	0.820	0.815	0.870	0.851	0.776	0.795	0.040
AKBL-en-2	0.734	0.819	0.799	0.832	0.808	0.806	0.793	0.042
AKBL-en-3	0.716	0.804	0.787	0.853	0.834	0.760	0.747	0.043
NTTMU-en-1	0.619	0.770	0.777	0.734	0.733	0.809	0.835	0.056
AKBL-en-1	0.613	0.772	0.755	0.656	0.649	0.936	0.945	0.065

[a]The system ID comprises the group ID (see Multimedia Appendix 1), the abbreviation of subtask (en indicates English subtask), and the system number from 1 to 3 since each group can submit three systems per subtask.

[b]The results are ordered by exact match accuracy.

Table 8. Performance in the Chinese subtask (6 participating systems and 2 baseline systems).

System ID[a]	Exact match[b]	F1		Precision		Recall		Hamming loss
		Micro	Macro	Micro	Macro	Micro	Macro	
NAIST-zh-2	0.880	0.920	0.906	0.899	0.887	0.941	0.925	0.019
NAIST-zh-3	0.878	0.919	0.904	0.899	0.885	0.940	0.924	0.019
NAIST-zh-1	0.877	0.918	0.904	0.899	0.887	0.938	0.921	0.020
TUA1-zh-3	0.786	0.860	0.844	0.772	0.760	0.970	0.971	0.037
SVM-unigram	0.780	0.858	0.843	0.831	0.815	0.888	0.883	0.034
TUA1-zh-1	0.773	0.853	0.838	0.766	0.753	0.963	0.965	0.039
SVM-bigram	0.767	0.850	0.835	0.824	0.806	0.878	0.876	0.036
TUA1-zh-2	0.719	0.824	0.809	0.712	0.710	0.978	0.982	0.049

[a]The system ID comprises the group ID (see Multimedia Appendix 1), the abbreviation of subtask (zh indicates Chinese subtask), and the system number from 1 to 3 since each group can submit 3 systems per subtask.

[b] The results are ordered by exact match accuracy.

Discussion

Principal Findings

One of the most valuable findings was that we could determine the best strategy for disease surveillance. The best system of the NAIST group had 2 characteristics: (1) cross-language features and (2) ensemble of multiple machine learning methods.

Cross-Language Features

For each language, the NAIST system utilized features from the other 2 languages. English and Chinese sentences were translated from a Japanese sentence, indicating that these 3 sentences shared the same symptom label set. Only the NAIST system focused on the property of this task's corpus and improved the accuracy from 0.767 to 0.823 in exact match.

Ensemble Methods

The NAIST system also utilized an ensemble method, which combines multiple methods to boost the classification accuracy. Although weak machine learning algorithms tend to be generally preferred to make an ensemble, the NAIST group created an ensemble consisting of strong machine learning methods: a hierarchical attention network and a deep convolutional neural network (CNN). The combination of methods varied the exact match accuracy of 0.836 at the minimum to 0.880 at the

maximum. In the near future, a technique to find a better combination needs to be developed.

Out of the 2 features, the cross-language feature is the unique feature of this task. Even if we discounted the cross-language feature, the NAIST ensemble method exhibited the best performance. As the multi-label classification is known as a complex task, the performance of straightforward approaches relying only on 1 method was relatively lower than that of the NAIST system.

Note that previous NTCIR medical tasks and MedNLP workshops [13-15] have shown that a rule-based approach is still competitive with machine learning approaches. One of the reasons for this was the small size of the corpus they used. Although the corpus size was also limited in this task, this result showed the advantage of complex machine learning, indicating the advancement of machine learning techniques.

Subtask-Based Comparison

The MedWeb task provided a cross-language corpus. Although this is another characteristic of this task, only 1 group (NAIST) challenged all subtasks, which was lesser than our expectation. The Japanese subtask had the highest participation (19 systems from 8 groups), whereas the Chinese subtask had the lowest participation (6 systems from only 2 groups), which was also lower than our expectation.

The performance varied depending on the subtasks. Figure 1 shows the distribution of the 3 metric scores of the systems in each subtask. For the Japanese subtask, the performance varied widely, relative to that of the other subtasks. Although the Chinese subtask had the lowest participation, their performance was relatively high. The 4 groups that participated in the Japanese subtask also challenged the English subtask, with better results, on average, in the English subtask. This indicates that the difficulty of classification in increasing order is Chinese, English, and Japanese. This is a surprising result because most of the groups came from Japan and must have been familiar with the Japanese NLP.

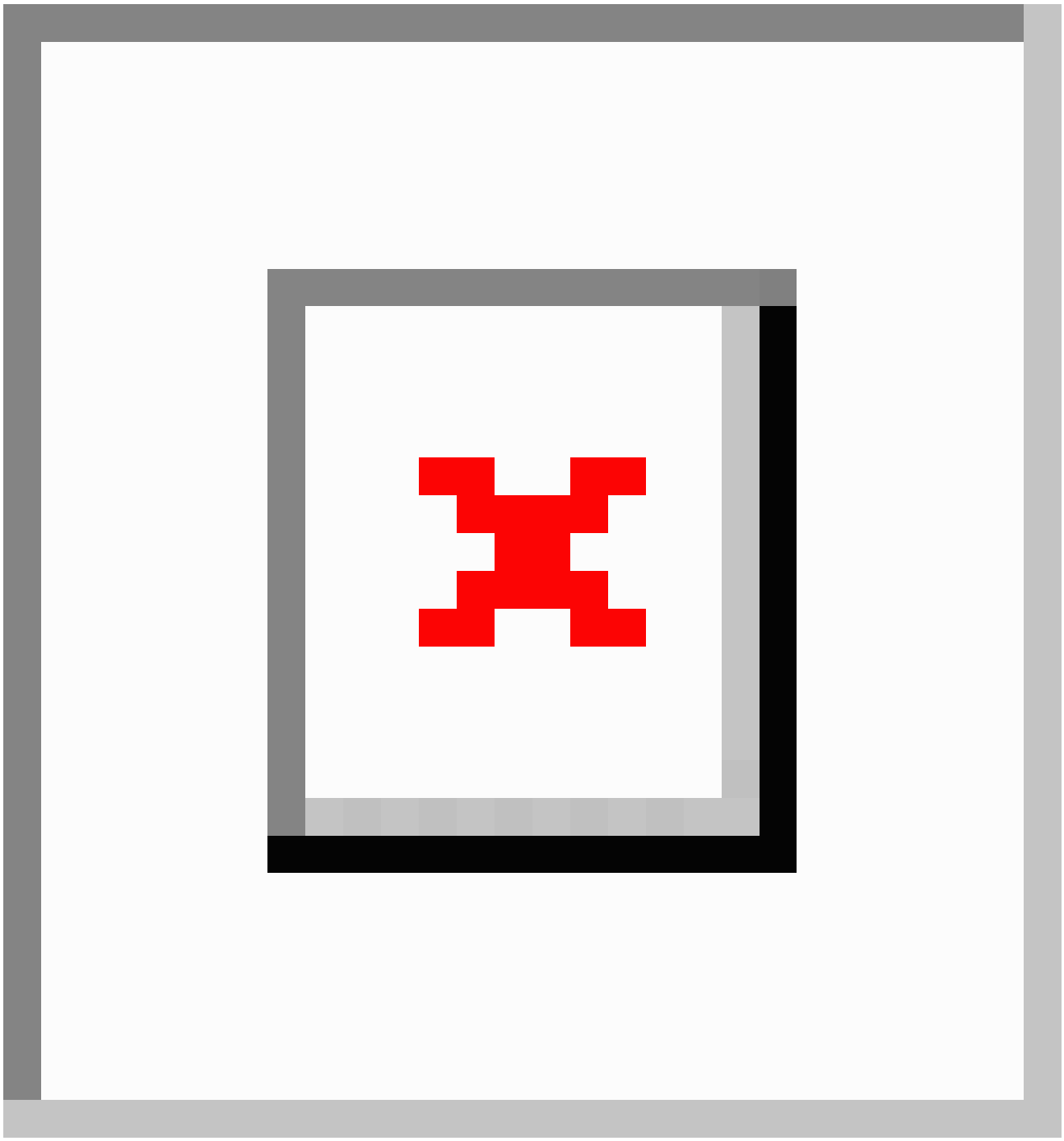
This indicates that the Chinese language has less ambiguity in clinical factuality analyses. Another possibility is that the process we used to generate the corpora had a language bias. For example, the translations from Japanese to English and Chinese might have reduced the ambiguity of the language in each case. To test for language bias, experiments based on different directions of translation are necessary. This is left for future work. Note that the baseline systems performed the best in the English subtask, indicating that the standard settings for SVM are effective in terms of classifying English tweets.

Limitations

The corpora provided by the MedWeb task have the following limitations. The first is the generating process of the corpora. For example, our pseudotweets do not include several tweet-specific features such as reply, retweet, hashtag, and URL. In addition, the translation process might bias the results. Although we asked translators to translate Japanese short messages without following standard English or Chinese as they could, some of them would be more formal than tweets.

Another limitation is the size of each corpus (1920 messages are used as training data and 640 messages are used as test data). Regardless of these limitations, we believe that this is a valuable attempt to generate and share a cross-language corpus consisting of multi-label pseudotweets.

Figure 1. Statistical summary of the performance of 3 evaluation metrics (A: Exact math accuracy, B: F1-micro, and C: Hamming loss) in each of the subtasks (ja: Japanese, en: English, and zh: Chinese). Note that higher scores are better in exact match accuracy and F1-micro, whereas lower scores are better in hamming loss. The bottom and top of a box are the first and third quartiles, the band inside the box is the median, and the dotted band inside the box is the mean. Dots on the right side of the box represent the distribution of values of participating systems.



Although our corpus has some limitations, we still believe it is helpful as a benchmark for tweet-based applications, because it is freely available and covers multiple languages.

Comparison With Prior Work

Currently, several shared tasks related to medical or health care have already been held. In the United States, the i2b2 tasks [3] were organized by the National Institute of Health [54] to enhance the ability of NLP tools to extract fine-grained information from clinical records. Specifically, i2b2 has provided sets of fully deidentified notes and proposed several challenges, such as deidentification and heart disease risk factor challenge, temporal relation challenge, conference challenge, relation challenge, medication challenge, obesity challenge, and deidentification and smoking challenge.

In addition, the TREC Medical Records Track (TREC2011-2012) [4] was established for the research community to focus on the problem of providing content-based access to free text fields of electronic health records. Then, Clinical Decision Support/Precision Medicine Tracks (TREC2014-2018) [5-9] were organized in TREC. The Clinical Decision Support Track focused on the retrieval of biomedical articles relevant for answering generic clinical questions about

medical records, and TREC Precision Medicine Track focused on a use case in clinical decision support, providing useful precision medicine-related information to clinicians treating cancer patients.

Furthermore, the CLEF eHealth [10] focused on NLP and information retrieval for clinical care in the European Union. In Japan, NTCIR Medical tasks and MedNLP workshops (MedNLP-1, MedNLP-2, and MedNLP-Doc) [11-16] were organized to promote and support the generation of practical tools and systems applicable in the medical industry, which will support medical decisions and treatments by physicians and medical staff. MedNLP-1 [11,14] aimed to retrieve important information (personal and medical) from the clinical text written in Japanese. MedNLP-2 [12,15] challenged to extract information from medical reports written by physicians and from past medical exams. MedNLP-Doc [13,16] proposed a task to guess the name of the disease (represented by the International Codes for Diseases [ICD]) from the provided medical records. However, to the best of our knowledge, the MedWeb is the first shared task for dealing with health-related social media data.

Due to the widespread use of the internet, considerable material concerning medical care or health has been made available on the Web, especially social media such as Twitter and Facebook. Furthermore, various Web mining techniques for utilizing the material have been developed. One of the most popular medical applications is disease surveillance, which aims to predict disease epidemics based on the use of disease-related terms. Particularly, influenza surveillance using social media has been extensively studied [17-27,55]. As most previous studies have relied on shallow textual clues in messages, such as the number of occurrences of specific keywords (eg, *flu* or *influenza*), there are several noisy messages. To filter out noisy tweets, a binary classifier has been employed. In contrast, the MedWeb has challenged a more difficult and practical task of performing a multi-label classification of cross-language user-generated messages.

Conclusions

This paper provided an overview of the NTCIR-13 MedWeb task, which was designed as a more generalized task for public surveillance, focusing on social media (such as Twitter). In particular, the task's goal was to classify symptom-related messages. This task had 2 characteristics: (1) multi-label (cold, cough, diarrhea, fever, hay fever, headache, flu, and runny nose) and (2) cross-language (Japanese, English, and Chinese). The results empirically demonstrated that an ensemble of multiple machine learning methods was effective in terms of classification of cross-language messages with multiple labels. We believe that the findings would be a foundation for future and deeper approaches for disease surveillance with social media data.

Acknowledgments

This work was supported by the Japan Agency for Medical Research and Development (Grant Number: JP16768699) and JST ACT-I (JPMJPR16UU). The authors appreciate annotators in Social Computing laboratory at Nara Institute of Science and Technology for their efforts on generating the corpus. The authors also greatly appreciate the NTCIR-13 chairs for their efforts on organizing the NTCIR-13 workshop. Finally, the authors thank all participants for their contributions to the NTCIR-13 MedWeb task.

Authors' Contributions

SW, MM, YK, TO, and EA organized the shared task; SW and EA created the data; SW and EA analyzed the results; and SW, MM, YK, TO, and EA prepared the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Organization of groups participating in MedWeb and statistics of result submissions. Note that it is listed in alphabetical order by Group ID, and ja, en, and zh correspond to Japanese, English, and Chinese subtasks, respectively.

[PDF File (Adobe PDF File), 20KB - jmir_v21i2e12783_app1.pdf]

References

1. Boonstra A, Versluis A, Vos JF. Implementing electronic health records in hospitals: a systematic literature review. BMC Health Serv Res 2014 Sep 4;14:370 [FREE Full text] [doi: 10.1186/1472-6963-14-370] [Medline: 25190184]
2. Hsiao CJ, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001-2012. NCHS Data Brief 2012 Dec(111):1-8 [FREE Full text] [doi: 10.1097/01.sa.0000451505.72517.a5] [Medline: 23384787]
3. i2b2: Informatics for Integrating Biology & the Bedside. URL: https://www.i2b2.org/index.html [accessed 2019-01-20] [WebCite Cache ID 75a8EIuUR]

4. Voorhees EM, Hersh W. National Institute of Standards and Technology. 2013 Jun 28. Overview of the TREC 2012 Medical Records Track URL: https://ws680.nist.gov/publication/get_pdf.cfm?pub_id=913781 [accessed 2019-01-22] [WebCite Cache ID 75cJJI0KS]
5. TREC Precision Medicine / Clinical Decision Support Track. URL: http://www.trec-cds.org/ [accessed 2019-01-21] [WebCite Cache ID 75aKpBWwa]
6. Simpson MS, Voorhees EM, Hersh W. Overview of the TREC 2014 Clinical Decision Support Track. In: NIST Special Publication 500-308: The Twenty-Third Text REtrieval Conference Proceedings (TREC 2014). 2014 Nov Presented at: TREC 2014; November 18-21, 2014; Gaithersburg, MD p. 1-8 URL: https://trec.nist.gov/pubs/trec23/papers/overview-clinical.pdf
7. Roberts K, Simpson MS, Voorhees EM, Hersh WR. Overview of the TREC 2015 Clinical Decision Support Track. In: NIST Special Publication: SP 500-319: Proceedings of the Twenty-Fourth Text REtrieval Conference (TREC 2015). 2015 Nov Presented at: TREC 2015; November 17–20, 2015; Gaithersburg, MD p. 1-12 URL: https://trec.nist.gov/pubs/trec24/papers/Overview-CL.pdf
8. Roberts K, Demner-Fushman D, Voorhees EM, Hersh WR. Overview of the TREC 2016 Clinical Decision Support Track. In: NIST Special Publication: SP 500-321: Proceedings of the Twenty-Fifth Text REtrieval Conference (TREC 2016). 2016 Nov Presented at: TREC 2016; November 15–18, 2016; Gaithersburg, MD p. 1-14 URL: https://trec.nist.gov/pubs/trec25/papers/Overview-CL.pdf
9. Roberts K, Demner-Fushman D, Voorhees EM, Hersh WR, Bedrick S, Lazar AJ, et al. Overview of the TREC 2017 Precision Medicine Track. In: NIST Special Publication: SP 500-324: Proceedings of the Twenty-Sixth Text REtrieval Conference (TREC 2017). 2017 Nov Presented at: TREC 2017; November 15–17, 2017; Gaithersburg, MD p. 1-13 URL: https://trec.nist.gov/pubs/trec26/papers/Overview-PM.pdf
10. CLEF eHealth Evaluation Lab. CLEF eHealth. URL: https://sites.google.com/site/clefehealth/home [accessed 2019-01-21] [WebCite Cache ID 75aIciOBv]
11. Morita M, Aramaki E, Kano Y, Miyabe M, Ohkuma T. Sociocom. About NTCIR-10 “Medical Natural Language Processing (MedNLP)” Pilot Task URL: http://sociocom.jp/~mednlp/medistj-en/index.html [accessed 2019-01-21] [WebCite Cache ID 75aJauZEE]
12. Aramaki E, Morita M, Kano Y, Ohkuma T. NTCIR11 MedNLP 2. URL: http://sociocom.jp/~mednlp/ntcir11/index-ja.html [accessed 2019-01-21] [WebCite Cache ID 75aJRWq5j]
13. Aramaki E, Morita M, Kano Y, Ohkuma T. MedNLPDoc. NTCIR MEDNLPDOC (MEDNLP-3) URL: https://sites.google.com/site/mednlpdoc/ [accessed 2019-01-21] [WebCite Cache ID 75aJ90OXu]
14. Morita M, Kano Y, Ohkuma T, Miyabe M, Aramaki E. Overview of the NTCIR-10 MedNLP Task. In: Proceedings of the 10th NTCIR Conference. 2013 Jun Presented at: NTCIR-10 Conference; June 18-21, 2013; Tokyo, Japan p. 696-701 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings10/pdf/NTCIR/MedNLP/01-NTCIR10-OV-MEDNLP-MoritaM.pdf
15. Aramaki E, Morita M, Kano Y, Ohkuma T. Overview of the NTCIR-11 MedNLP-2 Task. In: Proceedings of the 11th NTCIR Conference. 2014 Dec Presented at: NTCIR-11 Conference; December 9-12, 2014; Tokyo, Japan p. 147-154 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings11/pdf/NTCIR/OVERVIEW/01-NTCIR11-OV-MEDNLP-AramakiE.pdf
16. Aramaki E, Morita M, Kano Y, Ohkuma T. Overview of the NTCIR-12 MedNLPDoc Task. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies. 2016 Jun Presented at: NTCIR-12 Conference on Evaluation of Information Access Technologies; June 7-10, 2016; Tokyo, Japan p. 71-75 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings12/pdf/ntcir/OVERVIEW/01-NTCIR12-OV-MEDNLPDOC-AramakiE.pdf
17. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448 [FREE Full text] [doi: 10.1086/593098] [Medline: 18954267]
18. Culotta A. Towards Detecting Influenza Epidemics by Analyzing Twitter Messages. In: Proceedings of the First Workshop on Social Media Analytics.: ACM; 2010 Jul Presented at: SOMA'10; July 25-28, 2010; Washington DC, District of Columbia p. 115-122 URL: https://dl.acm.org/citation.cfm?id=1964874&dl=ACM&coll=DL
19. Aramaki E, Maskawa S, Morita M. Twitter Catches the Flu: Detecting Influenza Epidemics Using Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Jul Presented at: EMNLP'11; July 27–31, 2011; Edinburgh, Scotland, UK p. 1568-1576 URL: http://dl.acm.org/citation.cfm?id=2145432.2145600
20. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PLoS One 2011 May 4;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
21. Achrekar H, Gandhe A, Lazarus R, Yu S, Liu B. University of Massachusetts Lowell. 2012 Feb. Twitter improves Seasonal Influenza Prediction URL: http://www.cs.uml.edu/~hachreka/SNEFT/images/healthinf_2012.pdf [accessed 2019-01-22] [WebCite Cache ID 75cOHwa9E]
22. Lamb A, Paul M, Dredze M. The Association for Computational Linguistics. 2013. Separating Fact from Fear: Tracking Flu Infections on Twitter URL: http://www.aclweb.org/anthology/N13-1097 [accessed 2019-01-22] [WebCite Cache ID 75cOPDkio]

23. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-like illness surveillance on Twitter through automated learning of naïve language. PLoS One 2013;8(12):e82489 [FREE Full text] [doi: 10.1371/journal.pone.0082489] [Medline: 24324799]
24. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
25. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014 Oct 28;6(5):225-226 [FREE Full text] [doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117] [Medline: 25642377]
26. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EH, Olsen JM, et al. Using social media for actionable disease surveillance and outbreak management: a systematic literature review. PLoS One 2015 Oct;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
27. Iso H, Wakamiya S, Aramaki E. The Association for Computational Linguistics. 2016. Forecasting Word Model: Twitter-based Influenza Surveillance and Prediction URL: http://www.aclweb.org/anthology/C16-1008 [accessed 2019-01-22] [WebCite Cache ID 75cOYLTHS]
28. NTCIR. NTCIR-13 Conference URL: http://research.nii.ac.jp/ntcir/ntcir-13/conference.html [accessed 2019-01-21] [WebCite Cache ID 75aHh6YcI]
29. Aramaki E, Wakamiya S, Morita M, Kano Y, Ohkuma T. NTCIR-13 MedWeb. URL: http://mednlp.jp/medweb/NTCIR-13/ [accessed 2019-01-21] [WebCite Cache ID 75aHyBXJK]
30. Wakamiya S, Morita M, Kano Y, Ohkuma T, Aramaki E. Overview of the NTCIR-13: MedWeb Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 40-49 URL: https://tinyurl.com/y4w4sckt
31. Twitter. URL: https://twitter.com/ [accessed 2019-01-21] [WebCite Cache ID 75aIADN5x]
32. Facebook. URL: https://www.facebook.com/ [WebCite Cache ID 75aIfrC4W]
33. Sakai M, Tanioka H. Keyword-based Challenges at the NTCIR-13 MedWeb. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 50-51 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/02-NTCIR13-MEDWEB-SakaiM.pdf
34. Asakawa R, Akiba T. AKBL at the NTCIR-13 MedWeb Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 52-55 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/03-NTCIR13-MEDWEB-AsakawaR.pdf
35. Morita K, Takagi T. DrG at NTCIR-13: MedWeb Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 81-84 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/10-NTCIR13-MEDWEB-MoritaK.pdf
36. Sakishita M, Kano Y. Classification of Tweet Posters for Diseases by Combined Rule-Based and Machine Learning Method in NTCIR-13: MedWeb Twitter Task (Japanese Subtask). In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 62-64 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/05-NTCIR13-MEDWEB-SakishitaM.pdf
37. Iso H, Ruiz C, Murayama T, Taguchi K, Takeuchi R, Yamamoto H, et al. NTCIR13 MedWeb Task: Multi-label Classification of Tweets using an Ensemble of Neural Networks. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 56-61 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/04-NTCIR13-MEDWEB-IsoH.pdf
38. Ito M. NIL: Using scoring to analyse the ambiguous messages on the NTCIR-13 MedWeb task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 74 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/08-NTCIR13-MEDWEB-ItoM.pdf
39. Lin J, Dai H, Shao J. Principle Base Approach for Classifying Tweets with Flu-related Information in NTCIR-13 MedWeb Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 71-73 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/07-NTCIR13-MEDWEB-LinJ.pdf
40. Li C, Kang X, Ren F. Medweb Task: Identify Multi-Symptoms from Tweets Based on Active Learning and Semantic Information. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-8, 2017; Tokyo, Japan p. 75-80 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/09-NTCIR13-MEDWEB-LiC.pdf
41. Hang N, Kobayashi H, Quaresma P, Sawai Y. UE and Nikon at the NTCIR-13 MedWeb Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. 2017 Dec Presented at: NTCIR-13 Conference; December 5-9, 2017; Tokyo, Japan p. 65-70 URL: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/06-NTCIR13-MEDWEB-TranA.pdf

42. Twitter Developer Platform. Developer Agreement and Policy URL: https://developer.twitter.com/en/developer-terms/agreement-and-policy.html [accessed 2019-01-21] [WebCite Cache ID 75aHAkMpb]
43. Twitter Developer Platform. API Docs URL: https://developer.twitter.com/en/docs.html [accessed 2019-01-21] [WebCite Cache ID 75aHSWKV3]
44. NTCIR-13 MedWeb. 2018 Jul 23. NTCIR-13 MedWeb (Medical Natural Language Processing for Web Document) URL: http://research.nii.ac.jp/ntcir/permission/ntcir-13/perm-en-MedWeb.html [accessed 2018-08-23] [WebCite Cache ID 71sPdfSWe]
45. Aramaki E, Wakamiya S, Morita M, Kano Y, Ohkuma T. Figshare. 2018 Apr 02. NTCIR-13 MedWeb Annotation Corpus Guideline (English Ver 2.0) URL: https://figshare.com/articles/NTCIR-13_MedWeb_Annotation_Corpus_Guideline_English_Ver_2_0_/6072812 [accessed 2019-01-21] [WebCite Cache ID 75aGzBMsz]
46. Aramaki E, Wakamiya S, Morita M, Kano Y, Ohkuma T. Figshare. 2018 Apr 2. NTCIR-13 MedWeb Annotation Corpus Guideline (Japanese Ver 2.0) URL: https://figshare.com/articles/NTCIR-13_MedWeb_Annotation_Corpus_Guideline_Japanese_Ver_2_0_/6072821 [accessed 2019-01-21] [WebCite Cache ID 75aGfxygF]
47. Kudo T, Yamamoto K, Matsumoto Y. ACL Anthology Reference Corpus. 2014 Jul. Applying conditional random fields to Japanese morphological analysis URL: http://acl-arc.comp.nus.edu.sg/archives/acl-arc-090501d4/data/pdf/anthology-PDF/W/W04/W04-3230.pdf [accessed 2019-01-22] [WebCite Cache ID 75cOn2Nsj]
48. Bird S. NLTK: The Natural Language Toolkit. In: Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. 2006 Jul Presented at: COLING/ACL 2006; July 17–21, 2006; Sydney, Australia p. 69-72.
49. Natural Language Toolkit—NLTK 3.4 documentation. nltk.tokenize package URL: http://www.nltk.org/api/nltk.tokenize.html [accessed 2019-01-21] [WebCite Cache ID 75aGQkD2p]
50. Junyi S. GitHub. jieba URL: https://github.com/fxsjy/jieba [accessed 2014-01-27] [WebCite Cache ID 6MvXve9OT]
51. scikit-learn: Machine Learning in Python. URL: http://scikit-learn.org/stable/ [accessed 2018-11-04] [WebCite Cache ID 73eSO5vZc]
52. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Journal of Machine Learning Research. 2011 Oct 11. Scikit-learn: Machine Learning in Python URL: http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf [accessed 2019-01-22] [WebCite Cache ID 75cFEGrlb]
53. Zhang M, Zhou Z. A review on multi-label learning algorithms. IEEE Trans Knowl Data Eng 2014 Aug;26(8):1819-1837. [doi: 10.1109/TKDE.2013.39]
54. National Institutes of Health. Turning Discovery into Health URL: https://www.nih.gov/sites/default/files/about-nih/discovery-into-health/nih-turning-discovery-into-health.pdf [accessed 2018-03-30] [WebCite Cache ID 6yIXFi28d]
55. Wakamiya S, Kawai Y, Aramaki E. Twitter-based influenza detection after flu peak via tweets with indirect information: text mining study. JMIR Public Health Surveill 2018 Sep 25;4(3):e65 [FREE Full text] [doi: 10.2196/publichealth.8627] [Medline: 30274968]

Abbreviations

CLEF: Cross-Language Evaluation Forum for European Languages
CNN: convolutional neural network
i2b2: Informatics for Integrating Biology and the Bedside
ICD: International Codes for Diseases
MedNLP: Medical Natural Language Processing
MedWeb: Medical natural language processing for Web document
NLP: natural language processing
NTCIR: NII Testbeds and Community for Information access Research
SVM: support vector machine
TREC: Text Retrieval Conference

Edited by G Eysenbach; submitted 09.11.18; peer-reviewed by T Cruvinel, G Dini; comments to author 28.11.18; revised version received 12.12.18; accepted 13.12.18; published 20.02.19.

Please cite as:
Wakamiya S, Morita M, Kano Y, Ohkuma T, Aramaki E
Tweet Classification Toward Twitter-Based Disease Surveillance: New Data, Methods, and Evaluations
J Med Internet Res 2019;21(2):e12783
URL: http://www.jmir.org/2019/2/e12783/
doi:10.2196/12783
PMID:30785407

Original Paper

Health Care Professionals' Perceptions of Home Telemonitoring in Heart Failure Care: Cross-Sectional Survey

Ina Thon Aamodt[1,2], RN, MNSc; Edita Lycholip[3], RN, PhD; Jelena Celutkiene[3], MD, PhD; Anna Strömberg[4], RN, PhD; Dan Atar[5,6], MD, PhD; Ragnhild Sørum Falk[7], MSc, PhD; Thomas von Lueder[5], MD, PhD; Ragnhild Hellesø[2], RN, PhD; Tiny Jaarsma[8], RN, PhD; Irene Lie[1], RN, PhD

[1]Centre for Patient-Centered Heart and Lung Research, Department of Cardiothoracic Surgery, Oslo University Hospital Ullevål, Oslo, Norway

[2]Department of Nursing Science, Institute of Health and Society, University of Oslo, Oslo, Norway

[3]Clinic of Cardiac and Vascular Diseases, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University, Vilnius, Lithuania

[4]Division of Nursing, Department of Medical and Health Sciences, Linkoping University, Linkoping, Sweden

[5]Department of Cardiology B, Oslo University Hospital, Oslo, Norway

[6]Institute of Clinical Sciences, University of Oslo, Oslo, Norway

[7]Research Support Services, Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

[8]Division of Nursing, Department of Social and Welfare Studies, Linkoping University, Norrkoping, Sweden

Corresponding Author:
Ina Thon Aamodt, RN, MNSc
Centre for Patient-Centered Heart and Lung Research
Department of Cardiothoracic Surgery
Oslo University Hospital Ullevål
Building 63, Box 4956 Nydalen and Postboks 1130 Blindern, 0318
Oslo,
Norway
Phone: 47 48090883
Email: inamarieaamodt@gmail.com

Abstract

Background: Noninvasive telemonitoring (TM) can be used in heart failure (HF) patients to perform early detection of decompensation at home, prevent unnecessary health care utilization, and decrease health care costs. However, the evidence is not sufficient to be part of HF guidelines for follow-up care, and we have no knowledge of how TM is used in the Nordic Baltic region.

Objective: The aim of this study was to describe health care professionals' (HCPs) perception of and presumed experience with noninvasive TM in daily HF patient care, perspectives of the relevance of and reasons for applying noninvasive TM, and barriers to the use of noninvasive TM.

Methods: A cross-sectional survey was performed between September and December 2016 in Norway and Lithuania with physicians and nurses treating HF patients at either a hospital ward or an outpatient clinic. A total of 784 questionnaires were sent nationwide by postal mail to 107 hospitals. The questionnaire consisted of 43 items with close- and open-ended questions. In Norway, the response rate was 68.7% (226/329), with 57 of 60 hospitals participating, whereas the response rate was 68.1% (310/455) in Lithuania, with 41 of 47 hospitals participating. Responses to the closed questions were analyzed using descriptive statistics, and the open-ended questions were analyzed using summative content analysis.

Results: This study showed that noninvasive TM is not part of the current daily clinical practice in Norway or Lithuania. A minority of HCPs responded to be familiar with noninvasive TM in HF care in Norway (48/226, 21.2%) and Lithuania (64/310, 20.6%). Approximately half of the HCPs in both countries perceived noninvasive TM to be relevant in follow-up of HF patients in Norway (131/226, 58.0%) and Lithuania (172/310, 55.5%). For physicians in both countries and nurses in Norway, the 3 most mentioned reasons for introducing noninvasive TM were to improve self-care, to reduce hospitalizations, and to provide high-quality care, whereas the Lithuanian nurses described ability to treat more patients and to reduce their workload as reasons for introducing noninvasive TM. The main barriers to implement noninvasive TM were lack of funding from health care authorities or the Territorial Patient Fund. Moreover, HCPs perceive that HF patients themselves could represent barriers because of their physical or mental condition in addition to a lack of internet access.

Conclusions: HCPs in Norway and Lithuania are currently nonusers of TM in daily HF care. However, they perceive a future with TM to improve the quality of care for HF patients. Financial barriers and HF patients' condition may have an impact on the use of TM, whereas sufficient funding from health care authorities and improved knowledge may encourage the more widespread use of TM in the Nordic Baltic region and beyond.

(J Med Internet Res 2019;21(2):e10362) doi:10.2196/10362

KEYWORDS

nurses; physicians; perception; telemedicine; heart failure; self-care

Introduction

Background

Worldwide health care undergoes great changes where caring for chronically ill patients such as heart failure (HF) is expected to take place to a great extent in their homes [1,2]. To be able to perform safe and high-quality care, the use of telemonitoring (TM) is suggested to be widely implemented [3]. The Institute of Medicine defines telemonitoring as "monitoring patient status at a distance by the use of audio, video, and other telecommunications and electronic information processing technologies" [4]. Noninvasive TM detects decompensation at home, prevents unnecessary health care utilization, and decreases health care costs [5,6]. This is important because HF affects 26 million people worldwide, with a rapidly escalating prevalence in Europe and the United States because of an aging population as well as the improved treatment and survival of patients with cardiac disease [7-9]. HF is a complex, progressive clinical syndrome characterized by high mortality, high morbidity with high readmission rates (25% in Europe [7] and 27% in the United States [10]), and affecting quality of life [11]. At present, TM in HF care has tested numerous devices and systems [12]; however, implementation of noninvasive TM in daily HF clinical practice is scarce. Moreover, the evidence for noninvasive TM is not found sufficient to be part of follow-up care recommended by the European Society of Cardiology (ESC) guidelines or the American Heart Association guidelines [7,8]. Results from studies are not consistent regarding outcomes of TM. Some studies have shown that TM may improve survival, reduce HF-related hospitalizations, and improve quality of life compared with usual care or low access to care [2,13,14]; other studies have shown no improvements [15-17]. Furthermore, noninvasive TM studies have methodological weaknesses such as insufficiencies regarding large sample size, homogenous protocols, robust designs, a clear definition of noninvasive TM, and show a diversity of outcomes [5,18,19]. A program where patients monitor their condition at home and transmit information to external centers is not widely established. According to the third global survey on eHealth performed by The World Health Organization, only 22% of the responding countries use TM [3]. Furthermore, in countries with advanced information technology, it is identified that health care professionals (HCPs) are not familiar with using noninvasive TM in their daily follow-up care of HF patients [20]. According to HCPs' factors of success in using TM, the possibilities include improved diagnostics, improved communication with the patient, and improved support of patient centered care, whereas barriers include workflow and staff turnover [21]. Furthermore, HCPs raise questions on how to assess available technology, the value of using technology, and the evidence of effectiveness and knowledge [22]. TM has been reported to increase the workload for HCPs and the use of health care resources, which may be the reason it has not been widely implemented in clinical practice [23,24].

Objectives

Although the governments of Norway and Lithuania have approved the use of remote medical services in a new electronic health (eHealth) strategy, little is known on the actual use of TM in daily HF clinical care in the Nordic Baltic region [25,26]. Therefore, this study aimed to describe HCPs' (1) perceptions of and presumed experience with noninvasive TM in HF care, (2) perspectives of the relevance of and reasons for applying noninvasive TM, and (3) barriers to the use of noninvasive TM.

Methods

Study Design and Definition

We conducted a cross-sectional nationwide survey of noninvasive HF TM in Norway and Lithuania.

TM in the survey referred to noninvasive TM, which is implemented via internet-based personal devices monitoring body weight, blood pressure, heart rate, dyspnea, and other signs and symptoms that would reflect the actual volume status of HF patients. Patients use the devices in their home environment, and the generated data are transferred to health care providers over the internet. The information obtained is presumed to reflect the actual condition of HF patients (contrasted with patients' self-reports) and includes feedback to patients about their condition. Examples of TM implemented via only telephone, telephone support, telephone follow-up, or implantable devices or pacemakers were not considered in the survey. This definition was presented to the participants on the front page of the questionnaire (Multimedia Appendix 1).

Study Setting and Participants

Norway and Lithuania have 5 and 3 million inhabitants, respectively, and both countries are located in the northern part of Europe. Norway is a high-income country in which the health care system is funded by public sources [27]. Lithuania has undergone political and economic changes with a mixed health system predominantly funded by the National Health Insurance Fund (61% of funding in 2010) and supplemented by a substantial state contribution [28].

The inclusion criteria were nurses and physicians (ie, HCPs) currently working with HF patients in a hospital ward or in an outpatient clinic in Norway or in Lithuania. HCPs were recruited

from a list of all potential public and private hospitals caring for HF patients in Norway (N=60) and Lithuania (N=47). The former list was extracted from the Norwegian Heart Failure Registry and the latter from the National Insurance Funds list of hospitals in Lithuania. In addition to telephone follow-up, we contacted the head of each hospital ward and outpatient clinic by post for approval.

Data Collection and Questionnaire

Data were collected from September to December 2016 using a questionnaire mailed by post. It was sent to 784 physicians and nurses at 107 public and private hospitals providing HF care in Norway or Lithuania. A study researcher in each country (ITA and EL) made 1 phone call to remind the contacted individuals at each site (ward or outpatient clinic) about completing and returning the questionnaire.

Questionnaire

A 43-item questionnaire (Multimedia Appendix 1) for assessing HCPs' perceptions on the use of TM and potential use of TM was developed. It was based on the initial versions of a survey used in the Netherlands [29] and with HCPs in Japan and Sweden [20]. The questionnaire used for this study in Norway and Lithuania included open- and closed-ended response options, with additional questions on HCPs' characteristics, their perceptions of TM, and potential experiences with TM. Language and cultural adjustments were made from the preparation stage to the final report according to the Principles for Good Practice for the Translation and Cultural Adaption Process for Patient-Reported Outcome Measures [30]. Face validity (measuring target construction) and content validity (relevance, comprehensiveness, and balance) assessments were conducted [31] by 5 cardiologists and 10 nurses with expertise in daily clinical HF care in Norway or Lithuania. These professionals deemed that the questionnaire measured the intended HCPs' perceptions of TM and potential experiences with TM in HF care.

The questionnaire contained 3 main parts: (1) general questions about the participants and their experiences with information and communications technology (ICT) in general and TM in particular, (2) questions for users of TM, and (3) questions for nonusers of TM. In part 1 of the questionnaire, data on HCPs' characteristics were collected, with additional questions on education and competency in ICT. The participants with presumed experience and familiarity with TM were asked to respond yes or no to these questions. In parts 2 and 3 of the questionnaire, potential TM user and nonuser participants answered similarly detailed questions about TM. HCPs were asked what they considered to be good ways for performing follow-up of stable HF patients (eg, outpatient clinic, using noninvasive TM, and home visits by a nurse), and a follow-up by a general practitioner (GP) was added to the statement list.

In line with the recommendations by previous users of the survey regarding the low response rate by users with experience with TM [20], we added items to describe HCPs' perceptions of noninvasive TM: future purpose, criteria, relevance, and feasibility of TM in daily HF care in the participants' country presented as categories to be marked by the participant. HCPs were asked to rate how important 10 statements were for introducing TM in the care of HF patients (eg, offering higher-quality care and reducing costs) by importance level on a 10-point scale, ranging from *0* for not important to *10* for very important. HCPs responded to how long they considered appropriate for using TM by responding to statements of duration. Finally, HCPs' perceptions of funding responsibility and situations they thought inappropriate for TM were reported by additional open-ended questions to get an understanding of barriers to the implementation of noninvasive TM.

Data Analysis

Statistical Analysis

All data were analyzed using the Statistical Package for Social Sciences, version 24 (IBM Corp Released 2016 IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp). Descriptive analysis was presented as means with SDs or SE of the mean and median with interquartile range for continuous variables. Categorical variables are presented with numbers and percentages.

Summative Analysis of the Open-Ended Questions

Answers to the open-ended questions were translated from Norwegian and Lithuanian into English by 2 independent researchers (ITA and EL). A total of 44 double-spaced pages of transcripts were produced. First, responses were formulated like statements. Transcripts from both countries were thoroughly read to gain an understanding of the words or statements. In total, 4 authors (ITA, EL, JC, and IL) independently reread the responses to the open-ended questions before reaching consensus on categories and subcategories. Following Hsieh and Shannon, a summative content analysis was performed with numbers and percentages for the subcategories [32].

Ethical Consideration

All participants signed a written informed consent form before participation. The data protection officer at Oslo University Hospital, Oslo, Norway and University Hospital Santariskiu Klinikos, Vilnius, Lithuania granted consent to perform the study. The study was conducted in compliance with the principles of the Declaration of Helsinki.

Results

Participants

The overall response rate of the nurses and physicians was 68.4% (536/784) from 98 of the 107 contacted hospitals in Norway (57/60) and Lithuania (41/47). Characteristics and ICT competency of the participants are summarized in Table 1.

Table 1. Characteristics and information and communications technology competency of physicians and nurses in Norway (N=226) and Lithuania (N=310).

Characteristics	Norway		Lithuania	
	Physicians (n=63)	Nurses (n=163)	Physicians (n=137)	Nurses (n=173)
Gender, n (%)				
Female	16 (25)	151 (93.2)	107 (78.1)	171 (98.8)
Male	47 (75)	11 (6.8)	30 (21.9)	2 (1.2)
Age (range: 23-76 years), mean (SD)	48 (11)	45 (11)	51 (12)	46 (9)
Education degree				
PhD, n (%)	19 (30)	—[a]	37 (27.0)	—
Master, n (%)	12 (19)	13 (8.0)	100 (73.0)	14 (8.1)
Bachelor, n (%)	—	150 (92.0)	—	64 (37.0)
Other type of degree, n (%)	32 (51)	—	—	95 (54.9)
Post graduate experience (years), median (IQR[b])	19 (12-28)	16 (9-25)	28 (15-35)	26 (20-32)
Work time, n (%)				
Full-time	32 (51)	59 (36.2)	44 (32.1)	115 (66.5)
Part-time days/week	19 (30)	64 (39.3)	38 (27.7)	16 (9.2)
Part-time hours/week	10 (16)	30 (18.4)	54 (39.4)	35 (20.2)
Unreported or missing	2 (3)	10 (6.1)	1 (0.7)	7 (4.0)
Hospital level, n (%)				
University	11 (18)	25 (15.3)	47 (34)	88 (51.0)
Second or third	48 (76)	128 (78.5)	82 (60)	80 (46.2)
Private	4 (8)	7 (4.3)	8 (6)	3 (1.7)
Unreported or missing	0 (0)	3 (1.8)	0 (0)	2 (1.0)
ICT[c] competency				
Computer experience (years), median (IQR)	25 (20-30)	20 (16-25)	15 (10-20)	12 (10-17)
Operating system, n (%)	60 (95)	158 (96.9)	126 (92.0)	149 (86.1)
Programs, n (%)	61 (97)	153 (93.9)	109 (79.6)	124 (71.7)
Programming language, n (%)	11 (18)	43 (26)	18 (13)	22 (13)
Email, n (%)	63 (100)	162 (99.4)	137 (100)	163 (94.2)
Email mobile phone, n (%)	56 (89)	154 (94.5)	108 (78.8)	118 (68.2)
Internet, n (%)	63 (100)	162 (99.4)	137 (100)	167 (96.5)

[a]Participants did not have the degree in question.

[b]IQR: interquartile range.

[c]ICT: information and communications technology.

Among the responders, 28% and 44% worked as physicians in Norway and Lithuania, respectively. The majority of HCPs in Lithuania were female (90%), whereas in Norway, the majority of the physicians were males (75%). In Lithuania, about half of the nurses worked at a university hospital (88/173), whereas physicians more often worked in a second- or third-level hospital (82/137). In Norway, most HCPs worked at a second- or third-level hospital.

A total of 51% (32/63) of Norwegian physicians worked full-time with HF patients and one-third (44/137) of the Lithuanian physicians worked full-time in HF care. Among nurses, 36.2% (59/163) of Norwegian nurses worked full-time, whereas 66.5% (115/173) in Lithuania did. All participants had substantial experience with ICT, with a variance in use of programs such as Word, PowerPoint, or Excel and use of email on a mobile phone.

Experiences and Familiarity With Noninvasive Telemonitoring in Heart Failure Care

TM is not a part of routine clinical practice in HF care in Norway or Lithuania. None of the responding HCPs were using TM. Nevertheless, a minority of HCPs in both countries

confirmed to be familiar with TM in Norway (48/226, 21.2%) and in Lithuania (64/310, 20.6%).

Relevance for Follow-Up of Heart Failure Patients Today and in Future Care

As shown in Figure 1, HCPs responded that good ways of performing follow-up of stable HF patients were by a outpatient clinic, or GP as the most optimal way to follow up HF patients. HCPs in both countries supported the potential for internet-based TM in Norway (131/226, 58.0%) and in Lithuania (152/ 310, 49.0%). A nurse-led HF outpatient clinic was specifically commented by Norwegian HCPs as an item in the column *other*, not presented in Figure 1.

Although the HCPs in Norway and in Lithuania were nonusers of TM, they considered use of TM in future HF care to be relevant. They suggested TM is useful to monitor HF patients' physical condition signaling deterioration (in Norway: 187/226, 82.7% and in Lithuania: 225/310, 72.6%); to monitor the effect of the treatment and adjusting it remotely (140/226, 61.9% and 226/310, 72.9%); for patient education (104/226, 46.0% and 195/310, 62.9%) and remote drug titration (76/226, 33.6% and 190/310, 61.3%).

More than half of HCPs reported TM to be relevant in Norway (131/226, 58.0%) and in Lithuania (172/310, 55.5%) as shown in Table 2. The main responses to the open-ended item *other* were "I do not know," "I lack knowledge," and "there is a lack of evidence supporting the use of TM."

HCPs reported that daily feedback to HF patients using TM was more feasible (115/226, 50.9%) in Norway than in Lithuania (75/310, 24.2%), whereas the "I do not know" category was chosen more frequently in Lithuania (201/310, 64.8%) than in Norway (77/226, 34.1%).

Reasons to Consider Introducing Noninvasive Telemonitoring

The 3 most frequently mentioned reasons given by physicians and nurses in Norway and physicians in Lithuania for introducing TM to HF patients were to (1) improve patient self-care, (2) reduce hospitalizations, and (3) offer higher-quality care. The Lithuanian nurses' 3 most frequently mentioned reasons for introducing TM to HF patients were (1) to offer higher-quality care, (2) the ability to treat more patients, and (3) to reduce the workload of the outpatient clinic as shown in Table 3.

The statement "to introduce TM for health care authorities" had the lowest score for physicians and nurses in both countries.

Figure 1. Health care practitioners' (HCPs') opinion of good ways of performing follow-up of stable heart failure patients. HCPs in Norway (N=226) and Lithuania (N=310). More than 1 answer was possible.

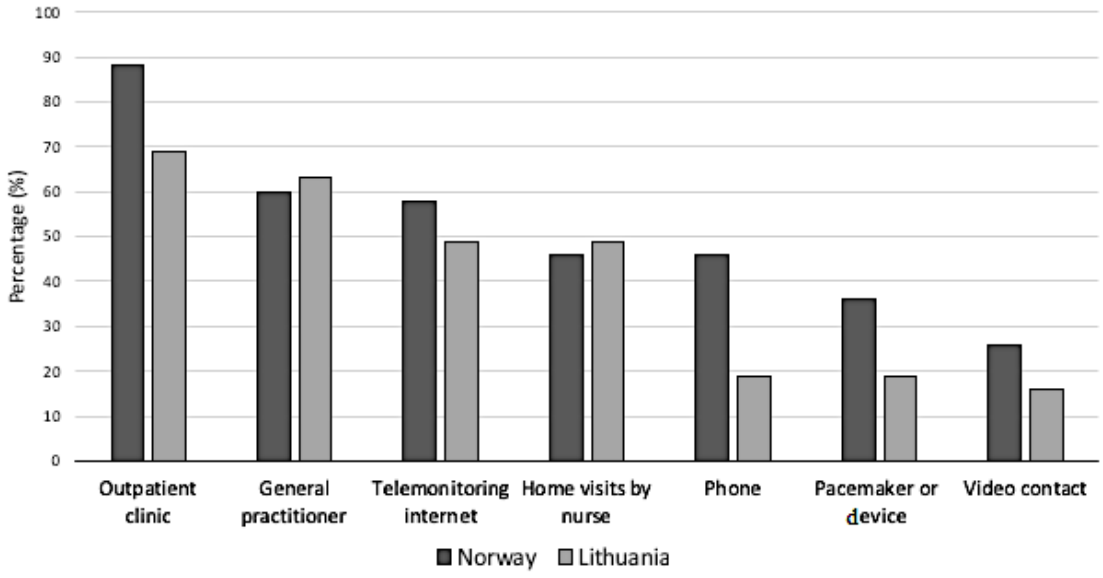


Table 2. Relevance of telemonitoring in Norway (N=226) and in Lithuania (N=310).

Relevance	Norway, n (%)	Lithuania, n (%)
Very relevant	44 (19.5)	42 (13.5)
Relevant	131 (58.0)	172 (55.5)
Not relevant	17 (7.5)	51 (16.5)
Other	24 (10.6)	33 (10.6)
Unreported or missing	10 (4.4)	12 (3.9)

Table 3. Physicians' and nurses' perception of important reasons for introducing telemonitoring into clinical practice in Norway (N=226) and Lithuania (N=310).

Reasons for introducing telemonitoring	Norway, mean score (SEM[a])	Lithuania, mean score (SEM)
Physicians	**n=63**	**n=137**
Reduce admissions or readmissions	6.85 (0.37)	8.08 (0.21)
Improve self-care of HF[b] patients	7.00 (0.34)	7.92 (0.20)
Offering higher-quality care	6.50 (0.42)	8.25 (0.20)
Ability to treat more patients	6.02 (0.40)	6.87 (0.25)
Improve adherence to HF guidelines	6.09 (0.39)	7.75 (0.20)
Reducing the workload on the HF outpatient clinic	5.64 (0.39)	7.80 (0.22)
Reducing costs	5.00 (0.38)	6.45 (0.29)
Our center is innovative	4.65 (0.38)	6.41 (0.29)
Implementing the vision of the hospital	3.98 (0.41)	6.23 (0.29)
Important for health authorities	3.81 (0.45)	3.58 (0.33)
Nurses	**n=163**	**n=173**
Reduce admissions or readmissions	8.11 (0.18)	7.46 (0.19)
Improve self-care of HF patients	8.03 (0.18)	7.36 (0.21)
Offering higher-quality care	7.84 (0.21)	7.69 (0.21)
Ability to treat more patients	7.22 (0.21)	7.62 (0.21)
Improve adherence to HF guidelines	6.92 (0.21)	6.91 (0.20)
Reducing the workload on the HF outpatient clinic	6.34 (0.23)	7.57 (0.19)
Reducing costs	6.52 (0.22)	6.41 (0.27)
Our center is innovative	6.10 (0.26)	6.42 (0.27)
Implementing the vision of the hospital	5.31 (0.25)	6.43 (0.25)
Important for health authorities	3.81 (0.25)	4.99 (0.30)

[a]SEM: standard error of the mean.

[b]HF: heart failure.

The criterion most often identified by all HCPs for treatment with TM was admission or readmissions in Norway (136/226, 60.2%) and Lithuania (217/310, 70.0%). HCPs in Lithuania more often mentioned New York Heart Association Functional Classification (191/310, 61.6%) than Norwegian participants (112/226, 49.6%) as a criterion for treatment with TM. The criterion patient education was supported by HCPs in Norway (116/226, 51.3%) and Lithuania (145/310, 46.8%), and the criterion adherence to medication was supported by HCPs in Norway (98/226, 43.4%) and Lithuania (169/310, 54.6%). Respondents were less interested in support and advice as a criterion for treatment with TM in Norway (93/226, 41.2%) and in Lithuania (68/310, 21.9%).

HCPs in both countries supported that TM should be used as long as necessary (Norway: 159/226, 70.4% and Lithuania: 183/310, 59.0%) or for unlimited time (Norway: 9/226, 3.9% and Lithuania: 79/310, 25.5%), with a few nonresponding in Norway (18/226, 8.0%).

Barriers to Implementing Noninvasive Telemonitoring

HCPs reported from a list of barriers to implementing the TM in the same order: lack of financing (Norway and Lithuania, 156/226, 69.0% and 277/310, 89.4%, respectively); lack of equipment (124/226, 54.9% and 252/310, 81.3%, respectively); lack of knowledge (87/226, 38.5% and 227/310, 73.2%, respectively); and lack of guidelines from health care authorities (39/226, 17.3% and 188/310, 60.6%, respectively). The HCPs reported other barriers as shortage of staff, security issues, and need of more documentation. The HCPs' own views of what they perceive as barriers to implement TM in both countries are shown in Table 4.

Table 4. Perception of health care professionals on barriers to implement telemonitoring in Norway (N=226) and Lithuania (N=310).

Barriers to implement telemonitoring	Norway, n (%)	Lithuania, n (%)
Financing[a]		
Health care authorities		
Regional health authorities, Ministry of health care services	123 (62.1)	87 (38.7)
Specialist health care services, Territorial Patient Fund	25 (12.6)	85 (37.8)
I do not know	29 (14.6)	23 (10.2)
Unreported or missing	28 (14.1)	85 (37.8)
Patients limitations[b]		
Mental and physical limitations		
Mental limitations	91 (48.4)	22 (13.9)
Acute or physical limitation	57 (30.3)	52 (32.9)
Age limitations	48 (25.5)	6 (3.8)
Knowledge limitations		
Technical skills	44 (23.4)	14 (8.9)
Compliance issues	14 (7.4)	11 (7.0)
I do not know	15 (8.0)	46 (29.1)
Unreported or missing	28 (12.4)	152 (49.0)

[a]In Norway, a total of 198 out of 226 (87.6%) participants responded to the open-ended question and a total of 225 out of 310 (72.6%) HCPs in Lithuania. More than 1 answer was possible.

[b]In Norway, a total of 188 out of 226 (83.2%) participants responded to the open-ended question and a total of 158 out of 310 (51.0%) HCPs in Lithuania. More than 1 answer was possible.

From the open-ended questions regarding funding, we found that health care authorities should be the main contributors of funding TM. Less than 5% of HCPs from both countries suggested other funding options such as from patients, the company involved, or the private sector. HCPs in Lithuania considered the European Union (EU) as a source of funding. Less than 5% of HCPs in Norway questioned whether there should be funding for TM. The barriers related to HF patients were their physical or mental condition, age, insufficient knowledge of technology, and adherence issues (eg, an acute HF condition, dementia, and cognitive or physical alterations caused by medication). Less than 5% of HCPs in both countries reported that limited access to health care services when using TM and patients with foreign language limitations were challenges to TM implementation. Access to the internet was a specific challenge mentioned by the HCPs in Lithuania.

Discussion

Principal Findings

Physicians and nurses working in Norwegian and Lithuanian hospitals are currently nonusers of noninvasive TM. However, HCPs in our study perceive noninvasive TM as a possibility in the future to improve the quality of care for HF patients at home. The findings of this study are in line with previously reported low use of TM in clinical practice [3,20]. In Norway, telemedicine was implemented in the 1990s; however, during a 5-year follow-up (2009-2013), the level of use was low compared with outpatient visits [33]. Moreover, Norway and Lithuania participated in a European survey that concludes that TM is not widely applied and has a potential for improving support to patients with chronic conditions in their home [34]. This is an example of health care providers lagging behind in implementing eHealth solutions as suggested by the eHealth Action Plan (2012-2020) from the European commission [35]. The eHealth position statement by the ESC [22] raises the same issues as our findings, namely, the problems with low awareness and use. HCPs' familiarity with noninvasive TM was low in both countries (21.2% in Norway and 20.6% in Lithuania). At HF conferences, research on the use of noninvasive TM in HF care is presented, which may be a reason why a minority of HCPs in both countries were familiar with the term and what it stood for.

Although noninvasive TM is currently not in use in Norway or Lithuania, HCPs perceive TM to be relevant in future HF care, in line with the health care authorities in both countries [25,26] as well as HCPs' expectations in Sweden and Japan [20]. All HCPs expected TM to improve the quality of care for HF patients, whereas improvements to patients' self-care were mainly mentioned by physicians in both countries and nurses in Norway. Self-care is a cornerstone in maintaining and managing life with a chronic disease such as HF [36], and HCPs have a responsibility in educating HF patients about self-care [7,37]. Nurses in Lithuania most often mentioned treating more HF patients and reducing their workload as reasons for implementing noninvasive TM. The differences in response by nurses can be explained by the ratio of nurses in the 2 countries, with 7.7 nurses per 1000 inhabitants in Lithuania and 17.7 nurses

per 1000 inhabitants in Norway [38]. Moreover, toward 2030, there is an increased need for human resources in health care globally [39] and this challenges the current preference of face-to-face follow-up of stable HF patients reported by our nurses and physicians. Furthermore, the Norwegian nurses work part-time at HF outpatient clinics, whereas physicians work full-time, for example, at hospital wards and HF outpatient clinics, which could explain why a nurse-led heart failure outpatient clinic was specifically mentioned by physicians and nurses in Norway. However, the least frequently mentioned reason for all HCPs was to implement TM because it is important for health care authorities. Therefore, it is important to consider the perceptions of nonusers of HCPs regarding the relevance and reasons for implementing noninvasive TM in future HF care in different countries.

A barrier to implementing noninvasive TM in HF clinical practice was funding and for health care authorities, as suggested by our HCPs, to be the main contributor as confirmed in previous studies [2,20,21,40]. In Norway, this is represented by national and specialized or municipal health care authorities and in Lithuania, by the Ministry of Health Care Services and the Territorial Patient Fund. A potential source to the differences in HCPs' responses is the financial situation, with Norway a high-income country and Lithuania a middle-income country [27,28]. The EU was mentioned as a potential source for funding TM by HCPs from Lithuania, as Lithuania is a member state and has received funding for health care services from the EU [41]. HCPs in Norway questioned if TM should be funded with the lack of evidence for noninvasive TM and not being a part of current HF guidelines [7]. Moreover, the cost of implementing TM in the management of HF is not clearly reported, varying from low-to-high cost depending on how costs are measured, for example, equipment, follow-up, or hospital admission [13,23,42]. Our participants' concern regarding shortage of staff is relevant as the workload for HCPs can increase [29,40] because nurses or physicians at hospitals, HF clinics, or telemedical centers mainly interpret the transmitted measurements [12]. Furthermore, not all the work done by nurses and physicians is visible, but the *invisible* work performed by HCPs impacts patients' ability to manage the use of TM in their home [43]. These differences call for additional funding in high- and middle-income countries to facilitate TM implementation.

The HF patient's condition, age, and insufficient knowledge of technology were barriers perceived by the HCPs to implementing noninvasive TM. Moreover, TM was mentioned as useful for HF patients in monitoring their HF signaling a worsening condition and reducing hospitalization. Our participants most often mentioned the criterion of readmission or admission to hospital when providing examples of participants eligible for TM. More recent noninvasive TM research with HF patients shows potential to reduce the risk of hospitalization [44] and identifies eligible HF patients [45]. Elderly HF patients without ICT competency can use new technology, and they describe better contact with HCPs [46]; however, this is not part of our findings. The lack of familiarity and high ICT competency among HCPs in our survey may be a reason why they perceived that elderly HF patients are not eligible for noninvasive TM. In Lithuania, HCPs presented a lack of internet access as a specific barrier, which may be a reason for their limitations in noninvasive TM, as the use of internet is essential in our TM definition [20,29]. To involve HCPs in an earlier phase of developing new technology is a way of making their contributions visible and acknowledged and may contribute to our participants finding noninvasive TM relevant for HF patients in Norway and Lithuania.

Strengths and Limitations

The fairly high response rate is a strength of the study. HCPs from both hospital wards, which discharge patients to their homes, and outpatient clinics, which see patients who live at home, were asked to participate. There are several limitations to our study. First, the self-reported questionnaire does not provide in-depth knowledge about HCPs' knowledge of TM. Second, it was not possible to sample HCPs' experiences with TM in HF clinical practice as none of the hospitals were using TM at the time of the survey. This was not anticipated. This shortfall implies that some questionnaire items need to be revised. Furthermore, a comprehensive assessment of the questionnaire's face and content validity may have strengthened the study.

Conclusions

HCPs in Norway and in Lithuania are currently nonusers of TM in daily follow-up of HF care; however, they perceive that a future use of TM is relevant to improve the quality of care for HF patients. From the perspective of physicians in both countries and Norwegian nurses, the main reason for introducing noninvasive TM in HF care was to improve patient's self-care. The nurses in Lithuania expected to treat more HF patients and reduce their workload by implementing TM. Financial barriers and HF patients' condition may have an impact on the use of TM, whereas sufficient funding from health care authorities and improved knowledge may encourage the more widespread use of TM in the Nordic Baltic region and beyond.

Acknowledgments

The authors wish to thank all participants in the study in Norway and in Lithuania. This study was supported by NordForsk Nordic Program on Health and Welfare (project #76015).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire used in this survey.

[PDF File (Adobe PDF File), 370KB - jmir_v21i2e10362_app1.pdf]

References

1. World Health Organization. Noncommunicable Diseases Progress Monitor 2017. Switzerland: WHO; 2017. URL: https://www.who.int/nmh/publications/ncd-progress-monitor-2017/en/
2. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015(10):CD007228. [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
3. World Health Organization. Global Diffusion Of eHealth - Making Universal Health Coverage Achievable: Report Of The Third Global Survey On eHealth (global Observatory For eHealth Series). Geneva: WHO; 2016. URL: https://www.who.int/goe/publications/global_diffusion/en/
4. Institute of Medicine (US) Committee on Evaluating Clinical Applications of Telemedicine. In: Field MJ, editor. Telemedicine: A Guide to Assessing Telecommunications in Health Care. Washington DC: National Academies Press; 1996. URL: https://www.ncbi.nlm.nih.gov/pubmed/20845554
5. von Lueder TG, Krum H. Current modalities for invasive and non-invasive monitoring of volume status in heart failure. Heart 2012 Jul;98(13):967-973. [doi: 10.1136/heartjnl-2011-301330] [Medline: 22572046]
6. Lainscak M, Blue L, Clark AL, Dahlström U, Dickstein K, Ekman I, et al. Self-care management of heart failure: practical recommendations from the Patient Care Committee of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2011 Feb;13(2):115-126. [doi: 10.1093/eurjhf/hfq219] [Medline: 21148593]
7. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, Authors/Task Force Members, Document Reviewers. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016 Dec 14;37(27):2129-2200. [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]
8. Yancy CW, Januzzi Jr JL, Allen LA, Butler J, Davis LL, Fonarow GC, et al. 2017 ACC Expert Consensus Decision Pathway for optimization of heart failure treatment: answers to 10 pivotal issues about heart failure with reduced ejection fraction: a report of the American College of Cardiology Task Force on Expert Consensus Decision Pathways. J Am Coll Cardiol 2018 Jan 16;71(2):201-230. [doi: 10.1016/j.jacc.2017.11.025] [Medline: 29277252]
9. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nat Rev Cardiol 2016 Dec;13(6):368-378 [FREE Full text] [doi: 10.1038/nrcardio.2016.25] [Medline: 26935038]
10. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. Nat Rev Cardiol 2011 Jan;8(1):30-41 [FREE Full text] [doi: 10.1038/nrcardio.2010.165] [Medline: 21060326]
11. Heo S, Moser DK, Lennie TA, Fischer M, Kim J, Lee M, et al. Changes in heart failure symptoms are associated with changes in health-related quality of life over 12 months in patients with heart failure. J Cardiovasc Nurs 2018;33(5):460-466. [doi: 10.1097/JCN.0000000000000493] [Medline: 29697502]
12. Brons M, Koudstaal S, Asselbergs FW. Algorithms used in telemonitoring programmes for patients with chronic heart failure: a systematic review. Eur J Cardiovasc Nurs 2018 Oct;17(7):580-588 [FREE Full text] [doi: 10.1177/1474515118786838] [Medline: 29954184]
13. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015 Mar 12;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
14. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, et al. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the Better Effectiveness After Transition -- Heart Failure (BEAT-HF) randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
15. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
16. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, TIM-HF Investigators. Telemedical Interventional Monitoring in Heart Failure (TIM-HF), a randomized, controlled intervention trial investigating the impact of telemedicine on mortality in ambulatory patients with heart failure: study design. Eur J Heart Fail 2010 Dec;12(12):1354-1362. [doi: 10.1093/eurjhf/hfq199] [Medline: 21098580]
17. Kotooka N, Kitakaze M, Nagashima K, Asaka M, Kinugasa Y, Nochioka K, HOMES-HF study investigators. The first multicenter, randomized, controlled trial of home telemonitoring for Japanese patients with heart failure: home telemonitoring study for patients with heart failure (HOMES-HF). Heart Vessels 2018 Aug;33(8):866-876. [doi: 10.1007/s00380-018-1133-5] [Medline: 29450689]

18. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
19. Piotrowicz E. The management of patients with chronic heart failure: the growing role of e-Health. Expert Rev Med Devices 2017 Apr;14(4):271-277. [doi: 10.1080/17434440.2017.1314181] [Medline: 28359169]
20. Kato NP, Johansson P, Okada I, de Vries AE, Kinugawa K, Strömberg A, et al. Heart failure telemonitoring in Japan and Sweden: a cross-sectional survey. J Med Internet Res 2015 Nov 13;17(11):e258 [FREE Full text] [doi: 10.2196/jmir.4825] [Medline: 26567061]
21. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
22. Cowie MR, Bax J, Bruining N, Cleland JG, Koehler F, Malik M, et al. e-Health: a position statement of the European Society of Cardiology. Eur Heart J 2016 Jan 1;37(1):63-66. [doi: 10.1093/eurheartj/ehv416] [Medline: 26303835]
23. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015(9):CD002098. [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
24. Varsi C, Ekstedt M, Gammon D, Børøsund E, Ruland CM. Middle managers' experiences and role in implementing an interactive tailored patient assessment eHealth intervention in clinical practice. Comput Inform Nurs 2015 Jun;33(6):249-257. [doi: 10.1097/CIN.0000000000000158] [Medline: 25988851]
25. The Norwegian Directorate of eHealth. National eHealth Strategy 2017-2022. 2017. URL: https://ehelse.no/Documents/Nasjonal%20e-helsestrategi%20og%20handlingsplan/Nasjonal%20e-helsestrategi%202017-2022%20%28PDF%29.pdf [accessed 2018-10-29] [WebCite Cache ID 73WvxmVMD]
26. Digital Health. 2007. Lithuania approves new eHealth strategy URL: https://www.digitalhealth.net/2007/07/lithuania-approves-new-ehealth-strategy/ [accessed 2018-10-29] [WebCite Cache ID 73WvP6YHc]
27. Ringard Å, Sagan A, Saunes IS, Lindahl AK. Norway: health system review. Health Syst Transit 2013;15(8):1-162 [FREE Full text] [Medline: 24434287]
28. Murauskiene L, Janoniene R, Veniute M, van Ginneken E, Karanikolos M. Lithuania: health system review. Health Syst Transit 2013;15(2):1-150 [FREE Full text] [Medline: 23902994]
29. de Vries AE, van der Wal MH, Nieuwenhuis MM, de Jong RM, van Dijk RB, Jaarsma T, et al. Health professionals' expectations versus experiences of internet-based telemonitoring: survey among heart failure clinics. J Med Internet Res 2013;15(1):e4 [FREE Full text] [doi: 10.2196/jmir.2161] [Medline: 23305645]
30. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for Translation and Cultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
31. Polit DF, Beck CT. Nursing Research: generating and assessing evidence for nursing practice. Philadelphia: Wolters Kluwer; 2017.
32. 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]
33. 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]
34. Melchiorre MG, Papa R, Rijken M, van Ginneken E, Hujala A, Barbabella F. eHealth in integrated care programs for people with multimorbidity in Europe: insights from the ICARE4EU project. Health Policy 2018 Dec;122(1):53-63 [FREE Full text] [doi: 10.1016/j.healthpol.2017.08.006] [Medline: 28899575]
35. European Commission. 2012. eHealth Action Plan 2012-2020: Innovative healthcare for the 21st century URL: https://ec.europa.eu/digital-single-market/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century [accessed 2018-10-29] [WebCite Cache ID 73WuHGBxa]
36. Riegel B, Jaarsma T, Strömberg A. A middle-range theory of self-care of chronic illness. ANS Adv Nurs Sci 2012;35(3):194-204. [doi: 10.1097/ANS.0b013e318261b1ba] [Medline: 22739426]
37. Riley JP, Astin F, Crespo-Leiro MG, Deaton CM, Kienhorst J, Lambrinou E, et al. Heart Failure Association of the European Society of Cardiology heart failure nurse curriculum. Eur J Heart Fail 2016 Jul;18(7):736-743 [FREE Full text] [doi: 10.1002/ejhf.568] [Medline: 27220672]
38. The Organisation for Economic Co-operation Development. 2018. Nurses URL: https://data.oecd.org/healthres/nurses.htm [accessed 2018-11-04] [WebCite Cache ID 73fo11GpD]
39. World Health Organization. Global strategy on human resources for health: Workforce 2030. Geneva: WHO; 2016. URL: https://www.who.int/hrh/resources/globstrathrh-2030/en/
40. Fairbrother P, Ure J, Hanley J, McCloughan L, Denvir M, Sheikh A, Telescot programme team. Telemonitoring for chronic heart failure: the views of patients and healthcare professionals - a qualitative study. J Clin Nurs 2014 Jan;23(1-2):132-144. [doi: 10.1111/jocn.12137] [Medline: 23451899]

41. Murauskiene L, Karanikolos M. The role of the European Structural and Investment Funds in financing health system in Lithuania: experience from 2007 to 2013 funding period and implications for the future. Health Policy 2017 Jul;121(7):727-730 [FREE Full text] [doi: 10.1016/j.healthpol.2017.04.012] [Medline: 28571667]
42. Pandor A, Thokala P, Gomersall T, Baalbaki H, Stevens JW, Wang J, et al. Home telemonitoring or structured telephone support programmes after recent discharge in patients with heart failure: systematic review and economic evaluation. Health Technol Assess 2013 Aug;17(32):1-207 [FREE Full text] [doi: 10.3310/hta17320] [Medline: 23927840]
43. Oudshoorn N. Diagnosis at a distance: the invisible work of patients and healthcare professionals in cardiac telemonitoring technology. Sociol Health Illn 2008 Mar;30(2):272-288 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01032.x] [Medline: 18290936]
44. Koulaouzidis G, Barrett D, Mohee K, Clark AL. Telemonitoring in subjects with newly diagnosed heart failure with reduced ejection fraction: from clinical research to everyday practice. J Telemed Telecare 2018 Jan 01:1357633X17751004 epub ahead of the print(forthcoming). [doi: 10.1177/1357633X17751004] [Medline: 29419343]
45. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Dec 22;392(10152):1047-1057 [FREE Full text] [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
46. Lind L, Carlgren G, Karlsson D. Old-and with severe heart failure: telemonitoring by using digital pen technology in specialized homecare: system description, implementation, and early results. Comput Inform Nurs 2016 Aug;34(8):360-368. [doi: 10.1097/CIN.0000000000000252] [Medline: 27223309]

Abbreviations

eHealth: electronic health
ESC: European Society of Cardiology
EU: European Union
GP: general practitioner
HCP: health care professional
HF: heart failure
ICT: information and communications technology
TM: telemonitoring

Edited by G Eysenbach; submitted 10.03.18; peer-reviewed by JM Weinstein, P Romano, NM Trofenciuc, I Mircheva; comments to author 17.08.18; revised version received 08.11.18; accepted 10.11.18; published 06.02.19.

Please cite as:
Aamodt IT, Lycholip E, Celutkiene J, Strömberg A, Atar D, Falk RS, von Lueder T, Hellesø R, Jaarsma T, Lie I
Health Care Professionals' Perceptions of Home Telemonitoring in Heart Failure Care: Cross-Sectional Survey
J Med Internet Res 2019;21(2):e10362
URL: http://www.jmir.org/2019/2/e10362/
doi:10.2196/10362
PMID:30724744

Original Paper

Cardiac Rehabilitees' Technology Experiences Before Remote Rehabilitation: Qualitative Study Using a Grounded Theory Approach

Marjo-Riitta Anttila[1], MSc; Heikki Kivistö[1], MSc; Arja Piirainen[1], PhD; Katja Kokko[1], PhD; Anita Malinen[2], PhD; Mika Pekkonen[3], MD, PhD; Tuulikki Sjögren[1], PhD

[1]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[2]Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland
[3]Peurunka Rehabilitation, Laukaa, Finland

Corresponding Author:
Marjo-Riitta Anttila, MSc
Faculty of Sport and Health Sciences
University of Jyväskylä
PO Box 35
Jyväskylä,
Finland
Phone: 358 40 805 4648
Fax: 358 14 260 2011
Email: marjo-riitta.m-r.anttila@jyu.fi

Abstract

Background: Even though technology is becoming increasingly common in rehabilitation programs, insufficient data are as yet available on rehabilitees' perceptions and experiences. It is important to understand their abilities when using technology for remote rehabilitation.

Objective: This is a qualitative study on technology experiences of persons affected by cardiovascular disease assessed before remote rehabilitation. The aim of the study was to explore rehabilitees' experiences and attitudes toward technology before 12 months of remote rehabilitation.

Methods: Qualitative interviews were conducted with 39 rehabilitees in four focus groups. The subjects were aged 34 to 77 years (average age 54.8 years) and 74% (29/39) of them were male. They had been diagnosed with coronary artery disease and were undergoing treatment in a rehabilitation center. The interviews were conducted between September 2015 and November 2016. Data were analyzed using Glaser's mode of the grounded theory approach.

Results: The result of the study was an "identifying e-usage" experience category, which refers to the rehabilitees' notions of the use of information and communication technologies (e-usage) in the process of behavior change. The main category comprises four subcategories that define the rehabilitees' technology experience. These subcategories are "feeling outsider," "being uninterested," "reflecting benefit," and "enthusiastic using." All rehabilitees expected that technology should be simple, flexible, and easy to use and learn. The results reflecting their technology experience can be used in e-rehabilitation programs. Rehabilitees who feel like outsiders and are not interested in technology need face-to-face communication for the major part of rehabilitation, while rehabilitees who reflect benefit and are enthusiastic about the use of technology need incrementally less face-to-face interaction and feel that Web-based coaching could offer sufficient support for rehabilitation.

Conclusions: The findings show that persons affected by heart disease had different experiences with technology and expectations toward counseling, while all rehabilitees expected technology to be easy to use and their experiences to be smooth and problem-free. The results can be used more widely in different contexts of social and health care for the planning of and training in remote rehabilitation counseling and education.

Trial Registration: ISRCTN Registry ISRCTN61225589; http://www.isrctn.com/ISRCTN61225589 (Archived by WebCite at http://www.webcitation.org/74jmrTXFD)

(J Med Internet Res 2019;21(2):e10985) doi:10.2196/10985

KEYWORDS

coronary disease; rehabilitees' experience; focus group; qualitative study; grounded theory; remote rehabilitation; e-health; e-rehabilitation; telerehabilitation; secondary prevention

Introduction

Cardiovascular diseases are the most common cause of death globally [1]. Cardiac rehabilitation is a means of secondary prevention intervention for cardiovascular diseases that includes efforts to reduce behavioral risks such as tobacco use, unhealthy diet, obesity, physical inactivity, alcohol use, and psychosocial problems such as depression [1,2]. However, many persons with coronary artery disease are not aware of opportunities to participate in rehabilitation programs or they choose not to participate in cardiac rehabilitation for a number of reasons, including living a long distance from a facility [3,4]. It is necessary to develop new methods of rehabilitation such as Web-based programs [5]. With global digitalization, rehabilitation increasingly uses technology. Remote rehabilitation programs use a range of remote technologies and Web-based applications. Remote counseling means professionally coached and monitored rehabilitation with a clearly defined beginning and end [6-8]. Digital eHealth tools include wireless digital devices like mobile phones and tablet computers, self-care and self-monitoring devices, video call services such as Skype for Business, wearable and ingestible sensors and various digital applications, and virtual reality made possible by robots and other forms of new technology [9]. Digitalization requires new attitudes and skills from rehabilitees and professionals [10].

The use of remote technology in cardiac rehabilitation has been studied mainly by quantitative methods [11]. Research has focused on the effectiveness [6] and usability [12] of technology-intensive interventions. Issues related to rehabilitees' physical activity [6,13,14] and lifestyle change [8,15] have been another focus area. Qualitative studies are a minority, and they have focused on experiences of participation in Facebook peer groups [16,17], mHealth [18], eHealth [19], or Web-based programs [20-22].

However, research has rarely looked at the role of remote technology in cardiac rehabilitation [14,23]. Research is needed to expand the understanding of the experiences of persons who use or have used remote technology and assess the pros and cons of this technology [7,8,11]. The aim of this study was to gain an understanding of cardiovascular rehabilitees' experiences with technology and of their attitudes toward technology.

Methods

Study Approach

We used a grounded theory approach in this study. Data were analyzed using Glaser's inductive grounded theory approach. We decided to apply a methodology proposed by the grounded theory approach because we found grounded theory useful in getting to understand the rehabilitees' subjective experiences for generating a substantive theory in a relatively new research area [24]. The focus was on finding out the rehabilitees' experiences and attitudes toward technology prior to using remote technology. The rehabilitees described in qualitative interviews their experiences with computer use, social media, and other applications of modern technology.

Recruitment

The interviews were conducted in 2015 and 2016 in Rehabilitation Centre Peurunka, Finland, where the Social Insurance Institution of Finland arranges regular cardiac rehabilitation courses. The study is a part of a remote technology in cardiac rehabilitation study registered at the ISRCTN Registry [ISRCTN61225589]. The Ethics Committee of the Central Finland Health Care District approved the study.

Participants

The participants were 39 rehabilitees (10 women and 29 men); 82% (32/39) of them had undergone coronary angioplasty and 10% (4/39) had undergone coronary artery bypass about 3 to 12 months prior to rehabilitation. Most subjects had a computer (23/27, 85%) and used the internet (25/27, 92%). Many had mobile phones (16/27, 59%) and tablets (10/27, 37%), and several used wrist activity trackers (10/27, 37%). These statistics are similar to those obtained during the testing of other European cardiac patient populations [25]. According to Glaser's inductive grounded theory approach, baseline information and characteristics of the subjects were collated later for this paper and were not taken into account in the analysis [24] (Table 1).

Data Collection

The total duration of rehabilitation was 15 days spread between three 5-day periods over a time span of 12 months. Rehabilitation took place in a rehabilitation center [26]. Qualitative interviews were conducted at the beginning of rehabilitation in 4 focus group discussions, each interview lasting 30 to 60 minutes, overall 156 minutes. The interviews were conducted by the same interviewer and they were informal and semistructured. The questions were mostly unplanned and spontaneous, but the interviewer also presented the same series of open-ended questions to all subjects. They were asked questions like, "Tell me about your experience with modern technology," and "What are your expectations of remote counseling?" The interviews were audio-recorded and transcribed for analysis. The transcripts were imported into ATLAS.ti (ATLAS.ti Scientific Software Development GmbH) computer software, which enables data storage, organization, and retrieval for analysis. The number of subjects was determined according to data saturation, which is a point at which no more new experiences of the topic could be elicited [24].

Table 1. Description of participants.

Characteristic	Under 55 years, n (%)	55 years and over, n (%)	Total, n (%)
Age (years)	20 (51)	19 (49)	39 (100)
Gender			
Male	17 (85)	12 (63)	29 (74)
Education			
Vocational or course-form school or other	13 (68)	14 (74)	27 (71)
College-level education	3 (16)	3 (16)	6 (16)
University of applied sciences	2 (11)	2 (11)	4 (11)
University	1 (5)	0 (0)	1 (3)
Time of operation			
0-3 months from rehabilitation	1 (5)	0 (0)	1 (3)
3-12 months from rehabilitation	13 (65)	11 (58)	24 (62)
Over 12 months from rehabilitation	4 (20)	6 (32)	10 (26)
No operations	2 (10)	2 (11)	4 (10)
Technology			
Internet, yes	12 (100)	13 (87)	25 (93)
Computer, yes	11 (92)	12 (80)	23 (85)
Tablet computer, yes	6 (50)	4 (27)	10 (37)
Mobile phone, yes	7 (58)	9 (60)	16 (59)
Physical activity tracker, yes	4 (33)	6 (40)	10 (37)

Data Analysis

The constant comparative model guided the data collection process. Data were collected through 4 informal interviews and analyzed using the constant comparative model. We collected and analyzed data concurrently, and as the analysis progressed the research question became more focused. First, we started substantive coding [24,27]. Subcategories were created in open coding. We analyzed incidents and compared them with other incidents, looking for similarities and differences and creating as many concepts as possible, coding substantively. Being theoretically sensitive, data were closely read and questioned. Next, we identified the properties and dimensions of each subcategory. Finally, we grouped the concept into subcategories creating as many subcategories as possible and then integrated the subcategories into the category. During this constant comparative model process, we recorded our ideas and notions, which helped us process the data. Data analysis was continued until the category was theoretically saturated [24].

The following example describes the creation of the "being uninterested" subcategory. The analysis began with open coding. We analyzed data on the diversity of the rehabilitees' experiences of technology use as well as their attitudes toward, and expectations for, remote counseling. This perspective expanded from their responses and debates. Constant comparison convinced us that all codes essentially described or explained how each rehabilitee used technology or what their attitude toward technology was. Next we named properties (1) using technology occasionally, (2) limiting to use, (3) challenging problem-free technology, and (4) activating empowerment counseling. We named the subcategory for this experience as "being uninterested." Textbox 1 shows an example of the process.

The constant comparison of properties resulted in hypotheses about relationships between the subcategories [28]. We continued to collect and analyze data until no new subcategories emerged and the subcategories were saturated; a category was thereby created and named "rehabilitees identifying e-usage in the process of behavior change."

Textbox 1. Creating category for e-usage identification.

Category:

- Identifying e-usage

Subcategory:

- Being uninterested

Property:

- Using technology occasionally
- Limiting to use
- Challenging problem-free technology
- Activating empowerment counseling

Concepts (code):

- The rehabilitee is not interested in using technology in his free time.
- Being bound to technology irritates; the rehabilitee does not want to use technology all the time.
- Ineffective technology worries the rehabilitee.
- The rehabilitee waits to be communicated with by email.
- The rehabilitee expects counseling to be a motivator and to spark interest

Text:

- I'm not terribly interested in that remote stuff, because...well, I use the computer at work every day....That email reading, I may go and check my mail once a week. It's not in a way...maybe it just isn't my thing...it's no trouble to surf on the internet in the evening. Only when I must. I can check email—if a bill has arrived, I can pay it there. [Participant 9, 44-year-old man, focus group 2]
- ...but I don't like that one's got to be, like, twenty-four hours a day available. [Woman, 59 years, focus group 4I've noticed sometimes that when I've been at it for some time, the machine has broken down in the middle of my work so no one could do anything. So, that is, of course, the downside of the thing... [Participant 4, 59-year-old woman, focus group 4]
- I guess now—real nice if sometimes one could be reached out to from there by email or something else... [Participant 9, 44-year-old man, focus group 2]
- I'm waiting for it and I'm truly interested, as if I were waiting for something like a spark. That it is something, something like, motivating... [Participant 56, 45-year-old man, focus group 3]

Results

The descriptors of the rehabilitees' prior technology experience are "feeling outsider," "being uninterested," "reflecting benefit," and "enthusiastic using." The category "identifying e-usage" describes the essence of the rehabilitees' experiences with using technology and identifying its usage (Figure 1). Individuals in the "feeling outsider" and "being uninterested" subcategories need more face-to-face counseling, while Web-based coaching is sufficient for the individuals in the "reflecting benefit" and "enthusiastic using" categories. These conclusions are based on the following results.

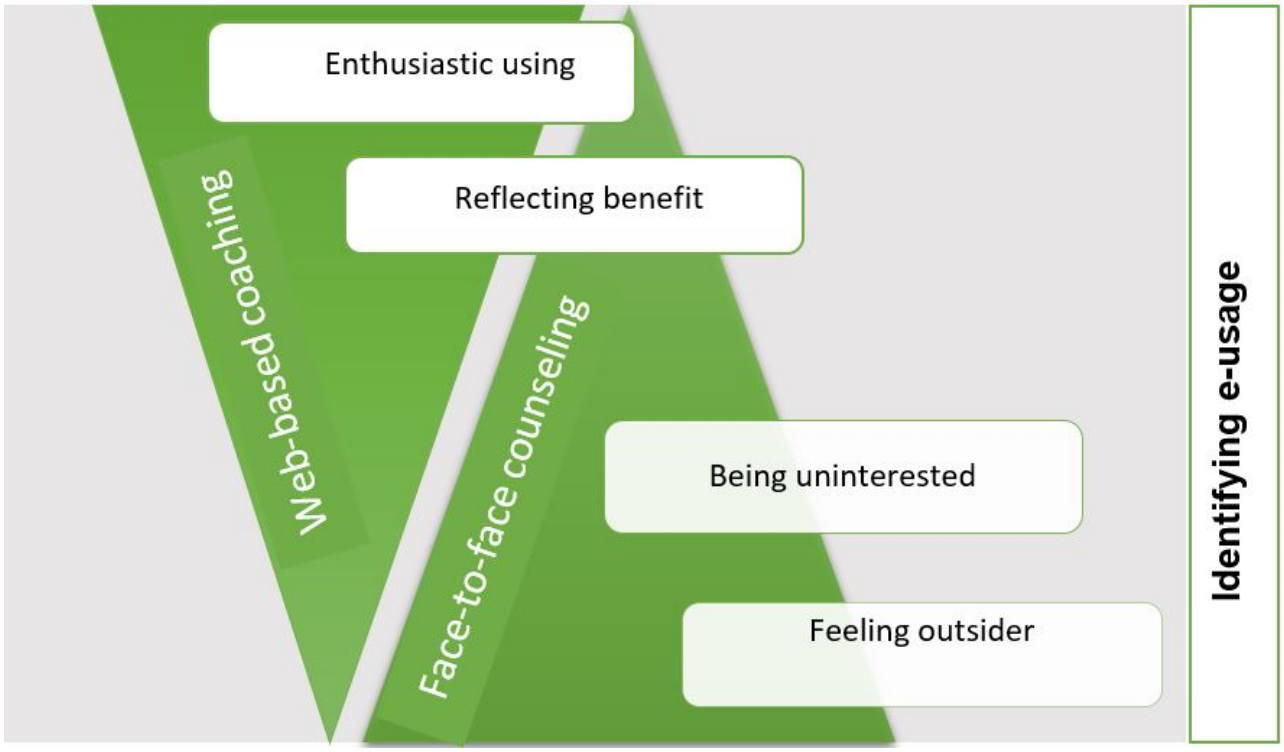
The first subcategory, "feeling outsider," consists of rehabilitees who fear that they do not have sufficient skills as computer users to participate in remote rehabilitation. They have not used the computer at all or they have only basic computer skills. In the following samples, two rehabilitees discuss "feeling outsiders."

> *Because I don't have a computer, I am a total outsider. So...because of this, I'm not so terribly interested. The only thing I know about this is that the self-tracking is great.* [Participant 39, 74-year-old man, focus group 1]

> *That technology hasn't really come...My wife taught the computer...supported, well, taught—so I went to the courses. And the kids did. I thought that if I'm still starting to tinker, there won't be enough hours in the day to learn.* [Participant 25, 60-year-old man, focus group 1]

These rehabilitees feel they do not have adequate computer skills, and few of them use computers at work. If they need help with technology, they request it from friends or family members. They feel they have no time to study computer use, and the English language is also difficult. Concerns about the impact of technology on health and security seem to be other reasons for avoiding computer use. In the following sample, a rehabilitee discusses his learning experiences with information technology.

Figure 1. Cardiac rehabilitees' different experiences of using technology and expectations of remote counseling.



> *Better that it leaves you, like if you go the bank computer, then everything gone. It doesn't help there. Still it doesn't. This isn't the only reason, but...I'm not interested. I've taken two computer courses, though. Last time I went to apply for a bus-driving license, and it was two weeks. And when I went and when I came back I was as dumb as I going there, and I come back...Of course, I know the time to buy one is coming, but I'm holding it off for as long as I can.* [Participant 58, 63-year-old man, focus group 4]

The following rehabilitee has not yet established expectations for counseling.

> *I'm not really sure...waiting to see what comes.* [Participant 64, 64-year-old, man, focus group 2]

For these rehabilitees, technology is something terrifying and almost incomprehensible. They are aware of its applications, such as Facebook, but these applications are foreign to them and therefore they feel like outsiders. Nevertheless, their positive expectations toward technology encounters are apparent although they do not expect anything amazing from Web-based counseling. They need guidance to support them in the use of technology (Figure 2). Figure 2 is a summary of the "identifying e-usage" category, which was created based on the rehabilitees' e-role and e-usage. The subjects in the "feeling outsider" subcategory regard themselves literally as outsiders and find technology fearsome; on the other hand, they look forward to overcoming this fear and expect adequate support.

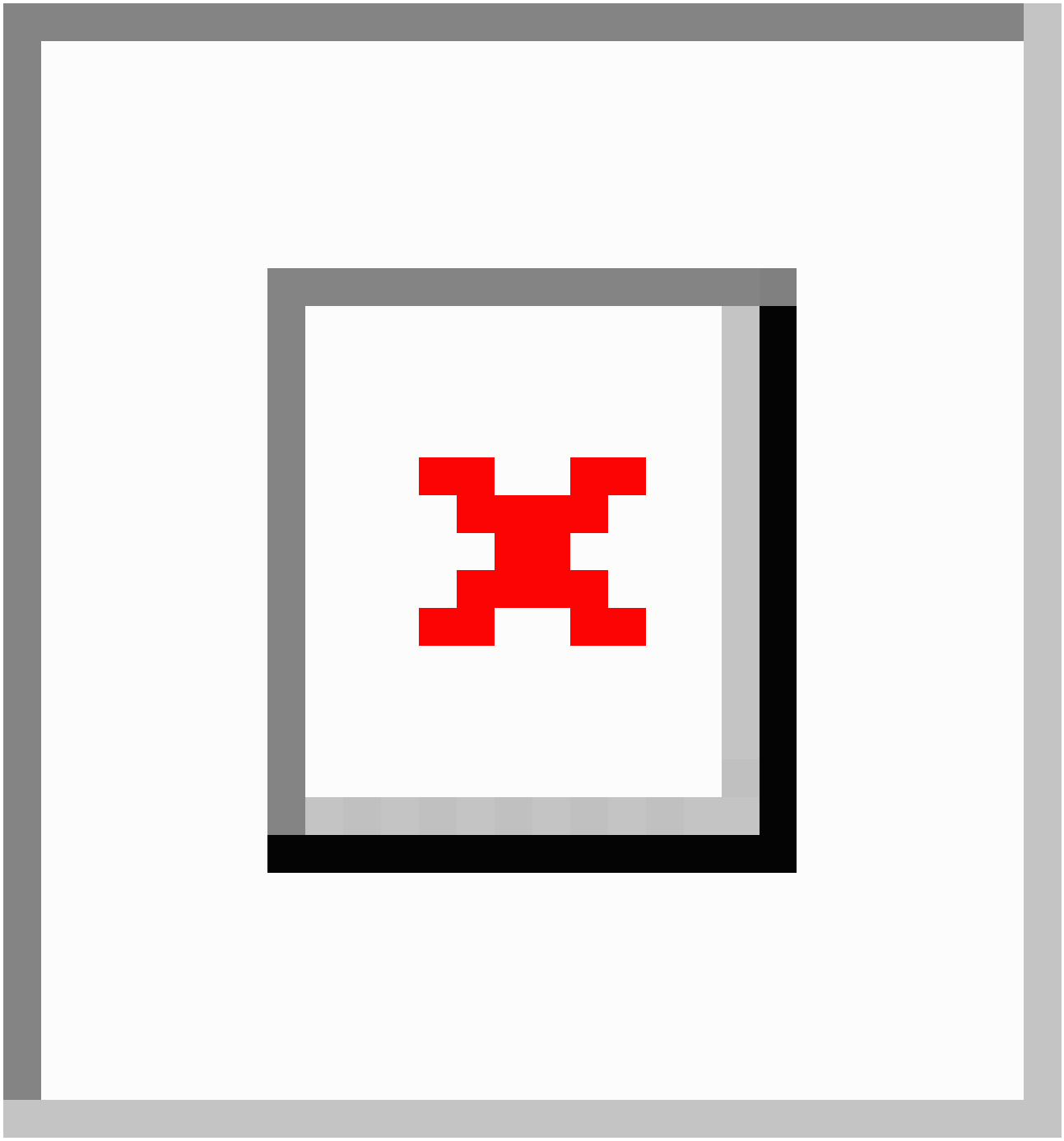
The second subcategory is "being uninterested." It is based on the experiences of rehabilitees who are conversant with technology and have experimented with social media. Their experiences are limited to necessary and occasional uses such as paying bills, renewing book loans, and reading emails. If they encounter a technical problem, their interest fades. They are not interested in using technology to connect socially via email and social media. They are worried about information security. In the following sample, a rehabilitee explains why he is not interested in technology.

> *I'm not terribly interested in that remote stuff, because...well, I use the computer at work every day...That email reading, I may go and check my mail once a week. It's not in a way...maybe it just isn't my thing...it's no trouble to surf on the internet in the evening. Only when I must. I can check email—if a bill has arrived, I can pay it there.* [Participant 9, 44-year-old man, focus group 2]

The following rehabilitee discusses problems related to technology and social media.

> *...I've noticed sometimes that when I've been at it for some time, the machine has broken down in the middle of my work so no one could do anything. So, that is, of course, the downside of the thing...* [Participant 5, 53-year-old woman, focus group 4]
>
> *But then what really irritates and frustrates me and just can't interest me—although I'm there in Facebook because my nephew forced me there.* [Participant 5, 53-year-old woman, focus group 4]

Figure 2. Identifying the e-usage category and subcategories of technology users.



The following rehabilitees discuss coaching.

> *I'm waiting for it and I'm truly interested, as if I were waiting for something like a spark. That it is something, something like, motivating, and...well...I can't say, but it like maybe not now for sure every week. If once a month, certainly something could come...a reminder.* [Participant 56, 45-year-old man, focus group 3]

> *When I could enter inputs in there, and if my own activities could be there, then I would be like a response: Is this the right or wrong direction, and...And that's when it's really somebody, something and someone monitoring what you're doing.* [Participant 41, 49-year-old woman, focus group 2]

These rehabilitees use technology occasionally and their daily use is limited. They value problem-free technology. They expect activity empowerment counseling, which should motivate and spark interest, but believe that technology demands a coach who would give feedback, assign weekly tasks, and issue regular reminders. This subcategory includes "occasionally," "limiting to use," "encouraging technology," and "supporting counseling" (Figure 2).

The subjects in the third subcategory, "reflecting benefit," maintain an interest in technology for only as long as they have an indispensable need for it in everyday life. In the following, a rehabilitee discusses his use of technology in free time.

> *Just like the pharmacy, like in that do I medications? In that case, is the prescription still valid? And like that, just in that way...Yeah, it is good to look...there are plenty of medicines left, and...Do I have to bring in, or order a new prescription? And other things, just in that way. What now happens every day or when it's needed so...so I don't go surfing on every webpage...Yes, with the children I use it, and with friends I like to connect over the internet.* [Participant 24, 65-year-old man, focus group 4]

The experiences of these persons fall into two groups. The subjects in the first group find technology difficult and need time to learn it. For example, they may find remembering passwords difficult. They need help to learn security procedures and computer usage.

> *And for paying bills I use it most, too. Some information when it's needed, well yes, I try to find it from there then. And if someone wants to find the frustrating side, well, those passwords frustrate me, because they always go missing...and a password has gotten lost, and...I can't go there anymore. Of course I can create a new password, but it is such a bother—just forget about it. And I read magazines in the computer, and...* [Participant 36, 68-year-old man, focus group 4]

Even though the use of technology is not a problem to the subjects in the second group, they still eschew technology. Since they see technology as something negative, they use it only when necessary—for example, to search for information. Some rehabilitees had experience in digital physical activity monitoring. In the following example, a rehabilitee discusses the usability of technology.

> *...But then the computer, when it runs all day—I don't want that. That's why I don't open the computer in the evening...Of course it's easy always that everything could be, like as easy-to-use as possible, because that's why I don't do it, when I could use it for remote technology. But it should be as easy-to-use as possible: it should be as automatic as possible, this thing. It, I think should be as flexible as possible.* [Participant 17, 57-year-old man, focus group 2]

In the next example, a rehabilitee discusses self-monitoring and coaching.

> *Let's put it in this way: I'm not actually now that way from being pushed, yeah. Yes it comes from my own desire. The main purpose is monitoring: it's for that. It's interesting to follow what happens if you change some exercise habits, and you can see from this, what changes have happened in the background. Very okay.* [Participant 17, 57-year-old man, focus group 2]

This subcategory, "reflecting benefit," emerged with four properties: "useful," "feeling technology useful," "easy-of-use technology," and "interactive tracking tool" (Figure 2). These rehabilitees expect technology to be easy to use and also expect it to enable communication. Essentially, they do not need a coach but they need tools for self-monitoring and helping to improve their health.

The subjects in the fourth subcategory, "enthusiastic using," show a positive attitude toward technology and have used it for a long time in a variety of ways, both in everyday life and at work through mobile phones, tablet computers, and desktop computers. In the following examples, enthusiastic users discuss the use of technology.

> *Well, laptops are always open less when you have a smartphone. In other words, I read those emails easily on my phone. Therefore I don't turn on the laptop.* [Participant 8, 61-year-old woman, focus group 2]

> *More there is, of course, invoice writing and information retrieval, but of course electrical diagrams, and...Sometimes some programming, logic, some small logic programming, and something like that.* [Participant 20, 64-year-old man, focus group 3]

These persons follow emails actively through mobile phones. Many of them have mobile health and exercise activity apps such as the Sports Tracker (Sports Tracking Technologies). They use social media such as Facebook and WhatsApp to keep in touch with friends and relatives. They do not consider internet problems particularly annoying and contact a specialist if they find this necessary. They are interested in technology and want to develop their technology skills. In the following example, rehabilitees describe what they expect from technology and from a coach.

> *I'm waiting and I'm interested. Yes, of course, this here now gives a little push in the pants. I'm already moving pretty well, that's what this thing around my arm tells me...Yeah...and then yes, I have the Sport Tracker on my phone, also. When I go somewhere, I tell it to draw a map, and I see the time and all that.* [Participant 66, 34-year-old man, focus group 3]

> *Modern opportunities. And if now, of course...from where soon could come a little spark, and that spark continues than exercise could begin. And it's really the same benefit. And then, of course, if nothing's heard from there. It sounds real good, and then reminders. Something like you can write comments, and...* [Participant 26, 61-year-old woman, focus group 2]

> *Maybe this is kind of a simple-enough device. When there's not anything amazing in here now, then owing to that, it's comfortable to use: It's not too complicated.* [Participant 15, 52-year-old man, focus group 1]

These rehabilitees use technology diversely and effortlessly, also for self-tracking. They expect Web-based intervention to be simple, motivating, easy to use, and interactive. They also expect coaches to give feedback if anything is missing. The attitude among the subjects in this subcategory is best described with the phrases "diversely" and "empowering self-tracking," and the best descriptors of expectations toward technology and counseling are, respectively, "smoothly functioning" and "getting feedback" (Figure 2).

Discussion

Principal Findings

The study shows that the diversity of the rehabilitees as technology users and their different needs for technology should be taken into account in rehabilitation planning. The four subcategories are "feeling outsider," "being uninterested," "reflecting benefit," and "enthusiastic using" (Figure 2). Some rehabilitees whose e-role is "outsider" or "occasionally using" need face-to-face communication for a large part of rehabilitation, while individuals whose e-role is "usefully" or "diversely" need incrementally less face-to-face interaction and feel that Web-based behavior change coaching will provide sufficient support for their rehabilitation. However, all rehabilitees hope that technology would be simple, flexible, and easy to use and learn, which would enable participation in an e-process. Participation in a remote rehabilitation program, as in this research, requires skills in areas such as Web-based log-ins and in reading and responding to tasks and messages. Activity self-monitoring requires downloading an activity tracker program to the computer and synchronizing the program with the computer.

The rehabilitees in the "feeling outsider" subcategory have a positive attitude toward technology, yet they do not see technology important for themselves. Their mindset supports previous research results that positive attitudes toward technology is a prerequisite for the uptake of technology [29,30] and eHealth [19]. At the beginning of remote rehabilitation, it is important to encourage rehabilitees' abilities to use various devices, since these abilities will make them more receptive to the use of technology [31] and enable participation in remote rehabilitation. Studies have also found that even though the digital skills of senior citizens have improved, they are still insufficient [32]. Rehabilitees' ability to use technology is also ensured by adequate internet technology support in remote rehabilitation [33]. Finally, it is important that apps and instructions are available in the user's own language.

The rehabilitees in the "being uninterested" subcategory were not interested in technology and eschewed its use. They felt that easy-to-use technology encourages technology use while technology that does not work frustrates, and they felt constant communication in social media irritating. They expect coach contact to maintain motivation during remote rehabilitation. However, when their expectations of technology are exceeded, the resulting experience is positive and pleasant [34,35] and maintains motivation [35], which has also been shown in previous research.

The rehabilitees in the "feeling outsider" and "being uninterested" subcategories need more face-to-face counseling during remote rehabilitation. The rehabilitees in the "feeling outsider" subcategory need supportive guidance in technology use, while the individuals in the "being uninterested" subcategory need to be motivated in order to create positive experiences. Maintaining a spark of interest and motivation requires a motivator and a coach who gives feedback, weekly assignments, and regular reminders [36]. Previous research has shown that interventions based on the behavior change theory may motivate more than those lacking a theoretical basis, but studies conducted on mobile cardiac rehabilitation have not specifically addressed behavior change strategies. Web-based interventions may provide an opportunity for real-time coaching [37,38], motivation, and engagement, allowing rehabilitees to achieve a meaningful behavior change [18]. The rehabilitees feel that they need an external motivator, and the importance of the behavior change theory should therefore be given an adequate emphasis in the planning of remote rehabilitation.

The rehabilitees in the "reflecting benefit" subcategory use technology daily, and technology challenges they encounter stem from technical problems and attitudes. They expect apps to be easy to use, secure, and in their own language. The perceived ease of technology use influences perceived usefulness and together these bolster their intention to use technology for a real purpose (usage behavior) [39]. The technology acceptance model has also been applied in the health care context [34,35]. In addition, perceived usefulness with a perceived value plays a role in the acceptance of technology [40]; for example, it provides personalized information, support, monitoring, and feedback [21,22]. The minimization of application risks increases trust in systems [21,33,35]. These rehabilitees' acceptance of technology increases when applications are easy to use and interactive, which has also been shown in previous qualitative studies.

Remote rehabilitation should enable social participation, such as peer group discussion and personalized feedback. These rehabilitees use social media as a means of communication and appreciate the possibility to interact. Social media, such as Facebook, Twitter, Pinterest, and Instagram, is part of their day-to-day life [41]. Social participation should be used in remote rehabilitation by granting them access to a peer group [16,17] and enabling problem-free peer group discussion on matters regarding the rehabilitation process. Health care rehabilitation applications allow users to receive information and interact since rehabilitees can receive assignments, record and review data, receive automated feedback, and connect with peers or health care professionals. All of these have been found to be important Web-based user experiences in previous studies [20-22,38]. These rehabilitees' experiences show little need for other services than automated feedback in the form of mainly interactive coaching, which gives a little push and supports a lifestyle change.

The rehabilitees in the "enthusiastic using" subcategory accept technology as an integral part of their everyday life. Statistics show that mobile phone use is increasing [42], and these rehabilitees use mobile phones actively. Mobile health apps are increasingly popular, and mobile phone users have downloaded mobile health apps [43]. They are keen users of sufficiently coach-supported Web-based intervention apps to boost motivation for physical activity [14,44,45]. Recent studies show that participants appreciate professional Web-based support [20]. The subjects of these studies had participated in Web-based e-rehabilitation, which reduces face-to-face interactions [45,46] and is particularly suited to an active user who adequately masters technology and is interest in it. Easy-to-use and smoothly functioning technology allows extensive personal activity and body function monitoring—in other words,

self-tracking. The recently termed quantified self notion has emerged to promote self-knowledge through numbers [9]. Health change coaching is based on the behavior change theory, motivational strategies, and communication techniques [20,46].

The rehabilitees' experiences with and attitudes toward technology provide information on how to implement a counseling theory and methods for the planning of remote rehabilitation. A coach should conduct individual risk factor assessment and management, exercise training, and self-management of modifiable risk factors and provide education and psychosocial support [38]. Professional health coaches can help rehabilitees increase self-direction, set specific goals, and take action to achieve and sustain health-supporting behaviors [38,47-49]. In addition to self-monitoring, an easy-to-use interface is a desirable feature in mobile apps for promoting physical activity. Examples of these interfaces are the integration of biosensors that collect information from body and life systems such as electrocardiogram, physical activity, heart rate, blood pressure, and blood glucose measurement [50]. Digital stethoscopes, thermometers, and weight scales [9] can also be used in remote rehabilitation.

The topic is important because remote rehabilitation is already being implemented and will continue to be implemented in the future due to increasing digitization [46,47]. Technology-related studies have shown that remote technology is most successful when it is simple and designed for easy understanding and easy use [30]. Easy-to-use technology also produces positive and successful experiences [49]. As rehabilitees' skills develop, they gain self-esteem and are empowered to expect positive, successful experiences [29,32]. Remote rehabilitation must take into account the abilities of each rehabilitee in learning, cognition, and motor and perceptual skills [29,33,51] and allow an individual ample time to master new skills [29]. In the future, the recognized needs and concerns of the subjects in the "feeling outsider," "being uninterested," "reflecting benefit," and "enthusiastic using" categories should be combined with previous research and taken into account. In this way, the acceptance and use of remote technology could be upped for more meaningful and effective rehabilitation. The results of this study can also be used in designing remote rehabilitation and health coach training programs. There is a lack of quality research on the experiences of coronary disease patients, and a need exists for mixed-methods research for the development of easy-to-use effective and meaningful welfare technology.

Limitations

There are weakness that need to be considered when interpreting the findings of this study. The subjects discussed their experiences of technology at the beginning of rehabilitation, and everyone was given the opportunity to share his or her experiences. The interviewer created an accepting atmosphere and encouraged silent participants. Despite this, it is possible that the participants were trying to please the group when answering the questions. On the other hands, there are advantages. The results have attracted interest and their relevance, credibility, and usefulness have been identified as important when implementing remote rehabilitation. They have been also used in comparative rehabilitation groups, in musculoskeletal system reconditioning, and in work ability rehabilitation.

Conclusions

The aim of this study was to explore in detail rehabilitees' experiences and attitudes toward technology. The results are the rehabilitees' technology experiences described as "feeling outsider," "being uninterested," "reflecting benefit," and "enthusiastic using," which relate to their e-usage. The results help providers and health workers to identify different technology users among potential rehabilitees and determine what use levels must be taken into account when developing remote rehabilitation. The category formed into four subcategories which define the rehabilitees' technology experience and attitude. The results can also be used more widely in different contexts of social and health care for the planning of and training in remote rehabilitation/e-rehabilitation counseling and education.

Conflicts of Interest

None declared.

References

1. World Health Organization. Cardiovascular diseases (CVD) URL: http://www.who.int/mediacentre/factsheets/fs317/en/ [accessed 2017-10-13] [WebCite Cache ID 6uBVaiDtW]
2. Balady GJ, Williams MA, Ades PA, Bittner V, Comoss P, Foody JM, et al. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2007 May 22;115(20):2675-2682 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.106.180945] [Medline: 17513578]
3. Kotseva K, Wood D, De Backer G, De Bacquer D, EUROASPIRE III Study Group. Use and effects of cardiac rehabilitation in patients with coronary heart disease: results from the EUROASPIRE III survey. Eur J Prev Cardiol 2013 Oct;20(5):817-826. [doi: 10.1177/2047487312449591] [Medline: 22718794]
4. Neubeck L, Freedman SB, Clark AM, Briffa T, Bauman A, Redfern J. Participating in cardiac rehabilitation: a systematic review and meta-synthesis of qualitative data. Eur J Prev Cardiol 2012 Jun;19(3):494-503. [Medline: 22779092]

5. Salminen A, Hiekkala S, Stenberg JH. Etäkuntoutus. Kansaneläkelaitoksen julkaisuja. Tampere: Juvenes Print; 2016.
6. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
7. Salminen A, Heiskanen T, Hiekkala S. Etäkuntoutuksen ja siihen läheisesti liittyvien termien määrittelyä. In: Inalminen AL, Hiekkala S, Stenberg JH, editors. Etäkuntoutus. Kansaneläkelaitoksen julkaisuja. Tampere: Juvenes Print; 2016:11.
8. Antypas K, Wangberg SC. An Internet- and mobile-based tailored intervention to enhance maintenance of physical activity after cardiac rehabilitation: short-term results of a randomized controlled trial. J Med Internet Res 2014;16(3):e77 [FREE Full text] [doi: 10.2196/jmir.3132] [Medline: 24618349]
9. Lupton D, Maslen S. Telemedicine and the senses: a review. Sociol Health Illn 2017 Dec;39(8):1557-1571. [doi: 10.1111/1467-9566.12617] [Medline: 29071731]
10. 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]
11. O'Reilly GA, Spruijt-Metz D. Current mHealth technologies for physical activity assessment and promotion. Am J Prev Med 2013 Oct;45(4):501-507 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.012] [Medline: 24050427]
12. Lloréns R, Noé E, Colomer C, Alcañiz M. Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Mar;96(3):418-425 [FREE Full text] [doi: 10.1016/j.apmr.2014.10.019] [Medline: 25448245]
13. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014;16(9):e186 [FREE Full text] [doi: 10.2196/jmir.3340] [Medline: 25217464]
14. Ravanne A, Rintala A, Hakala S, Sjögren T. Etäteknologian vaikuttavuus sydänkuntoutujien fyysiseen aktiivisuuteen. In: Rintala A, Hakala S, Sjögren T, editors. Etäteknologian vaikuttavuus liikunnallisessa kuntoutuksessa. Järjestelmällinen kirjallisuuskatsaus ja meta-analyysi. Sosiaali-ja terveysturvan tutkimuksia 145. Helsinki: Erweko; 2017:23-125.
15. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
16. Partridge SR, Grunseit AC, Gallagher P, Freeman B, O'Hara BJ, Neubeck L, et al. Cardiac patients' experiences and perceptions of social media: mixed-methods study. J Med Internet Res 2017 Sep 15;19(9):e323 [FREE Full text] [doi: 10.2196/jmir.8081] [Medline: 28916507]
17. Partridge SR, Gallagher P, Freeman B, Gallagher R. Facebook groups for the management of chronic diseases. J Med Internet Res 2018 Jan 17;20(1):e21 [FREE Full text] [doi: 10.2196/jmir.7558] [Medline: 29343460]
18. Miyamoto SW, Henderson S, Young HM, Pande A, Han JJ. Tracking health data is not enough: a qualitative exploration of the role of healthcare partnerships and mhealth technology to promote physical activity and to sustain behavior change. JMIR Mhealth Uhealth 2016 Jan 20;4(1):e5 [FREE Full text] [doi: 10.2196/mhealth.4814] [Medline: 26792225]
19. Peeters JM, Krijgsman JW, Brabers AE, Jong JDD, Friele RD. Use and uptake of ehealth in general practice: a cross-sectional survey and focus group study among health care users and general practitioners. JMIR Med Inform 2016 Apr 06;4(2):e11 [FREE Full text] [doi: 10.2196/medinform.4515] [Medline: 27052805]
20. Nordin C, Michaelson P, Eriksson MK, Gard G. It's about me: patients' experiences of patient participation in the web behavior change program for activity in combination with multimodal pain rehabilitation. J Med Internet Res 2017 Jan 18;19(1):e22 [FREE Full text] [doi: 10.2196/jmir.5970] [Medline: 28100440]
21. Pearson J, Walsh N, Carter D, Koskela S, Hurley M. Developing a web-based version of an exercise-based rehabilitation program for people with chronic knee and hip pain: a mixed methods study. JMIR Res Protoc 2016 May 19;5(2):e67 [FREE Full text] [doi: 10.2196/resprot.5446] [Medline: 27197702]
22. Lundgren J, Johansson P, Jaarsma T, Andersson G, Kärner Köhler A. Patient experiences of web-based cognitive behavioral therapy for heart failure and depression: qualitative study. J Med Internet Res 2018 Sep 05;20(9):e10302 [FREE Full text] [doi: 10.2196/10302] [Medline: 30185405]
23. Hakala S, Rintala A, Immonen J, Karvanen J, Heinonen A, Sjögren T. Effectiveness of physical activity promoting technology-based distance interventions compared to usual care: systematic review, meta-analysis and meta-regression. Eur J Phys Rehabil Med 2017 Dec;53(6):953-967. [doi: 10.23736/S1973-9087.17.04585-3] [Medline: 28466628]
24. Glaser BG. Theoretical Sensitivity: Advances in the Methodology of Grounded Theory. San Francisco: University of California; 1978.
25. Buys R, Claes J, Walsh D, Cornelis N, Moran K, Budts W, et al. Cardiac patients show high interest in technology enabled cardiovascular rehabilitation. BMC Med Inform Decis Mak 2016 Jul 19;16:95 [FREE Full text] [doi: 10.1186/s12911-016-0329-9] [Medline: 27431419]
26. Kelan avo- ja laitosmuotoisen kuntoutuksen standardi Sydänkurssi sydänsairautta sairastavien aikuisten kuntoutuskurssi, osittainen perhekurssi Voimassa. 2017. URL: http://www.kela.fi/standardit_hyvaksytyt-palvelukuvaukset [accessed 2017-01-26] [WebCite Cache ID 6wl5uwDmq]

27. Glaser B, Strauss A. Strategies for qualitative research. In: The Discovery of Grounded Theory: Strategies for Qualitative Research. New York: Routledge; 2017:105-107.
28. Glaser BG. Basic of Grounded Theory Analysis: Emergence Vs. Forcing. Mill Valley: Sociology Press; 1992:55.
29. Broady T, Chan A, Caputi P. Comparison of older and younger adults' attitudes toward and abilities with computers: implications for training and learning. Brit J Educ Technol 2010;41(3):485. [doi: 10.1111/j.1467-8535.2008.00914.x]
30. Torp S, Hanson E, Hauge S, Ulstein I, Magnusson L. A pilot study of how information and communication technology may contribute to health promotion among elderly spousal carers in Norway. Health Soc Care Community 2008 Jan;16(1):75-85. [doi: 10.1111/j.1365-2524.2007.00725.x] [Medline: 18181817]
31. Moon BC, Chang H. Technology acceptance and adoption of innovative smartphone uses among hospital employees. Healthc Inform Res 2014 Oct;20(4):304-312 [FREE Full text] [doi: 10.4258/hir.2014.20.4.304] [Medline: 25405067]
32. Näsi M, Räsänen P, Sarpila O. ICT activity in later life: Internet use and leisure activities amongst senior citizens in Finland. Eur J Ageing 2012 Jun;9(2):169-176 [FREE Full text] [doi: 10.1007/s10433-011-0210-8] [Medline: 28804417]
33. Arthanat S, Vroman KG, Lysack C. A home-based individualized information communication technology training program for older adults: a demonstration of effectiveness and value. Disabil Rehabil Assist Technol 2016;11(4):316-324. [doi: 10.3109/17483107.2014.974219] [Medline: 25512061]
34. Hung M, Jen W. The adoption of mobile health management services: an empirical study. J Med Syst 2012 Jun;36(3):1381-1388. [doi: 10.1007/s10916-010-9600-2] [Medline: 20878452]
35. Ortega Egea JM, Roman Gonzàlez MV. Explaining physicians' acceptance of EHCR systems: an extension of TAM with trust and risk factors. Comput Hum Behav 2011 Jan;27(1):319. [doi: 10.1016/j.chb.2010.08.010]
36. Knittle K, Nurmi J, Crutzen R, Hankonen N, Beattie M, Dombrowski SU. How can interventions increase motivation for physical activity? A systematic review and meta-analysis. Health Psychol Rev 2018 Feb 15:1-20. [doi: 10.1080/17437199.2018.1435299] [Medline: 29385950]
37. 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]
38. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013;2(6):e000568 [FREE Full text] [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
39. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
40. Kaasinen E. Technical Research Centre of Finland publication 566: User Acceptance of Mobile Service: Value, Ease of Use, Trust and Ease of Adoption. Helsinki: Valopaino; 2005:51.
41. Greenwood S, Perrin A, Duggan M. Social media update 2016. Washington: Pew Internet and American Life Project; 2016. URL: http://www.pewinternet.org/wp-content/uploads/sites/9/2016/11/PI_2016.11.11_Social-Media-Update_FINAL.pdf [accessed 2018-12-17] [WebCite Cache ID 74jBycGY1]
42. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. Washington: Pew Internet and American Life Project; 2016. URL: http://www.pewresearch.org/wp-content/uploads/sites/2/2016/02/pew_research_center_global_technology_report_final_february_22__2016.pdf [accessed 2018-12-17] [WebCite Cache ID 74jCEJC2L]
43. Fox S, Duggan M. Pew Internet. 2012 Nov 08. Mobile health 2012 URL: http://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf [accessed 2018-05-28] [WebCite Cache ID 6zkR4I2B9]
44. 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]
45. Brouwers RWM, Kraal JJ, Traa SCJ, Spee RF, Oostveen LMLC, Kemps HMC. Effects of cardiac telerehabilitation in patients with coronary artery disease using a personalised patient-centred web application: protocol for the SmartCare-CAD randomised controlled trial. BMC Cardiovasc Disord 2017 Dec 31;17(1):46 [FREE Full text] [doi: 10.1186/s12872-017-0477-6] [Medline: 28143388]
46. Oksman E, Linna M, Hörhammer I, Lammintakanen J, Talja M. Cost-effectiveness analysis for a tele-based health coaching program for chronic disease in primary care. BMC Health Serv Res 2017 Dec 15;17(1):138 [FREE Full text] [doi: 10.1186/s12913-017-2088-4] [Medline: 28202032]
47. Wolever RQ, Simmons LA, Sforzo GA, Dill D, Kaye M, Bechard EM, et al. A systematic review of the literature on health and wellness coaching: defining a key behavioral intervention in healthcare. Glob Adv Health Med 2013 Jul;2(4):38-57 [FREE Full text] [doi: 10.7453/gahmj.2013.042] [Medline: 24416684]
48. Brough C, Boyce S, Houchen-Wolloff L, Sewell L, Singh S. Evaluating the interactive web-based program, activate your heart, for cardiac rehabilitation patients: a pilot study. J Med Internet Res 2014;16(10):e242 [FREE Full text] [doi: 10.2196/jmir.3027] [Medline: 25359204]
49. Taylor A, Wade V, Morris G, Pech J, Rechter S, Kidd M, et al. Technology support to a telehealth in the home service: qualitative observations. J Telemed Telecare 2016 Jul;22(5):296-303. [doi: 10.1177/1357633X15601523] [Medline: 26362562]

50. Rawstorn JC, Gant N, Meads A, Warren I, Maddison R. Remotely delivered exercise-based cardiac rehabilitation: design and content development of a novel mhealth platform. JMIR Mhealth Uhealth 2016;4(2):e57 [FREE Full text] [doi: 10.2196/mhealth.5501] [Medline: 27342791]
51. Ellis D, Allaire JC. Modeling computer interest in older adults: the role of age, education, computer knowledge, and computer anxiety. Hum Factors 1999 Sep;41(3):345-355. [Medline: 10665203]

Edited by G Eysenbach; submitted 08.05.18; peer-reviewed by E Kärnä, S Lin; comments to author 01.09.18; revised version received 24.10.18; accepted 26.10.18; published 07.02.19.

Please cite as:
Anttila MR, Kivistö H, Piirainen A, Kokko K, Malinen A, Pekkonen M, Sjögren T
Cardiac Rehabilitees' Technology Experiences Before Remote Rehabilitation: Qualitative Study Using a Grounded Theory Approach
J Med Internet Res 2019;21(2):e10985
URL: http://www.jmir.org/2019/2/e10985/
doi:10.2196/10985
PMID:30730298

Original Paper

What Affects the Completion of Ecological Momentary Assessments in Chronic Pain Research? An Individual Patient Data Meta-Analysis

Masakatsu Ono[1], PhD; Stefan Schneider[1], PhD; Doerte U Junghaenel[1], PhD; Arthur A Stone[1,2], PhD

[1]Center for Self-Report Science, Center for Economic and Social Research, University of Southern California, Los Angeles, CA, United States
[2]Department of Psychology, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Masakatsu Ono, PhD
Center for Self-Report Science
Center for Economic and Social Research
University of Southern California
635 Downey Way
Los Angeles, CA, 90089
United States
Phone: 1 213 821 8862
Fax: 1 213 821 2716
Email: masakatsu.ono@manchester.ac.uk

Abstract

Background: Ecological momentary assessment (EMA) involves repeated sampling of people's current experiences in real time in their natural environments, which offers a granular perspective on patients' experience of pain and other symptoms. However, EMA can be burdensome to patients, and its benefits depend upon patients' engagement in the assessments.

Objective: The goal of this study was to investigate factors affecting EMA-completion rates among patients with chronic pain.

Methods: This individual patient data meta-analysis was based on 12 EMA datasets that examined patients with chronic noncancer-related pain (n=701). The EMA-completion rates were calculated on a daily basis for each patient. Multilevel models were used to test the following predictors of completion rates at different levels: within-patient factors (days into the study and daily pain level), between-patient factors (age, sex, pain diagnosis, and average pain level per person), and between-study EMA design factors (study duration, sampling density, and survey length).

Results: Across datasets, an EMA-completion rate of 85% was observed. The strongest results were found for the between-patient factor age: Younger respondents reported lower completion rates than older respondents (P=.002). One within-patient factor, study day, was associated with completion rates (P<.001): over the course of the studies, the completion rates declined. The two abovementioned factors interacted with each other (P=.02) in that younger participants showed a more rapid decline in EMA completion over time. In addition, none of the other hypothesized factors including gender, chronic pain diagnoses, pain intensity levels, or measures of study burden showed any significant effects.

Conclusion: Many factors thought to influence the EMA-completion rates in chronic pain studies were not confirmed. However, future EMA research in chronic pain should note that study length and young age can impact the quality of the momentary data and devise strategies to maximize completion rates across different age groups and study days.

(J Med Internet Res 2019;21(2):e11398) doi:10.2196/11398

KEYWORDS

chronic pain; completion rate; compliance rate; ecological momentary assessment; experience sampling; IPD meta-analysis

Introduction

Ecological momentary assessment, or EMA [1] (also known as experience sampling), is gaining increasing attention in medical research and research on chronic pain, mainly due to its ability to capture real-time data that reflect the dynamics of patients' experiences in their natural environment. The methodology involves prompting participants several times per day to answer questions about their current pain and symptoms. This facilitates

coverage of people's experiences under a full range of momentary contexts, circumvents potential recall bias, and enhances the ecological validity of assessments [1,2]. Despite the advantages of EMA, the quality of these data depends on adequate completion rates of the sampling protocol [3]. Specifically, systematic differences in completion patterns according to characteristics of the situation, person, or EMA protocol could lead to input of data that are not missing at random, which could severely bias the results of subsequent analyses [4]. Modern statistical approaches for handling missing EMA data (maximum likelihood estimation or multiple imputation) rely on the assumption that data are missing at random (MAR). For the MAR assumption to be met, variables that are predictive of missing values need to be included in the analysis or imputation model [5]. Therefore, identification of factors facilitating or reducing EMA-completion rates is important.

Although few studies have examined predictors of completion rates in EMA, the available evidence suggests that male sex, engagement in behaviors that draw attention away from participation (eg, drinking alcohol or exercising), and long participation in the protocol may be associated with low completion rates [4,6-9]. However, this research was based on nonpatient samples (eg, college students, healthy adults, and drug or tobacco users). To our knowledge, only two original studies have examined factors associated with EMA-completion rates in patients with chronic pain. Aaron and colleagues [10] examined the completion rates among patients with temporomandibular disorder and found that demographic and medical characteristics were not related to the number of missed EMA surveys; interestingly, participants with high negative mood and high stress tended to have low completion rates. Okifuji et al [11] found no differences in the missing response rates according to patient characteristics (eg, age, pain, and fatigue), but the missing response rates increased over the course of the 30-day EMA study. In addition, Morren and colleagues [12] used meta-analytic procedures to examine the EMA-completion rates in pain research and reported overall high completion rates (average, 83%) among papers that provided the completion rates (36 of 48 studies). In addition, their results suggested that studies with older patients, shorter EMA surveys, participation manuals, alarm functions, and financial incentives had higher average completion rates.

In this study, we used an individual patient data (IPD) meta-analytic approach to investigate predictors of EMA-completion rates in patients with chronic pain [13-15]. There are several advantages to the use of IPD meta-analysis. First, in contrast to traditional meta-analysis, we utilize and synthesize raw data of several EMA studies, which allows examination of predictors that are not reported in the published article or only reported as summary statistics. This avoids potential biases resulting from inferences about individuals made from group-level summary data ("ecological fallacy"). Second, with IPD meta-analysis, it is possible to standardize outcome definitions across studies; this is especially a concern because the conceptualization and quality of reporting of EMA-completion rates varies widely between published studies [16]). Third, IPD allows examination of predictors of completion rates across multiple levels of analysis, including features of the study protocol (study-level predictors), participant-level features (person-level predictors), temporal features (within-person changes over time), and the construct under investigation (ie, pain intensity).

The overall objective of our IPD meta-analysis was to determine the profile of the study level, participant level, and situational features that affect completion rates in EMA research on chronic pain. At a study level, we hypothesized that long EMA questionnaires, long study durations, and high sampling densities may be associated with low completion rates. On the basis of prior related literature, at the participant level, we hypothesized that male sex and young age may be associated with low completion rates [6-8,12]. Further, we hypothesized that completion rates would decrease over time [11] due to participant fatigue or reduced motivation in the later phase of the study. Finally, we hypothesized that chronic pain diagnosis or patients' pain intensity might predict completion of EMA assessments.

Methods

Data Acquisition

The data for this study were sourced from a larger study utilizing secondary data analyses of preexisting EMA datasets for characterizing momentary pain experiences in patients with chronic noncancer-related pain [16]. The inclusion criteria were study sample of at least 30 adult patients (excluding studies with pediatric patients) and administration of a minimum of 3 EMA pain intensity prompts per day for at least 4 days with a fixed- or random-assessment schedule that was assessed via electronic diaries, mobile phones, or interactive voice responses. Studies using paper diaries were excluded because of problems (eg, back-filling and forward-filling) that can undermine the validity of the estimated completion rates [17,18]. Observational studies and clinical trials were included, but clinical trials were limited to no-intervention or baseline assessment periods. EMA pain assessments needed to focus on monitoring momentary pain intensity. Studies that used EMA exclusively as an intervention trigger (eg, just-in-time adaptive interventions [19]) were excluded.

Eligible datasets were identified through a systematic literature search conducted in October 2016 using PubMed and Web of Science databases with the following search terms: ([“Ecological Momentary Assessment” or “Experience Sampling” or “Electronic Diary” or “Electronic Diaries” or “Electronic Interview” or “Electronic Interviews” or “Interactive Voice Response” or “Intensive Diaries” or “Ambulatory Monitoring” or “Ambulatory Assessment”] and “Pain”).

Analysis Strategy

Completion rates were calculated as the percentage of EMA prompts completed (relative to the number of prompts received) for each person and day of the study. For EMA protocols with a fixed sampling scheme (6 studies), we considered the number of prompts received as the number of prompts specified by the protocol (and reported in respective articles). For EMA protocols with a random-sampling scheme (4 studies), we obtained the

specific number of executed prompts from the datasets, because this number could vary across days (this was especially the case when studies allowed the number of momentary prompts to vary according to patients' waking hours). Given the proportional nature of the completion rates, we tested our models with both the original and arcsine-transformed scores [20]. In this paper, we report the results based on the original completion rates, because the analyses yielded nearly identical results.

To examine changes in completion rates over time, the study day was coded as a within-person (day-level) predictor variable. Data for patient-level predictors of EMA-completion rates—age and sex—were taken directly from the databases. Patients' chronic pain diagnosis was coded as osteoarthritis, rheumatoid arthritis, fibromyalgia, or other diagnoses. The following features of the EMA protocol were coded as study-level predictors of EMA completion rates: Study duration was coded as the total number of days of the EMA protocol, EMA sampling density was coded as the average number of EMA prompts received per day, and EMA survey length was coded as the number of EMA items presented at each prompt.

We used the momentary pain intensity ratings available in each dataset to examine whether EMA-completion rates were associated with pain intensity at the day, person, or study level. The number of scale points used to measure momentary pain differed across studies (range, 5-101 points), and the pain ratings were converted, so that the ratings were on a 101-point scale. For the conversion, we used the following equation: New rating=100*(original rating+0.5)/(number of scale points). Because momentary pain ratings were not assessed at the time of the missed EMA prompts, the analyses were based on averages of the nonmissing pain ratings, as per a previous analysis [10]. Specifically, study- and person-level averages of all available pain ratings were calculated to examine whether studies or patients who, on an average, reported higher pain levels showed lower (or higher) completion rates. Additionally, daily average pain levels were computed for each patient (and within-person centered) to examine whether day-to-day variations in pain in a given patient were related to daily completion rates.

Our multilevel models subsequently incorporated day-level (Level 1), person-level (Level 2), and study-level (Level 3) predictors of EMA-completion rates. Specifically, a model without predictors (Step 1) was followed by analyses of the day-level predictors, which examined changes in daily completion rates over time (ie, over the course of the EMA sampling protocols; Step 2). Step 3 added patient-level predictors including age, sex, and chronic pain diagnosis. Step 4 added the following study-level predictors: study duration, EMA sampling density, and EMA survey length. In the final step (Step 5), day-, patient-, and study-level averages of pain intensity were added as predictor variables of completion rates on each of the 3 levels. Analyses were conducted using maximum likelihood parameter estimation in Mplus, version 8 [21]. Values of $P<.05$ were considered statistically significant.

Power Calculations

Statistical power in multilevel models depends on several factors such as the sample sizes at each level of analysis and the intraclass correlations due to the clustering effect or observation dependence of lower-level units nested in higher-level units. We conducted a power analysis using Monte Carlo simulation [22] to determine the minimum effect sizes that would be detectable with 80% power (alpha=.05), given the sample sizes and intraclass correlations of the data analyzed in the present study (see Sample Size and Design Characteristics section below). The minimum detectable effect sizes were 0.04 at Level 1 and 0.1 at Level 2, corresponding with small effects following the Cohen [23] conventions. Owing to the limited sample size at Level 3 (the study level), the minimum detectable effect size at this level was 0.7, which was a large effect.

Results

Results of the Literature Search

Our literature search identified 20 eligible databases from 37 articles (Figure 1). Authors of these articles were contacted by the research team. Original patient data were received for 10 of the 20 databases: Nine datasets were not received because the authors did not respond to the request or declined to provide the data or because the data were no longer available, and one dataset was not included because it provided only partial data without information on demographic predictor variables. One database consisted of three substudies comprising independent patient samples with different EMA sampling designs. These were separated into three datasets; thus, a total of 12 independent datasets were included in the analyses.

Sample and Design Characteristics

Characteristics of the participants and studies are summarized in Table 1. Overall, our analyses included 7956 study days from a total of 701 patients. The study duration ranged from 4 to 28 days. The number of prompts per day ranged from 3 to 12 prompts, and the number of items per prompt ranged from 6 to 63 items.

Descriptive Results of Ecological Momentary Assessment Completion Rates

The distribution of average daily completion rates by individuals is presented in Figure 2. The average completion rate was 85%, with daily completion rates <70% for 13% of the patients, <80% for 27% of the patients, and <90% for 60% of the patients. Initial multilevel models without predictor variables showed that 58% of the total variance in completion rates was attributable to within-person (day-to-day) variation: 26% to reliable differences between patients and 15% to differences among studies. Thus, investigation of predictors from different levels was warranted.

Predictors of Ecological Momentary Assessment Completion Rates

Findings for the multilevel model predicting completion rates are summarized in Table 2. On the within-person level, we found a significant linear decline in daily completion over time (b=–2.29, $P<.001$); on average, completion rates decreased by approximately 2.0% per week of EMA sampling. On the between-person level, age showed a curvilinear relationship with completion rates (b_{age_linear}=1.77, P=.002; $b_{age\_quadratic}$=–0.62, P=.009). Completion rates were highest

among older patients (age≥60 years), and younger patients showed less completion (Figure 3). In addition, we found that the linear term of age significantly moderated the magnitude of changes in completion rates over time (b=0.56, P=.02). Younger patients had steeper declines over time as compared to older patients (Figure 4). Other patient characteristics—gender and chronic pain diagnosis—were not significant predictors in the model. On the between-study level, study duration, sampling density, and survey length were not significantly related to completion rates. Similarly, pain intensity levels were not significantly related to day-to-day variation in within-patient, between-patient, or study-level differences in completion rates.

Figure 1. Flow diagram describing the identification of databases. EMA: ecological momentary assessment.

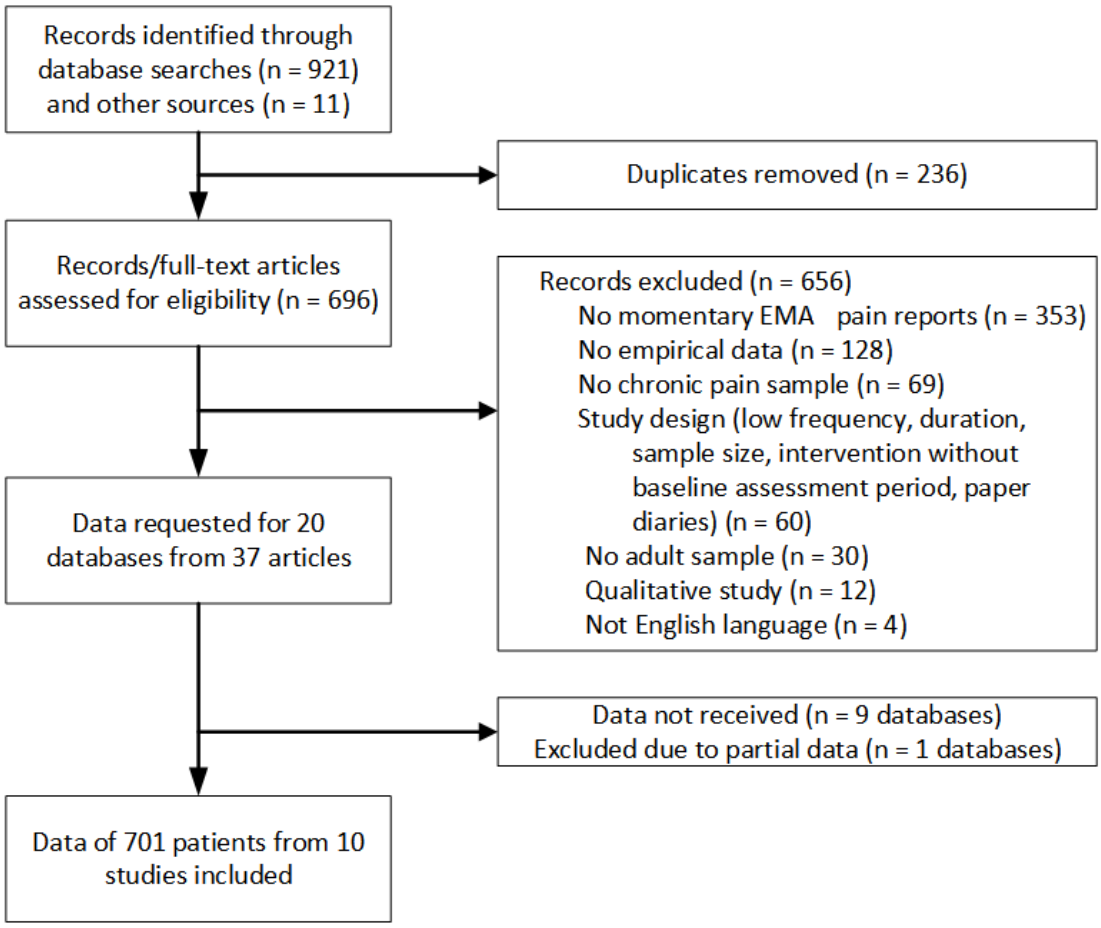


Table 1. Participant and study level descriptive characteristics.

Characteristics	Statistics	Reference
Participant level (n=701)		
Pain intensity, mean (SD), range	42.00 (20.54), 1.88-92.71	—[a]
Age (years), mean (SD), range	48.70 (13.08), 19-80	—
Female, n (%)	469 (67)	—
Diagnosis, n (%)		
Osteoarthritis	175 (25)	—
Rheumatoid arthritis	71 (10)	—
Fibromyalgia	73 (10)	—
Mixed or others	386 (55)	—
Study level (n=10), mean (SD), range		
Sample size	70.10 (22.75), 31-115	—
Study purpose, n (%)		
Within-person processes	5 (50)	[24-28]
Intervention to reduce pain	1 (10)	[29]
Methodological (eg, recall bias)	4 (40)	[30-33]
Study duration (days), mean (SD), range	12.30 (8.71), 4-28	—
4-7, n (%)	5 (50)	[24,28,29,31,33]
8-14, n (%)	3 (30)	[25,26,32]
15-28, n (%)	2 (20)	[27,30]
Sampling density (number of prompts per day)[b], n (%)	6.25 (2.60), 3-12	—
3-5	5 (42)	[19,24,27,29,32]
6-8	5 (42)	[25,26,30-32]
9-12	2 (17)	[32,33]
Items per prompt, n (%)	24.60 (18.93), 6-63	—
6-10	2 (20)	[24,31]
11-20	5 (50)	[25,29,30,32,33]
21-63	3 (30)	[26-28]

[a]Not applicable.

[b]Values for sampling density are based on n=12 datasets (one study contained three datasets with different numbers of prompts by design).

Figure 2. Distribution of patient-level average ecological momentary assessment completion rates.

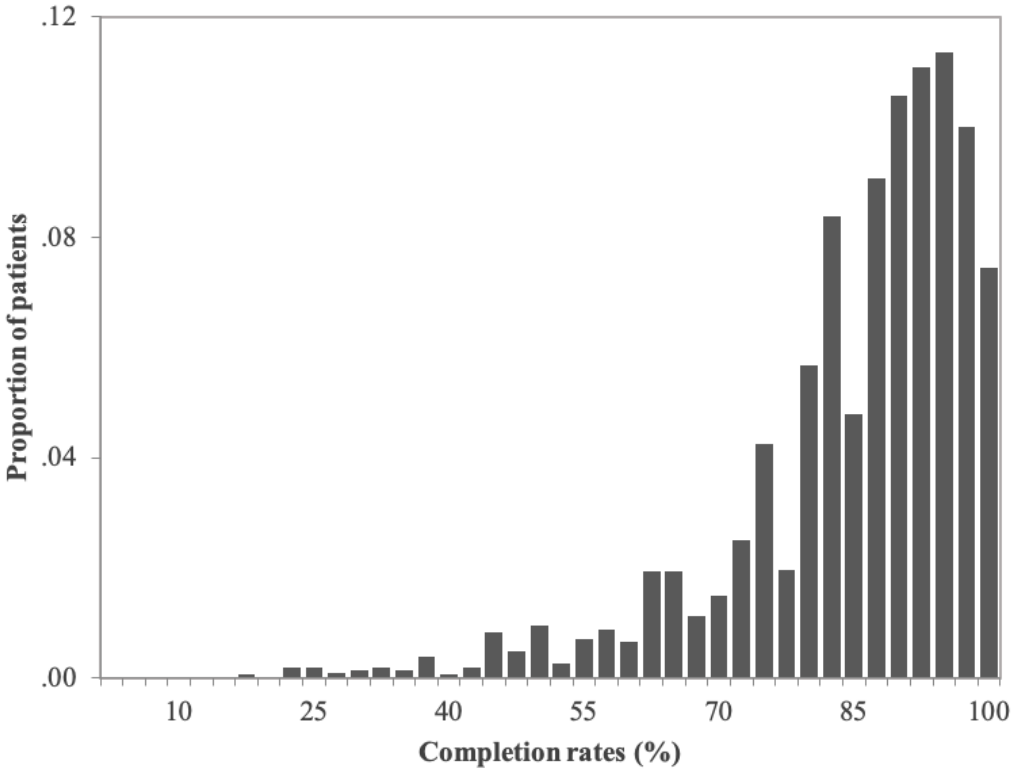


Table 2. Three-level multilevel model of predictors of ecological momentary assessment completion.

Predictors	Estimate	Standard error	*P* value
Fixed effects			
Intercept	88.61	16.26	<.001
Level 1 (day)			
Study days[a]	–2.29	0.32	<.001
Daily average pain[b]	–0.23	0.20	.25
Level 2 (patient)			
Age			
Linear	1.77	0.24	.002
Quadratic[c]	–0.62	0.24	.009
Female sex	–1.29	1.14	.26
Diagnosis			
Osteoarthritis	0.23	1.84	.90
Rheumatoid arthritis	0.99	2.17	.65
Fibromyalgia	0.64	1.76	.72
Age×study days	0.56	0.23	.02
Patient average pain[d]	0.03	0.24	.89
Level 3 (study)			
Duration (number of days)	0.26	0.36	.48
Density (number of prompts per day)	–0.16	1.06	.88
Lengths (number of items)	–0.17	0.16	.27
Study average pain	–0.57	3.78	.88
Random effects			
Level 1 (day)			
Within-person residual	257.12	4.47	<.001
Level 2 (patient)			
Intercept	105.61	9.20	<.001
Slope			
Study day	3.39	1.56	.03
Daily average pain	3.59	1.02	<.001
Level 3 (study)			
Intercept	58.46	25.09	.02
Parameters	23	—[e]	—
–2log likelihood	67872.34	—	—
AIC[f]	67918.35	—	—
BIC[g]	68078.85	—	—

[a]Study day was coded in weekly units.

[b]Daily pain was within-person centered.

[c]Age was centered at 50 years.

[d]Patient-level pain was within-study centered.

[e]Not applicable.

[f]AIC: Akaike Information Criterion.

[g]BIC: Bayesian Information Criterion.

Figure 3. Scatter plot of average daily completion rates by age. An overlaying line graph represents average completion rates by patient age groups. For example, the average of the first group of patients in their 20s is indicated at the age of 25 years.

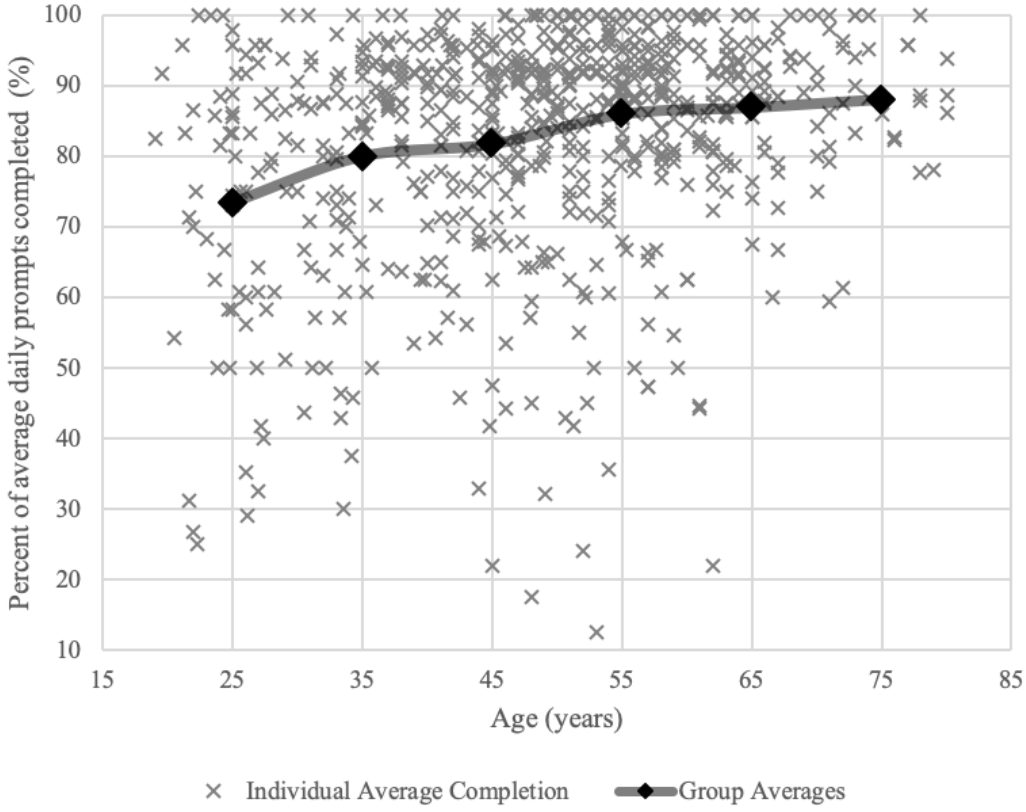
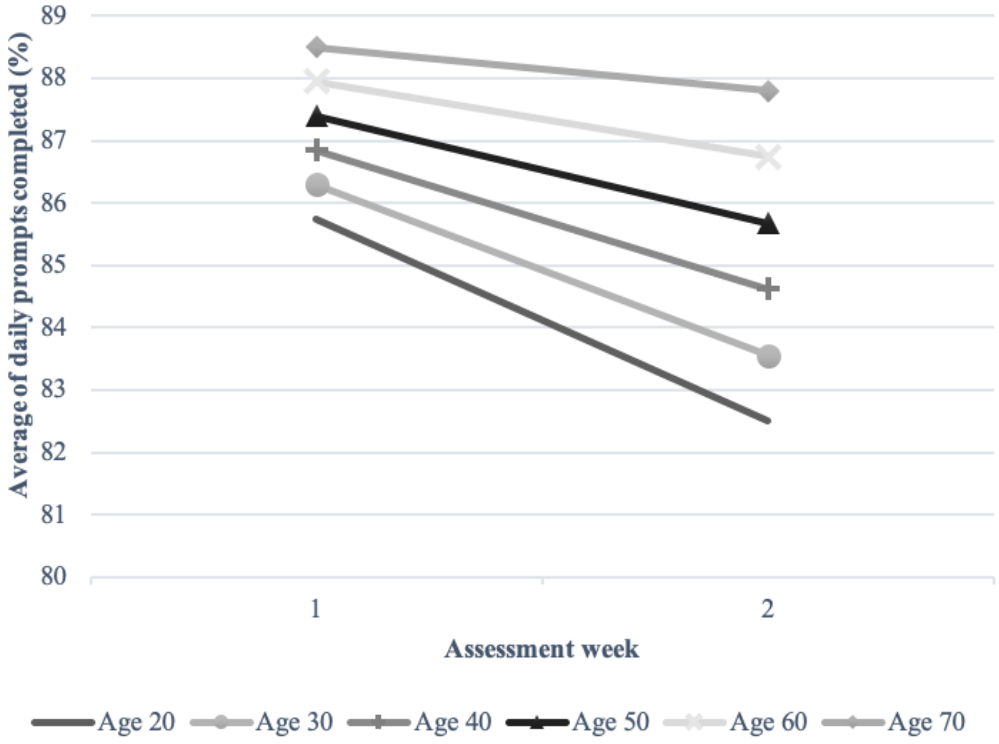


Figure 4. The cross-level interaction effect between age and study day.



Discussion

Principal Findings

The goal of this study was to determine whether completion rates in EMA studies on chronic pain differ systematically according to characteristics of the situation, person, or EMA protocol. Analyzing data of 701 patients from 10 studies, we found an average EMA-completion rate of 85%. Although this rate may seem high, this finding is consistent with that of other studies on chronic pain [12,16]. The results revealed lower

completion rates among younger patients than among older patients. In addition, we observed a decline in completion rates over the course of study protocols. The rate of this decline was moderated by respondent age: Younger patients showed a faster decline in completion rates than older patients.

Overview

The effect of age may be a result of fewer competing demands from elderly patients in their daily lives due to, for example, retirement, as compared to younger patients. In fact, evidence suggests that younger respondents are more prone to inattention and carelessness when completing online surveys than older respondents [34]. Similarly, the observed decline in completion rates over time may be due to survey fatigue or loss of motivation over the study days. Okifuji et al [11] recommend limiting the period of EMA assessments to 1 week, but this limit may not be desirable when, for example, evaluating responses to changes in treatment or adjustment to new medications using EMA. Research is needed to identify feasible ways to ensure sustained patient engagement in EMA assessments over time including the use of monetary or motivational incentives, close participant-researcher interactions over the course of the study, and the use of emerging data-collection strategies aimed at reducing participant burden, such as "microinteraction-based" EMA (a method developed to answer few EMA items very quickly) [35] or "measurement burst" designs (multiple brief EMA periods repeated over time) [36].

Our analyses did not support several *a priori* hypotheses. We did not detect any gender differences in the EMA-completion rates. In fact, previous findings of lower completion rates among men were primarily based on healthy samples [6-8], and studies on chronic pain did not identify such a pattern [10,11]. In addition, we found no evidence that either day-to-day variations in pain intensity or differences in pain intensity levels between participants or studies were systematically associated with EMA completion. This finding is especially important because an association between high pain levels and a low likelihood of responding to EMA surveys could severely undermine the validity of the EMA data collected to monitor patients' pain in everyday life. An important caveat of our analyses is that they were necessarily based on averages of pain from EMA prompts that were not missed by patients; considering that pain levels for missed EMA prompts are not known, the possibility that EMA prompts are more likely missed when patients are in higher (or lower) pain at the time of the prompt cannot be excluded. Our findings may have also resulted from the salience of pain experiences in the study samples. Participation in research wherein pain is of immediate relevance and intrinsic importance might have contributed to patients' motivation to complete the momentary assessments even in times of high pain intensity. This perspective aligns with our result of no differences in the EMA-completion rates according to the chronic pain diagnosis. The characteristics of the chronic pain experience can vary substantially between diagnoses; for example, high levels of fatigue and cognitive difficulties are more strongly associated with some diagnoses than with other pain diagnoses, and this might have contributed to differences in completion rates. However, the present results suggest that more complex symptomatology per se might not preclude patients' engagement in the momentary assessments.

At the between-study level, we did not detect any associations between design features related to participant burden (overall length of a given study, sampling density, and number of EMA survey items) and completion rates, which was surprising. Given the intensive nature of EMA protocols, participant burden has often been viewed as a major factor contributing to noncompliance [7,12]. One possibility is that other study design factors (eg, frequent contacts with participants to keep them motivated) are more important than load for continued engagement in EMA protocols, and these factors should be studied in future research. Additionally, despite the sizeable number of participants included in our analyses, IPD meta-analysis can be negatively influenced by a low statistical power at the highest (between-study) level of analysis, which may have limited our ability to detect effects based on study-level design features [37].

Limitations and Future Directions

A limitation of our study is the potential selectivity bias: We were able to include 10 of 20 eligible datasets. Inclusion of >90% of the eligible studies in IPD meta-analyses has been suggested as an ideal target [38], although, in practice, many IPD meta-analyses include <80% of the eligible datasets [39]. To evaluate the potential for selection bias in the data available for the present analyses, we examined the pooled average completion rate reported for eligible studies that were not included in the analyses; of 10 studies, 7 provided average completion rates in the published reports. The weighted average completion rate in these studies was 78.2%, suggesting a potential upward bias of completion rates in the data that were available for our analyses.

When calculating completion rates, we relied on a fixed number of EMA prompts, unless studies employed a variable prompting schedule based on patients' waking hours. As such, we assumed that each participant in those studies consistently received the same number of prompts. However, this assumption may have sometimes been violated due to the potential of malfunctioning of data-collection devices or limitations in their configuration capability (ie, prompting during sleep). Thus, some of the calculated daily completion rates may have underestimated participants' actual completion rates.

Our study is also limited by the number and types of predictor variables that were consistently available across the different datasets. Data on additional predictors such as negative affect, disability status, and stress levels were not consistently available but are undoubtedly candidates for understanding EMA completion in chronic pain. Similarly, EMA data-capturing methods differ among many intricate dimensions that could be theoretically important predictors of EMA completion rates, including whether participants should be allowed to delay the assessment (ability to "snooze" or "suspend" an assessment), how fast a participant is expected to start an assessment (ie, the time window during which an assessment stays open for completion), the frequency of contact between the research team and the participant, the availability of reminders to complete the assessment (and the type and frequency of reminders), and

the expectation of individualized feedback at the end of the study. Finally, our findings may not be generalizable to patients with other illnesses or healthy populations.

Conclusions

In summary, our IPD meta-analysis showed no evidence to suggest that EMA-completion rates in chronic pain differ by medical diagnoses; gender; EMA study design features related to participant burden; or variations in pain levels across days, patients, or studies. These findings support the use of EMA data-collection methods for careful assessment of patients' pain and other experiences. Future EMA research in chronic pain should note that study length and young age can affect the quality of the momentary data and devise strategies to maximize EMA-completion rates across different age groups and study days.

Acknowledgments

We would like to thank the investigators who shared their EMA datasets with us for the purpose of this study: Joan E Broderick, PhD; Stephen Bruehl, PhD; Geert Crombez, PhD; Ivan Huijnen, PhD; Hanne Kindermans, PhD; Madelon Peters, PhD; Mark Litt, PhD; Dylan Smith, PhD; and Joshua M Smyth, PhD. This work was supported by a grant from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01 AR066200).

Conflicts of Interest

AAS is a Senior Scientist with the Gallup Organization and a consultant with IQVIA and Adelphi Values, Inc. The remaining authors have no conflict of interest to declare.

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
2. Hamaker EL, Wichers M. No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. Curr Dir Psychol Sci 2017 Feb 08;26(1):10-15. [doi: 10.1177/0963721416666518]
3. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
4. Courvoisier DS, Eid M, Lischetzke T. Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. Psychol Assess 2012;24(3):713-720. [doi: 10.1037/a0026733] [Medline: 22250597]
5. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychol Methods 2002;7(2):147-177. [Medline: 12090408]
6. Messiah A, Grondin O, Encrenaz G. Factors associated with missing data in an experience sampling investigation of substance use determinants. Drug Alcohol Depend 2011;114(2-3):153-158. [doi: 10.1016/j.drugalcdep.2010.09.016] [Medline: 21075563]
7. Sokolovsky AW, Mermelstein RJ, Hedeker D. Factors predicting compliance to ecological momentary assessment among adolescent smokers. Nicotine Tob Res 2014 Mar;16(3):351-358 [FREE Full text] [doi: 10.1093/ntr/ntt154] [Medline: 24097816]
8. Silvia PJ, Kwapil TR, Eddington KM, Brown LH. Missed beeps and missing data: Dispositional and situational predictors of nonresponse in experience sampling research. Soc Sci Comput Rev 2013;31(4):471-481. [doi: 10.1177/0894439313479902]
9. McLean DC, Nakamura J, Csikszentmihalyi M. Explaining System Missing. Social Psychological and Personality Science 2017 May 16;8(4):434-441. [doi: 10.1177/1948550617708015]
10. Aaron LA, Mancl L, Turner JA, Sawchuk CN, Klein KM. Reasons for missing interviews in the daily electronic assessment of pain, mood, and stress. Pain 2004;109(3):389-398. [doi: 10.1016/j.pain.2004.02.014] [Medline: 15157700]
11. Okifuji A, Bradshaw DH, Donaldson GW, Turk DC. Sequential analyses of daily symptoms in women with fibromyalgia syndrome. J Pain 2011;12(1):84-93 [FREE Full text] [doi: 10.1016/j.jpain.2010.05.003] [Medline: 20591745]
12. Morren M, van Dulmen S, Ouwerkerk J, Bensing J. Compliance with momentary pain measurement using electronic diaries: A systematic review. Eur J Pain 2009;13(4):354-365. [doi: 10.1016/j.ejpain.2008.05.010] [Medline: 18603458]
13. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 2010 Feb 05;340:c221. [doi: 10.1136/bmj.c221] [Medline: 20139215]
14. Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. Psychol Methods 2009;14(2):81-100 [FREE Full text] [doi: 10.1037/a0015914] [Medline: 19485623]
15. Hussong AM, Curran PJ, Bauer DJ. Integrative data analysis in clinical psychology research. Annu Rev Clin Psychol 2013;9:61-89 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185522] [Medline: 23394226]
16. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological Momentary Assessment Methodology in Chronic Pain Research: A Systematic Review. J Pain 2018 Jul;19(7):699-716. [doi: 10.1016/j.jpain.2018.01.006] [Medline: 29371113]
17. Broderick JE, Schwartz JE, Shiffman S, Hufford MR, Stone AA. Signaling does not adequately improve diary compliance. Ann Behav Med 2003;26(2):139-148. [doi: 10.1207/S15324796ABM2602_06] [Medline: 14534031]

18. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. BMJ 2002;324(7347):1193-1194 [FREE Full text] [Medline: 12016186]
19. Heron KE, Smyth JM. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. Brit J Health Psych 2010;15:1-39. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]
20. Warton DI, Hui FKC. The arcsine is asinine: The analysis of proportions in ecology. Ecology 2011;92(1):3-10. [Medline: 21560670]
21. Muthén LK, Muthén BO. Mplus user's guide, 8th edition. Los Angeles, CA: Muthén & Muthén; 1998.
22. Muthén LK, Muthén BO. How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. Structural Equation Modeling: A Multidisciplinary Journal 2002 Oct;9(4):599-620. [doi: 10.1207/S15328007SEM0904_8]
23. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
24. Bruehl S, Liu X, Burns JW, Chont M, Jamison RN. Associations between daily chronic pain intensity, daily anger expression, and trait anger expressiveness: an ecological momentary assessment study. Pain 2012 Dec;153(12):2352-2358 [FREE Full text] [doi: 10.1016/j.pain.2012.08.001] [Medline: 22940462]
25. Viane I, Crombez G, Eccleston C, Devulder J, De Corte W. Acceptance of the unpleasant reality of chronic pain: effects upon attention to pain and engagement with daily activities. Pain 2004 Dec;112(3):282-288. [doi: 10.1016/j.pain.2004.09.008] [Medline: 15561383]
26. Huijnen IPJ, Verbunt JA, Roelofs J, Goossens M, Peters M. The disabling role of fluctuations in physical activity in patients with chronic low back pain. Eur J Pain 2009;13(10):1076-1079. [doi: 10.1016/j.ejpain.2008.12.008] [Medline: 19181547]
27. Peters ML, Sorbi MJ, Kruise DA, Kerssens JJ, Verhaak PF, Bensing JM. Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. Pain 2000;84(2-3):181-192. [Medline: 10666523]
28. Smyth JM, Zawadzki MJ, Santuzzi AM, Filipkowski KB. Examining the effects of perceived social support on momentary mood and symptom reports in asthma and arthritis patients. Psychol Health 2014;29(7):813-831. [doi: 10.1080/08870446.2014.889139] [Medline: 24568534]
29. Litt MD, Shafer DM, Ibanez CR, Kreutzer DL, Tawfik-Yonkers Z. Momentary pain and coping in temporomandibular disorder pain: exploring mechanisms of cognitive behavioral treatment for chronic pain. Pain 2009 Sep;145(1-2):160-168 [FREE Full text] [doi: 10.1016/j.pain.2009.06.003] [Medline: 19553018]
30. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. Pain 2008;139(1):146-157 [FREE Full text] [doi: 10.1016/j.pain.2008.03.024] [Medline: 18455312]
31. Smith DM, Parmelee PA. Within-day variability of fatigue and pain among African Americans and non-Hispanic Whites with osteoarthritis of the knee. Arthritis Care Res 2016;68(1):115-122 [FREE Full text] [doi: 10.1002/acr.22690] [Medline: 26315851]
32. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. Pain 2003;104(1-2):343-351. [Medline: 12855344]
33. Stone AA, Broderick JE, Schneider S, Schwartz JE. Expanding options for developing outcome measures from momentary assessment data. Psychosom Med 2012;74(4):387-397. [doi: 10.1097/PSY.0b013e3182571faa] [Medline: 22582336]
34. Schneider S, May M, Stone AA. Careless responding in internet-based quality of life assessments. Qual Life Res 2017:1-12.
35. Intille S, Haynes C, Maniar D, Ponnada A, Manjourides J. µEMA: Microinteraction-based Ecological Momentary Assessment (EMA) Using a Smartwatch. Proc ACM Int Conf Ubiquitous Comput 2016 Sep;2016:1124-1128 [FREE Full text] [doi: 10.1145/2971648.2971717] [Medline: 30238088]
36. Sliwinski M. Measurement-Burst Designs for Social Health Research. Social Pers Psych Compass 2008 Jan;2(1):245-261. [doi: 10.1111/j.1751-9004.2007.00043.x]
37. Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. Psychol Methods 2004;9(4):426-445. [doi: 10.1037/1082-989X.9.4.426] [Medline: 15598097]
38. Sterne JAC, Egger M, Moher D, editors. Addressing reporting biases. In: Cochrane Handbook for Systematic Reviews of Interventions. Chichester: Wiley & Sons Ltd; 2008:297-333.
39. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. BMJ 2012 Jan 03;344:d7762. [doi: 10.1136/bmj.d7762] [Medline: 22214758]

Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
EMA: ecological momentary assessment
IPD: individual patient data
MAR: missing at random

Edited by G Eysenbach; submitted 25.06.18; peer-reviewed by D Gromala, J Priebe, G Lewis; comments to author 13.09.18; revised version received 21.10.18; accepted 26.10.18; published 05.02.19.

Please cite as:
Ono M, Schneider S, Junghaenel DU, Stone AA
What Affects the Completion of Ecological Momentary Assessments in Chronic Pain Research? An Individual Patient Data Meta-Analysis
J Med Internet Res 2019;21(2):e11398
URL: https://www.jmir.org/2019/2/e11398/
doi:10.2196/11398
PMID:30720437

Original Paper

Online Information-Seeking About Potential Breast Cancer Symptoms: Capturing Online Behavior With an Internet Browsing Tracking Tool

Afrodita Marcu[1], PhD; Cecile Muller[2], PhD; Emma Ream[1], PhD; Katriina L Whitaker[1], PhD

[1]School of Health Sciences, Faculty of Health & Medical Sciences, University of Surrey, Guildford, United Kingdom

[2]School of Psychology, Faculty of Health & Medical Sciences, University of Surrey, Guildford, United Kingdom

Corresponding Author:
Afrodita Marcu, PhD
School of Health Sciences
Faculty of Health & Medical Sciences
University of Surrey
Duke of Kent Building
Guildford, GU2 7XH
United Kingdom
Phone: 44 0148 36 82515
Email: afrodita.marcu@surrey.ac.uk

Abstract

Background: People engage in health information-seeking online when experiencing unusual or unfamiliar bodily changes. It is not well understood how people consult the internet for health information after the onset of unfamiliar symptoms and before receiving a potential diagnosis and how online information-seeking can help people appraise their symptoms. This lack of evidence may be partly due to methodological limitations in capturing in real time the online information-seeking process.

Objective: We explored women's symptom attribution and online health information-seeking in response to a hypothetical and unfamiliar breast change suggestive of cancer (nipple rash). We also aimed to establish the feasibility of capturing in real time the online information-seeking process with a tool designed to track participant online searches and visited websites, the Vizzata browser tracker.

Methods: An online survey was completed by 56 cancer-free women (mean age 60.34 [SD 7.73] years) responding to a scenario asking them to imagine noticing a red scaly rash on the nipple. Participants were asked to make symptom attributions when presented with the scenario (T1) and again after seeking information online (T2). The online tracking tool, embedded in the survey, was used to capture in real time participant search terms and accessed websites.

Results: The tracking tool captured the search terms and accessed websites of most of the participants (46/56, 82%). For the rest (10/56, 18%), there was evidence of engagement in online information-seeking (eg, medical terminology and cancer attribution at T2) despite their searching activity not being recorded. A total of 25 participants considered cancer as a potential cause for the nipple rash at T1, yet only one of these used cancer as a search term. Most participants (40/46, 87%) used rash-related search terms, particularly nipple rash and rash on nipple. The majority (41/46, 89%) accessed websites containing breast cancer information, with the National Health Service webpage "Paget disease of the nipple" being the most visited one. At T2, after engaging in the internet search task, more participants attributed the nipple rash to breast cancer than at T1 (37/46, 66% vs 25/46, 45%), although a small number of participants (6/46) changed from making a cancer attribution at T1 to a noncancer one at T2.

Conclusions: Making a cancer attribution for an unfamiliar breast change did not necessarily translate into cancer-termed searches. Equally, not all internet searches led to a cancer attribution. The findings suggest that online information-seeking may not necessarily help women who experience unfamiliar breast cancer symptoms understand their condition. Despite some technical issues, this study showed that it is feasible to use an online browser tracking tool to capture in real time information-seeking about unfamiliar symptoms.

(J Med Internet Res 2019;21(2):e12400) doi:10.2196/12400

KEYWORDS

breast cancer; health information; internet search; online information seeking

Introduction

Seeking health information online is a ubiquitous activity, enabled by advances in Web 2.0 design (which includes user-generated content), developments in search engines, proliferation of social media, and wide ownership of mobile phones, tablets, and computers [1]. The Google search engine is commonly used in online health searches [2]. Of the world population, 54.5% has access to the internet, with Europe and North America having the highest internet access rates, 85.2% and 95%, respectively [3]. In the United Kingdom, the context of this research, 94.8% of the population is connected to the internet [3], which means that the vast majority of people in the United Kingdom have the opportunity to consult the internet for health information. Indeed, many do so: the website of the UK National Health Service Choices, which offers information on symptoms, causes, and treatment for most common diseases, reportedly receives about 15 million visits per month [4].

People search for health information online to self-diagnose [5] or complement information received from the family doctor [6]. Online searching for health-related information often arises on symptom manifestation and people's appraisal of them, known as the symptom appraisal interval (ie, the interval when people notice bodily changes, interpret them as symptoms of illness, and decide whether they warrant medical attention) [7,8]. This can be important for conditions where timeliness of presentation is an important contributor to treatment outcome. Cancer is one exemplar; prompt help-seeking for symptoms suggestive of cancer can be key to detecting cancer at an early stage and ensuring it is potentially curable [9,10]. Arguably, it is important for members of the public to have access to accurate health information online that is able to aid symptom appraisal and prompt early and appropriate help-seeking. However, this potential rests on a number of tacit assumptions which may not necessarily be met: first, that lay people experiencing bodily changes can articulate these into symptoms and into effective search terms; second, that lay people have the ability to locate online the most relevant and reliable sources of health information; and third, that, upon finding the relevant information, people can interpret it appropriately and apply it to themselves. One way to observe these processes and their implications for symptom appraisal and early symptomatic presentation is to use tools that record online information-seeking behavior in relation to a given health condition [11-13]. These tools have enriched recent research in the field of consumer online health information-seeking.

Breast cancer is one of the most frequently searched for cancer topics online [14-16], as evidenced by data from English-speaking countries including the United States, Canada, Australia and the United Kingdom [17]. Desire for online information about breast cancer appears to be affected by media coverage [18] and Breast Cancer Awareness months [19,20]. However, the high search volume may also reflect its relatively high incidence: in the United Kingdom, female breast cancer is the most common cancer among women, with around 55,000 cases diagnosed each year [21], while globally over 2 million women are diagnosed annually [22]. Indeed, it has been noted that the volume of cancer-related online searches reflects estimated cancer incidence and mortality within a given country [17,18,23], which would suggest that people are likely to seek information on breast cancer when they or someone they know receive a cancer diagnosis. However, online information sources on breast cancer can vary in their quality and completeness [24,25], and little is known about how women interpret or act on information found online in the event of breast-related symptoms.

Our previous studies on how symptomatic women make sense of symptoms indicative of breast cancer and intend to seek medical help suggested that some women do consult the internet for information about breast-related symptoms but are rather ambivalent about its value and lacking in confidence about how to appraise the information gathered [26,27]. While we have researched, in an online survey, how women attribute hypothetical breast changes to breast cancer and express preferences to seek help [28,29], we have not explored how women's symptom attribution might translate into online information-seeking strategies or how these actions might change as a result of online searches. Therefore, in this study we extended the work we conducted in the context of breast cancer and used a browser tracking tool, Vizzata (Vizzata Limited) [30], to explore how women seek information online when presented with a symptom scenario consistent with a breast cancer symptom. We examined whether information-seeking enabled them to interpret the symptom accurately as potentially relating to breast cancer.

Another aim of our study was to determine the efficacy of a browser tracking tool to remotely capture online information-seeking behavior within a more traditional survey setting. Within the context of this study, efficacy represented the ability of the tool to record information on participant search terms and websites visited, the latter recorded as Web addresses or URLs. There are arguably two aspects to efficacy here: first, the technical ability of the tool to accurately record participant online behavior during the search task (the search terms entered and the addresses of the opened Web links). Second, the validity of the remote tool and of real-time behavior-tracking tools in general to capture behavior that can provide meaningful insights into participant reasoning processes during information-seeking.

Previous research that has captured participant online information-seeking in response to health scenarios has used, for example, screen-capture video software (eg, Camtasia, TechSmith Corporation) but often in conjunction with think-aloud tasks that documented participant reasoning during the search process [12,13]. Other studies have employed individual interviews followed by a video-recorded online search task [11,31] or face-to-face observation of the use of interactive information menus where information-seeking was operationalized as the number of links accessed [32]. Overall, these studies where interviews or think-aloud tasks accompanied

or supplemented the online search have usually been conducted in face-to-face individual sessions in conditions mirroring lab settings (eg, private conference rooms, libraries, university offices) [12,13,31]. While these studies enable rich, in-depth data capture, they lack naturalism and may not necessarily mirror how people engage in online information-seeking in the context of their everyday settings (eg, at home or workplace, on a tablet or laptop). Therefore, to increase ecological validity, we used a tracking software tool that would capture remotely participant online search activities without the need for the researchers' presence. Our approach is in line with more recent methods of capturing online information-seeking behavior, which, for example, employ internet browser extensions to log search terms and accessed website addresses [33]. In summary, our main objectives were:

1. To test the efficacy of the Vizzata browser tracking tool to capture participant online information-seeking in response to a scenario describing a hypothetical breast change (ie, a nipple rash)
2. To explore symptom attribution before and after engaging in the online search task and describe the process of online information-seeking for the hypothetical nipple rash: the search terms used and the websites accessed

Methods

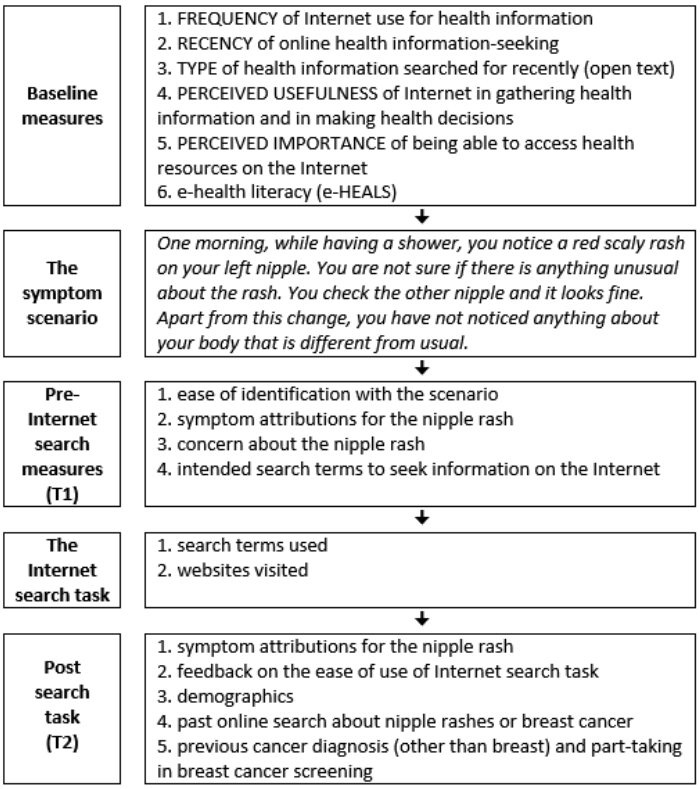
Study Design

The symptom scenario depicted a nipple rash—a lesser known symptom of breast cancer [34]—and was presented to participants prior to asking them to engage in the internet search task (see Figure 1).

To examine how women sought information online in relation to a hypothetical breast change indicative of breast cancer, we used the online survey platform Vizzata [30]. The browser tracking tool is an add-on feature of the Vizzata survey software and was designed to track and record search terms, URLs, and time spent on websites.

In the browser tracking section of our study, we included DuckDuckGo (duckduckgo.com) as the default search engine and instructed participants to use it within the survey internet window when performing the internet search task. DuckDuckGo looks similar to the Google search engine interface and performs similar functions and was selected because it was compatible with the browsing data capture function of the tracking tool. The key feature of DuckDuckGo is that it does not track users' online activity. This offers a potential advantage to studies on online information-seeking as the search results during the study are less likely to be influenced by users' past search history or other previous online activity.

Figure 1. Study protocol.



Recruitment Strategy

The study received ethical approval from the University of Surrey Ethics Committee (reference: UEC/2016/041/FHMS). A total of 125 British women were recruited via a market research company from among their online panel members between December 2016 and January 2017. The inclusion criteria were female gender, based in the United Kingdom, aged 50 years or older, with education levels ranging from no formal education to university degree or higher. The exclusion criterion was past or current breast cancer diagnosis, with the questions on breast cancer diagnosis being embedded among other health conditions (eg, diabetes) so that participants would not be aware of the focus of the study.

Procedure

Participants were invited to the study by email invitation generated from the Vizzata platform, which informed them that the study explored how people search for health information on the internet. Figure 1 presents the study protocol followed by participants. Participants were asked to answer a few baseline questions about their use of the internet for health information before reading the symptom scenario. Then they were instructed to use the search engine to find out more about the hypothetical bodily change and to visit as many sites as they wished and then return to the final part of the survey (see Multimedia Appendix 1). A disclaimer at the start of the search task informed participants that their internet activity would be recorded only for the duration of the study and the study tracking tool would not be able to access any personal data stored in their browsers (eg, passwords).

Materials and Measures

Baseline Measures

At the start of the survey (Time 1, henceforth T1), participants completed measures of frequency of internet use for health information (1 = never, 4 = very often), perceptions of the importance (1 = not important at all, 4 = very important) of the internet to provide health information, usefulness (1 = not useful at all, 4 = very useful) of the internet to provide health information and to help make decisions about one's health, and recency of online health information-seeking (7 = this week, 6 = last week, 5 = this month, 4 = last month, 3 = a few months ago, 2 = last year, 1 = can't remember). Participants also indicated through open text what health information they had recently searched for, which was subsequently coded as 1 if participants mentioned cancer, breast cancer, nipple, or nipple rash, 2 if participants did not mention these terms, or 3 if they couldn't remember. We assessed eHealth literacy using the eHEALS Literacy Scale developed by Norman and Skinner [35] (Cronbach alpha = .93), who have defined eHealth literacy as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [36].

The Symptom Scenario

We included a scenario describing the appearance of a nipple rash, a largely unfamiliar symptom of breast cancer [34]. We had developed and tested the scenario with women from a range of educational backgrounds during individual cognitive think-aloud interviews (n=10) and 3 focus groups (n=19) in our past research on symptom appraisal and help-seeking intentions for breast changes among cancer-free women [28]. A subsequent online survey with 961 British women (aged 47 to 92 years) confirmed that nipple rash was a significantly less familiar symptom of breast cancer than an armpit lump [28]. We chose a relatively unfamiliar symptom to increase the variability in how women interpret and seek information for the symptom.

Symptom Attribution Before the Internet Search

After reading the scenario, participants completed measures assessing ease of identification with the scenario (1 = very difficult, 4 = very easy), what they thought the symptom might be (in open text), and concern about the symptom (1 = not at all, 5 = extremely). Given that search engines have predictive text (autocompletion) that can influence people's formulation or choice of search terms, we also asked participants, prior to the start of the internet search task (T1), to indicate what search terms they intended to use in relation to the scenario.

The Internet Search Task

After reading the scenario, participants were required to engage in the online information-seeking task. Participants were instructed to search for information as long as they wished or until they were satisfied with the information found, and once they had finished searching they should click on the GO BACK TO SURVEY button to return to the survey and answer further questions. The data gathered during the information-seeking task included the search terms entered in the search engine and the websites visited by participants, which were saved in an Excel (Microsoft Corp) file by the Vizzata software. The viewed content was later coded yes or no as comprising breast cancer information.

Post Search Task

Next, participants were asked to make new symptom attributions as at T1. They also left feedback on the difficulty of the search task (1 = difficult, 2 = neither easy nor difficult, or 3 = easy) and indicated whether they had searched online for information on nipple rashes or breast cancer in the past. At the end of the survey, participants completed various demographics measures (ie, age, ethnicity, highest education level, marital status, and employment) and indicated past participation in breast cancer screening (yes/no/can't remember/not applicable). The participants also indicated whether they had any family/friend history of breast cancer (immediate family member/other family member/close friend); these options were not mutually exclusive, and participants could check as many as applied.

Results

Sample Characteristics

In total, 56 of the 125 initial panel members who were eligible and invited to the study completed the survey. Figure 2 presents a detailed description of the participants who were included and excluded from the study.

Figure 2. Participant recruitment and data collection flow diagram.

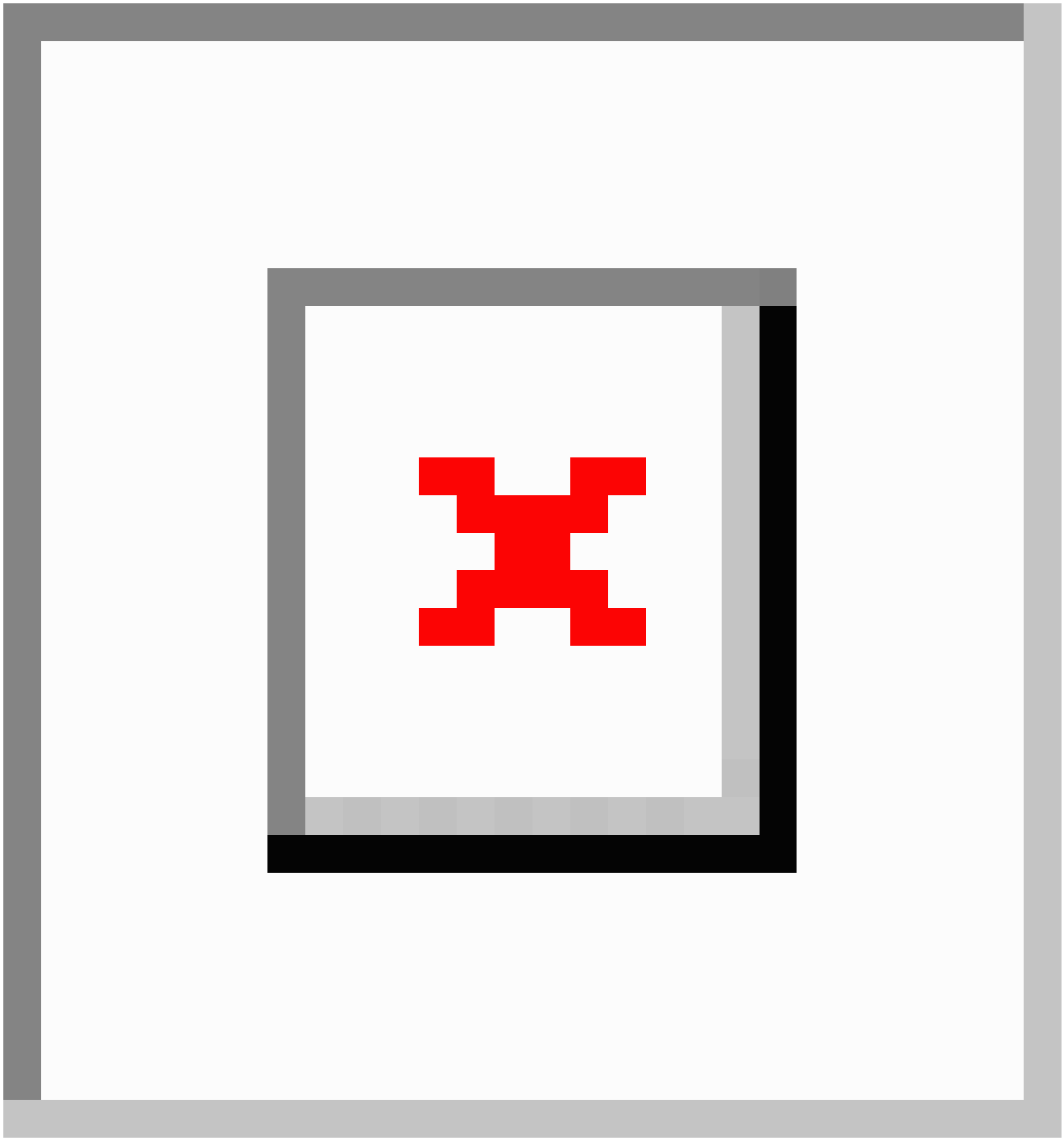


Table 1. Participant demographic characteristics (N=56).

Characteristics	Value, n (%)
Highest education level	
No formal qualifications	12 (21)
Education below degree level	20 (36)
Education at degree level or above	24 (43)
Relationship status	
Married/living with partner/in civil partnership	35 (63)
Single/never married/divorced/separated/widowed	21 (38)
Employment status	
Retired	26 (46)
Employed part-time/full-time/self-employed	21 (38)
Unemployed/homemaker/not working because of disability	9 (16)
Ethnicity	
White British	54 (96)
Other	2 (4)
Cancer screening participation	
Breast cancer (eligible sample aged 50-70 years, n=49)	45 (92)
Cervical cancer	51 (91)
Bowel cancer (eligible sample aged 60-74 years, n=25)	16 (64)
Family/friend history of breast cancer	
None	18 (32)
Immediate family member (parent/sibling/child)	4 (7)
Other family member	14 (25)
Close friend	31 (55)

The majority of the women who took part in this study were White British (54/56, 96%), retired (26/56, 46%), in a relationship (35/56, 63%), and aged on average 60 years (mean 60.34 [SD 7.73] years; age range: 50-78 years). Most participants had participated in breast cancer screening (45/56, 92%), and most (38/56, 68%) had a family/friend history of breast cancer (immediate family, close friend, or a nonimmediate family member with a diagnosis of breast cancer). Summarized demographic details are presented in Table 1.

eHealth Literacy and Recent Health Information-Seeking Online

Participants had relatively high levels of eHealth literacy (eHEALS) with regard to using the internet for health information (range: 2.13 to 5.00, median 3.88, mean 3.80, SD 0.67). The majority of the participants valued the internet as a useful source of health information (see Table 2). The majority (33/56, 59%) had searched for health information relatively recently, from last month to as recently as this week, yet none of their open-ended answers contained breast cancer or nipple rash as the focus of their recent online searches. This enabled us to assume that there would be no breast cancer search recency effects on the participants' present searches. At the end of the study, some participants indicated they had searched in the past, although not recently, for breast cancer information online (20/56, 36%) and/or for nipple rash information (3/56, 5%).

Efficacy of the Online Browser Tracking Tool

We assessed the efficacy of the tool (ie, its ability to capture online information-seeking behavior) by examining the number of participants who were able to complete the internet search task and the participant feedback on the browser tracking tool as an embedded element of the survey.

Table 2. Participant baseline use of the internet for health information (N=56).

Characteristics	Value, n (%)
Frequency of internet use to find information about health	
Occasionally	45 (80)
Often or very often	10 (18)
Never	1 (2)
The last time internet was used for health information	
This week/last week/this month/last month	33 (59)
A few months ago/last year	21 (38)
Can't remember	2 (4)
Past online searches for nipple rash	
Yes	3 (5)
No	52 (93)
Can't remember	1 (2)
Past online searches for breast cancer	
Yes	20 (36)
No	36 (64)
Internet useful to gather health information	
Very useful	14 (25)
Useful	41 (73)
Not useful	1 (2)
Internet useful to help make decisions about health	
Very useful	5 (9)
Useful	47 (84)
Not useful	4 (7)
Important to access health resources on the internet	
Very important	17 (30)
Important	31 (55)
Not important	8 (14)

Data Captured and Missing Data

As described in the flow diagram (Figure 2), 83% (64/77) of responders completed the survey. For 28% (18/64) of those, no search terms or visited websites (URLs) were saved during the internet search task. We inspected the incomplete data of these respondents to understand reasons for noncompletion and assess whether participants had engaged in information-seeking without this being captured; we compared the symptom attribution before (T1) and after (T2) the internet search task and examined their feedback on the internet browser tracking element. Eight of these 18 participants did not complete the internet search or had connectivity problems, as indicated by their comments (eg, "none of the sites would open" and "I was putting it into the correct space but nothing was happening?"). These 8 participants did not provide any information after the internet search task that would have indicated successful engagement in information-seeking, therefore we excluded them from further analyses.

However, 10 participants of the 18 whose data on search terms and URLs were missing were retained for further analyses because they provided feedback on the information-seeking task and sufficiently meaningful data for comparison of symptom attribution between T1 and T2, which indicated they had indeed engaged in online information-seeking despite this not being captured. We included these participants in the final sample because the focus of our study was not solely on what websites the participants visited but also on how the process of searching for health information on the internet had influenced symptom appraisal and symptom attribution at T2. One participant, for example, whose search terms or URLs were not recorded, reported finding the information on the National Health Service (NHS) site useful, particularly about Paget disease, and made symptom attributions at T2 that included cancer (in contrast to T1): "Eczema. Insect bite. Psoriasis. Paget disease – breast cancer. Mastitis."

The missing data on search terms and URLs and the feedback of these 10 participants whose data were incomplete yet

meaningful suggest that some may have misunderstood the task instructions (eg, installed the DuckDuckGo search engine and conducted searches in separate windows away from that of the survey), as one participant's feedback indicates:

> *Downloading search engine a bit tricky. Search itself easy.*

Other participants may have had problems with the internet connection:

> *...kept losing connection to search engine.*

In other cases the participants may have used internet browsers other than the recommended Google Chrome, which may have been incompatible with the Vizzata tracking tool, or may have used devices other than PC/laptop, which may have reduced the survey usability:

> *I struggled to find the "go back to survey" button as I'm using my phone.*

Feedback on the Search Task

Overall, participants left positive feedback on the search task. When returning to the survey, the participants left comments as to whether the internet search task had been easy to do. We coded the open-ended responses of the participants into easy (47/56, 84%), difficult (2/56, 4%), and neither easy nor difficult (7/56, 13%). The feedback suggests that the majority of the participants did not encounter significant problems with the DuckDuckGo search engine or with the Vizzata tracking tool and that any missing data pertaining to the search terms or the accessed websites may have been due to reasons other than participant-related factors. Some participants encountered but overcame technical issues, such as slow internet connection:

> *Yes [easy], just a bit slow.*

> *...sometimes the search engine said oops problem in searching but on the whole it came up with lots of different sites to use for more research.*

Symptom Attribution

Symptom Attribution and Related Measures at T1

Most participants who completed the survey (48/56, 86%) found the scenario fairly or very easy to imagine themselves in, and most (34/56, 61%) indicated that they would be only a little bit or moderately concerned about the nipple rash.

At both T1 and T2, we coded the symptom attributions as "cancer," "environmental," "physical," or "don't know" in line with our previous research [28]. We coded as missing attribution those responses that contained neither a symptom attribution nor a "don't know" response (eg, "a symptom worth further investigation"). If a participant made multiple attributions in their responses (eg, "an ordinary rash, an allergic reaction, cancer"), we coded each attribution separately. The majority of participants (30/56, 54%) made a single symptom attribution, and just under half (25/56, 45%) made a cancer attribution. Attributing the symptom to physical causes (eg, eczema) was common at T1 (32/56, 57%).

Intended Search Terms

We coded the intended search terms as 1 = breast or skin cancer, 2 = nipple rash or rash on nipples, 3 = rash on skin/breast, 4 = other skin-related conditions, and 5 = nipple or breast changes. One "not sure" response was coded as missing data. The most frequently intended search terms were rash-related phrases (nipple rash, nipple red rash, red scaly rash on nipple), mentioned by 86% of the participants (48/56). Only 3 participants intended to use breast cancer and 1 participant, skin cancer on the breast (all had made a cancer attribution at T1).

Search Terms Used

The actual search terms were coded as follows: 1 = breast or skin cancer, 2 = nipple rash or rash on nipples, 3 = rash on breast, 4 = nipple or breast change, and 5 = other. Only 2 participants (2/46, 4%) included cancer in their search terms. Of the participants who had made a cancer attribution at T1, 4% (1/25) used cancer in their search terms. The majority (40/46, 87%) used rash-related search terms, particularly nipple rash and rash on nipple (see Table 3 for a list of all the search terms used). It is noteworthy that while the scenario specified "a red scaly rash on your left nipple," some participants (3/56, 5%) added sensory terms (eg, itchy) which were not in the original scenario; this could have influenced the returned results and thus their interpretation of the symptom.

A very small number of participants (5/46, 11%) changed their search terms during the internet search task, which reflects their interpretation of the information found and their reappraisal of the symptom in light of the information. For example, participant #57 started with "changes in nipples," changed to "breast cancer symptoms," and then more specifically "breast cancer symptoms NHS." Participant #50 started with "red scaly rash on nipple" and later changed to "What are the chances of red scaly rash on nipple being cancer?" Other participants changed their search terms by focusing on the physical aspects of the hypothetical symptom (eg, from "red scaly rash on nipple" to "itchy rash on nipple" [participant #70]) or made their searches less specific, changing from "scaly rash on nipple" to "scaly rash" (participant #142). However, the vast majority of participants (41/46, 89%) did not change their search terms during the search task.

Websites Visited

A total of 9% (5/46) of the complete cases did not go beyond the first page with search results returned by the search engine. For these participants, the number of websites visited was coded as zero but not as missing data because these participants were nonetheless able to view some preview of the search results and thus be exposed to some information rather than none at all. The returned results pages included Web links mentioning Paget disease of the nipple or links to information on breast cancer. Three of the five participants who did not go beyond the first search results page changed from not suspecting cancer at T1 to making a cancer attribution at T2, which supports the view that information exposure can have an effect.

Overall, the number of websites visited ranged from 0 to 6 (mean 1.96, SD 1.30, median 2, mode 1). We coded the content viewed by the participants as containing cancer-related content (yes/no) (eg, information on breast cancer or Paget disease of

the breast). The majority of participants whose browsing data were saved (41/46, 89%) viewed websites containing breast cancer information, in particular Paget disease (see Table 4 with a list of the websites visited). Excluding the 10 participants whose Web browsing data were not saved, there was no significant association between making a cancer attribution at T1 and accessing cancer-specific websites ($\chi^2_{1,46}$=0.004, *P*=.95. Furthermore, the participants who made a cancer attribution at T1 did not visit more websites than those who did not make a cancer attribution at T1 ($t_{1,44}$=0.647, *P*=.52).

Table 3. The range of search terms used (N=46).

Search terms used	Participants using search term, n (%)
Red scaly rash on nipple	9 (20)
Nipple rash	7 (15)
Breast rash	2 (4)
Itchy rash on nipple	2 (4)
Rash on nipple	2 (4)
Scaly rash on nipple	2 (4)
Ask NHS[a] questions	1 (2)
Breast cancer symptoms	1 (2)
Breast cancer symptoms NHS	1 (2)
Breast changes	1 (2)
Breast health red scaly rash	1 (2)
Changes in nipples	1 (2)
Itchy red breast rash	1 (2)
Nipple rash scaly	1 (2)
Nipple red rash	1 (2)
Rash on breast area	1 (2)
Rash on female breast	1 (2)
Rash on nipple area	1 (2)
Rash on nipples problems	1 (2)
Red breast rashes	1 (2)
Red rash nipple	1 (2)
Red rash on nipple	1 (2)
Red scaly nipple rash	1 (2)
Red scaly rash	1 (2)
Red scaly rash around nipple	1 (2)
Red scaly rash in nipple	1 (2)
Red scaly rash on breast	1 (2)
Red scaly rash on left nipple	1 (2)
Scaley rash on nipples *(sic)*	1 (2)
Scaly rash	1 (2)
Scaly rash on nipple NHS	1 (2)
Scaly red nipple rash	1 (2)
Scaly red rash on nipple	1 (2)
Sore nipples	1 (2)
What are the chances of red scaly rash on nipple being cancer	1 (2)
What should I do if I suspect I have itchy nipples	1 (2)

[a]NHS: National Health Service.

Table 4. The range of websites (domains) visited (n=41).

Web domains visited	Breast cancer content (yes/no)	Participants visiting the website, n (%)
www.nhs.uk	Yes	21 (51)
www.skinsight.com	Yes	8 (20)
www.rightdiagnosis.com	Yes	7 (17)
www.webmd.boots.com	Yes	7 (17)
symptoms.rightdiagnosis.com	No	6 (15)
www.healthcentral.com	Yes	4 (10)
www.mayoclinic.org	Yes	4 (10)
healthunlocked.com	No	3 (7)
www.breastcancer.org	Yes	3 (7)
www.zocdoc.com	No	3 (7)
www.cancerresearchuk.org	Yes	2 (5)
www.healthhype.com	Yes	2 (5)
www.healthline.com	Yes	2 (5)
about-cancer.cancerresearchuk.org	Yes	1 (2)
breastcancernow.org	Yes	1 (2)
en.m.wikipedia.org	Yes	1 (2)
en.wikipedia.org	Yes	1 (2)
healthguides.healthgrades.com/treating-psoriatic-arthritis/	No	1 (2)
www.cancer.gov	Yes	1 (2)
www.everydayhealth.com	Yes	1 (2)
www.macmillan.org.uk	Yes	1 (2)
www.medhelp.org	Yes	1 (2)
www.nationalbreastcancer.org	Yes	1 (2)
www.phaa.com	Yes	1 (2)
www.webmd.com	No	1 (2)

Table 5. Attributions for the nipple rash at T1 and T2 (n=56).

Type of attribution	T1 (before internet search task), n (%)	T2 (after internet search task), n (%)
Cancer	25 (45)	37 (66)
Physical	26 (46)	32 (57)
Environmental	16 (29)	3 (5)
Don’t know	7 (13)	4 (7)
Missing attribution	2 (4)	2 (4)

Symptom Attribution at T2

After viewing information online, the majority (31/56, 55%) of participants made a single symptom attribution at T2. More participants made a cancer attribution at T2 (37/56, 66%) compared to T1 (25/56, 45%), although some participants changed from making a cancer attribution at T1 to a noncancer one at T2 (6/56, 11%). There was a noticeable increase in medical terms used to make symptom attributions at T2 compared to T1, showing the exposure to formal medical terms during the online search: Paget disease (22 vs 1), eczema (14 vs 8), dermatitis (12 vs 3), and mastitis (3 vs 0) (see Table 5 for a summary of symptom attributions at T1 and T2). There was no significant association between viewing websites with breast cancer content and making a cancer attribution at T2 ($\chi^2_{1,46}$=1.92, *P*=.31).

Discussion

Principal Findings

Regarding the efficacy of the browser tracking tool to capture online information-seeking, the results are rather mixed and

point to challenges in measuring information-seeking remotely and designing reliable and user-friendly tools. In the majority of cases, the Vizzata tracking tool captured the search terms participants entered in the search engine and the websites they accessed, yet there were instances where participants seemed to encounter technical problems. Close inspection of the incomplete data of the 18 participants, of which 10 were retained in the study, enabled us to envisage a number of reasons why the search and browsing data were not saved. First, it is possible that some participants may have misunderstood the task instructions and installed the DuckDuckGo search engine, thus conducting the search within a browsing window outside that of the survey and preventing their online behavior from being recorded. Unfamiliarity with the DuckDuckGo search engine or with internet search tasks may have contributed to this. These potential errors on the part of participants suggest that online tracking tools need to be intuitive and user-friendly in order to be used effectively without the researcher's facilitation. Using a training session with a mock search prior to the actual study to familiarize the participants with the search task, as other researchers have done, could be a useful way to safeguard against misunderstandings or technical problems [12,13].

Second, it is possible that, despite the instructions, some participants may have used other internet browsers than the recommended Google Chrome that may have had features incompatible with the Vizzata tracking tool (eg, blocking tracking tools by default). Third, the tracking tool itself may not be sensitive enough to deal with the software complexities required to record online searching and browsing. However, we cannot make a direct comparison of the Vizzata tracking tool and other methods used in similar studies (eg, the Camtasia screen capture video-recording software [12,13]), and it is likely that each software package or method has its own advantages and disadvantages [37]. As has been noted before, researchers aiming to capture online behavior need to be realistic about the capabilities and limitations of each tool and decide on their use according to factors such as ease of participant recruitment, naturalistic setting, data accuracy, and need to minimize technical complexity [37,38].

Furthermore, the failure to capture some participants' online searching behavior (either for participant- or software-related reasons) also points to the limits of recording participant online behavior remotely. Our research had a limited scope of testing the capabilities of the Vizzata tracking tool ahead of a larger survey. The results suggest that more sophisticated study designs with additional measures (eg, think-aloud tasks or individual interviews) may be necessary to produce a more accurate and detailed picture of online health information-seeking behavior. A number of researchers have noted that capturing naturalistic online information-seeking behavior is a complex process that can be challenging [33,37,38], and it may well be that facing technical problems (delay in loading webpages or slow speed of custom tools that track online behavior) is not unusual in this type of study [38]. For example, researchers conducting a think-aloud task in conjunction with a search task using either the Google search engine or WebMD's Symptom Checker reported that participants found it difficult to navigate the programs and troubleshoot them after receiving error messages [11]. Further challenges noted in the literature relate to developing adequate tools that can record online behavior remotely such as search terms, visualization of Web pages, scrolling through search engine results pages, time spent on different Web pages, and gathering, where necessary, enough contextual information to enrich the data collected remotely and help interpret it [38].

Regarding symptom attribution, at T1 less than half of the participants (25/56) attributed the symptom (red scaly rash on nipple) to cancer, which supports the finding that nipple rash is an unfamiliar symptom of breast cancer [28,34]. While 25 participants thought the symptom could be cancer, only 1 of these used cancer in their search terms. The fact that not all cancer attributions at T1 translated into cancer-related search terms suggests that the participants' searches were inductive (ie, driven by symptoms) rather than deductive (ie, driven by diagnosis assumptions such as cancer). The results suggest that suspecting cancer at T1 did not necessarily translate into a hypothesis-testing approach to health information-seeking as found in other studies [12]. In line with similar research on online health information-seeking in response to symptoms, the participants' search strategies suggest an evidence-gathering approach to information-seeking rather than a hypothesis-testing one [12].

During the search task, some participants did not go beyond the first page of results, a behavior observed in other studies of health information-seeking [13]. However, the majority of participants accessed websites containing breast cancer information, in particular Paget disease of the breast. Half of these participants viewed the webpage of the UK NHS presenting information on Paget disease. Yet, not all participants interpreted the nipple rash as a symptom of breast cancer despite coming across information linking the symptom to the disease. Nonetheless, seeking online information did have an impact, as 12 more participants made a cancer attribution at T2 compared to T1. Some participants maintained their cancer attribution from T1, while others changed from noncancer at T1 to cancer at T2, although a small number of participants also changed from making a cancer attribution at T1 to a noncancer one at T2.

However, cancer attributions made in light of information-seeking should not be viewed as categorical interpretations of the symptom but rather as a possibility among many. While some participants made a single attribution at T2, others made more than one attribution as they envisaged the possibility that the symptom could be eczema or other physical condition (eg, anything from just a rash to something really serious like a sign of cancer [participant #70]). Importantly, our previous research [28] found that attributing nipple rash to cancer (as at least one possible cause) was associated with increased likelihood of medical help-seeking, which is an important behavior for earlier diagnosis of cancer [9].

Limitations

This study was conducted on a relatively small sample of participants (n=56) which limits the generalizability of our findings. However, small samples are not unusual in surveys

that aim to capture online information-seeking behavior. Sample sizes in these studies vary from 20 [31] to 54 [33] and 78 [12].

The majority of participants did not refine their search terms or consult many websites. As noted in other studies [31], the hypothetical nature of the symptom scenario may have demotivated participants to sustain in-depth information-seeking. It could well be that when people are acutely experiencing troubling symptoms they may be more thorough in their seeking and interpretation of online health information.

Another limitation of this study is that we were not able to explore how the participants made sense of the information found and what made them change (or not) their symptom attribution after engaging in information-seeking. Exploring how people engage with the information found online (eg, what websites seem trustworthy and why) could have provided more insight into how the participants interpreted online health information in relation to the given symptom. As has been pointed out in similar research, examining online health information-seeking requires attending to the cognitive and perceptual processes that are involved in conducting and interpreting an internet search [11].

Last, it is noteworthy that participants' search terms closely mirrored the symptom described in the scenario, "red scaly rash on nipple" and "nipple rash" being the most frequently used ones. This could be due to the fact that the symptom scenario was presented in text that made the symptom explicit. It is possible that had the symptom been presented in visual rather than textual form the participants may have interpreted it differently and used a wider range of search terms. Further research in the field of online health information-seeking is needed to explore how symptom attribution varies according to textual versus visual scenarios.

Conclusions

This exploratory study revealed that, despite some technical limitations, it is possible to capture the process of online information-seeking in relation to possible cancer symptoms. This work has potential for impact, both in terms of developing methodology to understand real-world issues and furthering the research agenda on understanding responses to cancer symptoms and engagement in health information-seeking online.

Acknowledgments

This study was funded by a seed grant from the Faculty Research Support Fund, Faculty of Health and Medical Sciences, University of Surrey.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of instructions on how to conduct the information-seeking task.

[PDF File (Adobe PDF File), 56KB - jmir_v21i2e12400_app1.pdf]

References

1. Pang P, Chang S, Pearce JM. Understanding different search approaches. 2014 Presented at: Proceedings of the 18th Pacific Asia Conference on Information Systems; 2014; Chengdu URL: https://aisel.aisnet.org/pacis2014/229
2. Fox S, Duggan M. Health Online 2013. Washington: Pew Internet and American Life Project URL: http://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2018-12-23] [WebCite Cache ID 74sbsnhYG]
3. Internet World Stats. 2017. URL: https://www.internetworldstats.com/ [accessed 2018-10-02] [WebCite Cache ID 72ryog1g9]
4. Gann B. Giving patients choice and control: health informatics on the patient journey. Yearb Med Inform 2012;7:70-73. [Medline: 22890344]
5. Thomson MD, Siminoff LA, Longo DR. Internet use for prediagnosis symptom appraisal by colorectal cancer patients. Health Educ Behav 2012 Oct;39(5):583-588 [FREE Full text] [doi: 10.1177/1090198111423941] [Medline: 21990571]
6. Diviani N, van den Putte B, Meppelink CS, van Weert JCM. 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] [Medline: 26817407]
7. Walter F, Webster A, Scott S, Emery J. The Andersen Model of Total Patient Delay: a systematic review of its application in cancer diagnosis. J Health Serv Res Policy 2012 Apr;17(2):110-118 [FREE Full text] [doi: 10.1258/jhsrp.2011.010113] [Medline: 22008712]
8. Scott SE, Walter FM, Webster A, Sutton S, Emery J. The model of pathways to treatment: conceptualization and integration with existing theory. Br J Health Psychol 2013 Feb;18(1):45-65. [doi: 10.1111/j.2044-8287.2012.02077.x] [Medline: 22536840]
9. Lyratzopoulos G, Abel G. Earlier diagnosis of breast cancer: focusing on symptomatic women. Nat Rev Clin Oncol 2013 Sep;10(9):544. [doi: 10.1038/nrclinonc.2012.126-c1] [Medline: 23881030]

10. Webber C, Jiang L, Grunfeld E, Groome PA. Identifying predictors of delayed diagnoses in symptomatic breast cancer: a scoping review. Eur J Cancer Care (Engl) 2017 Mar;26(2):e12483. [doi: 10.1111/ecc.12483] [Medline: 26950652]
11. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]
12. Perez SL, Paterniti DA, Wilson M, Bell RA, Chan MS, Villareal CC, et al. Characterizing the processes for navigating internet health information using real-time observations: a mixed-methods approach. J Med Internet Res 2015;17(7):e173 [FREE Full text] [doi: 10.2196/jmir.3945] [Medline: 26194787]
13. Macias W, Lee M, Cunningham N. Inside the mind of the online health information searcher using think-aloud protocol. Health Commun 2018 Dec;33(12):1482-1493. [doi: 10.1080/10410236.2017.1372040] [Medline: 28952804]
14. McHugh SM, Corrigan M, Morney N, Sheikh A, Lehane E, Hill ADK. A quantitative assessment of changing trends in internet usage for cancer information. World J Surg 2011 Feb;35(2):253-257. [doi: 10.1007/s00268-010-0830-8] [Medline: 20972679]
15. Quinn EM, Corrigan MA, McHugh SM, Murphy D, O'Mullane J, Hill ADK, et al. Breast cancer information on the internet: analysis of accessibility and accuracy. Breast 2012 Aug;21(4):514-517. [doi: 10.1016/j.breast.2012.01.020] [Medline: 22349349]
16. Nguyen SKA, Ingledew P. Tangled in the breast cancer web: an evaluation of the usage of web-based information resources by breast cancer patients. J Cancer Educ 2013 Dec;28(4):662-668. [doi: 10.1007/s13187-013-0509-6] [Medline: 23832398]
17. Foroughi F, Lam AK, Lim MSC, Saremi N, Ahmadvand A. Googling for cancer: an infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. JMIR Cancer 2016 May 04;2(1):e5 [FREE Full text] [doi: 10.2196/cancer.5212] [Medline: 28410185]
18. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001-2003. J Med Internet Res 2005 Jul 1;7(3):e36 [FREE Full text] [doi: 10.2196/jmir.7.3.e36] [Medline: 15998627]
19. Fazeli DS, Carlos RC, Hall KS, Dalton VK. Novel data sources for women's health research: mapping breast screening online information seeking through Google trends. Acad Radiol 2014 Sep;21(9):1172-1176 [FREE Full text] [doi: 10.1016/j.acra.2014.05.005] [Medline: 24998689]
20. Sugrue R, Carthy E, Kelly ME, Sweeney KJ. Science or popular media: what drives breast cancer online activity? Breast J 2018 Mar;24(2):189-192. [doi: 10.1111/tbj.12864] [Medline: 28744998]
21. Breast Cancer Statistics 2018.: Cancer Research UK URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer [accessed 2018-10-02] [WebCite Cache ID 72rpPQwlL]
22. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep 12:1 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
23. Phillips CA, Barz LA, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship between state-level google online search volume and cancer Incidence in the United States: retrospective study. J Med Internet Res 2018 Jan 08;20(1):e6 [FREE Full text] [doi: 10.2196/jmir.8870] [Medline: 29311051]
24. Ream E, Blows E, Scanlon K, Richardson A. An investigation of the quality of breast cancer information provided on the internet by voluntary organisations in Great Britain. Patient Educ Couns 2009 Jul;76(1):10-15. [doi: 10.1016/j.pec.2008.11.019] [Medline: 19179036]
25. Warren E, Footman K, Tinelli M, McKee M, Knai C. Do cancer-specific websites meet patient's information needs? Patient Educ Couns 2014 Apr;95(1):126-136. [doi: 10.1016/j.pec.2013.12.013] [Medline: 24447523]
26. Marcu A, Black G, Vedsted P, Lyratzopoulos G, Whitaker KL. Educational differences in responses to breast cancer symptoms: a qualitative comparative study. Br J Health Psychol 2017 Feb;22(1):26-41. [doi: 10.1111/bjhp.12215] [Medline: 27680898]
27. Marcu A, Black G, Whitaker KL. Variations in trust in Dr Google when experiencing potential breast cancer symptoms: exploring motivations to seek health information online. Health, Risk & Society 2018 Nov 27;20(7-8):325-341. [doi: 10.1080/13698575.2018.1550742]
28. Marcu A, Lyratzopoulos G, Black G, Vedsted P, Whitaker KL. Educational differences in likelihood of attributing breast symptoms to cancer: a vignette-based study. Psychooncology 2016 Dec;25(10):1191-1197. [doi: 10.1002/pon.4177] [Medline: 27218858]
29. Davies H, Marcu A, Vedsted P, Whitaker KL. Is lower symptom recognition associated with socioeconomic inequalities in help-seeking for potential breast cancer symptoms? Psychooncology 2018 Feb;27(2):626-632. [doi: 10.1002/pon.4557] [Medline: 28940942]
30. Barnett J, Fife-Schaw C, Shepherd R, Timotijevic L, Fletcher J, Fletcher D. Online Deliberative Engagement: A Pilot Study. A report for the Wellcome Trust. London: The Wellcome Trust; 2008.
31. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. J Am Med Inform Assoc 2008;15(4):484-495 [FREE Full text] [doi: 10.1197/jamia.M2449] [Medline: 18436912]

32. von Wagner C, Semmler C, Good A, Wardle J. Health literacy and self-efficacy for participating in colorectal cancer screening: the role of information processing. Patient Educ Couns 2009 Jun;75(3):352-357. [doi: 10.1016/j.pec.2009.03.015] [Medline: 19386461]
33. 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 Hum Behav 2017;69:256-267. [doi: 10.1016/j.chb.2016.12.032]
34. Linsell L, Forbes LJL, Burgess C, Kapari M, Thurnham A, Ramirez AJ. Validation of a measurement tool to assess awareness of breast cancer. Eur J Cancer 2010 May;46(8):1374-1381. [doi: 10.1016/j.ejca.2010.02.034] [Medline: 20335018]
35. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
36. 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]
37. Denvir C. Remote control: evaluating the potential of virtual desktops as a data collection tool in studies exploring how people use the Internet. Int J Soc Res Methodol 2016 Nov 09;20(5):533-546. [doi: 10.1080/13645579.2016.1253196]
38. Kellar M, Hawkey K, Inkpen K, Watters C. Challenges of capturing natural Web-based user behaviors. Int J Hum Comp Interaction 2008 May;24(4):385-409. [doi: 10.1080/10447310801973739]

Abbreviations

NHS: National Health Service

Edited by G Eysenbach; submitted 16.10.18; peer-reviewed by S Kummer, G Lyratzopoulos, H Yu; comments to author 21.11.18; revised version received 05.12.18; accepted 09.12.18; published 06.02.19.

Please cite as:
Marcu A, Muller C, Ream E, Whitaker KL
Online Information-Seeking About Potential Breast Cancer Symptoms: Capturing Online Behavior With an Internet Browsing Tracking Tool
J Med Internet Res 2019;21(2):e12400
URL: http://www.jmir.org/2019/2/e12400/
doi:10.2196/12400
PMID:30724741

Original Paper

Developing a Decision-Aid Website for Breast Cancer Surgery: An Action Research Approach

Yu-Ting Hung[1,2], RN; Ching-Fang Wu[3], RN, MSN; Te-Hsin Liang[4], PhD; Shin-Shang Chou[5], RN, PhD; Guan-Liang Chen[6], RN, MSN; Pei-Ni Wu[7], BA; Guan-Rong Su[8], ME; Tsuey-Huah Jang[5], MSN; Chang-Yi Liu[9], RN, MS; Ching-Yen Wang[5], RN, MSN; Ling-Ming Tseng[10,11*], MD; Shuh-Jen Sheu[12*], RN, PhD

[1]School of Nursing, National Yang-Ming University, Taipei, Taiwan

[2]Public Health Center of Taoyuan District, Department of Public Health, Taoyuan City Government, Taoyuan, Taiwan

[3]Infinity Clinic, Taipei, Taiwan

[4]Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan

[5]Taipei Veterans General Hospital, Taipei, Taiwan

[6]National Taiwan University Hospital, Taipei, Taiwan

[7]Cathay General Hospital, Taipei, Taiwan

[8]AdvancedTEK International Corp, Taipei, Taiwan

[9]Infusion Treatment Center at Cancer Center South Bay, Stanford Healthcare, San Jose, CA, United States

[10]Comprehensive Breast Health Center & Division of General Surgery, Department of Surgery, Taipei Veterans General Hospital, Taipei, Taiwan

[11]Department of Surgery, School of Medicine, National Yang-Ming University, Taipei, Taiwan

[12]Institute of Community Health Care, School of Nursing, National Yang-Ming University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Shuh-Jen Sheu, RN, PhD
Institute of Community Health Care
School of Nursing
National Yang-Ming University
Nursing Building Room 404
Number 155, Section 2, Linong Street
Taipei, 11221
Taiwan
Phone: 886 2 28267271
Fax: 886 2 28229973
Email: sjsheu@ym.edu.tw

Abstract

Background: Patients with early-stage breast cancer have numerous options when choosing the type of breast surgery method to be applied. Each of these options lead to a similar long-term survival rate, but result in significant differences in appearance, function, cost, recurrence rate, and various other relevant considerations. However, the time available for detailed communication with each patient is often limited in clinics, which puts these women under great psychological stress and can hinder their surgery-related decision making.

Objective: The objective of this study was to develop a multipurpose surgery decision-making website providing medical information, psychological support, and decision-related simulation for women during breast cancer surgery-related decision making.

Methods: Using the 4 steps of action research, which involve multigroup teamwork via regular team meetings, the following were performed: (1) Planning: searching, analyzing, and evaluating health websites to consensually decide the major infrastructure; (2) Action: work was performed simultaneously in 4 groups, which consisted of medical information collection and editing, patient interviews and data extraction, webpage content design, and programming to create or host the website; (3) Evaluation: the website was tested by clinical experts and focus groups of former breast cancer patients to assess its effectiveness and pinpoint appropriate improvements; and (4) Reflection: constant dialogue was conducted between the various participants at each step, which was used as the foundation and motivation of next plan-action-evaluation-reflection circle.

Results: Using the action research approach, we completed the development of our website, which includes the following: (1) "Woman's Voice"—an animated comic depicting the story of a female breast cancer patient with interspersed questions for the users that will help them better empathize with the experience; (2) "Cancer Information Treasure House"—providing breast cancer surgery-related information through text, tables, pictures and a presentation video; (3) "Decision-making Simulator"—helping patients think through and check the pros and cons of the different surgical options via visual-based interactions including "Stairs Climbing" and "Fruit of Hope"; and (4) "Recommended Links"—providing reliable websites for further reference. Additionally, we have further improved the website based on the feedback received from postsurgery breast cancer patients and clinicians. We hope to continue improving to better meet both the patients' and health providers' needs and become a practical decision-making aid for patients undergoing breast cancer surgery.

Conclusions: We have created the first breast cancer surgery decision-making assistance tool in Taiwan using a "Web-based" and multifunctional website design. This site aims to provide health care knowledge, psychological healing, and emotional support functions, as well as decision-making capability enhancement simulations. We look forward to assisting breast cancer patients in their decision-making process and expect our website to increase patient's autonomy and improve their communication with clinicians.

(J Med Internet Res 2019;21(2):e10404) doi:10.2196/10404

KEYWORDS

breast cancer; surgery-related decision making; website; action research

Introduction

In Asia, as in other parts of the world, Cancer is one of the major health concerns. Different types of cancer have different incidence rate according to gender, age, and other demographic factors. Among the different types, breast cancer has the highest incidence rate among the female population of Taiwan. According to the 2017 report of the Taiwan Health Promotion Administration Ministry of Health and Welfare, the incidence rate of breast cancer was 125.64 per 100,000. This means that in 2015, the year for which the latest statistics are available, 14,801 women were diagnosed with breast cancer. Among these women, roughly 80% had been diagnosed with early-stage breast cancer [1]. The key point is the fact that patients with early-stage breast cancer have more than one treatment choice, but unfortunately, they do not necessarily have all the information needed to make an informed decision regarding the kind of treatment best suited to their particular case. For instance, lumpectomy and mastectomy are often recommended for most patients with early-stage breast cancer [2]. In addition, breast reconstruction following mastectomy is another choice. It is worth noting that there is a general consensus among the medical community that the long-term survival rate for this disease is very similar across these 3 different types of surgery [3-6]. Thus, if the tumor size and location allow, patients are able to choose their preferred surgical intervention.

Patients' decisions about breast cancer surgery are multifactorial, and their understanding of the disease and its treatment affects their decision making [7]. Our initial empirical observations, which were later confirmed, suggested that 4 important factors hinder Taiwanese breast cancer patients' ability to make well-informed decisions regarding the kind of treatment best suited to their particular case. First, there is a lack of easily available information. Studies have shown that when facing decision making regarding breast cancer surgery, patients often require more information to decide which type of breast surgery is suitable for them [8,9]. However, this requirement for information is mostly unmet; they do not fully understand the different surgical procedures that might apply to their case [10,11]. Second, there are negative emotions resulting from their personal experience of facing the diagnosis. Pretreatment symptoms in newly diagnosed patients include not only cognitive but also emotional aspects; patients often experience a variety of negative emotions due to a link between cancer and death [12-14], in particular, high levels of anxiety and uncertainty [14,15]. Third, another key factor hindering the patients' ability to make a well-informed decision regarding the kind of surgery most suitable for them is the limited clinic time available or, in other words, the relatively short time span of their appointments with their physician. Finally, there is a significant cultural factor that also affects the decision-making process of patients with breast cancer in Taiwan, namely local taboos regarding certain medical conditions. Talking about breast cancer, as well as other conditions related to the intimate parts of one's body, is not encouraged in Chinese culture. Therefore, women tend to be hesitant even when talking to their physicians, which affects their requesting of further information regarding their condition, disclosing their opinion, and showing their personal feelings when required.

Decision aids are tools intended to help patients make an informed decision, and they do this by providing them with better knowledge and a general overview of all the options available to them with the pros and cons of each one. By adding clarity and congruence between decision and personal values, decision aids help reduce conflict during the decision-making process, as well as helping patients break away from passiveness [16]. A key aspect of this research is the creation of a website that relies on a set of audiovisual and digital processes to aid patients make an appropriate decision regarding breast cancer. Compared with the traditional alternatives, the advantages of digital and audiovisual tools are that they provide better and more vivid visual effects. By emphasizing interactive features and varied visual stimuli, patients are able to obtain first-hand and immediate information. The digital nature of this site will facilitate the ability for the information to be updated frequently, which will provide the patients with the most reliable knowledge

available at a given moment [17]. In a face-to-face consultation between a physician and a patient, the use of computer-based decision aids can increase the efficiency of counseling [18].

Therefore, this study aims to develop a multipurpose surgery decision-making website that acts as a decision-aid tool by providing medical information, psychological support, and decision simulation for patients with breast cancer who are undertaking the decision-making process regarding future surgery.

Methods

Research Design

In this research, we applied an action research approach, which is defined as "a disciplined process of inquiry conducted by and for those taking the action. The primary reason for engaging in action research is to assist the 'actor' in improving and/or refining his or her actions." [19] Compared with conventional research methods, one of the key differences is the fact that the main goal of action research is to generate knowledge and solve problems through direct involvement of the actors [20]. In recent years, action research has used tackled topics such as (1) Cross-border Peer Health Educator Programs and Latina/Family Experiences; (2) Psychosocial Needs Post-Genetic Cancer Risk Assessment; and (3) finding positive effects in breast cancer patient/family education [21-23]. However, a limited number of breast cancer-related studies are involved in the decision-making and research processes. Therefore, the ultimate goal of this study is to argue in favor of the benefits to patients and families of a Web-based platform that assists patients with breast cancer in the decision-making process; we concluded that an action research approach is the most suitable way of conducting this investigation.

To achieve our goal, the research was designed to be performed by multiple groups, which were brought together to form an effective work team. Meetings were conducted on a regular basis, namely every 2 weeks during a semester and weekly during the summer and winter breaks (Table 1 provides a detailed account of the 90 meetings that were conducted). The aim was to develop a well-coordinated teamwork system that allowed the whole team to be able to review and adjust their assignments appropriately. During these regular meetings, each group reported on their progress, shared their experiences, and provided mutual feedback. Through the accumulated knowledge and experiences of each group in their designated area, we were able to complete the construction of the proposed website, which is shown in the conceptual framework presented in Figure 1.

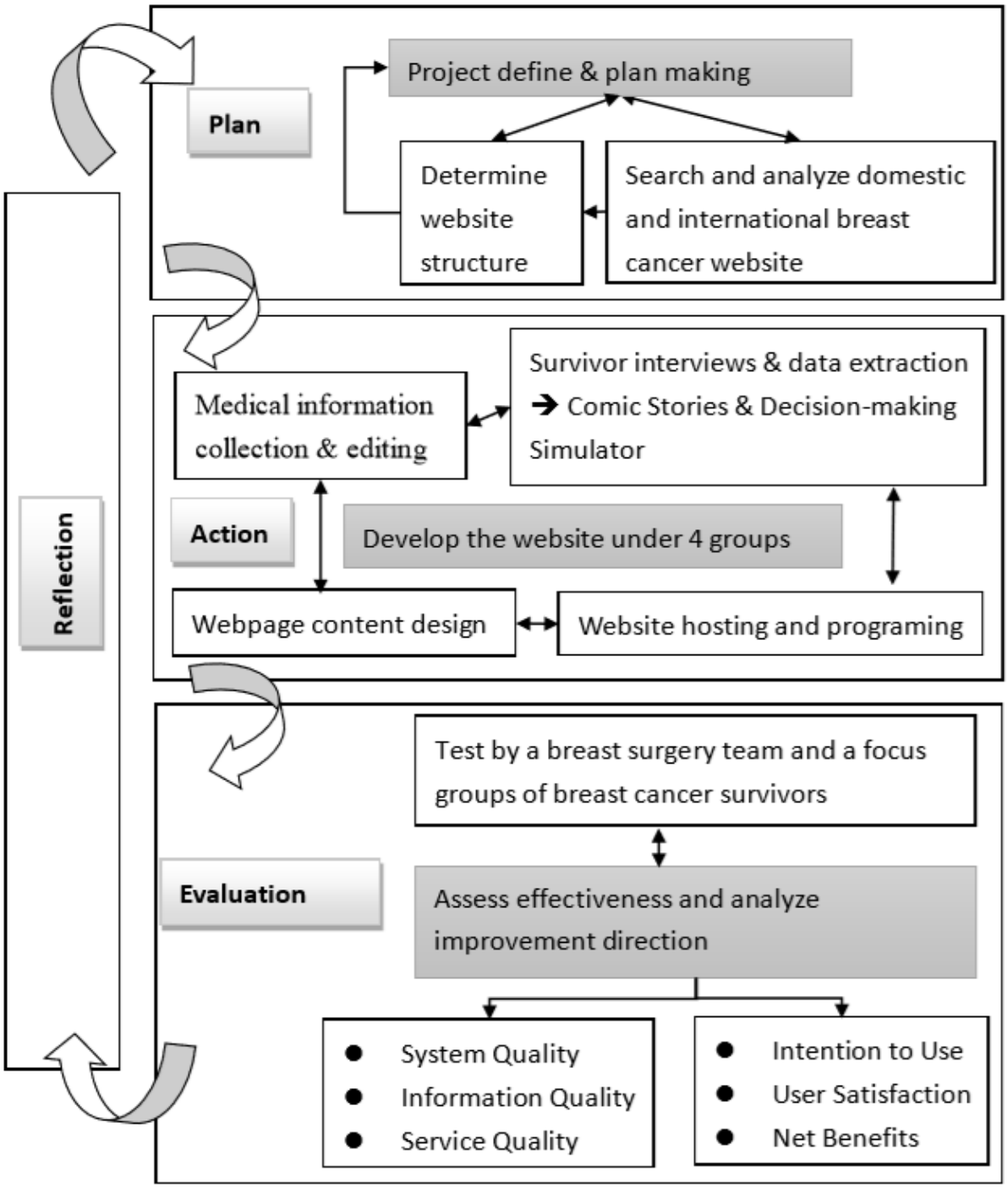
Step 1: Plan

The research team included the participation and expertise of a leading academic scholar (Prof Shuh-Jen Sheu), as well as involvement of significant contributions from several nursing doctoral and master degree candidates, an information technology (IT) engineer, and 3 computer graphics designers; in total, there were 13 team members. During this stage of the project, team members were tasked with the collection and analysis of information related to existing dedicated breast cancer websites. Based on the gathered information and the research conclusions drawn from this information, together with the assistance of the IT specialist, the team undertook the task of creating a website that is central to this investigation.

Table 1. Important meeting dates.

Year	Meeting dates	Times (n)
2010	09/21, 10/7, 10/21, 11/16, 12/7, 12/28	6
2011	01/20, 03/31, 04/19, 05/09, 05/30, 06/20, 06/30, 07/07, 07/15, 07/21, 07/28, 08/11, 08/18, 08/24, 09/15, 09/29, 10/25, 10/27, 10/31, 11/10, 11/24, 11/25	22
2012	01/16, 02/23, 03/08, 03/22, 04/05, 04/19, 05/03, 05/17, 05/31, 06/14, 06/28, 11/29, 12/13	13
2013	01/17, 01/29, 02/05, 02/25, 03/11, 04/01, 04/22, 05/23, 05/30, 06/06, 09/27, 10/11, 11/08, 11/22, 12/06, 12/27,	16
2014	01/10, 02/19, 03/17, 03/31, 05/26, 06/09, 06/25, 07/09, 07/16, 08/06, 08/20, 09/01, 09/09, 09/19, 09/26, 10/09, 10/16, 11/13, 10/30, 11/27, 12/11, 12/16, 12/18, 12/25	24
2015	01/20, 01/27, 02/03, 02/10, 03//03, 3/6, 3/10, 3/13, 3/24	9

Figure 1. Conceptual framework of the action research.



Step 2: Action

At this stage of the study, we formed 4 group work teams, each of which tackled different parts of the investigation. Through regular exchange between the 4 groups, various processes within the target areas were presented, analyzed, and corrected by the groups when necessary.

The mission of the 4 groups was as follows:

- *Group 1: The Collecting and Editing of Medical Information*. This group was responsible for gathering and reviewing of existing breast cancer surgery literature with the aim of creating easy-to-understand formats and simple graphics renderings.
- *Group 2: Patients Interviews and Data Extraction*. This group was responsible for interviewing women who had undergone breast cancer surgery in order to collect information related to their decision-making experiences. The gathered information was then analyzed through qualitative content analysis to understand patients' mindsets vis-à-vis the decision-making process; these were then used to identify any possible patterns. The mindset and possible patterns would become an important feature of the

Animated Comics and the Animated Decision-making Simulator on our website.

- *Group 3: Webpage Content Design*. This group was responsible for webpage design, which included the Animated Comics, Simulation Games, Medical Information Display, and Relevant Links Page. This content was created using various graphics software including Adobe Illustrator CS6 (Adobe Incorporated) and Photo Impact X3 (Ulead Systems Incorporated).
- *Group 4: Programming and Website Hosting*. This group was responsible for informatics work related to program coding and the hosting of the website. The tech team used ActionScript technology for front-end animation programs and J2EE technology for back-end animation programming. The operating system was Ubuntu 10 (Canonical Ltd), while the network server was Apache Tomcat 7 (Apache Software Foundation). The Web-based machine was hosted by the Amazon's EC2 Cloud Service (Amazon.com Inc). The internet domain name was accessed through Seed net (Software Engineering Environment Development Network). In addition, various program development tools were used including Eclipse Java EE IDE (Eclipse Foundation) for Web Developers, Oracle Java SE Development Kit 7 (Oracle Corporation), and Flash Develop 4 (Flash Develop Team) .

Step 3: Evaluation

On completion of the preliminary website, we used a multifaceted approach to evaluate the website. The participants in the evaluation consisted of 9 former breast cancer patients; 12 clinical staff members comprising surgeons, case managers, nurse specialists, and oncology radiologists; and 19 research team members, including academic scholars, doctoral students, master students, undergraduate students, information engineers, and graphic designers. The website evaluation was conducted using three different processes: individual interviews and focus groups with former breast cancer patients, individual and group consultations with the clinical staff, and an internal evaluation via regular team meetings. Any relevant information was recorded and logged via minutes of meeting.

Step 4: Reflection

Reflection consisted of the constant dialogue among team members during every stage of the investigation process. Team members communicated with each other and shared their views from their diverse perspectives. This allowed the researchers to review their role in retrospective during this action research and to develop different patterns of thinking related to the health care needs of patients with breast cancer during the surgery-related decision-making process. This, in turn, became the foundation and motivation for the next plan-action-evaluation-reflection circle.

Results

Patients Interviews and Data Extraction

This survey took place at two different medical centers in Taipei City from September 2010 through January 2012. A total of 31 women who were being treated for breast cancer were interviewed, including 11 who underwent mastectomy, 15 who underwent lumpectomy, and 5 who underwent mastectomy followed by immediate breast reconstruction. Of these patients, 20 were interviewed 1 day before their surgery, 2 while they were under chemotherapy, and 9 after they finished chemotherapy, which had started 6-12 months earlier. Based on the results of the Qualitative Content Analysis, we concluded that "the impact factors for surgery-related decision making," "the decision-making modes related to the 3 types of breast surgery," and "the mindset after deciding to undergo breast cancer surgery" were the most suitable areas that could function as the basis of our webpage content and help with developing the Animated Comics and Decision-making Simulator.

Webpage Content Design

After a broad collection of information and its analysis by the various work groups, we concluded that the information available on breast cancer surgery-related decision making and on the psychological support of patients with breast cancer on Taiwanese websites was insufficient for local patients with breast cancer. Based on this premise, we adopted a multifaceted approach to design and construct a website that was more interactive and individualistic; the aim being to fulfill this information need. Our final proposal was a breast cancer surgery decision-making support website based on a decision-making assistant, together with appropriate psychological support. The contents of the website are as follows.

Website Structure

Upon entering the welcome page, patients are shown a thematic map entitled "Tale of the Breast Country." Underneath this map, our webpage follows a logical structure where the 4 main thematic parts with their corresponding submenus are displayed as the user clicks on them. The 4 main themes are Women's Voice, Cancer Information Treasure House, the Decision-making Simulator, and Recommend Links (Figure 2).

Thematic Map

In addition to displaying a simple text menu on the top of the page, we also created a thematic map (Figure 3) using colorful flash animations to help users understand the main themes and the whole structure of the website.

A Tale of Breast Country

We created an animated story "Tale of the Breast Country," which describes how women in Breast Country were attacked by the evil of Adenocarcinoma, which led to the development of breast cancer. The main character is the female warrior Maya (primarily adopted from legendary Greek Amazon who cut away her left breast to use her bow and arrows more effectively) who stands up for the inhabitants of the country and seeks ways to fight this problem. Her adventures are a descriptive process of the various possible approaches to dealing with breast cancer. The different paths that her journey takes form the basis of part of the website (Figure 4).

Women's Voice

We created an animated story based on the results obtained from analyzing our patient interviews, and these focused on "the mindset after deciding to opt for breast cancer surgery." This

story was named the "Story of Shu-Jun" and tells the story of Shu-Jun (a common Chinese female first name), who is a woman full of despair and anxiety after realizing she had developed cancer. She represents the various emotional reactions that women tend to show when diagnosed with this condition, such as fear of death and concerns about treatment options. At the end of the story, our heroine finds the courage to face the disease. Overall, 32 animated videos are displayed covering 8 different topics, which means that each topic consists of 4 animated comics. These 4 animated comics are constructed on the basis of the parameters of "the mindset after deciding breast cancer surgery." Each comic is accompanied by a female narration and soft background music; they are interspersed with questions designed to help patients improve their reaction to matters that create uncertainty and that need psychological adjustment; the aim is to lead patients toward a more natural expression of their personal and psychological feelings (Figure 5).

Cancer Information Treasure House

We gathered and analyzed a wide array of information that might be helpful to patients during the breast cancer surgery decision-making process, and this information was edited into 7 topics that are presented in a Chinese homophonic or symbolic way. These include the following: basic knowledge of breasts and breast cancer, risk factors, diagnostic examinations to identify breast cancer, different stages of breast cancer, various treatments available for breast cancer, various factors related to the prognosis of breast cancer, and care after breast cancer surgery. In addition, we provide presentation videos as different choices available using text and pictures covering the 7 topics (Figure 6).

Decision-Making Simulator

"Stairs Climbing" and "Fruits of Hope" are the 2 breast surgery simulation tools available on our platform; they are intended to help patients to think through the process from different perspectives. These tools were crafted on the basis of the qualitative analysis of the interviews conducted with the postsurgery breast cancer patients. The qualitative analysis was conducted focusing on two key aspects, namely "the impact factors of surgery-related decision making" and "decision-making modes regarding the 3 types of breast surgery."

In "Stairs Climbing" (Figure 7), the left part shows the instructions and 24 questions related to "the factors that have an impact on the surgery-related decision-making process." On the right side, there are 3 stairs with 3 characters, which represent the 3 types of breast cancer surgery. By choosing Strongly Agree, Agree, No Opinion, Disagree, and Strongly Disagree, the 3 characters change their hierarchical position on the stairs according to the user's answer choices. Following this logic, the character that climbs to the top at the end of the questionnaire represents the surgery that seems to be preferred by the patient.

"Fruits of Hope" (Figure 8) is the second tool designed on the basis of the qualitative analysis of the interviews with postsurgery breast cancer patients. This tool was specifically based on the interview's responses regarding the "decision-making modes of the 3 types of breast surgery." Nine aspects of the surgical outcomes are applied as follows: personal feeling, recurrence rate, postoperative appearance, body balance, clothing choices, sex and intimacy, therapeutic category, clinic numbers, and economic considerations. There are fruits of different sizes and colors on a "Hope Tree" that reflect the consequences of the different surgeries. Users can choose different surgical outcomes by clicking on the pictures to pick the fruits. The fruits are then counted by color, and the result is shown at the end of the simulation. Users are told if the choice can meet their needs at the outcome level.

To avoid conflicts or confusion after using the tools, detailed instructions are provided, at the beginning of the process, during the process, and at the end of the simulation page. It is worth remembering that the main purpose of these 2 simulations is to assist patients in need of further information and to allow them to make well-informed decisions, rather than replacing entirely the decision-making process or making the decision for them. Furthermore, it is clearly stated that the results of the simulation only reflect an individual's values and preferences. Patients are reminded that it is still necessary to consult with their surgeon before making a final decision.

Related Links

As the main purpose of this work is the creation of a decision-making aid website for patients with breast cancer, we have also collected a list of related local and foreign websites that will be able to help patients as reference points and allow them to make comparisons between options.

Local websites are as follows:

- Taiwan Breast Cancer Foundation
- Taiwan Breast Cancer Alliance
- Taichung Kaihuai Association
- Taipei warm Association
- Taiwan Breast Reconstruction Society
- Kaohsiung Heart to Hand Association
- Global Chinese Breast Cancer Organizations Alliance
- HER2 Club
- Breast Center of Taiwan Adventist Hospital

Foreign websites are as follows:

- Breast Cancer Care
- Breastcancer.org
- Bresdex

Figure 2. Website structure.

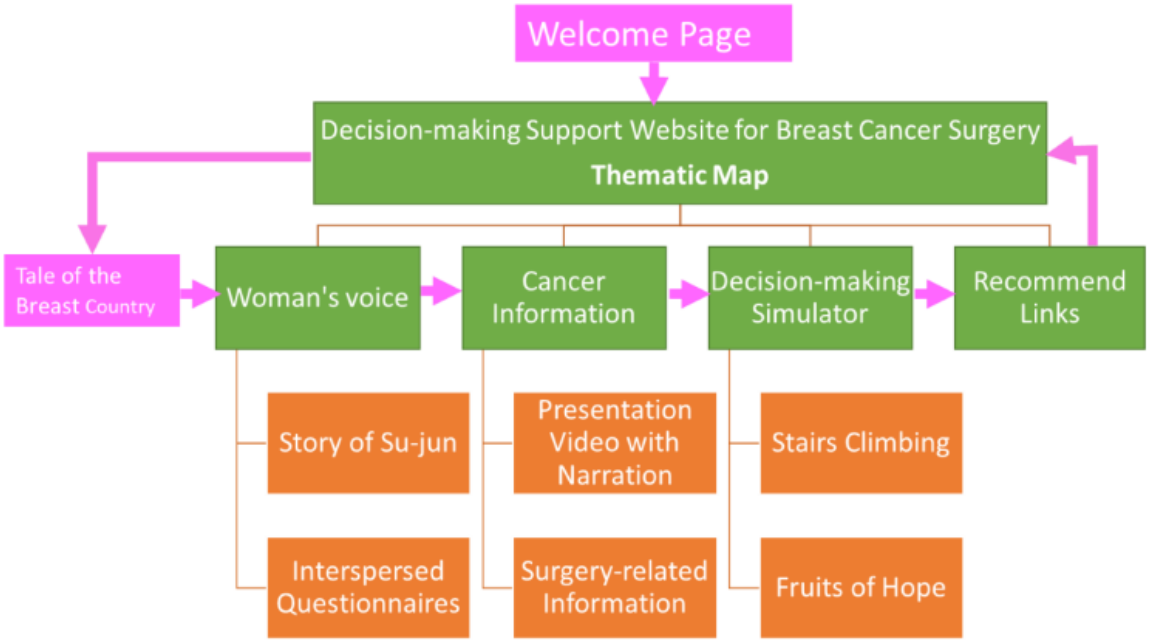


Figure 3. Thematic map.

Figure 4. Tale of Breast Country.

Figure 5. Four examples of women's voice.

Figure 6. Cancer information treasure house.

Figure 7. Stairs Climbing.

Figure 8. Fruits of Hope.

Table 2. A list of clinical staff members who participated in the study.

Office unit	Date of participation	Form of participation	Job title
Medical center	04/17/2015	Breast cancer multispecialty medical team joint discussion	3 Surgeons; 1 General surgery chief resident; 1 Radiation oncology physician; 2 Breast cancer case manager; 1 Radiologist; 1 Pathology physician
Medical center	08/2010-12/2014	Participated in research team meetings	1 Surgical nurse practitioner
Medical center	01/27/2015	Participated in the focus group	1 Surgical nurse practitioner
Medical center	10/19/2014	Participated in research team meeting	1 Breast cancer case manager
Regional hospital	07/15/2015	Individual consultation	1 Assistant of head of the surgical department

Website Evaluation

To further evaluate the functioning and effectiveness of our website, 3 main approaches were used during the evaluation process. First, we performed internal assessments through regular team meetings. Second, we conducted several interviews with 3 individual patients and with focus groups consisting of 7 patients. We tested our website's logics and functions with these former breast cancer patients about their experiences regarding the process of deciding the type of breast cancer surgery best suited for them. The clarification and verification took place in one medical center in Taipei City, from December 2014 through March 2015. Third, we formed a group consisting of 12 clinical staff members, namely surgeons, case managers, nurse specialists, and oncology radiologists. We then demonstrated our website and conducted individual and group consultations with this medical team. Table 2 lists the clinical staff partaking in this group.

Discussion

Principal Findings

Past studies in western literature have highlighted the importance of the principles of informed consent and patient autonomy; however, the medical literacy of patients, their preferences, as well as a close relationship (mutual communication and understanding) between patients with breast cancer and their surgeon, are often key challenges for women while making a meaningful decision regarding their impending breast surgical choice. Data analysis of the first encounter of patients with their surgeon has indicated the high level of reassurance patients obtain from the authority and expertise they see in the surgeon figure [24]. From a clinician's point of view, this situation is of decisive importance to patients with breast cancer; however, we found that what the patients want and what the clinicians deem necessary in terms of the website content can often be conflicting. For instance, some clinicians believe that what a patient needs during decision making is medical knowledge as opposed to reading comics, while almost all participants in breast cancer support groups felt very touched after watching the Story of Shu-Jun because they can relate the animated comic to their own experiences. In addition, the question helped them express their current feelings, and the last comic topic "revival" made them feel calmer and full of hope when facing upcoming treatment. This work applies action research, which aims to bridge the gap between patients and clinicians and caters more toward the feelings and needs of a patient's mind and body than previous approaches.

Studies have often demonstrated the positive effects of Web-based decision-making tools [25]. Considering this, our research is aimed to more closely identify with and assure patients during the dynamic relationship that is patient-doctor communications. During the process of developing the website, the requirement to balance patients' needs and clinicians' professional concerns became a problem that needed a solution. After conducting a thorough literature review and discussing the problem during the research meetings of the work groups, we finally came to an agreement regarding the actions needed. First, the team would closely collaborate with professionals in medical centers; this would allow the updating and releasing of medical information onto the website quickly to ensure consistency of the information. Second, the medical team would evaluate the website before referring it to patients. In parallel, we emphasized the fact that the site had an assisting role and it was not intended to replace the professional clinician's assessment. Furthermore, we would continue to evaluate the effectiveness of the Web-based tool for women who are newly diagnosed breast cancer using experiences by designing more comprehensive quantitative and qualitative research to put in place subsequent relevant modifications and various promotional plans. This would continue the cycle of action research into the future based on any new results.

We consider that the decisions of patients regarding breast cancer surgery are affected by multiple factors, including the reliability of the information available, availability of information, patients' emotional needs, the fact that clinical counseling time is limited, and the effect of patients' culture on the process. Therefore, to meet the needs of early-stage breast cancer-related decision-making process, this study was designed to use a multifaceted approach.

Web-based breast cancer decision aids have become increasingly common in recent years [26-30]. The research design we used was an action research approach. The core ideas of applying action research involve practical reflection and real performance during the investigation process, critique liberation, building investigation strategies, and research innovation [31]. Compared with the development of medical information websites, in general, our Web-based breast cancer decision aid development process has been more systematic and sophisticated and has improved on previous sites primarily by learning about the patients' perspectives and needs. To get closer to the experience of women, we primarily used qualitative interviews and content

analysis. In addition, we used a multifaceted approach to verify the meaning and functioning of the website. These distinct approaches had the same objectives, which was to meet the needs and special conditions associated with local Taiwanese cases. This study serves as a reference for describing how to develop decision support tools for women with breast cancer by considering their own corresponding culture and the region in which they live.

Limitations

We faced many challenges during website construction for this study; they included language, communication, information engineering, and art design, as well as how to train the decision-making ability. In addition to the type of tumor and its stage being different for each patient, opinions and expert advice may also differ between clinicians and hospitals. Detailed individual treatment information, such as recurrence rate, survival rate, treatment plan, and reconstructive decisions, is still dependent on the physician's assessment, which is also dependent on their experience and personal opinions, rather than what is described on the site. Hence, we can only recommend patients to find more information and discuss their personal issues with their surgeon directly. Finally, the technique of website hosting, programming, and patterns design depends on external technical support. The action research cycles provide an opportunity to modify and confirm this repeatedly, as it is quite a repetitive and lengthy process for the professionals involved. This also means that maintaining long-term collaborations with various professionals to maintain these cycles requires a considerable budget. Therefore, it has been suggested that more students and teachers in IT engineering and design-related departments should be trained and invited to work together in this area.

This paper is focused on the construction part, and further detailed quantitative and qualitative studies on the website effectiveness will be the next step. As our search and analysis of the website are limited to a Chinese version, we are, therefore, unable to explore the advantages and disadvantages related to non-Chinese and non-English literature and websites. However, this study can serve as a significant foundation and reference point for anyone who is interested in the future in the planning, promotion, operation, and long-term management of similar websites.

Conclusions

This study aims to build up a dedicated decision-aid website with a good structure and useful content so that it can help women who are facing the dilemma of choosing the right breast cancer surgery for them. As the first surgery decision-making assistance website in Taiwan, the content of this site provides not only health care knowledge but also psychological healing functionality and decision-making simulation. Using the action research approach, we gathered breast cancer survivors, clinicians, nursing scholars, nursing students, information engineers, and computer graphics designers together. With their combined knowledge and expertise, this team managed effectively to shorten the distance between theory and practice. After a rigorous process of investigation, analysis, and testing, we have successfully completed the development of our surgery decision-making website for patients with breast cancer. Through testing of our website via the focus group and further interviews, we have concluded that this kind of work can indeed help patients with early-stage breast cancer make a more informed decision regarding the type of surgery they would prefer to undergo. We look forward to assisting patients with breast cancer in their decision-making process. We expect that this will increase the autonomy of these patients and improve communication with their clinicians.

Acknowledgments

We appreciate the funding provided by the Medical Scholarship Foundation In Memory of Professor Albert Ly-Young Shen and the Ministry of Science and Technology of the Republic of China (2000-2012: The exploration and promotion of suspected breast cancer women's medical autonomy and decision-making for receiving a surgery, NSC 99-2629-H-010-001MY2; 2013-2015: Construction and exploration of breast surgery decisions website interactive content—animated comic and decisions games design, NSC 102-2629-H-010-001MY2) and the input of all the team members who took part in this research. Special thanks to Ya-Ning Zhan, Han-Yi Tsai, and Chi-Shan Tsai for their assistance in the research process. We also thank the NYMU-NTUST Joint Research Program (107TY02, The Exploration of Constructing Breast Surgery Decision Aide Website on Public Network Platform and Using Experience) for providing funds, and Professor Kung-Jeng Wang for his support.

Conflicts of Interest

None declared.

References

1. Health Promotion Administration, Ministry of Health and Welfare. 2017. Cancer Registry Annual Report, 2015 Taiwan URL: https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=269&pid=8084 [WebCite Cache ID 6xuUXPe8S]
2. Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, et al. NCCN Guidelines Insights: Breast Cancer, Version 1.2017. J Natl Compr Canc Netw 2017 Dec;15(4):433-451. [Medline: 28404755]
3. Black DM, Hunt KK, Mittendorf EA. Long term outcomes reporting the safety of breast conserving therapy compared to mastectomy: 20-year results of EORTC 10801. Gland Surg 2013 Aug;2(3):120-123 [FREE Full text] [doi: 10.3978/j.issn.2227-684X.2013.06.01] [Medline: 25083471]

4. Chen H, Wu K, Wang M, Wang F, Zhang M, Zhang P. A standard mastectomy should not be the only recommended breast surgical treatment for non-metastatic inflammatory breast cancer: A large population-based study in the Surveillance, Epidemiology, and End Results database 18. Breast 2017 Oct;35:48-54. [doi: 10.1016/j.breast.2017.06.002] [Medline: 28649032]
5. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med 2002 Oct 17;347(16):1233-1241. [doi: 10.1056/NEJMoa022152] [Medline: 12393820]
6. Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med 2002 Oct 17;347(16):1227-1232. [doi: 10.1056/NEJMoa020989] [Medline: 12393819]
7. Mac Bride MB, Neal L, Dilaveri CA, Sandhu NP, Hieken TJ, Ghosh K, et al. Factors associated with surgical decision making in women with early-stage breast cancer: a literature review. J Womens Health (Larchmt) 2013 Mar;22(3):236-242. [doi: 10.1089/jwh.2012.3969] [Medline: 23428286]
8. Kimiafar K, Sarbaz M, Shahid Sales S, Esmaeili M, Javame Ghazvini Z. Breast cancer patients' information needs and information-seeking behavior in a developing country. Breast 2016 Aug;28:156-160. [doi: 10.1016/j.breast.2016.05.011] [Medline: 27318171]
9. Littlechild S, Barr L. Using the Internet for information about breast cancer: a questionnaire-based study. Patient Educ Couns 2013 Sep;92(3):413-417. [doi: 10.1016/j.pec.2013.06.018] [Medline: 23891419]
10. Liao M, Chen S, Chen S, Lin Y, Hsu Y, Hung H, et al. Changes and predictors of unmet supportive care needs in Taiwanese women with newly diagnosed breast cancer. Oncol Nurs Forum 2012 Sep;39(5):E380-E389. [doi: 10.1188/12.ONF.E380-E389] [Medline: 22940517]
11. Puts MTE, Papoutsis A, Springall E, Tourangeau AE. A systematic review of unmet needs of newly diagnosed older cancer patients undergoing active cancer treatment. Support Care Cancer 2012 Jul;20(7):1377-1394. [doi: 10.1007/s00520-012-1450-7] [Medline: 22476399]
12. Mastrovito RC. Emotional considerations in cancer and stroke. In cancer. N Y State J Med 1972 Dec 01;72(23):2874-2877. [Medline: 4508407]
13. Chappy SL. Women's Experience With Breast Biopsy. AORN Journal 2004 Nov 01;80(5):885-901. [doi: 10.1016/s0001-2092(06)60511-5]
14. Liao M, Chen M, Chen S, Chen P. Uncertainty and anxiety during the diagnostic period for women with suspected breast cancer. Cancer Nurs 2008;31(4):274-283. [doi: 10.1097/01.NCC.0000305744.64452.fe] [Medline: 18600114]
15. Liu L, Li C, Tang ST, Huang C, Chiou A. Role of continuing supportive cares in increasing social support and reducing perceived uncertainty among women with newly diagnosed breast cancer in Taiwan. Cancer Nurs 2006;29(4):273-282. [Medline: 16871094]
16. 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]
17. Sheehan J, Sherman KA. Computerised decision aids: a systematic review of their effectiveness in facilitating high-quality decision-making in various health-related contexts. Patient Educ Couns 2012 Jul;88(1):69-86. [doi: 10.1016/j.pec.2011.11.006] [Medline: 22185961]
18. Ventura F, Öhlén J, Koinberg I. An integrative review of supportive e-health programs in cancer care. European Journal of Oncology Nursing 2013 Aug;17(4):498-507. [doi: 10.1016/J.EJON.2012.10.007]
19. Sagor R. Guiding School Improvement with Action Research. Alexandria, Va: ASCD; 2000:A.
20. Stringer E. Action Research. Thousands Oak, CA: Sage Publications; 2013.
21. Crawford J, Frisina A, Hack T, Parascandalo F. A Peer Health Educator Program for Breast Cancer Screening Promotion: Arabic, Chinese, South Asian, and Vietnamese Immigrant Women's Perspectives. Nurs Res Pract 2015;2015:947245 [FREE Full text] [doi: 10.1155/2015/947245] [Medline: 25810922]
22. Espenschied CR, MacDonald DJ, Culver JO, Sand S, Hurley K, Banks KC, et al. Closing the loop: action research in a multimodal hereditary cancer patient conference is an effective tool to assess and address patient needs. J Cancer Educ 2012 Jun;27(3):467-477 [FREE Full text] [doi: 10.1007/s13187-012-0373-9] [Medline: 22610836]
23. Macdonald DJ, Deri J, Ricker C, Perez MA, Ogaz R, Feldman N, et al. Closing the loop: an interactive action-research conference format for delivering updated medical information while eliciting Latina patient/family experiences and psychosocial needs post-genetic cancer risk assessment. Fam Cancer 2012 Sep;11(3):449-458 [FREE Full text] [doi: 10.1007/s10689-012-9535-5] [Medline: 22678665]
24. Beesley H, Goodfellow S, Holcombe C, Salmon P. The intensity of breast cancer patients' relationships with their surgeons after the first meeting: Evidence that relationships are not 'built' but arise from attachment processes. Eur J Surg Oncol 2016 May;42(5):679-684. [doi: 10.1016/j.ejso.2016.02.001] [Medline: 26944364]
25. Jimbo M, Kelly-Blake K, Sen A, Hawley ST, Ruffin MT. Decision Aid to Technologically Enhance Shared decision making (DATES): study protocol for a randomized controlled trial. Trials 2013 Nov 11;14:381 [FREE Full text] [doi: 10.1186/1745-6215-14-381] [Medline: 24216139]

26. Sherman KA, Kilby CJ, Shaw L, Winch C, Kirk J, Tucker K, et al. Facilitating decision-making in women undergoing genetic testing for hereditary breast cancer: BRECONDA randomized controlled trial results. Breast 2017 Dec;36:79-85. [doi: 10.1016/j.breast.2017.10.001] [Medline: 29031121]
27. Roberto A, Colombo C, Candiani G, Giordano L, Mantellini P, Paci E, et al. Personalised informed choice on evidence and controversy on mammography screening: study protocol for a randomized controlled trial. BMC Cancer 2017 Jun 19;17(1):429 [FREE Full text] [doi: 10.1186/s12885-017-3428-9] [Medline: 28629329]
28. Coe AM, Ueng W, Vargas JM, David R, Vanegas A, Infante K, et al. Usability Testing of a Web-Based Decision Aid for Breast Cancer Risk Assessment Among Multi-Ethnic Women. AMIA Annu Symp Proc 2016;2016:411-420 [FREE Full text] [Medline: 28269836]
29. Manne SL, Topham N, D'Agostino TA, Myers VS, Kirstein L, Brill K, et al. Acceptability and pilot efficacy trial of a web-based breast reconstruction decision support aid for women considering mastectomy. Psychooncology 2016 Dec;25(12):1424-1433. [doi: 10.1002/pon.3984] [Medline: 26383833]
30. Elkin EB, Pocus VH, Mushlin AI, Cigler T, Atoria CL, Polaneczky MM. Facilitating informed decisions about breast cancer screening: development and evaluation of a web-based decision aid for women in their 40s. BMC Med Inform Decis Mak 2017 Dec 21;17(1):29 [FREE Full text] [doi: 10.1186/s12911-017-0423-7] [Medline: 28327125]
31. Greenwood D, Levin M. Introduction to Action Research: Social Research for Social Change, Second Edition. Thousand Oaks, CA: Sage Publications; 2007.

Abbreviations

IT: information technology

Edited by G Eysenbach; submitted 11.05.18; peer-reviewed by D Young-Afat, Z Zhou; comments to author 05.09.18; revised version received 19.09.18; accepted 25.09.18; published 04.02.19.

Please cite as:
Hung YT, Wu CF, Liang TH, Chou SS, Chen GL, Wu PN, Su GR, Jang TH, Liu CY, Wang CY, Tseng LM, Sheu SJ
Developing a Decision-Aid Website for Breast Cancer Surgery: An Action Research Approach
J Med Internet Res 2019;21(2):e10404
URL: https://www.jmir.org/2019/2/e10404/
doi:10.2196/10404
PMID:30714941

Original Paper

The Digital Education to Limit Salt in the Home Program Improved Salt-Related Knowledge, Attitudes, and Behaviors in Parents

Durreajam Khokhar[1], M Human Nutrition; Caryl Anne Nowson[1], PhD; Claire Margerison[1], PhD; Madeline West[1], BSc (Hons); Karen J Campbell[1], PhD; Alison Olivia Booth[1], PhD; Carley Ann Grimes[1], PhD

Institute for Physical Activity and Nutrition Research, Deakin University, Waurn Ponds, Geelong, Australia

Corresponding Author:
Carley Ann Grimes, PhD
Institute for Physical Activity and Nutrition Research
Deakin University
Deakin University Locked Bag 20000
Waurn Ponds, Geelong, 3220
Australia
Phone: 61 39244 6223
Fax: 61 39244 6017
Email: carley.grimes@deakin.edu.au

Abstract

Background: Currently, Australian children and adults are eating too much salt, increasing their risk of cardiovascular-related conditions. Web-based programs provide an avenue to engage the parents of primary schoolchildren in salt-specific messages, which may positively impact their own salt-related knowledge, attitudes, and behaviors (KABs).

Objective: This pilot study aimed to determine whether parents' salt-related KABs improved following participation in the Digital Education to LImit Salt in the Home (DELISH) Web-based education program.

Methods: The DELISH program was a 5-week, home-delivered, Web-based intervention, with a pre- and posttest design, targeting schoolchildren aged 7 to 10 years and their parents. Parents received weekly Web-based educational newsletters and text messages and completed online pre- and postprogram surveys assessing salt-related KABs. Upon completion of the program, all parents were also invited to complete an online evaluation survey. Changes in KABs outcomes were assessed using McNemar tests and paired *t* tests.

Results: Of the 80 parents that commenced the program, 73 parents (mean age 41.0, SD 7.0 years; 86% (63/73) females) completed both pre- and postsurveys. Overall, mean score for salt-related knowledge improved (+3.6 [standard error (SE) 0.41] points), and mean behavior score also improved (+4.5 [SE 0.61] points), indicating a higher frequency of engaging in behaviors to reduce salt in the diet, and mean attitude score decreased (−0.7 [SE 0.19] points), representing lower importance of using salt to enhance the taste of food (all $P<.001$). Following participation, the proportion of parents aware of the daily salt intake recommendation increased from 40% (29/73) to 74% (54/73) ($P<.001$), and awareness of bread as the main source of salt increased from 58% (42/73) to 95% (69/73) ($P<.001$). The proportion of parents who agreed that salt should be used in cooking to enhance the flavor of food decreased from 30% (22/73) to 11% (8/73) ($P=.002$) and the proportion who agreed that sodium information displayed on food labels was difficult to understand decreased from 52% (38/73) to 32% (23/73) ($P=.009$). There was a reduction in the proportion of parents who reported adding salt during cooking (55% [40/73] vs 41% [30/73]; $P=.03$) and at the table (32% [23/73] vs 18% [13/73]; $P=.002$). Of the 16 parents who completed the evaluation survey, 75% (12/16) enjoyed the program, and all parents found the newsletters to be useful. Almost all parents (15/16, 94%) agreed that the DELISH program would be useful to other parents.

Conclusions: The improvement in salt-related KABs in the DELISH program indicates the potential for online technology, to disseminate simple salt reduction education messages to families with primary school–aged children. Future work should seek to improve the quality of data collected by including a larger sample size and a control group to integrate the program within the school setting to enable wider dissemination.

(J Med Internet Res 2019;21(2):e12234) doi:10.2196/12234

KEYWORDS

dietary sodium; knowledge; attitude; behavior; parent; internet; family; Australia

Introduction

Salt intakes among Australian adults [1] and children [2] exceed recommendations, and this has deleterious implications on cardiovascular health. It is well established that high salt intakes raise blood pressure in adults, [3] and when salt intake is reduced, there is a significant reduction in blood pressure [4]. A similar effect has been observed in children, with higher salt intakes associated with higher blood pressure [5,6]. This is of concern as blood pressure tracks across the life course, with higher levels in childhood associated with higher levels in adolescence and adulthood [7,8].

The early years of life are a crucial time of rapid physical growth and the development of eating behaviors, which aid in laying strong foundations for future food preferences and eating patterns [9,10]. During this period, children acquire knowledge and experience about what, when, and how much to eat. Parents have a strong influence on their children's eating behaviors through provision of food and modeling [9,11-13]. Moreover, parents' level of nutrition-related knowledge and attitudes toward diet informs the quality of foods purchased, prepared, and made accessible to their children [14-16]. There is consistent evidence of a positive association between parent and child intakes of fat, fruits, and vegetables [17,18]. Targeting parents' nutrition-related knowledge and attitudes to diet is likely to be an important means by which a child's diet can be improved [17-19].

Recent national data indicates that in 2016-17, 97% of households with children aged under 15 years had access to the internet in Australia [20]. This widespread access provides an opportunity for the dissemination of nutrition-related information via online methods [21,22]. Recent Australian evidence suggests that parents are interested in participating in online healthy life-style programs and want information delivered via websites, smartphone apps, and text messages [23]. Furthermore, parents want online programs that are easy to use, engaging, practical, and directly involve their children [23].

Parents have previously been included in school-based salt reduction education programs [24,25]. These interventions have demonstrated some improvements in children's knowledge regarding salt and health and food label reading [24,25]; however, these studies were conducted over 30 years ago. To date, there have been no Web-based family education programs specifically targeting salt reduction. Therefore, we developed and pilot tested the Digital Education to LImit Salt in the Home (DELISH) program, a 5-week, family-based salt-reduction education program targeting schoolchildren aged 7 to 10 years and their parents in the state of Victoria, Australia. The main outcomes relating to data collected from children in the DELISH program indicated positive changes in children's salt-related knowledge, self-efficacy, and reported discretionary salt use behaviors; however, there was no change in children's daily salt intake measured via 24-hour urinary sodium excretion [26].

In this study, we assessed changes in salt-related knowledge, attitudes, and behaviors (KABs) among the parents of primary schoolchildren who participated in the DELISH program.

Methods

Study Design

The DELISH study was a 5-week, home-delivered, Web-based salt reduction intervention targeting schoolchildren aged 7 to 10 years (grades 2-4) and their parents. This was a single-arm study with a pre- and posttest design. Full details of the methodology have been described elsewhere [27]. Ethics approval was granted by the Deakin University Human Research Ethics Committee (Project No. HEAG-H 37_2016) and the Department of Education and Early Childhood Development, Victoria State Government (Project No. HEAG-H 91/2015). Written informed consent from parents and assent from the child was obtained before participation in the program.

Inclusion Criteria

Parents with a child aged 7 to 10 years (grade 2-4) who attended a participating government primary school in the state of Victoria, Australia, were invited to participate in the study. Parents were required to have an email address to receive intervention materials, and their child was required to have access to a computer or an iPad with internet to view the child-specific intervention materials.

Recruitment of Study Participants

Children and their parents were recruited via primary schools. To conduct this, a Web-based school locator search engine was used to identify all schools (N=40) with enrollments for primary schoolchildren in the Greater Geelong area of Victoria. Following this, based on the school postcode and corresponding Index of Relative Socioeconomic Advantage and Disadvantage, schools were grouped into tertiles of socioeconomic status (SES) (ie, low, mid, and high). Schools were then randomly selected across tertiles of SES and invited to participate (n=38) [28]. Approximately 2 to 3 schools from each tertile were invited to participate in each round of invites sent out. The principal of each school was sent an email invitation to participate in the study, with a follow-up courtesy call reminder. Once the school principal accepted the invitation, school classroom teachers (grades 2-4) were provided with information about the study. An information session was held with children in grades 2 to 4, and study packs, including a plain language brochure and consent form for parents, were distributed. For logistical reasons related to data collection, we excluded schools that had a low level of interest from children (<9 returned consent forms).

Sample Size Calculations

The sample size was based on the primary outcome of a 20% reduction in dietary salt intake (1.2 g/day) in children postprogram participation, measured via 24-hour sodium excretion [27]. On the basis of this, it was estimated that 122

children and 122 parents (1 parent per child) would be recruited across 6 schools.

Key Behavioral Messages

Overall, 3 key behavioral messages were incorporated in the DELISH program:

1. *Stop* using the salt shaker at the table and during cooking.
2. *Switch* to lower salt foods by checking food labels (primarily focusing on breads, cereals, and cheese).
3. *Swap* processed salty foods (ie, processed meats, pizza, burgers, and savory sauces) for healthier alternatives.

The rationale for the behavioral messages and selection of targeted food groups has previously been described [27].

The program focused on constructs that have previously been shown to be related to dietary intake in children, such as self-efficacy [29], intentions (ie, goal setting) [30,31], reinforcements [32], and knowledge [33]. Strategies to address the intervention content were mapped to behavior change theory and cognitive theory constructs [34]. Behavior change techniques included in the intervention were selected based on the mode of delivery (ie, Web-based, with no face-to-face contact), techniques that had previously been used in effective interventions targeting children's eating behaviors [35,36], as well as strategies that had previously been effective in reducing salt intakes. For example, in adults and children, effective strategies to reduce salt intake included providing education on reading sodium information included on food labels, cooking recipes with spices and herbs, information on selecting low-sodium foods when eating out, and goal setting [37-40].

Parent Newsletters, Study Website, and Short Messaging Service Text Messages

Weekly newsletters were available online, with access to newsletters via a hyperlink sent to parents' email address, the study website, or via text messages (described below). The newsletters were available on the website for the duration of the study and could be downloaded in PDF format. Examples of some sections of the parent newsletters are shown in Figure 1, and the content of material delivered to parents is shown in Textbox 1. Educational material covered in each newsletter complemented the content delivered to the child for the relevant week of the program. Furthermore, parents were able to access extra resources (ie, video for reading food labels, supermarket cheat sheets to help find foods with less salt, and information on using herbs and spices) using the additional hyperlinks embedded within the newsletters. The content of the newsletters was reviewed by a dietitian and tested for language, layout, and graphics with 2 mothers of primary school–aged children.

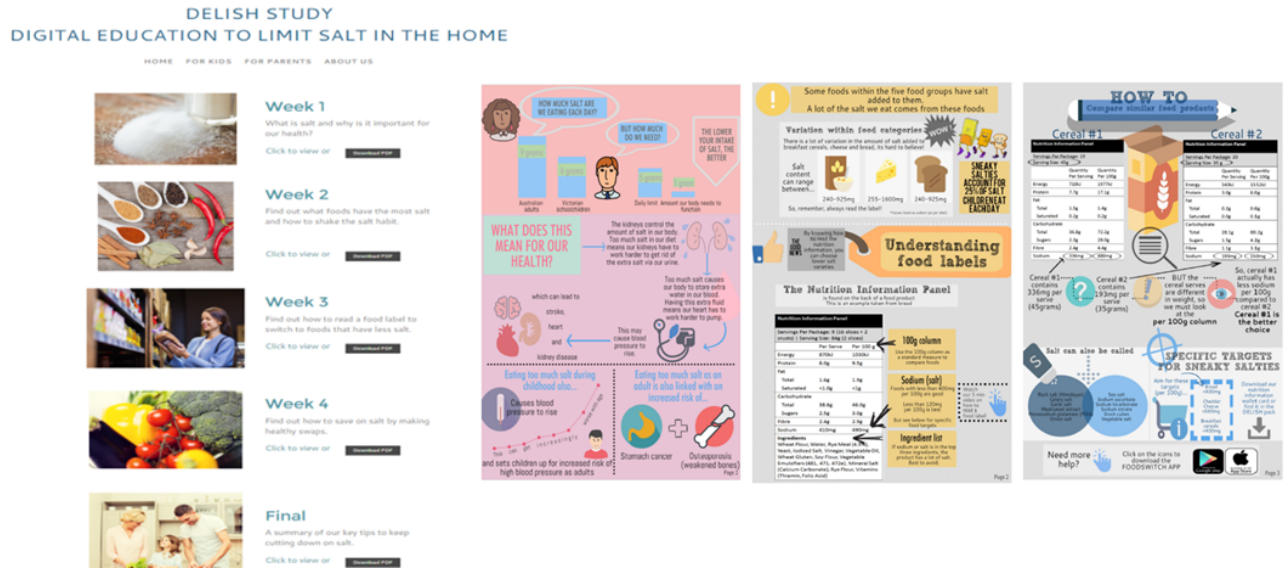
Study resources for both children and parents were accessible via a password-protected study website during the intervention (Figure 1). The website was updated each week with relevant resources. The parents' section contained access to weekly newsletters and additional resources such as the key messages of the program, information on how to read and interpret food labels, and healthy recipe ideas. Recipe resources including existing recipes from health agencies, such as the National Heart Foundation of Australia and World Action on Salt and Health, were used, and if necessary, instructions for modifying the recipe to reduce salt were provided.

Parents received 2 to 3 reminder SMS (short messaging service) text messages during each week of the intervention utilizing the Telstra Desktop Messaging service. These messages aimed to engage parents with education materials and provide information related to weekly key messages and goal setting. Example SMS messages included "Cutting down on salt, can help keep our blood pressure levels healthy and protect our hearts" and "Did you know? Bread, cereal and cheese account for 25% of salt children eat each day! Cut down on salt by checking the food label."

Web-Based Sessions for Children

Each week, parents received an email with access to the corresponding Web-based session for their child/children. These sessions were designed using the e-learning software Articulate Storyline 2 (Articulate Global, Inc). A detective theme was selected for the storyline of the Web-based sessions and included comic strips, characters to introduce key concepts, activities and games, video content, and sound effects. Each detective case file targeted the key learning objectives related to the behavioral messages of the intervention and would have to be solved to be awarded a reward badge.

Figure 1. Digital Education to LImit Salt in the Home (DELISH) study website and example of parental newsletters.



Textbox 1. Content of material delivered to parents during the Digital Education to LImit Salt in the Home (DELISH) program.

Week 1

- Newsletter key messages
  - Health consequences of eating too much salt
  - Recommended daily salt intake
  - Difference between salt and sodium
- Plus
  - 3 × SMS text messages

Week 2

- Newsletter key messages
  - Stop using salt during cooking and at the table
  - Main sources of salt in the diet (and top 7 sources of salt in children's diets)
- Plus
  - 3 × SMS text messages
- Extra resource
  - Additional information sheet on various herbs and spices as flavoring for cooking

Week 3

- Newsletter key messages
  - How to read sodium information on food labels and select foods with lower sodium content
  - Choosing "reduced sodium" or "no added salt" foods
  - Specific sodium targets when selecting breads, cheese, and breakfast cereals
- Plus
  - 3 × SMS text messages
- Extra resource
  - Video resource on how to read a food label (accessed via the study website or link on newsletter)

Week 4

- Newsletter key messages
  - Swap processed foods high in salt with healthier alternatives
  - Eat a variety of foods from the 5 food groups as per Australian Guide to Healthy Eating
  - Healthy lunch box and dinner swap ideas with lower salt content
- Plus
  - 3 × SMS text messages

Week 5

- Newsletter key messages
  - Key messages re-emphasized from past 4 weeks
- Plus
  - 3 × SMS text messages

Survey Instrument

An online questionnaire, containing 34 questions (preprogram) and 23 questions (postprogram), assessed salt-related KABs in parents (Multimedia Appendix 1). The same parent was asked to complete both pre- and postprogram surveys. Parents were instructed to report their unique assigned identification number at the start of both the pre- and postprogram surveys. Both surveys comprised identical salt-related KAB questions. However, in the postprogram survey, the demographic information was omitted, with the exception of 2 questions on parents' date of birth and gender. This information was included as a safeguard to cross-check and match surveys in the event that parents incorrectly reported their identification number. Furthermore, it was used to cross-check whether the same parent completed the pre- and postprogram surveys. The survey questions were modeled on a previously validated salt survey [41] as well as those used in previous salt-related surveys [42-50]. The questionnaire was tested for readability by 5 parents of primary school–aged children of varying demographic backgrounds.

Test-Retest Reliability

The test-retest reliability of the questionnaire was assessed in a separate sample of 43 parents, mean age 41.3 (SD 5.1) years, 95% (41/43) female, and recruited via a Facebook advertisement and flyers. Out of the total 49 KAB items assessed, 8 items (8/49, 16%) showed no agreement/poor agreement (kappa≤0-.20), 10 items (10/49, 21%) showed fair agreement (kappa=.21-.40), and 31 items (31/40, 63%) items showed moderate to perfect agreement (kappa=.41-1.00; Multimedia Appendix 2). The instrument displayed good to excellent test-retest reliability for construct scores, intraclass correlation coefficients: knowledge .76, attitude .70, and behavior .84 (Multimedia Appendix 3).

Demographic Characteristics

Information on demographic characteristics (13 questions) was collected as well as self-reported height and weight, which was used to calculate the body mass index (BMI) and categorize participants into weight categories [51]. SES was based on educational attainment, defined as (1) low: some or no level of high school education, (2) mid: technical/trade certificate or diploma, and (3) high: university/tertiary qualification. Data were collected on the number and age range of children living in the household. Parents with more than 1 child aged less than 18 years could select more than 1 age category.

Knowledge Related to Salt Intake and Total Knowledge Score

Overall, 13 questions, consisting of 32 individual items, assessed parents' salt-related knowledge. Of these, 11 questions assessed declarative knowledge asking about the relationship between salt and sodium, current salt intake recommendations, main sources of dietary salt, and health conditions associated with excess salt intake. In addition, 2 questions assessed procedural knowledge, which included interpreting sodium information across 3 food labels to select a bread and pasta sauce with the lowest salt content. Out of the 13 questions, 8 questions were in the form of multiple choice, 2 questions used 5-point Likert scales ("certainly wrong" to "certainly true" or "strongly disagree" to "strongly agree"), and 3 questions were closed-ended questions ("yes," "no," or "don't know"). All correct responses for multiple-choice and closed-ended questions were scored as 1, whereas incorrect responses including "don't know" or "not sure" were assigned a score of 0. To differentiate lack of knowledge from knowledge held with low levels of confidence, the first of the Likert scale questions was scored as a 2 for "certainly true," 1 for "probably true," and 0 for incorrect answers, including "not sure" and "don't know" responses. Negative statements were reversed before scoring [41]. The second Likert scale question referred to parents' knowledge about whether eating too much salt during childhood may have harmful effects on children's health and was scored as a 1 for "strongly agree/agree" and 0 for all other responses. Individual items were summed to generate a total salt knowledge score ranging from 0 to 38, with higher scores indicative of a higher level of salt-related knowledge (Multimedia Appendix 1).

Attitudes Related to Salt Intake and Total Attitude Score About the Importance of the Taste of Salt in Food

Overall, 4 questions, consisting of 6 individual items, addressed salt-related attitudes among parents. A multiple-choice question was used to assess parents' perceptions of their own salt intake (ie, "I eat less salt than recommended," "I eat the right amount of salt," "I eat more salt than recommended," or "I don't know"), and a 6-point Likert scale was used for assessing parents' perception of their child/children's salt intakes ("far too much" to "don't know"). Another question was related to the parents' attitude toward the importance of their child/children consuming foods with lower amounts of salt, with responses measured on a 5-point Likert scale ("not important at all" to "very important"). On the last 5-point Likert scale ("strongly disagree" to "strongly agree") question, parents were provided with 3 salt-related attitude statements regarding difficulty in interpreting sodium information on food labels, salt as a flavor enhancer, and taste of low-salt foods. For attitude questions, the response options were dichotomized such that the question with the response options "strongly agree/agree" were collapsed into 1 category, whereas "strongly disagree/disagree/neither agree nor disagree" were collapsed into another category. Similarly, the response options "very important/important" were collapsed into 1 category and "not important at all/not important/neither important nor unimportant" were collapsed into another category. An attitudes score was based on 2 items that assessed the addition of salt in food for taste [41]. Scores were assigned from 0 for "strongly disagree" to 4 for "strongly agree" [41,52]. Potential attitude scores were 0 to 8, with higher scores indicative of a stronger attitude about the importance of salt in food to enhance taste. No other attitude items were scored, as these questions did not reflect a common construct and used different scales for response options, rendering it nonsensical to include these questions in the total attitude score.

Behaviors Related to Salt Intake and Total Behavior Score

Overall, 4 questions, consisting of 11 individual items, addressed salt-related behaviors among parents—2 regarding discretionary salt use and 2 regarding salt reduction actions. The first question

asked parents about their frequency of 3 discretionary salt use behaviors (ie, adding salt at the table, adding salt during cooking, and placing a salt shaker on the table at meal times). The second question was related to the child's use of salt at the table (parent proxy reported). Both questions used a 5-point Likert scale with responses ranging from "always" to "never." A discretionary salt use score was made, based on the parents' 3 discretionary salt use behaviors (eg, table salt, cooking salt, and salt shaker on the table). Responses were scored as 0 for "always" to 4 for "never," and the total score ranged from 0 to 12; higher scores indicated lower frequency of discretionary salt use.

In addition, 1 question asked parents whether they had taken any action (currently on preprogram survey and in the past 5 weeks postprogram survey) to reduce the amount of salt their child/children eat. A score of 1 was assigned to "yes" and 0 to "no." Those responding "yes" were provided with a free-form response option to specify the actions taken. The final question asked about the frequency of 6 specific actions taken (currently on preprogram survey and in the past 5 weeks on postprogram survey) to reduce salt intake at home. A total of 4 positive behaviors (eg, cook meals from scratch, use herbs and spices as flavoring for cooking, look at food labels to check salt/sodium content, and purchase foods labeled "low salt/reduced salt/sodium") were scored with a 0 for "never," 1 for "1 time/month," 2 for "2 to 3 times/month," 3 for "1 to 2 times/week," 4 for "3 to 4 times/week," 5 for "5 to 6 times/week," and 6 for "1 time/day or more." Scoring was reversed for the 2 negative statements (eg, provide your child/children with processed meats such as ham or salami for lunch and use ready-made sauces, marinades, or mixes [eg, pasta sauce] for cooking). For descriptive reporting, we summarized change as the proportion of parents who reported a positive, negative, or no change in their salt reduction–related behaviors postprogram participation (Multimedia Appendix 1). A subscore for salt reduction–related behaviors ranged from 0 to 37, with higher scores indicating more favorable changes to reduce salt in the diet. A total salt behavior score (discretionary salt use and salt reduction–related behavior) was created by combining the subscores for discretionary salt use and salt reduction–related behaviors, which ranged from 0 to 49. Higher scores indicated better adherence to targeted salt-related behaviors.

Process Evaluation

A number of process evaluation measures were used to determine the acceptability of the DELISH program among parents and their child/children. Metrics regarding the number of page views and visitors to the study website were collated, along with the number of views for each week's online newsletters. At the completion of the program, all parents were invited via email (with 2 reminders) to complete an anonymous 13-question online evaluation survey. The survey assessed the acceptability of materials and any factors that might have hindered engagement in the program. The survey included a selection of Likert scale, closed-ended ("yes", "no", "maybe", or "don't know"), and open-ended questions (Multimedia Appendix 4).

Data Analysis

All data obtained from the questionnaires were analyzed using STATA SE (StataCorp LP) version 15.0. Descriptive statistics were performed for demographic variables with continuous (mean and SD or SE) and categorical variables (n and %). McNemar tests were performed to assess the change in proportion of parents' KAB survey responses at preprogram and postprogram. Paired *t* tests were used to assess the change in mean KAB construct scores between pre- and postprogram. Cohen *d* was calculated to determine the effect size for the change in each construct score. The normality of the KAB scores was determined using histograms and was deemed normal. A value of $P<.05$ was considered statistically significant.

Results

Demographic Characteristics

From the 38 schools invited, 5 government primary schools from the Greater Geelong area of Victoria agreed to participate in the DELISH program (school response rate 13% [5/38]). One school was excluded because of a low level of interest from children (n=2) wishing to participate in the study. Across the 4 participating schools, 98 parents (response rate 12% [4/98]) agreed to participate. A total of 80 parents commenced the DELISH program and 73 parents completed both the pre- and postprogram surveys. Moreover, 2 parents were excluded as the same parent (ie, different date of birth and gender reported in pre- and postprogram surveys) did not complete pre- and postprogram KAB surveys, and 5 parents were excluded as they only completed the preprogram survey. The majority of parents were females, and the mean age of the sample was 41 (SD 7) years (Table 1). Parents were predominantly born in Australia and spoke English as their primary language. Over half of the parents were classified as either overweight or obese. Overall, there was a reasonable spread of socioeconomic backgrounds (Table 1).

Table 1. Demographic characteristics of parents participating in the Digital Education to LImit Salt in the Home (DELISH) program.

Characteristics	Statistics
Gender (n=73), n (%)	
Male	10 (14)
Female	63 (86)
Age (years), mean (SD)	41.0 (7.0)
Country of birth (n=73), n (%)	
Australia	62 (85)
New Zealand	2 (3)
Other[a]	9 (12)
Language spoken (n=73), n (%)	
English	71 (97)
Other[b]	2 (3)
Socioeconomic status (SES)[c] (n=73)[d] , n (%)	
High SES	31 (43)
Mid SES	22 (31)
Low SES	19 (26)
Body mass index (kg/m^2)[e], mean (SD)	27.4 (6.6)
Weight category (n=73), n (%)	
Underweight	2 (3)
Healthy weight	27 (39)
Overweight	26 (37)
Obese	15 (21)
Main grocery shopper (n=73), n (%)	
Yes	56 (77)
No	4 (5)
I share the responsibility	13 (18)
Main meal preparer (n=73), n (%)	
Yes	56 (77)
No	5 (7)
I share the responsibility	12 (16)
Number of children living in household (n=73), n (%)	
1	7 (10)
2	38 (52)
3	19 (19)
4	7 (10)
5 or more	2 (3)
Number of children in age groups[f] (n=102), n (%)	
0 to 1 years	4 (3.9)
2 to 4 years	11 (10.8)
5 to 12 years	72 (70.6)
13 to 17 years	14 (13.7)
Over 18 years	1 (1.0)

Characteristics	Statistics
Diagnosed with cardiovascular related condition (n=73), n (%)	
Yes	10 (14)
No	63 (86)
Reported conditions included (n=10), n (%)	
Stroke	1 (10)
High blood pressure	7 (70)
Other[g]	2 (20)
Taking medication to control high blood pressure[h] (n=7), n (%)	
Yes	3 (43)
No	4 (57)

[a]Includes United Kingdom, Italy, Greece, China, Vietnam, Lebanon, and others, which include Fiji, Germany, Spain, and United States of America.

[b]Includes Italian, Greek, Cantonese, Mandarin, Arabic, Vietnamese, German, Spanish, Tagalog, and others, which includes Punjabi and Thari.

[c]SES: socioeconomic status.

[d]SES based on level of education; n=72 as 1 parent responded "prefer not to answer" in the level of education question.

[e]n=70 as 2 parents responded "I don't know" to the height question, and 1 parent responded "I don't know" to the weight question.

[f]n=102 as parents could select more than 1 age category.

[g]This was a free-form option, and responses included left ventricular noncompaction and high blood pressure during pregnancy.

[h]Question only presented to those that reported being diagnosed with high blood pressure, that is, n=7.

Table 2. Change in mean salt-related knowledge, attitudes, and behaviors scores pre- and postprogram (n=73).

Construct score	Preprogram, mean (SE[a])	Postprogram, mean (SE)	Change[b], mean (SE)	*P* value[c]	Effect size (Cohen *d*)
Knowledge	24.6 (0.52)	28.2 (0.50)	+3.6 (0.41)	*<.001*	1.02
Attitude	3.2 (0.19)	2.5 (0.17)	−0.7 (0.19)	*<.001*	0.44
Behavior	30.2 (0.62)	34.7 (0.71)	+4.5 (0.61)	*<.001*	0.87
Discretionary salt use behavior[d]	8.04 (0.31)	9.19 (0.24)	+1.1 (0.21)	*<.001*	0.65
Salt reduction-related behavior	18.9 (0.49)	22.0 (0.59)	+3.1 (0.48)	*<.001*	0.76

[a]SE: standard error.

[b]Change assessed via paired *t* test.

[c]Italicized values represent significance at $P<.05$.

[d]Includes parents' reported use of table salt, cooking salt, and salt shaker placed on table.

Salt-Related Knowledge, Attitudes, and Behaviors

After participation in the DELISH program, parents' mean scores for salt-related knowledge and behaviors (both discretionary salt use and salt reduction–related behaviors) significantly increased, indicating improvements in knowledge and adherence to targeted salt reduction behaviors. Conversely, mean attitude score significantly decreased, representing lower importance of using salt to enhance the taste of food (Table 2).

In regards to individual knowledge items, after participating in the program, the proportion of parents responding correctly improved in some items (10/32), whereas all other knowledge items remained unchanged (Table 3). For example, the proportion of parents aware of the dietary recommendations for salt in children and adults increased after participating in the program, but there was no change in the proportion of parents aware that excess salt intakes damage health overall and damages children's health. Parents' awareness of the link between excess salt intakes and high blood pressure was high at baseline (93%, [68/73]) and did not significantly change post intervention. However, there was an improvement in the proportion of parents aware of the link between excess salt intakes and kidney disease, heart disease/attack, stroke, and stomach cancer. With the exception of ham and cheddar cheese, there were no changes in the proportion of parents who correctly identified foods with added salt. However, a greater proportion of parents were aware that bread is a principal source of salt (Table 3). Post intervention, there was no change in the proportion of parents aware of most common salt-related misconceptions such as drinking more water neutralizes salt in the diet and reducing salt causes leg cramps; however, a greater proportion of parents were aware of the misconception that sea salt is better than table salt (Table 3).

Table 3. Proportion of parents correctly responding to salt-related knowledge items pre- and post Digital Education to LImit Salt in the Home (DELISH) program participation (n=73).

Knowledge questions[a]	Preprogram corrects, n (%)	Postprogram corrects, n (%)	*P* value[b,c]
Relationship between salt and sodium: salt contains sodium	28 (38)	32 (44)	.37
How much salt do Australians eat: far too much/too much	65 (89)	71 (97)	.07
Main source of salt in the Australian diet: salt from processed foods	69 (95)	70 (96)	>.99
Daily salt intake recommendation: 5 g/day	29 (40)	54 (74)	*<.001*
Eating too much salt could damage your health: correct response is yes	69 (95)	72 (99)	.37
Link between excess salt intake and blood pressure: correct response is yes	68 (93)	72 (99)	.22
Link between excess salt intake and kidney disease: correct response is yes	49 (67)	60 (82)	*.02*
Link between excess salt intake and heart disease/heart attack: correct response is yes	66 (90)	73 (100)	*.02*
Link between excess salt intake and stroke: correct response is yes	52 (71)	40 (96)	*<.001*
Link between excess salt intake and stomach cancer: correct response is yes	24 (33)	38 (52)	*.006*
Sea salt is better than table salt: correct response is probably wrong/certainly wrong	20 (27)	35 (48)	*<.001*
Fast foods are high in salt: correct response is probably true/certainly true	70 (96)	73 (100)	.25
Cutting down on salt causes leg cramps: correct response is probably wrong/certainly wrong	42 (58)	44 (60)	.66
Salt is naturally present in fresh food: correct response is probably true/certainly true	36 (49)	45 (62)	.07
Drinking more water can neutralize salt in the diet: correct response is probably wrong/certainly wrong	29 (40)	35 (48)	.24
Bread is one of the main sources of salt in Australians' diets: correct response is probably true/certainly true	42 (58)	69 (95)	*<.001*
Identify foods with added salt			
Ham (Yes)	64 (88)	71 (97)	*.04*
Tomato sauce (Yes)	71 (97)	73 (100)	.50
White rice (boiled; No)	54 (74)	51 (70)	.63
Beef steak (No)	52 (71)	57 (78)	.18
Mixed fresh vegetables (No)	66 (90)	64 (88)	.75
Bread (Yes)	72 (99)	73 (100)	>.99
Sausages (Yes)	69 (95)	71 (97)	.50
Corn flakes (Yes)	58 (79)	62 (85)	.39
Cheddar cheese (Yes)	57 (78)	71 (97)	*<.001*
Sausage roll (Yes)	71 (97)	72 (99)	>.99
Yoghurt (No)	45 (62)	42 (58)	.65
Correctly choose the lowest bread sodium content	53 (73)	59 (81)	.24

Knowledge questions[a]	Preprogram corrects, n (%)	Postprogram corrects, n (%)	*P* value[b,c]
Correctly choose the lowest pasta sauce sodium content	73 (100)	71 (97)	.50
Knowledge of recommended sodium content of bread: 400 mg/100 g	29 (40)	25 (34)	.43
Eating too much salt during childhood may have harmful effects on children's health: strongly agree/agree	59 (81)	64 (87)	.30
Salt intake recommendation for children aged 7 to 10 years: 5 g/day	22 (30)	39 (53)	*<.001*

[a]Correct responses to the knowledge items are provided after each of the italicized statements.

[b]McNemar test.

[c]Italicized values represent significance at $P<.05$.

With regard to individual attitude items, post program, parents were less likely to agree that salt should be used in cooking to enhance the flavor of food (30% [22/73] pre- and 11% [8/73] postprogram; $P=.002$); however, no change was observed in the proportion who agreed with the statement that in general, low-salt food tastes bad (12% [9/73] pre- and 10% [7/73] postprogram; $P=.75$). A lower proportion of parents agreed that sodium information displayed on food labels is difficult to understand (52% [38/73] pre- and 32% [23/73] postprogram; $P=.009$). However, there was no change in the proportion of parents who reported that it is important for their child/children to consume foods with lower amounts of salt (81% [59/73] pre- and 88% [64/73] postprogram; $P=.27$).

After participation in the program, parents were less likely to report engaging in all 4 discretionary salt use behaviors assessed (ie, parent adding salt to the food at the table and during cooking, placing a salt shaker on the table, and child adding salt to food at the table; Table 4).

After the program, more parents reported that they were taking action to reduce the amount of salt their child/children consumed (55% [40/73] pre- and 85% [62/73] postprogram; $P<.001$). The specific actions taken to reduce salt consumption were similar at both time points, with the most commonly reported actions including purchasing low-salt products, not adding salt during cooking and to food at the table, checking sodium content on food labels, cooking with fresh ingredients, providing fresh fruits and vegetables as an alternative to salty snacks, and limiting processed food.

Following the intervention, there were 3 salt reduction–related behaviors for which about half of the parents reported making a positive behavior change, these included checking food labels for salt/sodium content, purchasing foods labeled "no added salt," "salt reduced," or "reduced sodium," and providing their child/children with processed meats for lunch (Table 5). Comparatively, fewer parents reported positive changes for the other 3 salt reduction–related behaviors (Table 5). The frequency with which parents engaged in each individual salt reduction–related behavior pre- and postprogram (ie, 1 time/month and 2-3 times/month, etc) is detailed in Multimedia Appendix 5.

Parent Educational Material Metrics

In week 1, there was a total of 75 views for the first parent newsletter. There was a slight decline in total views for week 2 (67 views) and a further decline by week 3 (31 views). In week 2, an additional information sheet was made available to parents, which highlighted the use of various herbs and spices as flavoring for cooking. This sheet was viewed 8 times. In week 4, the total views increased to 55 views and then declined to 35 views for the final week 5 newsletter, which provided the overall key messages from the DELISH program. The DELISH website had a total of 120 page views in week 1, 87 in week 2, 48 in week 3, 34 in week 4, and 8 in week 5.

Postprogram Evaluation Survey

A total of 16 parents (response rate 22%) completed the postprogram evaluation survey. Overall, the majority (75%, 12/16) of parents reported that they enjoyed the program. Just over one-third (37%, 6/16) of parents reported viewing all 5 newsletters. All parents (n=16) reported that the information presented in the weekly newsletters was useful. The majority (88%, 14/16) of parents reported that the right number of text messages were sent throughout the duration of the program, and most (94%, 15/16) agreed that the text messages were helpful in prompting them to the information materials. Moreover, 44% (7/16) of parents reported visiting the website once a week throughout the program and 69% (11/16) of parents reported finding the information included on the website to be useful. The majority (88%, 14/16) of parents agreed that the education materials included in the program helped to reduce the amount of salt their children consumed. Overall, 81% of parents (13/16) agreed that the time required to complete the weekly activities was manageable and appropriate. Almost all parents (94%, 15/16) agreed that the DELISH education program would be useful to other parents and 69% (11/16) reported that they would recommend the DELISH program to others. A total of 44% of parents (7/16) reported that a barrier prevented them from viewing the weekly online education materials, with the most commonly reported barrier being time pressures. With regard to aspects of the program that parents liked, the most commonly reported was the engaging format for children (fun for children).

Table 4. Discretionary salt use reported by parents pre- and post Digital Education to LImit Salt in the Home (DELISH) program participation (n=73).

Behaviors	Always/usually/sometimes, n (%)		*P* value[a,b]
	Preprogram	Postprogram	
How often do you add salt to food at the table?	23 (32)	13 (18)	*.002*
How often do you add salt to food during cooking?	40 (55)	30 (41)	*.03*
How often do you place a salt shaker on the table at meal times?	26 (36)	14 (19)	*.003*
How often does your child/children add salt to their meal at the table?	14 (19)	5 (7)	*.01*

[a]McNemar test.

[b]Italicized values represent significance at $P<.05$.

Table 5. Proportion of parents reporting positive, negative, or no change in their salt reduction–related behaviors following participation in the Digital Education to LImit Salt in the Home (DELISH) program (n=73).

Behavior[a]	Statistics, n (%)
Provide your child/children with processed meats such as ham or salami for lunch	
Positive change in behavior post program	33 (45)
Negative change in behavior post program	13 (18)
No change in behavior post program	27 (37)
Cook meals from scratch with fresh ingredients	
Positive change in behavior post program	14 (19)
Negative change in behavior post program	12 (16)
No change in behavior post program	47 (64)
Use herbs and spices as flavoring for cooking	
Positive change in behavior post program	18 (25)
Negative change in behavior post program	17 (23)
No change in behavior post program	38 (52)
Use ready-made sauces, marinades, or mixes (eg, pasta sauce) for cooking	
Positive change in behavior post program	23 (32)
Negative change in behavior post program	12 (16)
No change in behavior post program	38 (52)
Look at a food label to check the salt/sodium content of a food item	
Positive change in behavior post program	39 (53)
Negative change in behavior post program	5 (7)
No change in behavior post program	29 (40)
Purchase foods labeled “no added salt,” “salt reduced,” or “reduced sodium”	
Positive change in behavior post program	38 (52)
Negative change in behavior post program	17 (23)
No change in behavior post program	18 (25)

[a]For positive behaviors, a positive change was defined as parents moving from engaging in a less frequent category preprogram to engaging in a more frequent category postprogram for a particular salt reduction–related behavior (or from a more frequent preprogram to a less frequent category post program for negative behaviors). For positive behaviors, a negative change was defined as parents moving from engaging in a more frequent category preprogram to engaging in a less frequent category post program for a particular salt reduction–related behavior (or from a less frequent preprogram to a more frequent category post program for negative behaviors).

Discussion

Summary of Key Findings and Comparison With Previous Literature

Participation in the DELISH Web-based salt reduction education program resulted in an improvement in salt-related knowledge, attitudes, and self-reported behaviors among parents. Furthermore, parents reported that they found the content and the method of delivery of the program to be useful in helping to reduce salt at home with support of the usefulness of such a program to other families.

Post DELISH program participation, parents specifically showed an improvement in the knowledge of daily salt intake recommendation, both for adults as well as children and the links between excess salt intake and heart disease/attack, stroke, kidney disease, and stomach cancer, with the latter 2 in general being less well-known conditions [53-56]. Similarly, participation in the program resulted in an increased proportion of parents who knew that bread is one of the main sources of salt as well as the number of parents who knew that ham and cheddar cheese contained added salt. Both foods were key target foods covered in the program.

Although overall mean salt-related knowledge score improved, there were a number of individual knowledge items that did not change post program participation, reasons for which might be two-fold. First, 2 questions related to misconceptions about salt, which showed no change (ie, reducing salt causes leg cramps and drinking more water can neutralize salt) were taken from a previously validated salt knowledge questionnaire [41]. These areas were not covered within the DELISH education materials. Second, baseline knowledge was high (>80% correct) for a number of items, for example, processed foods being the main source of salt in the diet, high intake of salt among Australians, and link between excess salt and adverse health. Similarly, preprogram, a high proportion of parents (approximately 80% or more) could already identify some foods with added salt such as tomato sauce, bread, sausages, corn flakes, and sausage rolls, and most were able to use the sodium information displayed on nutrition information panels to correctly choose a lower salt bread and pasta sauce. With regard to the findings related to use of nutrition information panels to select lower salt bread/pasta sauce, the poor test-retest reliability of these questions should also be considered when interpreting these findings. It appears that these questions may not have been challenging enough for parents to answer, and future studies utilizing this survey should consider revising these questions. Other studies have reported that much lower proportions of consumers are able to use sodium information on food labels to select lower salt options. A study by Grimes et al (n=474) found that less than half (42%) of Victorian consumers, aged 18 to 65 years, accurately used the nutrition information panel to rank 3 types of bread from lowest to highest salt content [46], whereas in a qualitative study involving 16 adult grocery shoppers, no participant had the background knowledge required to interpret information about salt on existing food labels, with most (10/16) only able to recognize that salt content was labeled as sodium [57]. Although our study showed no change in parents' ability to read food labels to choose lower salt foods, we did find that postprogram significantly fewer parents reported difficulty in understanding the sodium information displayed on food labels (52% pre- and 32% postprogram). This is further reflected in the proportion of parents who reported making a positive change in relation to specific salt reduction–related behaviors, whereby over half of the parents reported checking food labels for salt/sodium content and purchasing foods labeled "no added salt," "salt reduced," or "reduced sodium" compared with before the program.

After participation in the DELISH program, mean attitude score related to importance of salt in food for taste decreased, and it appears that the individual attitude item, which was the main driver of this, was the reduction in the proportion of parents agreeing that salt should be used in cooking to enhance the flavor of food. This may be a result of the education materials delivered to parents during the program to equip them with skills about using herbs and spices as alternatives to adding salt to flavor their food. There was no change in the proportion of parents who agreed that in general, low-salt food tastes bad, and the level of agreement for this attitude item was already low (approximately 12% agreement) at the start of the program. Previous studies have shown that Australian adults' perception about the importance of the taste of salt in foods is positively associated with their discretionary salt use behaviors [52]. Modifying parents' perception about the importance of the taste of salt in foods may be a potential means of inducing changes in salt use behaviors, which in turn might lead to the acceptance and preference of lower salt products [58,59]. With regard to the other attitude item, there was no change in parents' attitude toward their child consuming foods with lower amounts of salt; however, this proportion was already high (81%) at the start of the program and is perhaps not unexpected, given their agreement to participate in a salt reduction education program. In our previous findings in a larger sample of parents (n=837), a similar proportion of parents (70%) reported that limiting the amount of salt their child eats is important [60].

Parents showed an improvement in all discretionary salt-related behaviors post DELISH program, including reporting lower salt use at the table by their child/children postprogram. Although discretionary salt use is not the main source of salt in the Australian diet (accounting for 10%-15% of salt consumed [61,62]), this improvement highlights adherence to one of the program's key behavioral messages (ie, stop using the salt shaker on the table and during cooking). Avoiding adding salt during cooking and at the table were some of the more commonly reported actions taken by parents to help reduce salt use at home. Our findings are, therefore, indicative of the positive impact of educational messages on improving salt-related behaviors among families and are similar to a home-based salt reduction education program conducted in Finland [25]. In this Finnish study of 58 parents and their 13 year old children (n=58), 78% of parents reported reducing salt added to food prepared at home after a 5-hour course on reducing salt intake by changing food habits [25].

Despite the positive changes in reported discretionary salt use and salt reduction–related behaviors in parents, we have previously reported that there was no change in the amount of

salt children consumed post program as determined by 24-hour urinary sodium excretion [26]. Dietary salt reduction through behavior change alone may be difficult given that salt is ubiquitous in the food supply [63]. Therefore, a multifaceted salt reduction strategy, which includes regulatory or structural strategies that drive food reformulation and education strategies to improve nutrition labeling on packaged foods is needed to support parents in making better choices for their family [64]. The complexity of salt reduction has been highlighted in a pilot study by Lofthouse et al, which assessed the feasibility of adhering to a low sodium diet in 11 healthy adults who received nutritional counseling and support for the 4-week study [65]. Participants completed a 24-hour urine collection at baseline and follow-up, and although mean sodium intake reduced over the 4-week period, results from semistructured interviews with participants revealed that following a low-salt diet required extensive changes to the types of foods purchased and prepared [65]. In addition, the lack of control of the sodium content of foods consumed outside the home, the complexity and time consumption of deciphering sodium information of food labels, and identifying low-salt snacks, were some barriers to adherence [65]. Therefore, although salt reduction is feasible, substantial commitment and changes to both eating behaviors and to the food environment are required.

Limitations

Limitations of this study included the small sample size and recruitment of participants only from Geelong, a regional city in the state of Victoria, Australia. We also acknowledge the low (12%) response rate of overall parent participation in the program. Similarly, the low response rate (22%) of the postprogram evaluation survey should be interpreted with caution as it is possible that this survey might have been completed by those parents who were more engaged with the program. These limitations, coupled with an over-representation of females and those from high socioeconomic backgrounds, indicates that our findings cannot be generalized to the wider population. Self-reported behaviors by parents may also be susceptible to social desirability bias. Although the DELISH program was shown to be a feasible platform for the delivery of simple, salt reduction messages, a lack of control group means no causality between the intervention and KAB outcomes can be inferred. It is possible that other external environmental factors may have contributed to the improvement in salt-related KABs. Furthermore, participants in this study may have been more responsive to being educated about salt as they had completed the preprogram KABs survey, possibly priming them to behavior change. However, as this was a pilot study, we recommend that this study is replicated in the future to include a control group to improve the quality of data collected as well as the conclusions made about the effectiveness of such a program. A strength of the study was the use of a questionnaire developed for the intervention, which was based on a previously validated salt knowledge questionnaire. On the basis of the test-retest reliability of the questionnaire used in this study, those items that showed poor agreement (the link between salt and health in general and in children, awareness that fast foods are high in salt, identifying food items with added salt such as tomato sauce and without such as uncooked beef steak, comparing sodium information on food labels to choose a bread and pasta sauce with the lowest salt content, and awareness of the recommended sodium content of bread) should be interpreted with caution. These items may need to be revised if they are to be used in future surveys. However, 31 out of the total 49 items (63%) items showed moderate to perfect agreement and importantly, the instrument displayed overall good to excellent test-retest reliability for the KAB construct scores.

Conclusions

This study has shown that a home-delivered, Web-based education program delivered to parents of primary schoolchildren may improve salt-related KABs of these parents. In this program, parents were provided with simple resources to help them, and their children, to make positive salt-related changes in the home environment. This was the first study to develop and pilot test the feasibility of using online technology as a potential platform to disseminate simple salt reduction education messages to families with primary school–age children. Future work should explore opportunities to replicate this study with a control group and to integrate the program within the school setting for wider dissemination and to determine whether changes in salt-related knowledge, attitudes, and behaviors among parents can be sustained in the long term.

Acknowledgments

This study was funded by the Heart Foundation (Vanguard Grant Application 100574). During the development of the program, CAG was supported by a Heart Foundation Postdoctoral Fellowship (Award 100155). DK is supported by the Deakin University Postgraduate Research Scholarship, and CAG was supported by an Alfred Deakin Postdoctoral Fellowship. The authors would like to acknowledge the Department of Education and Training, Victoria State Government for their support in allowing recruitment and data collection to be conducted within the government school sector. They thank all participating schools and families. The authors wish to acknowledge research assistants Carrie Service and Maree Bouterakos for their contribution to the development of intervention resources, Anne Griffiths in her role as project coordinator, and all the parents who participated in the program with their children.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Digital Education to LImit Salt in the Home-parent knowledge, attitudes, and behaviors questionnaire.

[PDF File (Adobe PDF File), 237KB - jmir_v21i2e12234_app1.pdf]

Multimedia Appendix 2

Test-retest reliability results of parents for knowledge, attitude and behavior items using Kappa coefficients (n=43).

[PDF File (Adobe PDF File), 47KB - jmir_v21i2e12234_app2.pdf]

Multimedia Appendix 3

Test-retest reliability results of parents for total construct scores using intraclass correlation co-efficients (n=43).

[PDF File (Adobe PDF File), 36KB - jmir_v21i2e12234_app3.pdf]

Multimedia Appendix 4

Digital Education to LImit Salt in the Home-parent postprogram evaluation survey.

[PDF File (Adobe PDF File), 27KB - jmir_v21i2e12234_app4.pdf]

Multimedia Appendix 5

Change in the frequency of engaging in salt reduction-related behaviors pre- and postprogram participation (n=73).

[PDF File (Adobe PDF File), 29KB - jmir_v21i2e12234_app5.pdf]

References

1. Land MA, Neal BC, Johnson C, Nowson CA, Margerison C, Petersen KS. Salt consumption by Australian adults: a systematic review and meta-analysis. Med J Aust 2018 Feb 05;208(2):75-81. [doi: 10.5694/mja17.00394] [Medline: 29385968]
2. Grimes CA, Riddell LJ, Campbell KJ, Beckford K, Baxter JR, He FJ, et al. Dietary intake and sources of sodium and potassium among Australian schoolchildren: results from the cross-sectional Salt and Other Nutrients in Children (SONIC) study. BMJ Open 2017 Oct 30;7(10):e016639 [FREE Full text] [doi: 10.1136/bmjopen-2017-016639] [Medline: 29084791]
3. Mente A, O'Donnell MJ, Rangarajan S, McQueen MJ, Poirier P, Wielgosz A, PURE Investigators. Association of urinary sodium and potassium excretion with blood pressure. N Engl J Med 2014 Dec 14;371(7):601-611. [doi: 10.1056/NEJMoa1311989] [Medline: 25119606]
4. He FJ, Li J, Macgregor GA. Effect of longer term modest salt reduction on blood pressure: Cochrane systematic review and meta-analysis of randomised trials. Br Med J 2013;346:f1325 [FREE Full text] [doi: 10.1136/bmj.f1325] [Medline: 23558162]
5. He FJ, Marrero NM, Macgregor GA. Salt and blood pressure in children and adolescents. J Hum Hypertens 2008 Jan;22(1):4-11. [doi: 10.1038/sj.jhh.1002268] [Medline: 17823599]
6. Leyvraz M, Chatelan A, da Costa BR, Taffé P, Paradis G, Bovet P, et al. Sodium intake and blood pressure in children and adolescents: a systematic review and meta-analysis of experimental and observational studies. Int J Epidemiol 2018 Jun 27;47(6):1796-1810. [doi: 10.1093/ije/dyy121] [Medline: 29955869]
7. Chen X, Wang Y. Tracking of blood pressure from childhood to adulthood: a systematic review and meta-regression analysis. Circulation 2008 Jun 24;117(25):3171-3180 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.730366] [Medline: 18559702]
8. Cheng S, Xanthakis V, Sullivan LM, Vasan RS. Blood pressure tracking over the adult life course: patterns and correlates in the Framingham heart study. Hypertension 2012 Dec;60(6):1393-1399 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.112.201780] [Medline: 23108660]
9. Savage JS, Fisher JO, Birch LL. Parental influence on eating behavior: conception to adolescence. J Law Med Ethics 2007;35(1):22-34 [FREE Full text] [doi: 10.1111/j.1748-720X.2007.00111.x] [Medline: 17341215]
10. Birch L, Savage JS, Ventura A. Influences on the development of children's eating behaviours: from infancy to adolescence. Can J Diet Pract Res 2007;68(1):s1-56 [FREE Full text] [Medline: 19430591]
11. Birch LL, Davison KK. Family environmental factors influencing the developing behavioral controls of food intake and childhood overweight. Pediatr Clin North Am 2001 Aug;48(4):893-907. [doi: 10.1016/S0031-3955(05)70347-3] [Medline: 11494642]
12. Scaglioni S, Salvioni M, Galimberti C. Influence of parental attitudes in the development of children eating behaviour. Br J Nutr 2008 Feb;99(S1):S22-S25. [doi: 10.1017/S0007114508892471] [Medline: 18257948]

13. Burrows T, Warren JM, Collins CE. The impact of a child obesity treatment intervention on parent child-feeding practices. Int J Pediatr Obes 2010;5(1):43-50. [doi: 10.3109/17477160902957158] [Medline: 19437180]
14. Gibson EL, Wardle J, Watts CJ. Fruit and vegetable consumption, nutritional knowledge and beliefs in mothers and children. Appetite 1998 Oct;31(2):205-228. [doi: 10.1006/appe.1998.0180] [Medline: 9792733]
15. Gable S, Lutz S. Nutrition socialization experiences of children in the Head Start program. J Am Diet Assoc 2001 May;101(5):572-577. [doi: 10.1016/S0002-8223(01)00143-2] [Medline: 11374352]
16. Adamo KB, Brett KE. Parental perceptions and childhood dietary quality. Matern Child Health J 2014 May;18(4):978-995. [doi: 10.1007/s10995-013-1326-6] [Medline: 23817727]
17. Holsten JE, Deatrick JA, Kumanyika S, Pinto-Martin J, Compher CW. Children's food choice process in the home environment. A qualitative descriptive study. Appetite 2012 Feb;58(1):64-73. [doi: 10.1016/j.appet.2011.09.002] [Medline: 21986183]
18. Knowlden AP, Sharma M. Systematic review of family and home-based interventions targeting paediatric overweight and obesity. Obes Rev 2012 Jun;13(6):499-508. [doi: 10.1111/j.1467-789X.2011.00976.x] [Medline: 22221298]
19. Golan M. Parents as agents of change in childhood obesity--from research to practice. Int J Pediatr Obes 2006;1(2):66-76. [doi: 10.1080/17477160600644272] [Medline: 17907317]
20. Australian Bureau of Statistics. Household Use of Information Technology, Australia, 2016-17 URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0 [accessed 2018-09-16] [WebCite Cache ID 72UDsjes7]
21. Burrows TL, Khambalia AZ, Perry R, Carty D, Hendrie GA, Allman-Farinelli MA, et al. Great 'app-eal' but not there yet: a review of iPhone nutrition applications relevant to child weight management. Nutr Diet 2015 Apr 16;72(4):363-367. [doi: 10.1111/1747-0080.12184]
22. An JY, Hayman LL, Park YS, Dusaj TK, Ayres CG. Web-based weight management programs for children and adolescents: a systematic review of randomized controlled trial studies. ANS Adv Nurs Sci 2009;32(3):222-240. [doi: 10.1097/ANS.0b013e3181b0d6ef] [Medline: 19707091]
23. Burrows T, Hutchesson M, Chai LK, Rollo M, Skinner G, Collins C. Nutrition interventions for prevention and management of childhood obesity: what do parents want from an eHealth program? Nutrients 2015 Dec 15;7(12):10469-10479 [FREE Full text] [doi: 10.3390/nu7125546] [Medline: 26694456]
24. Luepker RV, Perry CL, Murray DM, Mullis R. Hypertension prevention through nutrition education in youth: a school-based program involving parents. Health Psychol 1988;7(Suppl):233-245. [doi: 10.1037/0278-6133.7.Suppl.233] [Medline: 3243221]
25. Pietinen P, Ruotsalainen P, Tanskanen A, Puska P. Sodium intake reduction in volunteer families by using a salt substitute and nutrition counselling. Ann Nutr Metab 1981;25(6):371-380. [doi: 10.1159/000176518] [Medline: 7332315]
26. Grimes CA, Booth A, Khokhar D, West M, Margerison C, Campbell KJ, et al. Digital Education to Limit Salt in the Home (DELISH) program improves knowledge, self-efficacy, and behaviors among children. J Nutr Educ Behav 2018 Jun;50(6):547-554. [doi: 10.1016/j.jneb.2018.04.002] [Medline: 29886898]
27. Grimes CA, Booth A, Khokhar D, West M, Margerison C, Campbell K, et al. The development of a web-based program to reduce dietary salt intake in schoolchildren: study protocol. JMIR Res Protoc 2017 May 31;6(5):e103 [FREE Full text] [doi: 10.2196/resprot.7597] [Medline: 28566266]
28. Australian Bureau of Statistics. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/8C5F5BB699A0921CCA258259000BA619?opendocument [accessed 2018-09-17] [WebCite Cache ID 72UYtozmx]
29. Cerin E, Barnett A, Baranowski T. Testing theories of dietary behavior change in youth using the mediating variable model with intervention programs. J Nutr Educ Behav 2009;41(5):309-318. [doi: 10.1016/j.jneb.2009.03.129] [Medline: 19717113]
30. Cullen KW, Baranowski T, Smith SP. Using goal setting as a strategy for dietary behavior change. J Am Diet Assoc 2001 May;101(5):562-566. [doi: 10.1016/S0002-8223(01)00140-7] [Medline: 11374350]
31. McClain AD, Chappuis C, Nguyen-Rodriguez ST, Yaroch AL, Spruijt-Metz D. Psychosocial correlates of eating behavior in children and adolescents: a review. Int J Behav Nutr Phys Act 2009;6:54 [FREE Full text] [doi: 10.1186/1479-5868-6-54] [Medline: 19674467]
32. Lowe CF, Horne PJ, Tapper K, Bowdery M, Egerton C. Effects of a peer modelling and rewards-based intervention to increase fruit and vegetable consumption in children. Eur J Clin Nutr 2004 Mar;58(3):510-522. [doi: 10.1038/sj.ejcn.1601838] [Medline: 14985691]
33. van Stralen MM, Yildirim M, te Velde SJ, Brug J, van Mechelen W, Chinapaw MJ, ENERGY-consortium. What works in school-based energy balance behaviour interventions and what does not? A systematic review of mediating mechanisms. Int J Obes (Lond) 2011 Oct;35(10):1251-1265 [FREE Full text] [doi: 10.1038/ijo.2011.68] [Medline: 21487398]
34. 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]
35. Hendrie GA, Brindal E, Corsini N, Gardner C, Baird D, Golley RK. Combined home and school obesity prevention interventions for children: what behavior change strategies and intervention characteristics are associated with effectiveness? Health Educ Behav 2012 Apr;39(2):159-171. [doi: 10.1177/1090198111420286] [Medline: 21984691]

36. Golley RK, Hendrie GA, Slater A, Corsini N. Interventions that involve parents to improve children's weight-related nutrition intake and activity patterns-what nutrition and activity targets and behaviour change techniques are associated with intervention effectiveness? Obes Rev 2011 Feb;12(2):114-130. [doi: 10.1111/j.1467-789X.2010.00745.x] [Medline: 20406416]
37. Ireland DM, Clifton PM, Keogh JB. Achieving the salt intake target of 6 g/day in the current food supply in free-living adults using two dietary education strategies. J Am Diet Assoc 2010 May;110(5):763-767. [doi: 10.1016/j.jada.2010.02.006] [Medline: 20430138]
38. Anderson CA, Cobb LK, Miller ER, Woodward M, Hottenstein A, Chang AR, et al. Effects of a behavioral intervention that emphasizes spices and herbs on adherence to recommended sodium intake: results of the SPICE randomized clinical trial. Am J Clin Nutr 2015 Sep;102(3):671-679 [FREE Full text] [doi: 10.3945/ajcn.114.100750] [Medline: 26269371]
39. Cotter J, Cotter MJ, Oliveira P, Cunha P, Polónia J. Salt intake in children 10-12 years old and its modification by active working practices in a school garden. J Hypertens 2013 Oct;31(10):1966-1971. [doi: 10.1097/HJH.0b013e328363572f] [Medline: 24107730]
40. He FJ, Wu Y, Ma J, Feng X, Wang H, Zhang J, et al. A school-based education programme to reduce salt intake in children and their families (School-EduSalt): protocol of a cluster randomised controlled trial. BMJ Open 2013;3(7):- [FREE Full text] [doi: 10.1136/bmjopen-2013-003388] [Medline: 23864214]
41. Sarmugam R, Worsley A, Flood V. Development and validation of a salt knowledge questionnaire. Public Health Nutr 2014 May;17(5):1061-1068. [doi: 10.1017/S1368980013000517] [Medline: 23507427]
42. Arcand J, Mendoza J, Qi Y, Henson S, Lou W, L'Abbe MR. Results of a national survey examining Canadians' concern, actions, barriers, and support for dietary sodium reduction interventions. Can J Cardiol 2013 May;29(5):628-631 [FREE Full text] [doi: 10.1016/j.cjca.2013.01.018] [Medline: 23489574]
43. The Secretariat of the Australian Division of World Action on Salt and Health. 2007 Survey of Australian Consumer Awareness and Practices Relating to Salt Report URL: http://www.awash.org.au/wp-content/uploads/2012/10/AWASH_ConsumerSurveyReport_2007_05_15.pdf [accessed 2018-09-17] [WebCite Cache ID 72UdyqakU]
44. Consensus Action on Salt and Health. TNS Public Opinion Survey Summary Report 2010 URL: http://www.actiononsalt.org.uk/media/action-on-salt/news/surveys/2010/public-opinion/Salt-Awareness-Week-2010-presentation.pdf [accessed 2018-09-17] [WebCite Cache ID 72UaPoakR]
45. Food Standards Australia New Zealand. Consumer Attitudes Survey 2007 URL: https://www.foodstandards.gov.au/publications/documents/Consumer%20Attitudes%20Survey.pdf [accessed 2018-09-17] [WebCite Cache ID 72UeFKSMj]
46. Grimes CA, Riddell LJ, Nowson CA. Consumer knowledge and attitudes to salt intake and labelled salt information. Appetite 2009 Oct;53(2):189-194. [doi: 10.1016/j.appet.2009.06.007] [Medline: 19540891]
47. New Zealand Food Safety Authority. Salt Consumer Survey URL: https://www.foodsafety.govt.nz/elibrary/industry/salt-survey.pdf [accessed 2018-09-17] [WebCite Cache ID 72UacaYqU]
48. Newson RS, Elmadfa I, Biro G, Cheng Y, Prakash V, Rust P, et al. Barriers for progress in salt reduction in the general population. An international study. Appetite 2013 Dec;71:22-31 [FREE Full text] [doi: 10.1016/j.appet.2013.07.003] [Medline: 23891557]
49. Papadakis S, Pipe AL, Moroz IA, Reid RD, Blanchard CM, Cote DF, et al. Knowledge, attitudes and behaviours related to dietary sodium among 35- to 50-year-old Ontario residents. Can J Cardiol 2010 May;26(5):e164-e169 [FREE Full text] [doi: 10.1016/S0828-282X(10)70384-6] [Medline: 20485696]
50. World Health Organization. WHO Steps Instrument (Core and Expanded) URL: http://www.who.int/ncds/surveillance/steps/instrument/STEPS_Instrument_V3.1.pdf [accessed 2018-09-17] [WebCite Cache ID 72Uck56jT]
51. World Health Organization. Body Mass Index- BMI URL: http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi [accessed 2018-09-17] [WebCite Cache ID 72UYEwTwx]
52. Sarmugam R, Worsley A, Wang W. An examination of the mediating role of salt knowledge and beliefs on the relationship between socio-demographic factors and discretionary salt use: a cross-sectional study. Int J Behav Nutr Phys Act 2013 Feb 19;10:25 [FREE Full text] [doi: 10.1186/1479-5868-10-25] [Medline: 23418906]
53. Wyllie AM, Brown R. Internet. Salt consumer survey URL: https://www.foodsafety.govt.nz/elibrary/industry/salt-survey.pdf [accessed 2018-09-17] [WebCite Cache ID 72UeYDTeW]
54. Nasreddine L, Akl C, Al-Shaar L, Almedawar MM, Isma'eel H. Consumer knowledge, attitudes and salt-related behavior in the Middle-East: the case of Lebanon. Nutrients 2014 Nov 13;6(11):5079-5102 [FREE Full text] [doi: 10.3390/nu6115079] [Medline: 25401502]
55. Kim MK, Lopetcharat K, Gerard PD, Drake MA. Consumer awareness of salt and sodium reduction and sodium labeling. J Food Sci 2012 Sep;77(9):S307-S313. [doi: 10.1111/j.1750-3841.2012.02843.x] [Medline: 22957915]
56. Grimes CA, Kelley S, Stanley S, Bolam B, Webster J, Khokhar D, et al. Knowledge, attitudes and behaviours related to dietary salt among adults in the state of Victoria, Australia 2015. BMC Public Health 2017 Dec 30;17(1):532 [FREE Full text] [doi: 10.1186/s12889-017-4451-0] [Medline: 28558745]
57. McLean R, Hoek J. Sodium and nutrition labelling: a qualitative study exploring New Zealand consumers' food purchasing behaviours. Public Health Nutr 2014 May;17(5):1138-1146. [doi: 10.1017/S1368980013001079] [Medline: 23651916]

58. Bertino M, Beauchamp GK, Engelman K. Long-term reduction in dietary sodium alters the taste of salt. Am J Clin Nutr 1982 Dec;36(6):1134-1144. [doi: 10.1093/ajcn/36.6.1134] [Medline: 7148734]
59. Blais CA, Pangborn RM, Borhani NO, Ferrell MF, Prineas RJ, Laing B. Effect of dietary sodium restriction on taste responses to sodium chloride: a longitudinal study. Am J Clin Nutr 1986 Aug;44(2):232-243. [doi: 10.1093/ajcn/44.2.232] [Medline: 3728360]
60. Khokhar D, Nowson CA, Margerison C, Bolam B, Grimes CA. Knowledge and attitudes are related to selected salt-specific behaviours among Australian parents. Nutrients 2018 Jun 04;10(6):- [FREE Full text] [doi: 10.3390/nu10060720] [Medline: 29867025]
61. James WP, Ralph A, Sanchez-Castillo CP. The dominance of salt in manufactured food in the sodium intake of affluent societies. Lancet 1987 Feb 21;1(8530):426-429. [Medline: 2880223]
62. Brown IJ, Tzoulaki I, Candeias V, Elliott P. Salt intakes around the world: implications for public health. Int J Epidemiol 2009 Jun;38(3):791-813 [FREE Full text] [doi: 10.1093/ije/dyp139] [Medline: 19351697]
63. Cobb LK, Appel LJ, Anderson CA. Strategies to reduce dietary sodium intake. Curr Treat Options Cardiovasc Med 2012 Aug;14(4):425-434 [FREE Full text] [doi: 10.1007/s11936-012-0182-9] [Medline: 22580974]
64. Kumanyika S. Behavioral aspects of intervention strategies to reduce dietary sodium. Hypertension 1991 Jan;17(1):I190-I195. [doi: 10.1161/01.HYP.17.1_Suppl.I190]
65. Lofthouse C, Te Morenga L, McLean R. Sodium reduction in New Zealand requires major behaviour change. Appetite 2016 Oct 01;105:721-730. [doi: 10.1016/j.appet.2016.07.006] [Medline: 27395412]

Abbreviations

BMI: body mass index
DELISH: Digital Education to LImit Salt in the Home
KABs: knowledge, attitudes, and behaviors
SE: standard error
SES: socioeconomic status
SMS: short messaging service

Edited by G Eysenbach; submitted 17.09.18; peer-reviewed by J Zhang, Y Wu; comments to author 12.10.18; revised version received 28.10.18; accepted 31.10.18; published 25.02.19.

Please cite as:
Khokhar D, Nowson CA, Margerison C, West M, Campbell KJ, Booth AO, Grimes CA
The Digital Education to Limit Salt in the Home Program Improved Salt-Related Knowledge, Attitudes, and Behaviors in Parents
J Med Internet Res 2019;21(2):e12234
URL: http://www.jmir.org/2019/2/e12234/
doi:10.2196/12234
PMID:30801255

Original Paper

Design and Development of a Person-Centered Patient Portal Using Participatory Stakeholder Co-Design

John Kildea[1,2,3,4], PhD; John Battista[5]; Briana Cabral[2]; Laurie Hendren[3,6], PhD; David Herrera[6], BSc; Tarek Hijal[1,3,5], MD; Ackeem Joseph[7], MSc

[1]Gerald Bronfman Department of Oncology, McGill University, Montreal, QC, Canada
[2]Medical Physics Unit, McGill University, Montreal, QC, Canada
[3]Cancer Research Program, Research Institute of the McGill University Health Centre, Montreal, QC, Canada
[4]Department of Biomedical Engineering, McGill University, Montreal, QC, Canada
[5]Division of Radiation Oncology, McGill University Health Centre, Montreal, QC, Canada
[6]School of Computer Science, McGill University, Montreal, QC, Canada
[7]Department of Medical Physics, McGill University Health Centre, Montreal, QC, Canada

Corresponding Author:
John Kildea, PhD
Gerald Bronfman Department of Oncology
McGill University
Suite 720
5100 de Maisonneuve Boulevard West
Montreal, QC,
Canada
Phone: 1 5149341934 ext 44154
Email: john.kildea@mcgill.ca

Abstract

Background: Patient portals are increasingly accepted as part of standard medical care. However, to date, most patient portals provide just passive access to medical data. The use of modern technology such as smartphones and data personalization algorithms offers the potential to make patient portals more person-centered and enabling.

Objective: The aim of this study is to share our experience in designing and developing a person-centered patient portal following a participatory stakeholder co-design approach.

Methods: Our stakeholder co-design approach comprised 6 core elements: (1) equal coleadership, including a cancer patient on treatment; (2) patient preference determination; (3) security, governance, and legal input; (4) continuous user evaluation and feedback; (5) continuous staff input; and (6) end-user testing. We incorporated person-centeredness by recognizing that patients should decide for themselves their level of medical data access, all medical data should be contextualized with explanatory content, and patient educational material should be personalized and timely.

Results: Using stakeholder co-design, we built, and are currently pilot-testing, a person-centered patient portal smartphone app called Opal.

Conclusions: Inclusion of all stakeholders in the design and development of patient-facing software can help ensure that the necessary elements of person-centeredness, clinician acceptability, and informatics feasibility are achieved.

(J Med Internet Res 2019;21(2):e11371) doi:10.2196/11371

KEYWORDS

patient portals; patient participation; telemedicine; software design

Introduction

Context

A patient portal, in its most basic form, is a secure extension of a health care institution's electronic medical record (EMR) that is accessible to patients [1,2]. It provides patients with access to some or all of their personal health information (PHI) within the EMR. Patient portals are generally positively reviewed by patients and clinicians and have been associated with improved patient engagement, patient empowerment, and patient satisfaction [3-11].

The argument for sharing PHI with patients via a patient portal is clear. However, the best way to do so is not obvious [12,13]. According to Prey et al [14], similar barriers to enabling patient access to clinical data are encountered throughout the world, highlighting the importance of reporting strategies used for patient portal design, development, and adoption. Despite this, few reports describing how patient portals were designed and developed and few detailed evaluations of their usability are available in the literature to provide guidance for creating new software.

Furthermore, in the present era of mobile devices and using more advanced computing technologies, patient portals can offer much more than just passive sharing of PHI with patients via the Web [15]. They can provide personalized educational material that explains the PHI, the patient's condition, and treatment options; waiting room tools such as mobile check-in with call-in for appointments and waiting time estimates; communication tools such as patient-reported outcome (PRO) questionnaires and secure patient-clinician messaging; and other assisting functionalities on a variety of devices, including smartphones. As such advanced patient portals account for the fact that patients (care recipients) are people with complex needs that extend beyond just the immediate delivery of care, and as described in Rigby et al [15], they help equalize the patient-clinician knowledge balance by incorporating the patient as an equal member of their own care team, we will refer to them as *person-centered patient portals*. Person-centered needs include the ability to plan ahead and know one's position in a waiting list, to feel in control of one's own care, to understand one's treatment options, and to share in all decision making about one's care.

With all of the above in mind, this paper describes the approach we used to design and develop a person-centered patient portal, which we call *participatory stakeholder co-design*, involving patients and health care providers. The resulting patient portal, called Opal [16] represents a real-world example of a patient-facing electronic health (eHealth) project that was designed and built from scratch within the health care system. Opal supports personalized information, including PHI with appropriate explanations, appointment schedules with appointment-specific advice, mobile appointment check-in and call-in functionality, questionnaires, educational material automatically tailored to health condition and treatment plan, and an autonomous module specific to an institution's patients' committee. In our estimation, the success of the Opal project to date is due in large part to our participatory stakeholder co-design approach. Success in this context is measured in terms of the project's ability to go from conception through design and development to use by patients in a pilot release, having navigated and survived the various legal, logistical, and cultural hurdles thrown up by the health care system.

Background

As is the case for most health care technology, the development of a person-centered patient portal is complex, involving many stakeholders and numerous organizational layers that each may determine the success or failure of the initiative [15,17-19]. Numerous authors have pointed out the importance of employing user participatory design in software development [20-22], including eHealth projects [23-28] such as patient portals [29-32]. Indeed, according to a frequently cited report by the Standish Group International [33], user involvement in the design process is the number one reason why software projects succeed or, conversely, the lack thereof is the main reason why projects fail.

In the case of design of a person-centered patient portal, as was the aim of this study, the patient is the user, and patient involvement brings the design process into the realm of *patient co-design*, which itself is increasingly recognized as an integral component of sustainable quality improvement in health care [34-36].

Patient *co-design* is often confused and conflated with *patient-centered design* and *person-centered design*. However, these 3 concepts have distinctly different meanings [34,37,38], and it is worth stressing the difference between them. In *co-design*, patients help identify the process or project that needs to be designed (or redesigned) based on their personal experience, and they codrive the effort in partnership with the clinical team. In effect, as we learned from the presently described project, the effort is really *participatory stakeholder co-design*, our preferred term, since both patients and clinicians are equal stakeholders in the final result and they actively participate in all aspects of the design process. In *patient-centered* design, the design team strives to ensure that the needs of the patient are centermost. However, the project being designed may or may not have been identified by patients, and the design process may or may not involve actual patients. Similarly, in *person-centered design*, the design team works to ensure that the needs of the patient, as a whole person and an equal partner in their care [37], rather than simply a passive recipient, are foremost. Again, however, the design process may be person-centered without having the patient in the room and fully involved in the project. The design process is only *participatory stakeholder co-design* if all stakeholders are fully and equally involved.

In the context of a person-centered patient portal, stakeholder co-design ensures that the necessary elements of person-centeredness, clinician acceptability, and informatics feasibility are accounted for, with each element being necessary for the ultimate success of the portal.

As described in this paper, we believe that the uniqueness of the Opal patient portal is that it was designed from the ground

up, in a noncommercial environment, to be person-centered using a participatory stakeholder co-design approach.

Setting

Our center, the Cedars Cancer Centre, is a new comprehensive cancer center within the McGill University Health Centre (MUHC), a large academic teaching hospital with an affiliated Research Institute (RI-MUHC) in Montreal, Quebec, Canada. The Cedars Cancer Centre was formed in 2015 when the previously disparate cancer services (radiotherapy, medical oncology, surgical oncology, and supportive care) of the MUHC were brought together under one roof.

Our patient portal development team followed our participatory stakeholder co-design approach from the very beginning. The team was formed in 2014 when LH (who is also a professor of computer science) was receiving radiotherapy for breast cancer under the direction of radiation oncologist TH. Following a conversation regarding the computational needs of radiation oncology, TH introduced LH to JK, a medical physicist with whom he was already collaborating on a number of custom eHealth projects. The 3 decided to colead a health informatics research collaboration, with a vision to use participatory stakeholder co-design to improve the experience and outcomes of people receiving care at our cancer center. The 3 coleads (patient who happens to be a computer scientist, radiation oncologist, and medical physicist) have led the research effort as equals since the beginning, each bringing their own unique but complementary expertise to the leadership table. LH, as a patient, was able to provide the patient perspective and access to other patients through the patients' committee of the cancer center. LH, as a computer scientist, was able to contribute computational expertise and access to computer science students. JK, as an academic medical physicist, was able to provide health informatics know-how, facilitate access to the EMR, and access to medical physics students. TH, as a radiation oncologist and acting chief of the Division of Radiation Oncology, was able to provide his clinical expertise, help identify the needs and concerns of staff, manage necessary change within radiation oncology, and provide access to senior hospital management.

Overall, 2 areas for improvement were quickly identified by LH from her perspective as a person who was experiencing cancer treatment: (1) the *pain* of waiting for health care services and (2) the lack of PHI and relevant educational material provided to patients. A potential solution to first area of improvement was proposed in the form of a *patient app* to provide patients with personalized waiting time predictions determined using timestamp data and machine learning algorithms. Institutional funding was secured and the project began in the summer of 2015. As discussions around the functionality of the patient app evolved, it became clear that it could also serve as a patient portal and, as such, it could address second area of improvement, the lack of PHI and relevant educational material provided to people receiving care at the institution.

Objectives

Opal represents a real-world example of a patient-facing eHealth initiative that was designed and developed from within the health care system using an approach that we have called *participatory stakeholder co-design*, with a cancer patient who is also a computer scientist, clinician, and medical physicist equally coleading the effort. Our participatory stakeholder co-design approach was fundamental to the success of the software from all perspectives—patient, clinician, and informatics. The purpose of this paper is to share how we used this participatory stakeholder co-design approach, and the key elements that we identified within our approach, to create a patient portal that we believe is both person-centered and useful.

Methods

Context

At the outset of this project, a patient portal was not available either within our institution or within Quebec's public health care system. We scanned the commercial marketplace but were not able to identify an existing solution that provided our most basic need—multilingual support (French and English)—nor our person-centeredness requirements such as waiting room management tools (needed to reduce the pain of waiting), automated access to medical notes within any EMR, and automated personalization of information delivery. Accordingly, a custom-developed solution was pursued.

Insights From the Literature

We studied the literature pertaining to the design of a patient portal, including the typical features that portals contain, the PHI elements that they offer, and the layouts that are used. We were particularly interested in publications detailing patient portals in oncology and for smartphones (our initial focus), but our search was not limited to these areas. Our intention was to inform ourselves regarding previous work, recommended best practices, areas where we could improve upon using our participatory stakeholder co-design approach, and challenges that our approach might address. To remain updated, our literature search was repeated many times over the course of the project.

We found 3 reports in the literature that provided particularly useful guidance on the presentation of patient data:

- Ahern et al [39] split patient portal services into 3 main categories: (1) information and transactions, (2) expert care, and (3) self-care and community. Transactions are bureaucratic in nature, such as viewing or creating or changing appointments, filling out forms, and requesting information. Expert care includes access to clinical services such as secure messaging, remote monitoring, and PRO questionnaires with clinician feedback. Self-care and community relates to relevant educational material and access or referral to services and social networks that provide support.
- A study by Tang et al [40] contains definitions and recommendations regarding data that an Electronic Health Record-tethered Personal Health Record (essentially a patient portal) should contain. Of note, it points out that to be useful to patients, health data should be accompanied by tools that help the patient understand and act on them,

and data presentation should be adapted to the individual to optimize potential benefits.

- Baudendistel et al [12] point out how patients want to be able to track their long-term medical history and that the information provided must be accessible and filtered to the patient's specific situation. Furthermore, they found that both patients and providers found value in PROs.

We were unable to find any detailed literature on the design of mobile patient portal apps.

Participatory Stakeholder Co-Design

Our methodology for participatory stakeholder co-design comprised 6 key elements. These included (1) equal coleadership (patient who is also a computer scientist, clinician, and medical physicist); (2) patient preference determination; (3) security, governance, and legal input; (4) user evaluation and feedback; (5) continuous staff input; and (6) end-user testing. Figure 1 presents an approximate timeline for the Opal project showing how the various elements of our participatory stakeholder co-design methodology came together, culminating in the pilot release of the Opal smartphone app in Radiation Oncology at the Cedars Cancer Centre. We elaborate on each of the 6 elements below.

Element 1: Equal Coleadership

Our 3 coleads partnered in leading all aspects of the design and development process. At a practical level, this meant that all 3 were equally involved in all important decisions and in constant communication, usually by email, videoconference, or in person. Patient participation was facilitated by a parking pass (provided by virtue of our patient colead being part of the cancer center's patients' committee) and by associate membership of the institution's research institute (which also allowed LH to officially work on the project during her sabbatical year) but was complicated by disease progression and treatment. Videoconferencing was vital to ensuring full patient participation at all meetings.

Element 2: Patient Preference Determination

To obtain input from the wider population of people receiving cancer treatment (our initial focus), we conducted a voluntary convenience sampling survey within the waiting rooms of our cancer center. Participants were asked to provide basic demographic information (age and gender), state whether or not they use a smartphone, and if they would avail of a patient portal to access their PHI. A number of possible patient portal features were presented, and respondents were asked to rate their interest in having each using a 5-point Likert scale, ranging from "not at all interested" to "very interested." Patients were invited to complete the main part of the survey only if they had reported that they use a smartphone.

To determine participants' level of comfort with accessing their PHI, we listed 3 possible levels of PHI and asked them to select which one they would choose if given the choice. The 3 levels included:

1. I would like access to all of my medical record, including lab results, as soon as the information is available.
2. I would like access to all of my medical record, including lab results, after I have reviewed them with my doctor.
3. I would like access to just my appointment schedule and other need-to-know information.

We note that our survey was conducted after the development of our portal had begun and that it was conducted during 2 summer periods (with the help of summer students). The goal of the survey was to affirm patient preferences and verify that our software design was on the right track by sampling the wider patient population.

Element 3: Security, Governance, and Legal Input

Early into the development process (about 6 months), when a sufficiently developed prototype was available for demonstration, we engaged our institution's Security and Governance team to provide guidance regarding the security and confidentiality of patient data and compliance with applicable regulations. Guidance was provided by means of a number of in-person meetings, email correspondence, and ultimately, a detailed risk assessment report.

The technical cybersecurity aspects of the software were validated by an internal vulnerability assessment by the institution's Information Services security team and gray box penetration testing by an independent external consulting firm. Legal support (drawing up agreements and disclaimer forms) was provided by the institution's Legal team and the Business Development office of our affiliated research institute.

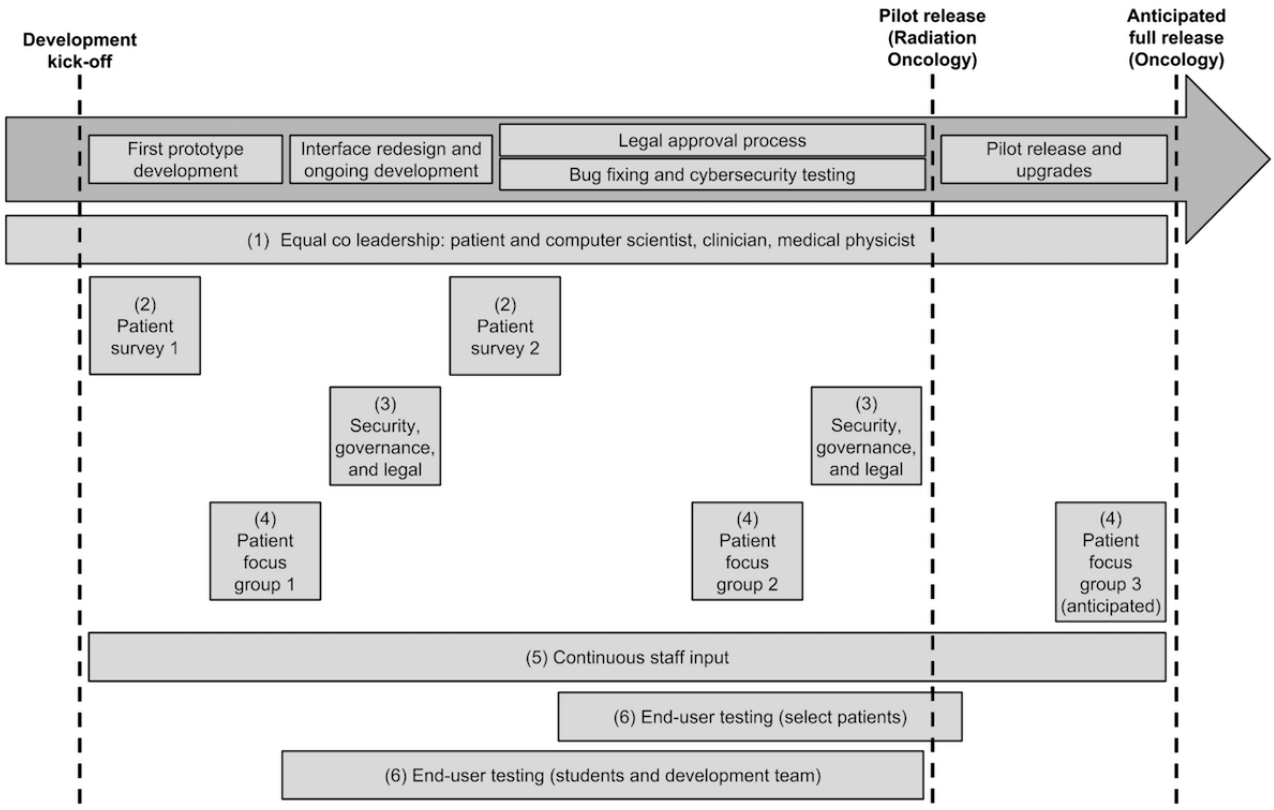
Element 4: User Evaluation and Feedback

Once a prototype version of our patient portal smartphone app was ready, we invited a number of patients to participate in a purposeful sample focus group to provide feedback regarding its features and usability. Our radiation therapy team helped us to identify a number of engaged patients who had finished their treatments. A total of 10 patients were identified and contacted and 3 ultimately participated. Those who declined to participate indicated that they were unavailable at the time and date chosen.

Without describing the smartphone app or divulging its features, we started the focus group by asking participants for their thoughts on what information they would like to see in a smartphone portal without regard to the technical, legal, or logistical challenges on providing such information (ie, in an ideal world scenario). We then demonstrated the prototype app and observed the participants using it. Finally, we went through each of the features of the prototype and sought feedback. Our focus group guide was developed by the project leadership and was designed to solicit maximum feedback about the prototype app and the ease with which patients were able to navigate its features. A redacted version of our guide is provided in Multimedia Appendix 1.

Shortly before the pilot release, we conducted a second focus group comprising 5 members of the cancer center's patients' committee. The goal of this focus group was to rehearse the registration process and anticipate initial real-world problems and questions.

Figure 1. Approximate timeline for the Opal project showing how the 6 elements of our participatory stakeholder co-design methodology came together, culminating in the pilot release of Opal in Radiation Oncology. Each element is numbered in the figure and explained in the text. Just over 3 years of co-design and development were undertaken between the development kick off (May 2015) and the pilot release (June 2018). Figure not to scale.



Element 5: Continuous Staff Input

Continuous staff input was ensured by our clinician colead. In addition, a radiation therapist (JB) participated in the design team during the first year of development. Over the course of the project, we also presented the planned features and functionality of the smartphone portal to various staff groups, ranging from on-the-floor care providers to senior management and the board of directors of the institution. The purpose of these presentations was threefold: (1) to ensure awareness and build buy-in at all levels, (2) to obtain staff feedback and address staff concerns, and (3) to seek support to continue development beyond the initial pilot release. During the final 2 months before the pilot release, a content development team comprising 2 radiation therapists, 2 medical physicists, an oncology nurse, a radiation oncologist, and an administrative assistant met weekly to ensure that the content of the app and the education material were *production ready*.

Element 6: End-User Testing

To help ensure usability and to identify bugs, we engaged volunteer testers as soon as the first prototype was ready. Overall, 2 types of volunteers were recruited: student testers (approximately 50 students in total) who used mock patient accounts, and real patients (including our colead) who had limited access to their PHI but full access to appointment schedules, waiting room management functionality, and available educational material. Our volunteer end-user testers provided invaluable feedback.

Development Approach

We employed an *Agile* development approach [41] in which we rapidly prototyped various features and iteratively integrated and tested them in the main product. Our approach was facilitated by the research environment of our academic hospital and our strong links with McGill University. Over 3 summers, we employed summer students to develop prototypes of various features and we engaged undergraduate and graduate students (typically 7-10 computer science and medical physics students) in term research projects to explore possible new features. A core team of developers worked on the official version of the smartphone app for pilot release. The core team met at least weekly (but generally more frequently) with the coleads to discuss user feedback and bug fixes and communicated daily in person and via email. During term, student meetings were held roughly once every 2 weeks. The Github Issues tool [42] was used to submit bugs, the Zenhub board [43] was used to track them, and the Crashlytics platform [44] was used to release frequent builds to volunteer testers. At the time of pilot release, the core team comprised 3 full-time developers (1 front end, 1 back end, and 1 full stack) and 2 part-time full-stack developers (2 days per week each, typically).

Pilot Release

A pilot release of our smartphone patient portal app, involving patients receiving radiotherapy at our center, is currently underway. We are taking a phased approach to the pilot, starting slowly with a small number of invited patients under the care of specific physicians. Our goal is to expand gradually to include all radiation oncology patients within 3 months, followed by a

full release to all patients at our cancer center. Results of the pilot study will be submitted for publication once complete.

Results

In this section, we present the results of our design and development research project that gave rise to Opal as a person-centered patient portal, and we detail the contributions provided by each element of our participatory stakeholder co-design methodology.

Participatory Stakeholder Co-Design

Element 1: Equal Coleadership

Equal coleadership ensured that the main stakeholder perspectives of the project were always represented at the leadership table. In turn, these perspectives ensured that the design and development of Opal incorporated the necessary elements of person-centeredness, clinician acceptability, and informatics feasibility. A nonexhaustive list of these elements, as identified over the course of the project, is provided in Multimedia Appendix 2.

Although full and equal coleadership was a fundamental element of our participatory stakeholder co-design approach, we found that this was often not clear to people outside of the immediate design and development team. External to the team, it was necessary to continuously insist that all 3 coleads were equal. The tendency in the hierarchical health care system is to assume that a project must be led by a clinician, and it is commonly assumed that the patient participant is a token member of the team. Overcoming this required persistence, it ensured that the team was truly coled and it broke new ground within the hospital by demonstrating the benefits of full and equal patient involvement. Equal coleadership was vital to the success of the project.

Element 2: Patient Preference Determination

A total of 361 patients participated in our voluntary convenience-sampled survey. Of these, 65.7% (237/361) said they did have a smartphone. Figure 2 presents the distribution of patients by age group and the percentage of patients in each age group who reported having a smartphone. It addresses a concern we heard often that many older cancer patients may not have smartphones. As expected, smartphone usage is lower for the older age groups. However, it is clear that a smartphone app would nevertheless have uptake across the age spectrum and so can be considered a potentially useful tool for all patients. We also note that many older patients come to appointments with younger caregivers.

With regard to patient preferences for access to their PHI using an app or portal, we found that the majority of respondents with a smartphone (63.8%, 148/232) would prefer to have access to all of their data immediately once they are available. The remainder of respondents were roughly split two-to-one in favor of accessing all their data after review with their physician or accessing just need-to-know information such as appointments and educational material. Figure 3 shows the breakdown of patient's PHI access preferences.

Figure 2. Distribution of the ages and smartphone usage of the 361 cancer patients who participated in our waiting room survey. Overall, 66% (237) of the respondents reported that they use a smartphone. These data demonstrate that a smartphone app would reach a broad patient population. The survey was conducted during the summer of 2016 and repeated during the summer of 2017 in the waiting rooms of our cancer center.

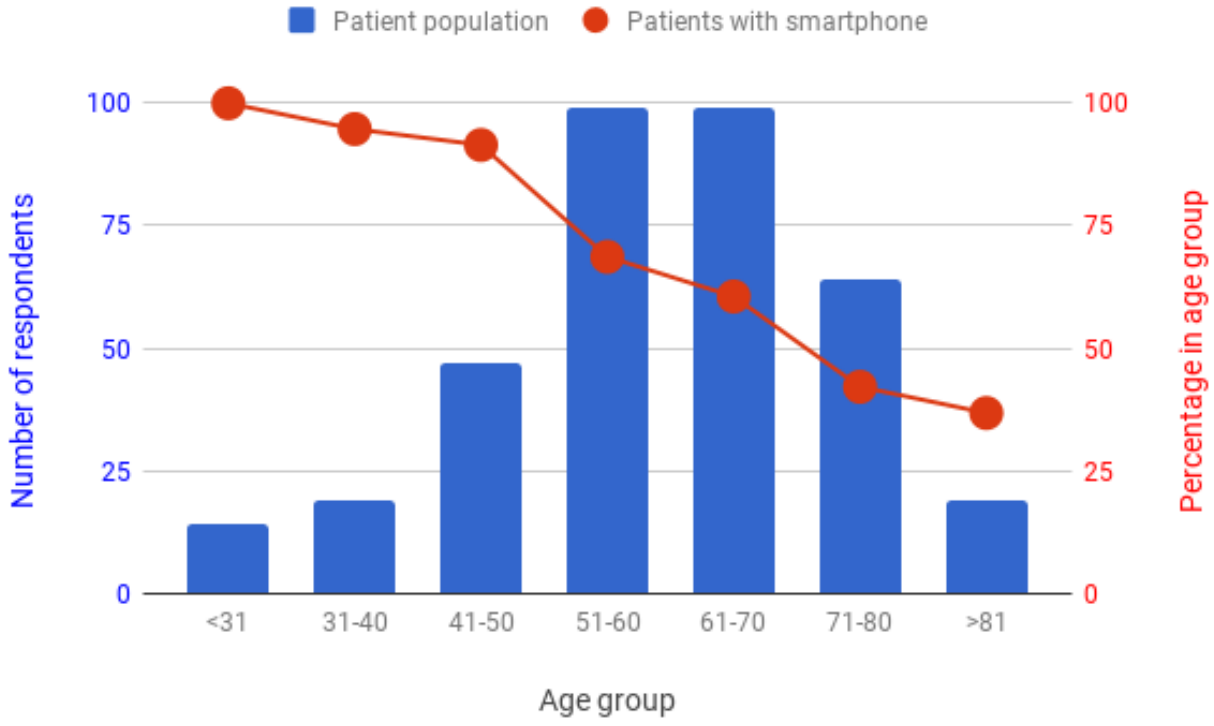


Figure 3. Patient preferences with regard to the personal health information (PHI) that they would like to access via an app or portal. We only included responses from patients who reported that they had a smartphone and who selected just 1 of the 3 options (n=232).

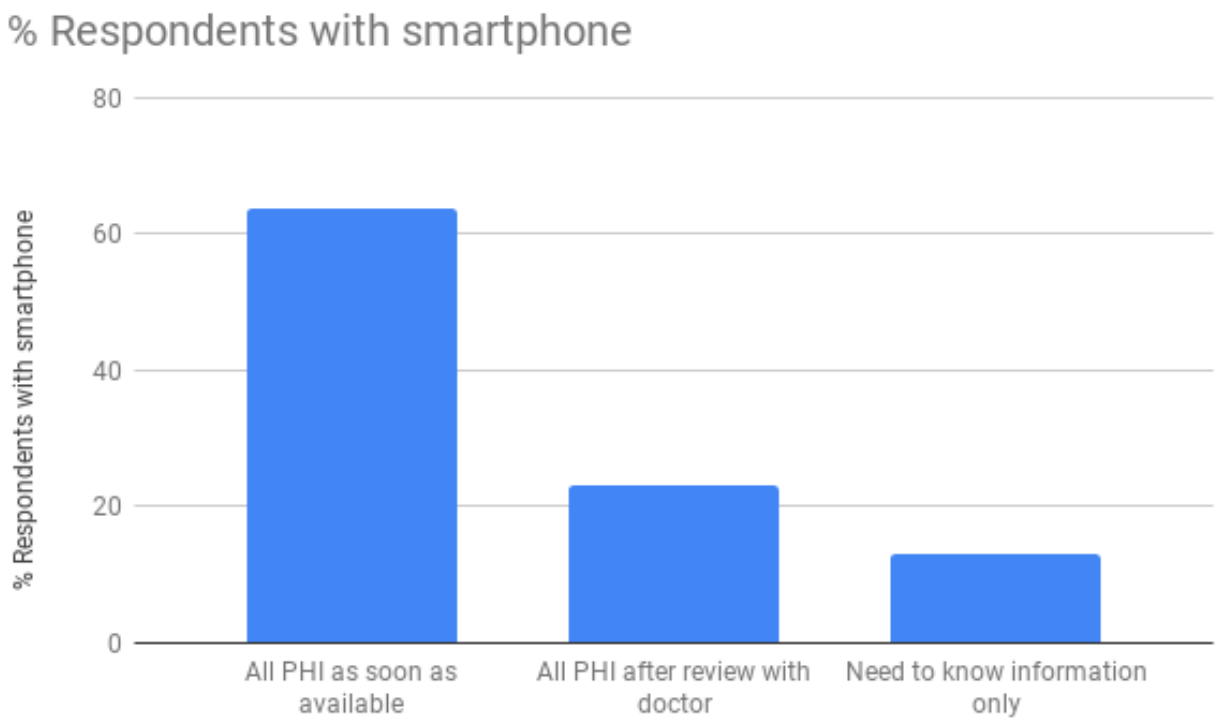


Table 1 presents the results of the main part of the survey and shows that the vast majority of patient respondents indicated that they were very interested in all of the possible features of an app or portal that we presented.

Element 3: Security, Governance, and Legal Input

The risk assessment report provided by our institution's Security and Governance team comprised 16 specific recommendations with regard to data security and confidentiality. As they may be of use to other teams developing patient-facing software, we have summarized these recommendations in Multimedia Appendix 3.

In our experience, the legal aspect of the project, which was mainly beyond our control, was the slowest and most frustrating component. The main time-consuming legal issues encountered pertained to ownership of intellectual property, liability, and the contents of the patient disclaimer form. As the health care system is a busy legal environment with many competing priorities, we found that our nonurgent legal needs took time to be completed and required constant follow-up. Access to a separate legal team for business development tasks was not possible but would have helped.

Element 4: User Evaluation and Feedback

Most of the features suggested by the participants at the start of the first focus group were already in our early prototype. These included hospital maps, radiotherapy treatment plan views with beam entry points, appointment schedules, waiting time estimates, and notifications. Some suggestions, such as an in-built feature to pay for hospital parking or a system to leave questions for call back (rather than navigating the hospital's phone system), were not possible to include in the first release but will be included in future versions. Participants found the initial user interface difficult to understand. In particular, the "hamburger" side menu that was used to provide access to the options available within the app was nonintuitive and needed to be explained. This finding provoked a complete redesign of the user interface. Figure 4 illustrates the evolution of the user interface based on focus group feedback.

Interestingly, we found that our focus group participants prefaced many of their responses with disclaimers such as "I doubt it is possible, but it would be nice if..." or "I wouldn't want to disturb my treating team by asking for this information, but it would be nice if I could see it myself in my own time."

Our second focus group was quite different to the first as the product was mature (pilot-release ready) and participants accessed their actual data. The focus group's main purpose was to evaluate the registration workflow, but it also provided important feedback regarding data and content. As the pilot release was centered in radiation oncology and only partially integrated with the medical oncology databases, patients receiving both radiotherapy and chemotherapy were confused that not all of their data were accessible. This motivated the development team to expand the pilot beyond radiation oncology as soon as possible.

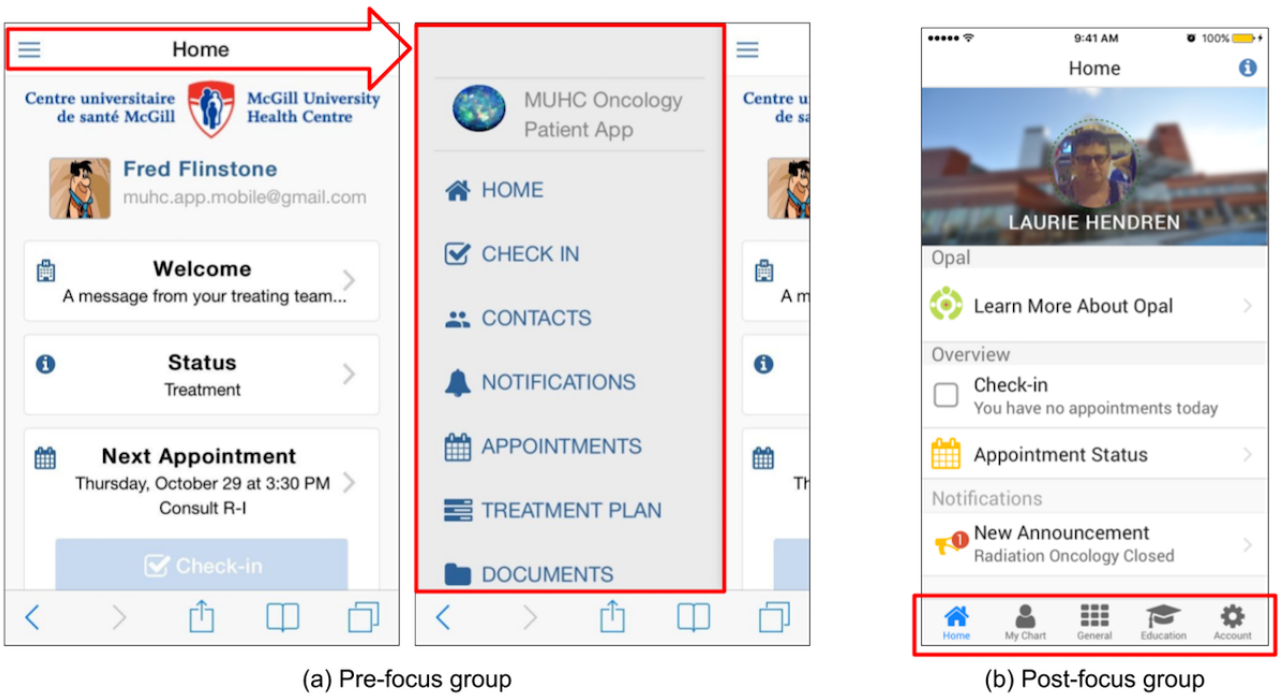
Table 1. Results from the main part of the patient survey regarding possible features of a patient app or portal. Participants were presented with possible features of an app or portal and asked to rate their interest in having them using a 5-point Likert scale, ranging from "1=not at all interested" to "5=very interested".

Possible feature	1, n (%)	2, n (%)	3, n (%)	4, n (%)	5, n (%)	*Pos*[a], n (%)
Your personal appointment schedule (N=267)[b]	11 (4.1)	9 (3.4)	13 (4.9)	21 (7.9)	213 (79.8)	234 (87.6)
Secure access to doctor's notes in your medical record (N=269)	20 (7.4)	11 (4.1)	9 (3.3)	33 (12.3)	196 (72.9)	229 (85.1)
Secure access to your personal laboratory results (N=267)	21 (7.9)	8 (3.0)	12 (4.5)	26 (9.7)	200 (74.9)	226 (84.6)
Educational material specific to your diagnosis (N=259)	16 (6.2)	7 (2.7)	17 (6.6)	54 (20.8)	165 (63.7)	219 (84.6)
Educational material specific to your phase of treatment (N=259)	19 (7.3)	9 (3.5)	15 (5.8)	54 (20.8)	162 (62.5)	216 (83.4)
Notifications sent to your phone to advise you that you are next in line to see your doctor or for treatment (N=267)	25 (9.4)	13 (4.9)	8 (3.0)	26 (9.7)	195 (73.0)	221 (82.8)
Personalized check-in and call-in for your appointments via your phone (N=269)	26 (9.7)	7 (2.6)	18 (6.7)	34 (12.6)	184 (68.4)	218 (81.0)
Contact information for your treating team (N=260)	21 (8.1)	10 (3.8)	21 (8.1)	37 (14.2)	171 (65.8)	208 (80.0)
A secure messaging system with your treatment team (N=262)	21 (8.0)	11 (4.2)	20 (7.6)	34 (13.0)	176 (67.2)	210 (80.2)
Questionnaires to describe your symptoms or side effects before each appointment (N=260)	18 (6.9)	12 (4.6)	29 (11.2)	37 (14.2)	164 (63.1)	201 (77.3)
Step-by-step status of your personal treatment planning while waiting at home before starting treatment (N=258)	26 (10.1)	15 (5.8)	20 (7.8)	36 (14.0)	161 (62.4)	197 (76.4)
Secure access to your personal radiotherapy treatment plan showing beam configuration and possible areas of your skin that might be affected (radiotherapy patients only; N=233)	36 (15.5)	10 (4.3)	20 (8.6)	29 (12.4)	138 (59.2)	167 (71.7)
Maps and hospital information (N=260)	22 (8.5)	23 (8.8)	32 (12.3)	35 (13.5)	148 (56.9)	183 (70.4)
Option to anonymously donate your medical data for research (N=260)	34 (13.1)	20 (7.7)	29 (11.2)	40 (15.4)	137 (52.7)	177 (68.1)
Parking information (N=255)	40 (15.7)	21 (8.2)	30 (11.8)	33 (12.9)	131 (51.4)	164 (64.3)

[a]*Pos*: the percentage of patients who rated their interest as 4 or 5, that is, the total number of patients who said they were positively "interested" in having the feature.

[b]N indicates the number of participants who answered each question. Rows are sorted by popularity of the feature offered.

Figure 4. Screenshots of the Opal app demonstrating how the user interface changed based on feedback received during the first focus group. (a) The pre-focus group hamburger menu (illustrated by the red arrow and outlined section) was replaced with (b) a simpler and more intuitive bottom-of-the-screen tab-based menu.



(a) Pre-focus group (b) Post-focus group

Although our focus groups contained only small numbers of participants, the experience was very beneficial. They affirmed to the team that we were on the right track in terms of design, and they allowed us to make a number of changes to improve usability. A third focus group, to be held during the pilot release, will be used to inform the final release with real-world end-user feedback.

Element 5: Continuous Staff Input

In general, staff involvement engendered frequent discussions around 2 important points: (1) which clinical data should be accessible by patients and when and (2) the fear of increased workload by staff. These findings illustrate the importance of engaging staff early in the design and development process to ensure buy-in.

Over the course of the project, as the maturing product was presented to diverse groups, we anecdotally noticed an increase in acceptance, and ownership, of the initiative. Nevertheless, concerns about patients accessing bad news before meeting with their clinician were, and continue to be, frequently raised. Having a motivated patient colead helped ensure that the person-centeredness of the project (ie, the patient, as an equal partner in their own care, should decide for themselves what level of information they want and when) was foremost in discussions. In an attempt to account for all viewpoints, our ultimate design allows for 3 levels of information provision corresponding to those described in our patient survey (Figure 3), with each patient selecting their personal preference at the time of registration. The need to avoid increased workload on staff was identified early on by having a clinician colead the project. As such, a fundamental design component of Opal is that it should be automated. Beyond the initial setup and ongoing maintenance of rules to provide contextualized data and personalized educational material to patients, individual clinicians should notice no change to their workflows but may find that their patients are better educated on their disease and have more precise questions.

Element 6: End-User Testing

In total, over 60 volunteers, mainly students with mock data but including 10 patients and/or family members with access to some of their personal data, provided continuous feedback during the development process. Students mainly provided feedback regarding usability, navigation problems, and technical bugs, whereas patients mainly provided feedback regarding missing data and data presentation. Numerous improvements were, and continue to be, made as a result of user testing.

Development Approach

Over the 3 years of software development to the pilot release, approximately 1500 commits were made to the master code base on Github by 16 unique contributors, with 39 additional branches of the patient-facing code developed by student contributors. Approximately 500 issues (bug reports) were created on Github, with 85 outstanding at the time of pilot release. These findings highlight the importance of a collaborative approach in building complex software such as a patient portal.

It was found to be difficult to find and recruit experienced full-time software developers. Being a small unknown development group within a Quebec hospital was a disadvantage when recruiting within the highly competitive marketplace for software developers in Montreal. Despite expensive advertising via the standard job-posting forums, our software developers were all ultimately recruited through word-of-mouth connections. Our access to McGill University students, in particular, paid summer students who were able to develop prototypes and prepare the groundwork for new modules within Opal, greatly facilitated our software development.

Pilot Release

In the pilot release of Opal, patients have access to their appointment schedules, some PHI (certain radiation oncology clinical notes, laboratory test results showing longitudinal trends, and radiotherapy treatment planning status information), personalized educational material tailored to diagnosis and stage of treatment, waiting room management tools, and PRO questionnaires. Consistent with our goal of a person-centered portal, all data are contextualized with detailed explanatory content to help ensure that they are useful and empowering. Educational and explanatory content were prepared by a multidisciplinary content development team in radiation oncology that built upon existing paper-based materials. The content development team setup rules for the automated provision of education material and tags to link PHI with explanatory content. Although a Web browser version of the Opal portal is in development, it was not included in the pilot release to focus available resources.

Screenshots of the Opal smartphone app for the pilot release are provided in Figure 5. Figure 6 illustrates how Opal provides contextualized data using 2 examples. In the first example, the patient's calendar shows a scheduled Computed Tomography (CT) simulation appointment. A link from this appointment opens up a full explanation of what the CT simulation is. In the second example, the platelet count blood test is linked to explanatory material at Lab Tests Online [45]. Table 2 lists the features and data included in the pilot release.

Currently, Opal belongs to the members of the initial design and development team and the RI-MUHC. Ultimately, an open-source front-end code base and a licensable back-end software are envisaged.

Figure 5. Screenshots of the Opal smartphone app provided in the pilot release, showing the Home view, the My Chart view, the radiotherapy treatment planning view, and the education material library.

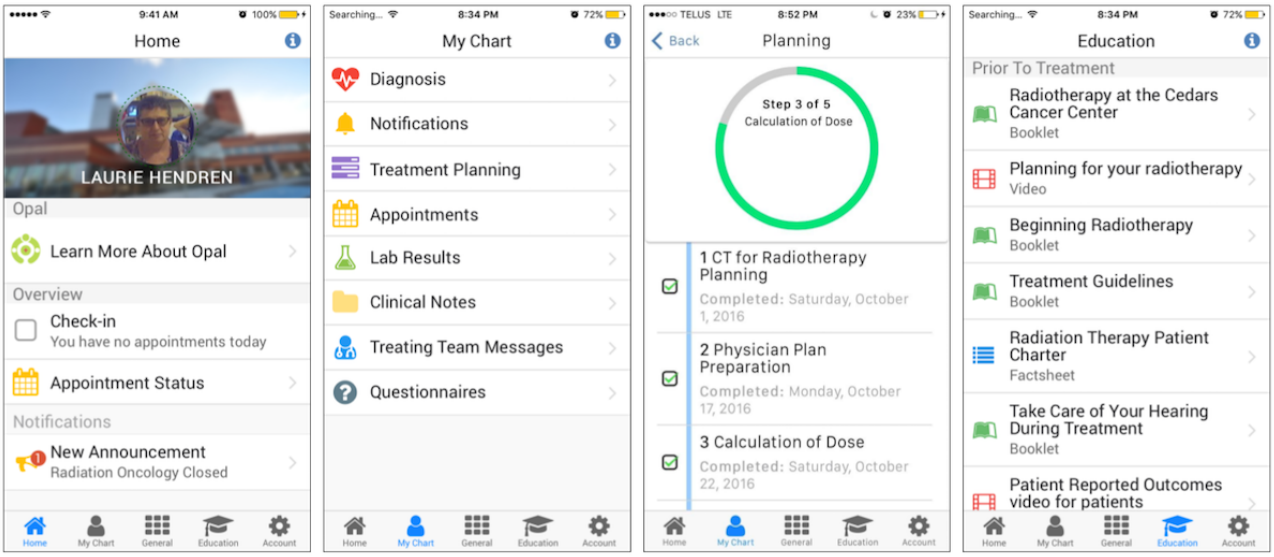


Figure 6. Two examples of how Opal contextualizes patient data. On the left, the Computed Tomography (CT) Simulation for Radiotherapy Planning appointment is linked to explanatory material about the CT Simulation procedure. On the right, the Platelet Count blood test results are linked to explanatory material at labtestsonline.org. The red arrows highlight the area in the left view that when tapped brings the user to the view on the right.

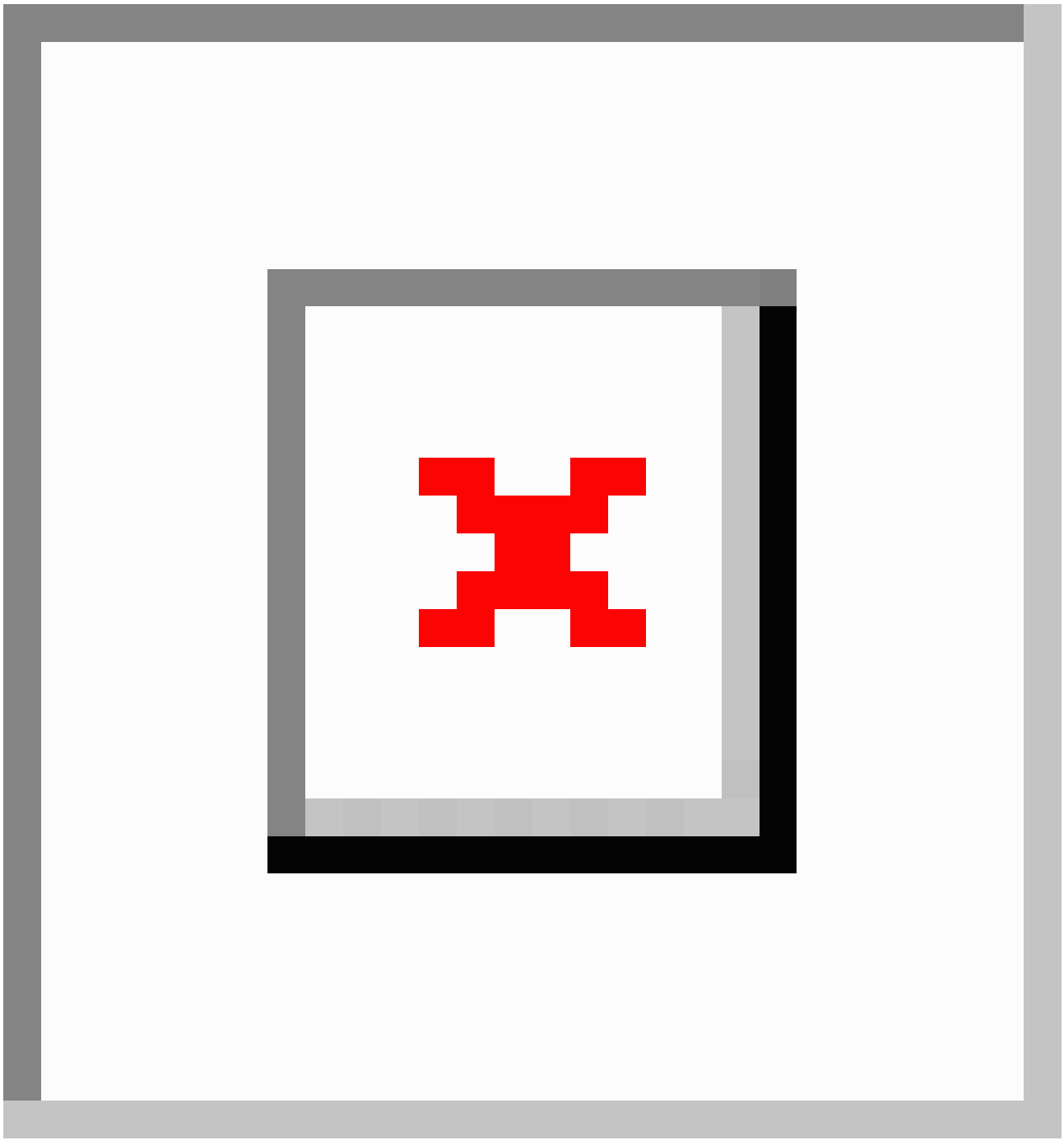


Table 2. Categories of information and features or functionality identified by patients and staff to be provided to patients via Opal. The majority of information provided to patients via Opal is personalized to their disease and phase of treatment. Information made available in the pilot release is indicated.

Category (menu/tab) and features or functionality	Personalized	Pilot release
Home screen or overview		
Next appointment	✓	✓
Notifications (eg, new document and new message)	✓	✓
Posts (messages from treating team and general hospital announcements)	✓	✓
Status of treatment or treatment planning	✓	✓
Waiting room management (check-in, call-in, and waiting time estimate)	✓	✓
My chart		
Diagnosis information	✓	✓
Notification archive	✓	✓
Appointment schedule with appointment location maps	✓	✓
Appointment change requests	✓	✗
Treatment or treatment planning information	✓	✓
Access to (selected) doctors' notes and nursing notes	✓	✓
Laboratory test results	✓	✓
Messages from treating team	✓	✓
Secure 2-way messaging with clinicians	✓	✗
Patient-reported outcome and satisfaction questionnaires	✓	✓
General information		
Phone directory and contact information (personalized on log-in)	✓	✓
General hospital announcements	✗	✓
Patient charter	✗	✓
Parking information	✗	✓
General hospital maps	✗	✓
Way finding	✗	✗
Leave feedback regarding app or portal	✗	✓
Facility to report bugs in the app or portal	✗	✓
Educational material (relevant and just-in-time)		
Videos	✓	✓
Booklets	✓	✓
Pamphlets or fliers	✓	✓
Account settings		
Language preference	✓	✓
Font size	✓	✓
Synchronization with phone's calendar	✓	✗
Facility to update demographic information in electronic medical record	✓	✗

Discussion

Principal Findings

In this project, we used participatory stakeholder co-design to create a person-centered patient portal smartphone app. Our approach engaged all stakeholders by having a patient (who is also a computer scientist) colead the project as an equal with a clinician (radiation oncologist) and an informatics expert (medical physicist) and by involving patients, health care providers, technical experts, and legal personnel throughout the design and development process. Looking back over the project, we identified that we had followed 6 elements for participatory stakeholder co-design: (1) equal coleadership (patient who is

also a computer scientist, clinician, and medical physicist); (2) patient preference determination; (3) security, governance, and legal input; (4) user evaluation and feedback; (5) continuous staff input; and (6) end-user testing. The benefits of participatory stakeholder co-design were clear. Our patient colead helped identify the requirements for person-centeredness, our clinical and informatics coleads strategized on their implementation within the existing hospital context, patient surveys and focus groups confirmed patient acceptability, and staff and institutional engagement ensured full stakeholder buy-in. In our estimation, the inclusion of all stakeholders in the design process is the best way to avoid the need for future redesigns and to ensure that the final product has wide acceptability.

End-user testing was achieved by the full involvement of our patient colead, partial involvement of a small cohort of volunteer tester patients who accessed some but not all of their PHI, and full involvement of a large group of enthusiastic tester students who developed and tested prototype software and accessed mock data. We found that it is important to engage end users with hands-on testing as soon as possible. We did not use wire-frame prototypes but with hindsight, we recognize that such prototyping at the beginning may have allowed us to avoid the user interface redesign provoked by our first focus group. We also maximized access to our active patient population through surveys and focus groups, and we engaged the very supportive patients' committee of our cancer center. The wider staff of our institution were engaged continuously by means of presentations and discussions, ranging from departmental rounds to a meeting with the board of directors. We found that the legal component of the project was the most time consuming and difficult to control. If we were to start the project again, we would engage the institution's legal team from the very beginning rather than waiting for a prototype to be ready.

True person-centeredness is difficult to achieve and, no doubt, continuous refinements to our software to ensure its person-centeredness will be required in future. Fundamental to achieving this is the involvement of patients now and into the future. Our attempt to maximize person-centeredness in our smartphone app includes 3 important considerations: (1) patients should decide for themselves their level of PHI access, (2) all PHI provided to patients should be contextualized with explanatory content so that they are useful and empowering, and (3) educational material should be personalized and tailored to the patient's immediate medical situation. These considerations are consistent with the findings of a study on patients' experience with cancer published recently by the Canadian Partnership Against Cancer [46].

The concerns of staff regarding increased workload because of the presence of a patient portal were addressed by designing the software to be automated. By using a rule-based system to send educational material and explanatory content to patients, a once-off setup is required by staff in each clinical setting. Thus, beyond the need for regular maintenance, the ongoing portal-related workload for staff should be minimal. Our multidisciplinary content development team in radiation oncology took care of the initial setup for Opal's pilot release. The team met for several hours each week for about 3 months to prepare the content. Patient sign-up for Opal during its pilot release is facilitated by staff from our cancer center's foundation. Ultimately, for a hospital-wide portal, sign-up should be facilitated by the hospital's admitting staff and may require additional personnel.

Comparison With Prior Work

It is well established that user involvement is a necessary component of successful software design [33]. Similarly, it is increasingly recognized that co-design is a necessary ingredient for successful quality improvement in health care [34-36]. Taken together, it is clear that for development of patient-facing software to be successful, all stakeholders, including patients and health care providers, must be involved in the design process. Achieving meaningful stakeholder involvement in practice is often difficult; however, clinicians are busy and may not have time to provide sufficient feedback, and active patients, although plentiful, are often not easy for development teams to access for reasons of confidentiality.

Our work, qualitatively and quantitatively, has shown that patients overwhelmingly feel that digital access to their PHI is important, with a majority of patients desiring immediate access to all of their data. These findings are consistent with experience reported elsewhere [47-49]. Our clinical staff, although supportive of the Opal initiative in general, tended to raise concerns that patients accessing their PHI in the absence of a clinician to provide support may result in anxiety and misunderstandings. Again, these findings are consistent with the literature [50-52]. Evidence from early adopter patient portals, however, tends to show that clinicians need not be too concerned [53]. For example, a pilot study on providing laboratory test results to patients in British Columbia found no difference in the levels of anxiety among patients who received their laboratory results online compared with a control group [54]. Similarly, early results from 7 clinics using the *myUHN* patient portal at the University Health Network in Toronto indicate no significant evidence of increased patient anxiety [55]. Of particular interest to our project in the context of cancer care, where laboratory results may contain alarming information, qualitative studies by Rexhepi et al [56] and Giardina et al [51] found that the majority of patients in their studies wanted immediate online access to their test results even if those results were alarming. In Rexhepi et al's [56] study, many patients expressed the belief that immediate access causes less anxiety than having to wait to discuss with a physician regardless of the nature of the results. These studies and others, for example, Ammenwerth et al's [11], have also shown that online access to PHI enables patients to better prepare for their consultations, which in turn improves clinician-patient communication, a goal of person-centered care. Throughout our participatory stakeholder co-design process, we found it useful to highlight these findings from other patient portal projects. Indeed, as our project matured, we noticed a softening in attitudes regarding sharing PHI with patients via Opal, a change we attribute to the involvement of all stakeholders in our design and development process.

Limitations

An important limitation of our quantitative findings is that they are specific to our context—a comprehensive cancer center in

Montreal. For example, our convenience-sampled patient survey was solely intended to provide input to the design and development team. As such, our quantitative results may contain biases to our setting that limit wider applicability. Our qualitative findings (ie, our description of the approach followed) are based on our retrospective analysis of the design and development strategy that worked for us. Another team following a broadly similar approach may describe it differently.

Conclusions and Future Work

We have designed and developed a patient portal smartphone app from within the Quebec health care system using a participatory stakeholder co-design approach, involving patients, staff, software developers, and students. Our methodology consisted of 6 core elements: (1) equal coleadership (patient who is also a computer scientist, clinician, and medical physicist); (2) patient preference determination; (3) security, governance, and legal input; (4) continuous user evaluation and feedback; (5) continuous staff input; and (6) end-user testing. Our final design adhered to principles of person-centeredness, recognizing that patients should decide for themselves their level of PHI access, all PHI should be contextualized with explanatory content, and educational material should be personalized and timely. As our project matured, and more and more stakeholders were engaged, we noticed an increase in the acceptance by clinical staff of the concept of sharing PHI with patients.

Future work will focus on evaluating the uptake and acceptability of our portal during its pilot release and expansion of its use beyond our cancer center to our general hospital and to additional health care institutions.

Acknowledgments

Financial support for this project was generously provided (in chronological order) by the Montreal General Hospital Cooperation and the McGill University Health Centre (MUHC; Challenge Q+ Initiative), by the Cedars Cancer Foundation, by the Montreal General Hospital Foundation, summer student scholarships (Collaborative Research and Training Experience Program of the Natural Sciences and Engineering Research Council of Canada (NSERC)–Medical Physics Research Training Network Grant #432290, NSERC Undergraduate Student Research Awards, McGill University Faculty of Science Undergraduate Research Awards, McGill University Faculty of Arts Undergraduate Research Internship Awards), the MUHC Cancer Care Mission Patients' Committee, and the MUHC Patient Education Committee. Support for and commitment to the project by the MUHC Department of Medical Physics (Mr William Parker) and by the McGill University Medical Physics Unit (Professor Jan Seuntjens) facilitated the software development resources and presence of students. Championing of the project by Dr Armen Aprikian and Professor Carolyn Freeman within the institution and by Mr Daniel Morin and Dr Jean Latreille at the provincial level is gratefully acknowledged. Support for the project by the Research Institute of the MUHC (in particular, Dr Costas Karatzas, Dr Bruce Mazer, and Dr Marie-Claude Gingras), and by the director of MUHC Information Services, Mr Hichem Motemem, is likewise gratefully acknowledged. Finally, we are indebted to the numerous students, patients, and family members who voluntarily tested Opal and provided input during its design and development.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group guidelines document. Used by the moderators of the first patient focus group that evaluated the early Opal prototype app.
[PDF File (Adobe PDF File), 120 KB - jmir_v21i2e11371_app1.pdf]

Multimedia Appendix 2
Elements of person-centeredness, clinician acceptability, and informatics feasibility identified during the development of Opal and facilitated by the stakeholder co-design approach that was followed.
[PDF File (Adobe PDF File), 42 KB - jmir_v21i2e11371_app2.pdf]

Multimedia Appendix 3
Security and governance recommendations for development of a patient portal as identified by our institution's Security and Governance team.
[PDF File (Adobe PDF File), 32 KB - jmir_v21i2e11371_app3.pdf]

References

1. 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]
2. Office of the US National Coordinator for Health Information Technology. 2016. What is a patient portal URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2018-06-20] [WebCite Cache ID 70K65Exx0]

3. Lin CT, Wittevrongel L, Moore L, Beaty BL, Ross SE. An internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
4. Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care 2007 Jul;13(7):418-424 [FREE Full text] [Medline: 17620037]
5. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
6. 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]
7. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
8. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
9. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2017 Apr 1;24(e1):e166-e172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
10. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct release of test results to patients increases patient engagement and utilization of care. PLoS One 2016 Jun 23;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
11. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
12. Baudendistel I, Winkler E, Kamradt M, Längst G, Eckrich F, Heinze O, et al. Personal electronic health records: understanding user requirements and needs in chronic cancer care. J Med Internet Res 2015;17(5):e121 [FREE Full text] [doi: 10.2196/jmir.3884] [Medline: 25998006]
13. Otte-Trojel T, de Bont A, Aspria M, Adams S, Rundall TG, van de Klundert J, et al. Developing patient portals in a fragmented healthcare system. Int J Med Inform 2015 Oct;84(10):835-846. [doi: 10.1016/j.ijmedinf.2015.07.001] [Medline: 26190049]
14. Prey JE, Polubriaginof F, Kuperman GJ, Tiase V, Collins SA, Vawdrey DK. International perspectives on sharing clinical data with patients. Int J Med Inform 2016 Feb;86:135-141. [doi: 10.1016/j.ijmedinf.2015.11.007] [Medline: 26631956]
15. Rigby M, Georgiou A, Hyppönen H, Ammenwerth E, de Keizer N, Magrabi F, et al. Patient portals as a means of information and communication technology support to patient- centric care coordination - the missing evidence and the challenges of evaluation. A joint contribution of IMIA WG EVAL and EFMI WG EVAL. Yearb Med Inform 2015 Aug 13;10(1):148-159 [FREE Full text] [doi: 10.15265/IY-2015-007] [Medline: 26123909]
16. Opal. Opal Health URL: http://www.opalmedapps.com/ [WebCite Cache ID 74nTJoHoa]
17. Beard L, Schein R, Morra D, Wilson K, Keelan J. The challenges in making electronic health records accessible to patients. J Am Med Inform Assoc 2012;19(1):116-120 [FREE Full text] [doi: 10.1136/amiajnl-2011-000261] [Medline: 22120207]
18. Jensen TB, Thorseng AA. Building National Healthcare Infrastructure: The Case of the Danish e-Health Portal. In: Aanestad M, Grisot M, Hanseth O, Vassilakopoulou P, editors. Information Infrastructures within European Health Care. New York: Springer; 2017:209-224.
19. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
20. Kujala S. User involvement: a review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
21. McConnell S. Rapid Development: Taming Wild Software Schedules. Seattle: Microsoft Press; Jul 12, 1996.
22. Robertson T, Simonsen J. Challenges and opportunities in contemporary participatory design. Des Issues 2012 Jul;28(3):3-9. [doi: 10.1162/DESI_a_00157]
23. Tang T, Lim ME, Mansfield E, McLachlan A, Quan SD. Clinician user involvement in the real world: designing an electronic tool to improve interprofessional communication and collaboration in a hospital setting. Int J Med Inform 2018 Feb;110:90-97 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.11.011] [Medline: 29331258]
24. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]
25. Yip MH, Phaal R, Probert DR. Stakeholder engagement in early stage product-service system development for healthcare informatics. In: Proceedings of PICMET '13: Technology Management in the IT-Driven Services (PICMET). 2013 Presented at: PICMET '13: Technology Management in the IT-Driven Services (PICMET); 28 July-1 Aug 2013; San Jose, CA, USA. [doi: 10.1080/10429247.2014.11432020]

26. Sligo J, Gauld R, Roberts V, Villa L. A literature review for large-scale health information system project planning, implementation and evaluation. Int J Med Inform 2017 Dec;97:86-97. [doi: 10.1016/j.ijmedinf.2016.09.007] [Medline: 27919399]
27. Kushniruk A, Nøhr C. Participatory design, user involvement and health IT evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]
28. Wilson EV. Patient-Centered E-Health Design. In: Patient-Centered E-Health. Hershey: IGI Global; 2008:10-25.
29. Ryan BL, Brown JB, Terry A, Cejic S, Stewart M, Thind A. Implementing and using a patient portal: a qualitative exploration of patient and provider perspectives on engaging patients. J Innov Health Inform 2016 Jul 4;23(2):848 [FREE Full text] [Medline: 27869582]
30. van Limburg M, Wentzel J, Sanderman R, van Gemert-Pijnen L. Business modeling to implement an eHealth portal for infection control: a reflection on co-creation with stakeholders. JMIR Res Protoc 2015 Aug 13;4(3):e104 [FREE Full text] [doi: 10.2196/resprot.4519] [Medline: 26272510]
31. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. Br Med J 2010;341:c5814 [FREE Full text] [Medline: 21081595]
32. Kooij L, Groen WG, van Harten WH. The effectiveness of information technology-supported shared care for patients with chronic disease: a systematic review. J Med Internet Res 2017 Jun 22;19(6):e221 [FREE Full text] [doi: 10.2196/jmir.7405] [Medline: 28642218]
33. The Standish Group.: The Standish Group; 1994. The Chaos Report (1994) URL: https://www.standishgroup.com/sample_research_files/chaos_report_1994.pdf [accessed 2018-06-20] [WebCite Cache ID 70K9BMjxv]
34. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care 2006 Oct;15(5):307-310 [FREE Full text] [doi: 10.1136/qshc.2005.016527] [Medline: 17074863]
35. Boyd H, McKernon S, Mullin B, Old A. Improving healthcare through the use of co-design. N Z Med J 2012 Jun 29;125(1357):76-87. [Medline: 22854362]
36. Conway J, Johnson B, Edgman-Levitan S, Schlucter J, Ford D, Sodomka P, et al. Institute for Patient- and Family-Centered Care. Bethesda, MD; 2006 Jun. Partnering with patients and families to design a patient-and family-centered health care system: a roadmap for the future: a work in progress URL: http://www.ipfcc.org/resources/Roadmap.pdf [accessed 2018-06-20] [WebCite Cache ID 70KBVwCSC]
37. Ekman I, Swedberg K, Taft C, Lindseth A, Norberg A, Brink E, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011 Dec;10(4):248-251. [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
38. Christensen T. Institute for Healthcare Improvement. 2017 Sep 14. The Evolution of Patient-Centered Care and the Meaning of Co-Design URL: http://www.ihi.org/communities/blogs/evolution-of-patient-centered-care-and-the-meaning-of-co-design [accessed 2018-06-21] [WebCite Cache ID 70LcLNsZb]
39. Ahern DK, Woods SS, Lightowler MC, Finley SW, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. Am J Prev Med 2011 May;40(5 Suppl 2):S162-S172. [doi: 10.1016/j.amepre.2011.01.005] [Medline: 21521591]
40. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
41. Beck K, Beedle M, Van Bennekum A, Cockburn A, Cunningham W, Fowler M, et al. Agile Manifesto. 2001. Principles behind the agile manifesto URL: http://agilemanifesto.org/principles.html [accessed 2018-12-03] [WebCite Cache ID 74P6ewrAe]
42. GitHub. URL: https://github.com/ [WebCite Cache ID 74niumWJ3]
43. ZenHub. URL: https://www.zenhub.com/ [WebCite Cache ID 74njl81IK]
44. Crashlytics. URL: https://try.crashlytics.com/ [WebCite Cache ID 74nkIJ60w]
45. Lab Tests Online. URL: https://labtestsonline.org/ [WebCite Cache ID 74nkmDdFA]
46. Canadian Partnership Against Cancer Reports. Toronto: The Canadian Partnership Against Cancer; 2018. Living with Cancer: A Report on the Patient Experience URL: http://www.systemperformance.ca/report/living-with-cancer-patient-experience/ [accessed 2018-12-04] [WebCite Cache ID 74P7Jcwmv]
47. Peacock S, Reddy A, Leveille SG, Walker J, Payne TH, Oster NV, et al. Patient portals and personal health information online: perception, access, and use by US adults. J Am Med Inform Assoc 2017 Apr 1;24(e1):e173-e177. [doi: 10.1093/jamia/ocw095] [Medline: 27413120]
48. Leonard KJ, Wiljer D. Patients are destined to manage their care. Healthc Q 2007;10(3):76-78 [FREE Full text] [Medline: 17626550]
49. MyUHN Patient Portal. 2015. How myUHN Patient Portal can improve care in your clinic: Results from the myUHN early adopter evaluation URL: http://www.uhn.ca/corporate/News/Documents/myUHN_infographic.pdf [accessed 2018-06-21] [WebCite Cache ID 70LcvIZXQ]
50. Grünloh C, Cajander Å, Myreteg G. "The record is our work tool!"-Physicians' framing of a patient portal in Sweden. J Med Internet Res 2016 Dec 27;18(6):e167 [FREE Full text] [doi: 10.2196/jmir.5705] [Medline: 27349531]

51. Giardina TD, Callen J, Georgiou A, Westbrook JI, Greisinger A, Esquivel A, et al. Releasing test results directly to patients: a multisite survey of physician perspectives. Patient Educ Couns 2015 Jun;98(6):788-796 [FREE Full text] [doi: 10.1016/j.pec.2015.02.011] [Medline: 25749024]
52. Miller DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
53. 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 [FREE Full text] [Medline: 12595402]
54. Mák G, Smith FH, Leaver C, Hagens S, Zelmer J. The effects of web-based patient access to laboratory results in British Columbia: a patient survey on comprehension and anxiety. J Med Internet Res 2015;17(8):e191 [FREE Full text] [doi: 10.2196/jmir.4350] [Medline: 26242801]
55. MyUHN Patient Portal. 2016. LEADing Practise Challenge: MyUHN Patient Portal URL: http://imaginenationchallenge.ca/wp-content/uploads/2017/02/myUHN-Patient-Portal.pdf [accessed 2018-12-03] [WebCite Cache ID 74P76n2Jm]
56. Rexhepi H, Åhlfeldt RM, Cajander Å, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: a qualitative study. Health Informatics J 2016 Jul 19;24(2):115-124. [doi: 10.1177/1460458216658778] [Medline: 27440056]

Abbreviations

CT: Computed Tomography
eHealth: electronic health
EMR: electronic medical record
MUHC: McGill University Health Centre
PHI: personal health information
PRO: patient-reported outcome
RI-MUHC: Research Institute-MUHC

Edited by G Eysenbach; submitted 22.06.18; peer-reviewed by P Dykes, SW Bae, J Apolinário-Hagen; comments to author 06.09.18; revised version received 31.10.18; accepted 03.11.18; published 11.02.19.

Please cite as:
Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, Joseph A
Design and Development of a Person-Centered Patient Portal Using Participatory Stakeholder Co-Design
J Med Internet Res 2019;21(2):e11371
URL: http://www.jmir.org/2019/2/e11371/
doi:10.2196/11371
PMID:30741643

Original Paper

Development of the Multidimensional Readiness and Enablement Index for Health Technology (READHY) Tool to Measure Individuals' Health Technology Readiness: Initial Testing in a Cancer Rehabilitation Setting

Lars Kayser[1*], MD, PhD; Sine Rossen[2,3*], MSc, PhD; Astrid Karnoe[4], MSc; Gerald Elsworth[5], PhD; Jette Vibe-Petersen[6], MD; Jesper Frank Christensen[2,3], MSc, PhD; Mathias Ried-Larsen[2,3], MSc, PhD; Richard H Osborne[1,5], PhD

[1]Department of Public Health, University of Copenhagen, Copenhagen, Denmark
[2]Centre of Inflammation and Metabolism, Rigshospitalet, Copenhagen, Denmark
[3]Centre for Physical Activity Research, Rigshospitalet, Copenhagen, Denmark
[4]Danish Multiple Sclerosis Society, Valby, Denmark
[5]Centre for Population Health Research, Faculty of Health, Deakin University, Geelong, Australia
[6]Copenhagen Centre for Cancer and Health, Municipality of Copenhagen, Copenhagen, Denmark
[*]these authors contributed equally

Corresponding Author:
Lars Kayser, MD, PhD
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
Copenhagen, DK-1014
Denmark
Phone: 45 28757291
Email: lk@sund.ku.dk

Abstract

Background: The increasing digitization of health care services with enhanced access to fast internet connections, along with wide use of smartphones, offers the opportunity to get health advice or treatment remotely. For service providers, it is important to consider how consumers can take full advantage of available services and how this can create an enabling environment. However, it is important to consider the digital context and the attributes of current and future users, such as their readiness (ie, knowledge, skills, and attitudes, including trust and motivation).

Objective: The objective of this study was to evaluate how the eHealth Literacy Questionnaire (eHLQ) combined with selected dimensions from the Health Education Impact Questionnaire (heiQ) and the Health Literacy Questionnaire (HLQ) can be used together as an instrument to characterize an individual's level of health technology readiness and explore how the generated data can be used to create health technology readiness profiles of potential users of health technologies and digital health services.

Methods: We administered the instrument and sociodemographic questions to a population of 305 patients with a recent cancer diagnosis referred to rehabilitation in a setting that plans to introduce various technologies to assist the individuals. We evaluated properties of the Readiness and Enablement Index for Health Technology (READHY) instrument using confirmatory factor analysis, convergent and discriminant validity analysis, and exploratory factor analysis. To identify different health technology readiness profiles in the population, we further analyzed the data using hierarchical and k-means cluster analysis.

Results: The confirmatory factor analysis found a suitable fit for the 13 factors with only 1 cross-loading of 1 item between 2 dimensions. The convergent and discriminant validity analysis revealed many factor correlations, suggesting that, in this population, a more parsimonious model might be achieved. Exploratory factor analysis pointed to 5 to 6 constructs based on aggregates of the existing dimensions. The results were not satisfactory, so we performed an 8-factor confirmatory factor analysis, resulting in a good fit with only 1 item cross-loading between 2 dimensions. Cluster analysis showed that data from the READHY instrument can be clustered to create meaningful health technology readiness profiles of users.

Conclusions: The 13 dimensions from heiQ, HLQ, and eHLQ can be used in combination to describe a user's health technology readiness level and degree of enablement. Further studies in other populations are needed to understand whether the associations between dimensions are consistent and the number of dimensions can be reduced.

(J Med Internet Res 2019;21(2):e10377) doi:10.2196/10377

KEYWORDS

health technology readiness; questionnaire; eHealth literacy, enablement

Introduction

Background

The modernization of health care systems and the introduction of technologies are changing how health care systems are being improved and designed, not only in regard to quality and safety, but also for reach and active reduction of health inequalities. To achieve the potential benefits of digitization involving patients, it is important to address the individual's digital readiness. Here, we define readiness in accordance with the Oxford dictionary [1] as *being prepared and willing*, where prepared is interpreted as the result of an individual's knowledge, skills, and attitudes, including trust and motivation, which should be considered as enabling factors as they expand an individual's possibilities to act [2].

Increasing digitization may benefit users and bridge geographical or social gaps [3]. Digitization may, however, impose new barriers for those with low literacy, who are reluctant to use technology [4,5] or have limitations due to physical or cognitive disorders, or who don't have the requisite skills and attitudes, including trust and motivation (ie, those who are not prepared and willing). Health technology has the potential to support those who have the resources and competence to take advantage of the ever-expanding market of public and private health services, and has the potential to identify, include, and empower those who are disadvantaged and may be underserved, if these barriers can be overcome. It is important to be able to differentiate between users with respect to their needs, resources, and health technology readiness in order to be able to provide stratified solutions that address these user differences [6]. In a recent systematic review on instruments generally used to evaluate eHealth interventions, Wakefield et al identified a need for psychometrically robust instruments that are concept based; include several aspects of use, including eHealth literacy; and can be used across platforms and technology [7]. In their viewpoint on eHealth literacy, Griebel et al called attention to a lack of "a well-founded theoretical basis and approaches to put eHealth literacy in a broader context" and "how to link measured levels of eHealth literacy to the development of eHealth services" [8].

To be able to effectively discriminate between users, there is an imperative to fully conceptualize and measure health technology readiness of users or potential consumers of technology-assisted health care services. Despite that a recent review identified 8 frameworks for eHealth readiness [9], we have not been able to identify robust, psychometrically sound instruments to assess eHealth or health technology readiness. The only 2 instruments we could identify were the Patient eHealth Readiness Questionnaire (PERQ) and the Service User Technology Acceptability Questionnaire (SUTAQ). PERQ is not conceptually based or a psychometric instrument but consists of specific questions in relation to internet use, social support, personal abilities, and economic barriers [10]. SUTAQ was developed to assess technology acceptance in the Whole System Demonstrator project and is conceptually and psychometrically robust. The context is the user's experience with a given technology, where an experience of reduced privacy and comfort predicts rejection of the technology, whereas an experience of benefits by the user lowers the likelihood of rejection [11].

Proposal for the Readiness and Enablement Index for Health Technology

Here, we propose a concept-based, psychometrically sound, validated instrument, the Readiness and Enablement Index for Health Technology (READHY), based on the concept of eHealth literacy supplemented with relevant scales from other instruments assessing aspects of self-management and social support: the eHealth Literacy Questionnaire (eHLQ) [12], the Health Education Impact Questionnaire (heiQ) [13], and the Health Literacy Questionnaire (HLQ) [14].

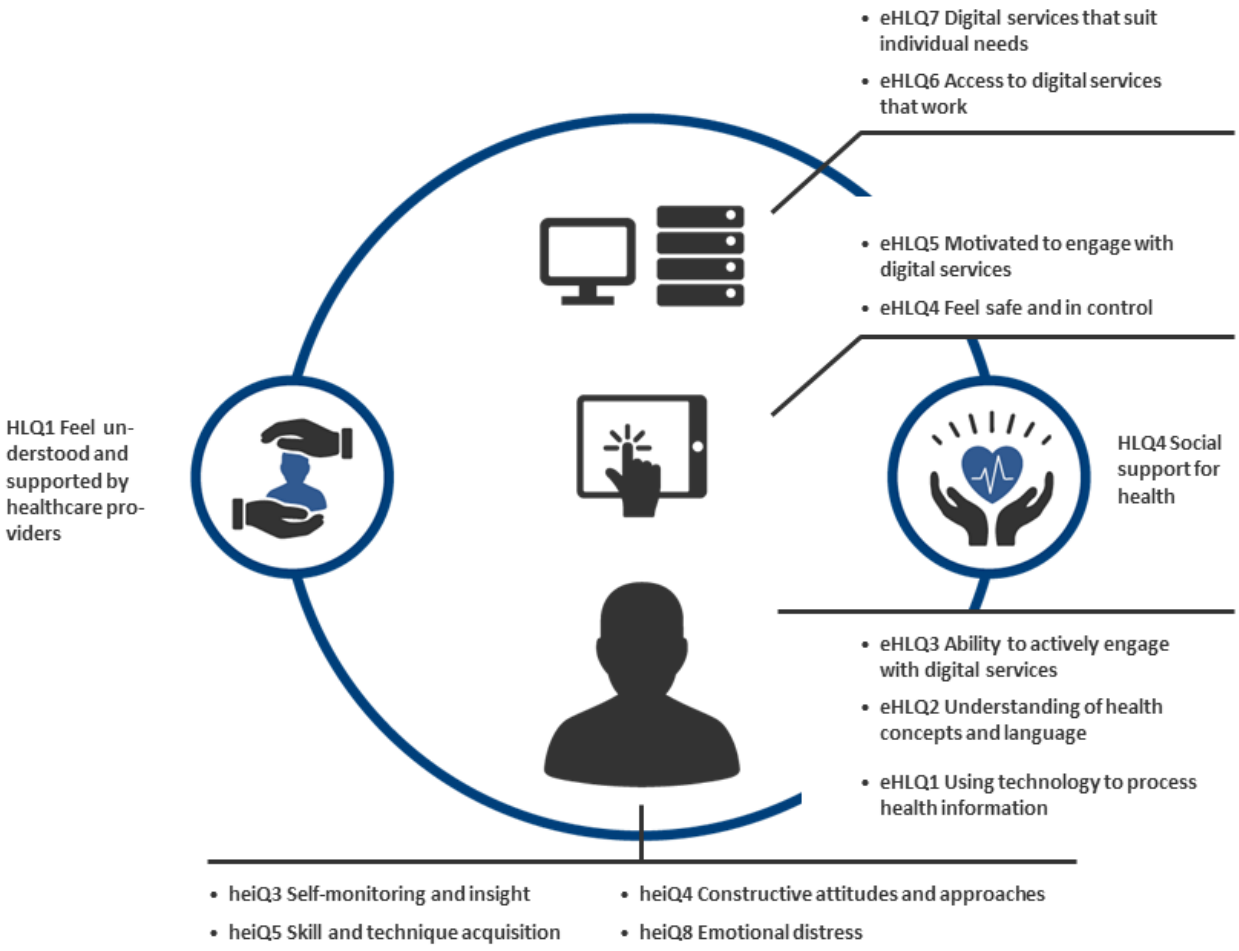
In the period of 2015 to 2017, we developed the eHLQ [12] based on our previously published eHealth literacy framework [15] to enable systematic assessment of eHealth literacy. The eHLQ describes the user's knowledge and skills, and their interaction and experiences with digital health services and technology. The dimensions are as follows: (1) using technology to process health information, (2) understanding of health concepts and language, (3) ability to actively engage with digital services, (4) feel safe and in control, (5) motivated to engage with digital services, (6) access to digital services that work, and (7) digital services that suit individual needs. The eHLQ provides the means to understand an individual's health technology readiness by addressing dimensions of knowledge and skills (eHLQ scales 1 to 3), user experiences (eHLQ6 and eHLQ7), and user trust (eHLQ4) and motivation (eHLQ5). However, Gilstad pointed out that eHealth literacy needs to be understood in a cultural, social, and institutional context [16], and May et al proposed that the burden of treatment may have a negative influence on the individual's mental state [17]. We recognize that these additional areas may be important to assist with fully understanding the degree to which individuals are able, prepared, and willing to use health technologies; therefore, supplementing the eHLQ with these additional areas should deepen our understanding of end users' readiness for full participation in health technology and to monitor how health technology potentially enable users.

We therefore decided to evaluate whether dimensions from the heiQ, which assesses self-management, and domains from the

HLQ, which assesses health literacy, including support from relatives, peers, and health professionals, could supplement the eHLQ. The heiQ comprehensibly evaluates the impact of health education interventions on self-management [13]. The heiQ has 8 dimensions: (1) health-directed activities, (2) positive and active engagement in life, (3) self-monitoring and insight, (4) constructive attitudes and approaches, (5) skill and technique acquisition, (6) social integration and support, (7) health services navigation, and (8) emotional distress. We considered the HLQ, as 2 of its 9 dimensions include the aspect of social context. The HLQ consists of 9 dimensions: (1) feeling understood and supported by health care providers, (2) having sufficient information to manage my health, (3) actively managing my health, (4) social support for health, (5) appraisal of health information, (6) ability to actively engage health care providers, (7) navigating the health care system, (8) ability to find good health information, and (9) understanding health information well enough to know what to do. All 3 instruments were developed using a concept mapping process and a validity-driven approach [18], in which all items are based on statements from users and clustered into concepts that are grounded in the users' collective experiences and knowledge.

To identify suitable dimensions from the heiQ and HLQ, covering social context, capabilities to handle the situation, and burden of disease and treatment, to add to the eHLQ to obtain a complete health technology readiness instrument, we (LK and AK) mapped the 24 dimensions of the eHLQ, heiQ, and HLQ and evaluated the content of the items for this purpose. In 2016, we identified possible dimensions to be included and used information from the literature and ongoing projects known to us. We identified 13 dimensions (Figure 1) from the 3 conceptually distinct instruments. The 7 eHLQ dimensions describe the attributes of the users (information and knowledge about their health and use of technology); the intersection between users and the technologies (their feeling of being safe and in control and their motivation); and users' experience of systems (they work, are accessible, and suits users' needs) [12]. From the heiQ, we selected heiQ3 (self-monitoring and insight), heiQ4 (constructive attitudes and approaches), heiQ5 (skill and technique acquisition), and heiQ8 (emotional distress), as evidence was available that they reflect intended outcomes relevant to educational or technological interventions (Multimedia Appendix 1). These candidate dimensions reflect an individual's capabilities to handle their condition and emotional response. From HLQ, we selected HLQ1 (feeling understood and supported by health care providers) and HLQ4 (social support for health), as they add knowledge about the interaction with and impact of social and health care provider networks.

Figure 1. The 13 dimensions of the Readiness and Enablement Index for Health Technology (READHY) (modified from [12]). The 7 eHealth Literacy Questionnaire (eHLQ) dimensions describe users' attributes; the intersection between users and technologies; and users' experience of systems. The 4 Health Education Impact Questionnaire (heiQ) dimensions add knowledge about the individuals' capabilities to handle their condition and emotional response. The 2 eHLQ dimensions add knowledge about individuals' social context (represented by the circle encompassing the individual and the individual's attributes).



Objective

The aim of this study was to evaluate this new unified instrument, READHY, with exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) and to suggest how the generated data can be used to assess the readiness of potential users of health technologies and digital health services, as well as their degree of enablement; that is, users' knowledge and skills, their self-management of disease, their perceptions and mindset, their experience with health technology systems, and an understanding of the extent to which users feel supported by relatives, peers, and health professionals.

Methods

Research Design, Setting, and Test Population

This study was part of a larger study investigating health technology readiness and motivation for training in the Copenhagen Centre for Cancer and Health, Copenhagen, Denmark. The study used a cross-sectional design with convenience sampling among patients referred to cancer rehabilitation in the Centre in the period August 2016 to July 2017. Questionnaires were administered in the period October 2016 to July 2017. We selected this setting because the organization is legally obligated to offer cancer rehabilitation and is interested in the development of a digital intervention for this context. Patients were either asked to participate in connection with their assessment consultation with their Centre contact person or contacted by telephone by a research project member.

Exclusion criteria were age less than 18 years, insufficient cognitive function, or inability to understand Danish. Individuals who could not be contacted by telephone were excluded after 3 attempts to reach them at varying times. The patient's contact person also had the opportunity to decline a potential participant's participation in the study for reasons not stated.

We considered 300 an adequate sample size for the CFA [19-21]. For example, Iacobucci [19], in reviewing the effect of sample size on convergence, numerical quality, and the typical fit indices used in CFA (and which are also available for EFA and exploratory structural equation modeling [ESEM] in Mplus), concluded that "It is of some comfort that SEM models can perform well, even with small samples (e.g., 50 to 100)." (pg 92). Similarly, Bandalos [20] and Forero et al [21] quoted studies where diagonally weighted least squares analysis

(as used in our investigation in the form of weighted least squares means and variance adjusted [WLSMV] in Mplus) performed acceptably at sample sizes larger than 200, while their own simulation studies suggested that, while larger sample sizes provide increasingly robust results, samples of approximately 300 typically produce relatively unbiased parameter estimates and standard errors with diagonally weighted least squares.

All participants answered a background information questionnaire, the Behavioral Regulation in Exercise Questionnaire, and the READHY instrument.

The READHY Instrument

The READHY instrument consists of 13 dimensions with a total of 65 items from the heiQ, HLQ, and eHLQ. All 13 scales reflect the conceptual dimensions and are rated on a Likert-type scale from 1 (strongly disagree) to 4 (strongly agree). The participants filled out the instrument in paper form. If the participant wished, a research project member was present to clarify questions or assist the participant. The READHY instrument was administered in the same order to all participants. The items were grouped according to the instrument they belong to and administered in the same sequence as in the original instruments. We calculated the overall score of each dimension as the mean of the 4 to 6 items constituting the dimension. Regarding missing items, if 50% or more of the items in a dimension were answered, we calculated an average for the dimension based on the filled-in items.

Statistics

Evaluation of the Properties of READHY Using Confirmatory Factor Analysis

We based factor analysis of the selected 13 heiQ, HLQ, and eHLQ scales on a sequence that followed and extended the general idea of "semiconfirmatory" factor analysis suggested by McDonald [22]. A semiconfirmatory factor analysis is partially confirmatory and partially exploratory. Its aim is to use a known set of homogeneous independent clusters of items as the basis for the exploration of the location of further items that are less well understood in relation to these established item clusters ([22], pg 165). In our case, we considered the 7 eHLQ dimensions as the starting basis for the analysis and aimed to explore the relationship of the selected additional heiQ and HLQ scales to this basic structure. We were, however, cognizant that some eHLQ scales are quite highly correlated [12] and may not be clearly distinguishable in a combined-factor structure. To this end, we incorporated an analysis of the convergent and discriminant validity of the 13 selected scales into the procedure to establish whether some might be combined in a more parsimonious integrated structure. We conducted all analyses with Mplus 8 (Muthén & Muthén), mostly using the ESEM program feature with polychoric correlations, WLSMV estimation appropriate for ordinal data, and geomin rotation when we applied EFA.

The sequence of analyses was as follows. (1) The first analysis was a CFA hypothesizing 13 factors, followed by an analysis of the convergent and discriminant validity of the 13 factors based on the CFA results, using Fornell and Larcker's criteria [23,24]. If the analysis of convergent and discriminant validity suggested that some of the scales were showing insufficient discriminant validity, a more parsimonious factor structure might yield a satisfactory fit to the data. (2) The CFA and analysis of convergent and discriminant validity was followed by a full EFA. When a parallel analysis and scree slope from this analysis clearly suggested that a more parsimonious factor solution might be a satisfactory account of the relationships between the 13 scales, we applied a series of EFA analyses to extract a smaller number of factors. (3) We then hypothesized an ESEM model based on the previous analyses. (4) A final CFA to confirm the revised factor structure and estimate its fit to the data completed the sequence.

We assessed model goodness-of-fit by the root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis fit index (TLI) [25]. We regard a well-fitting model as one where the RMSEA is less than .06 and the CFI and TLI are greater than .95, while we took a value of less than .08 for the RMSEA to indicate a reasonable fit [26-29].

Cluster Analysis

The READHY instrument is intended to characterize populations stratified by their level of health technology readiness. Inspired by the creation of personas in information technology systems development [30] and by the recently published model for the Optimising Health Literacy and Access (OPHELIA) process [31,32], we explored how dimensions can be used for modeling profiles or data to develop personas. To make interpretation easier, we reversed the scale heiQ8 (emotional distress) so that a high score meant less distress.

We applied a combined approach using hierarchical analysis to perform an exploratory evaluation and then the k-means method to evaluate the strength of the resulting number of clusters and to characterize the subgroups. We performed cluster analysis using IBM SPSS statistics version 22 (IBM Corporation). To identify the appropriate number of clusters, we used the method from the OPHELIA process; we performed the hierarchical approach using Ward's method for linkage [33] for a range of cluster solutions. The appropriate number of clusters for the dataset was guided by examining the agglomeration schedule to identify the demarcation point (Multimedia Appendix 2) and the dendrogram to identify when the variance of the dimensions within the clusters increased. We also compared the standard deviations for the group mean profiles of the different cluster solutions, as standard deviations greater than 0.6 could indicate that there are still significant subgroups within the cluster [32]. Guided by the findings of the hierarchical cluster analysis, we performed a k-means cluster analysis of the dataset. As opposed to hierarchical clustering methods where cases are consecutively added to existing clusters, the k-means algorithm constantly reassigns cases to clusters independently of former assignments to minimize the within-cluster variation [34,35]. To obtain a meaningful number of clusters in the given context of cancer survivors, we performed a range of k-means cluster (2-8) solutions, lying around the demarcation point of the hierarchical analysis, to identify the best fit. We evaluated the appropriateness of the k-means cluster solution by the number of dimensions with a standard deviation greater than 0.6 and

by examining the 1-way analysis of variance performed for each dimension to see whether there where variables not contributing to cluster separation (insignificant *F* values) [35]. We populated the various cluster solutions with sociodemographic data to understand which number of clusters made the most sense (from the clinical and sociodemographic sense) in the context of cancer survivors receiving rehabilitation. The final number of clusters decided on should reflect the given population and for what purpose the clusters will be used. In our context of cancer survivors receiving rehabilitation, possibly involving technology, we found a 4-cluster solution to be the most suitable for stratification, when taking both the statistical results and the sociodemographic data into consideration (S Rossen MSc PhD, unpublished data, 2018). We also present an 8-cluster solution to illustrate how clustering can be used for other purposes in explaining and identifying particularly vulnerable subgroups and their characteristics.

Ethics

The project complied with the Declaration of Helsinki and was approved by the Danish Data Protection Agency (2015-55-0630). Under Danish law, permission from an ethics committee was not required because biological material was not used in the study. All participants received oral and written information about the survey and were informed that their participation was voluntary, that they were ensured anonymity, and that all data would be handled confidentially. We obtained written informed consent from all participants.

Results

Study Participants

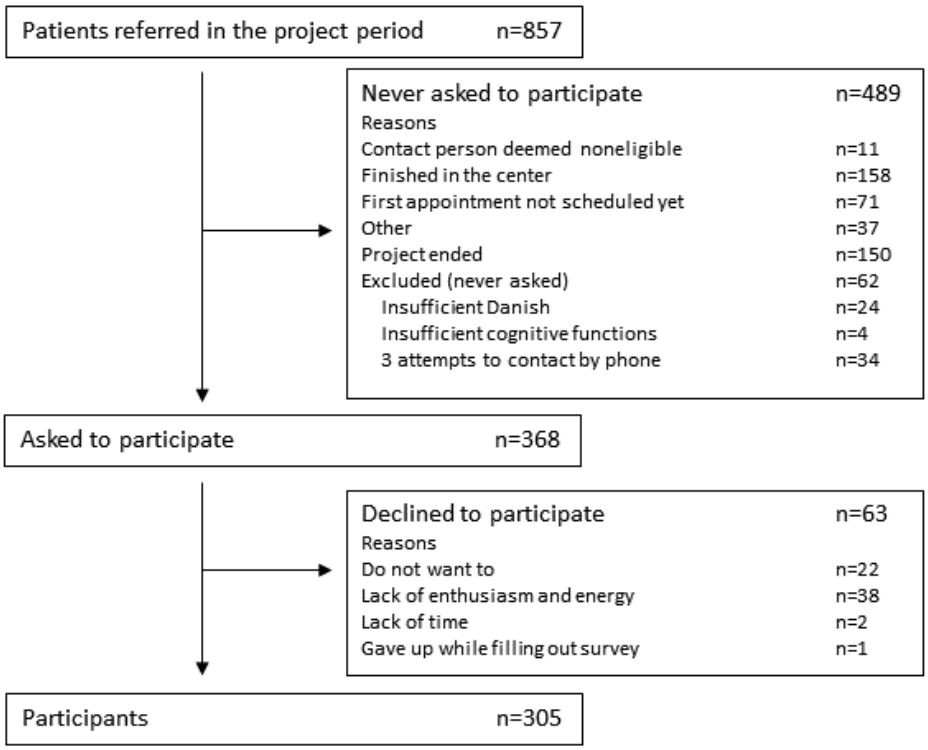
Of 857 patients referred in the project period, 368 were asked to participate. Of these, 63 declined to participate, resulting in 305 participants (see Figure 2 for study population flowchart). Participants had a mean age of 58 years (ranging from 18 to 90 years), 70.8% (216/305) were women, and 81.6% (249/305) owned a smartphone.

Evaluation of the Properties of READHY Using Confirmatory Factor Analysis

The 13-factor CFA demonstrated a reasonable fit for the 13 factors (RMSEA=.049, 95% CI 0.046-0.052; CFI=.94; TLI=.935). With the exception of heiQ3 and eHLQ6, all scales were well identified by the hypothesized items. Only item 10 cross-loaded between heiQ3 and heiQ8 (see Multimedia Appendix 3).

For 2 scales, heiQ3 and eHLQ6, the average variance extracted was less than .5 (Fornell and Larcker's first criterion [23]) and thus where convergent validity was questionable (Multimedia Appendix 4). Also, following Fornell and Larcker's second criterion, 9 pairs of scales showed insufficient discriminant validity where the average variance extracted of 1 or the other in the pair was less than the variance shared between the 2 (heiQ3 and heiQ5; heiQ3 and eHLQ2; eHLQ1 and eHLQ5; eHLQ1 and eHLQ 6; eHLQ2 and eHLQ6; eHLQ3 and eHLQ6; eHLQ5 and eHLQ6; eHLQ5 and eHLQ7; eHLQ and eHLQ7). eHLQ4 showed good discriminant validity, while eHLQ2 was problematic in relation to eHLQ6 because of the low convergent validity of eHLQ6.

Figure 2. Flowchart of the study population.



This suggests that, in this population, a more parsimonious model with a fit similar to the 13-factor model might be achieved. Therefore, we performed, an EFA. The parallel analysis, conducted with maximum likelihood estimation in Mplus, clearly suggested a 5-factor solution, while the scree slope plotted from the eigenvalues of the polychoric correlations used in the WLSMV analyses suggested that a 5- or 6-factor solution might be suitable (Multimedia Appendix 5). When we used close-fit criteria, the 6-factor solution was a satisfactory fit to the data (RMSEA=.046, 95% CI 0.043-0.049; CFI=.953; TLI=.943) with both the RMSEA and the CFI satisfying the prespecified cutoff values, whereas the fit of the 5-factor solution was less satisfactory (RMSEA=.052, 95% CI 0.049-0.055; CFI=.938; TLI=.927). The factor loading pattern from this solution generally paralleled the results of the convergent and discriminant analysis of the 13-factor CFA solution. Items from heiQ scales 4 and 8 loaded on 1 bipolar factor that contrasted the 2 constructs, while items from heiQ scales 3, 4, and 5 loaded on another factor. One resolution of this pattern would be to hypothesize 3 discrete heiQ factors consisting of heiQ4 items, heiQ8 items, and items from heiQ3 and heiQ5 combined. Items from HLQ1 and HLQ4 loaded on the same factor, while another factor was constituted only by secondary loadings from HLQ4 items. The eHLQ items separated into 1 large factor and a smaller factor constituted by eHLQ4, some eHLQ6 items, and, secondarily, 2 eHLQ7 items.

Given the lack of clarity in this solution, we extended the EFA analyses to 7 and 8 factors, and subsequently an 8-factor ESEM model. The ESEM model posited 3 factors comprising heiQ scales (heiQ4 and heiQ8 separately, and heiQ3 and heiQ5 combined) and 2 HLQ scales (HLQ1 and HLQ4 separately). These were CFA factors in that cross-loadings between them were not allowed. Additionally, the model posited 3 factors comprising eHLQ items that were fitted by an EFA component with geomin rotation. The fit of this 8-factor ESEM model was good by all close-fit criteria (RMSEA=.043, 95% CI 0.040-0.046; CFI=.953; TLI=.950). This model clearly confirmed heiQ4, heiQ8, and heiQ3/heiQ5 combined, and HLQ1 and HLQ4 as independent factors. It also suggested that eHLQ1, eHLQ3, eHLQ5, eHLQ6, and eHLQ7 items largely combined into 1 factor with items from eHLQ2 and eHLQ4 identifying separate factors. We conducted a final CFA analysis to confirm this 8-factor model (Multimedia Appendix 6). We required all items to load on only 1 factor as suggested by the ESEM analysis, with 1 exception: we allowed eHLQ item 33 to load on both the “omnibus” eHLQ factor and the eHLQ factor comprising eHLQ4 items. Fit was satisfactory given the extensive model constraints of only 1 allowed cross-loading and no residual correlations (RMSEA=.053, 95% CI 0.051-0.056; CFI=.927; TLI=.923). All hypothesized loadings, with 3 exceptions (including item 33), were greater than 0.4, and all but 6 loadings were greater than 0.5. Item 33, originally an eHLQ6 item was, in this sample, clearly associated with eHLQ4 items.

Cluster Analysis

The hierarchical cluster analysis suggested a 4-cluster solution to be appropriate, and we explored the characteristics of 3, 4, and 5 k-means cluster solutions. Based on the k-means evaluation, we concluded that in this dataset a 4-cluster solution was the best fit. The magnitude of the *F* values from the analysis of variance performed on each dimension indicated that, in particular, eHLQ dimensions discriminated between the clusters. The subgroups of the 4-cluster solution were distinct in their READHY profiles (Figure 3). The heiQ and HLQ scales distinguished between profiles 1 and 2 (high) on the one hand and profiles 3 and 4 (lower) on the other. The eHLQ scales identified 3 groups: profile 1 (high on all), profiles 2 and 3 (middle on all), and profile 4 (low on 5). Profile 4 was lowest on the 5 scales that seemed to be most clearly measuring motivation, access, and capability with digital health services. Profile 1 had the highest smartphone ownership, was the youngest, and had the lowest number of chronic conditions. Profile 4, with the lowest scores, was the oldest and least likely to own a smartphone (Table 1).

When we conducted clustering to identify individuals belonging to certain subgroups, an 8-cluster solution (Figure 4) revealed 2 profiles, 6 and 8, that scored low in the heiQ scales compared with the other clusters. Profile 8 was low on HLQ4 (social support for health) and all eHLQ scales, while profile 6 was in the middle range on HLQ and eHLQ scales. Profile 7 was very low on 4 of the 5 eHLQ scales associated most directly with eHealth motivation, access, and capability and middle to relatively high on heiQ and HLQ scales. Profile 1 was high in eHLQ scales and relatively high in the heiQ and HLQ scales. Profile 8 had a higher number of chronic conditions and was least likely to own a smartphone (Table 1). Profile 6 had a high smartphone ownership and an average of 1 additional chronic condition. Profile 7 was the oldest, had the highest number of chronic conditions, and was less likely to have a smartphone. Profile 1 was the youngest, had the lowest number of chronic conditions, and had the highest smartphone ownership.

Figure 3. Four health technology readiness profiles based on cluster analysis of questionnaires administered to 305 people with a recent diagnosis of cancer. Health Education Impact Questionnaire (heiQ) dimension 8 was reverse scored so that a high score means a low level of distress. eHLQ: eHealth Literacy Questionnaire; HCPs: health care providers; HLQ: Health Literacy Questionnaire; READHY: Readiness and Enablement Index for Health Technology.

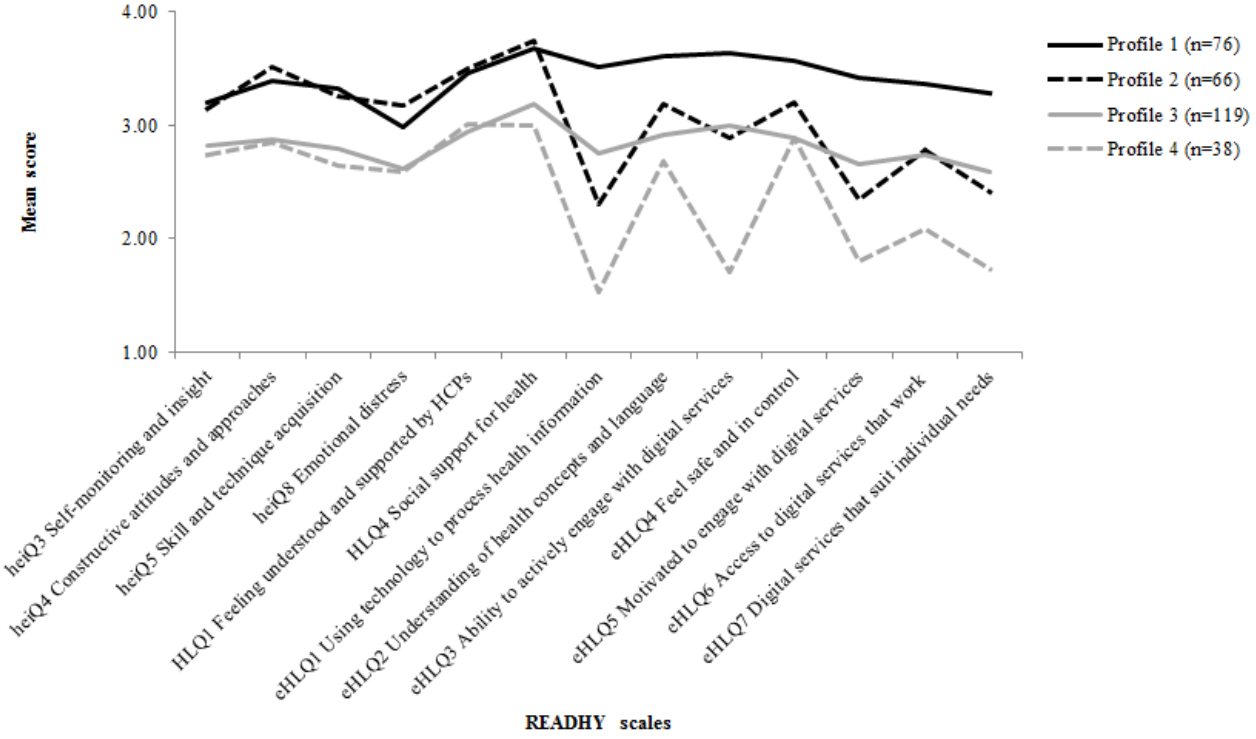
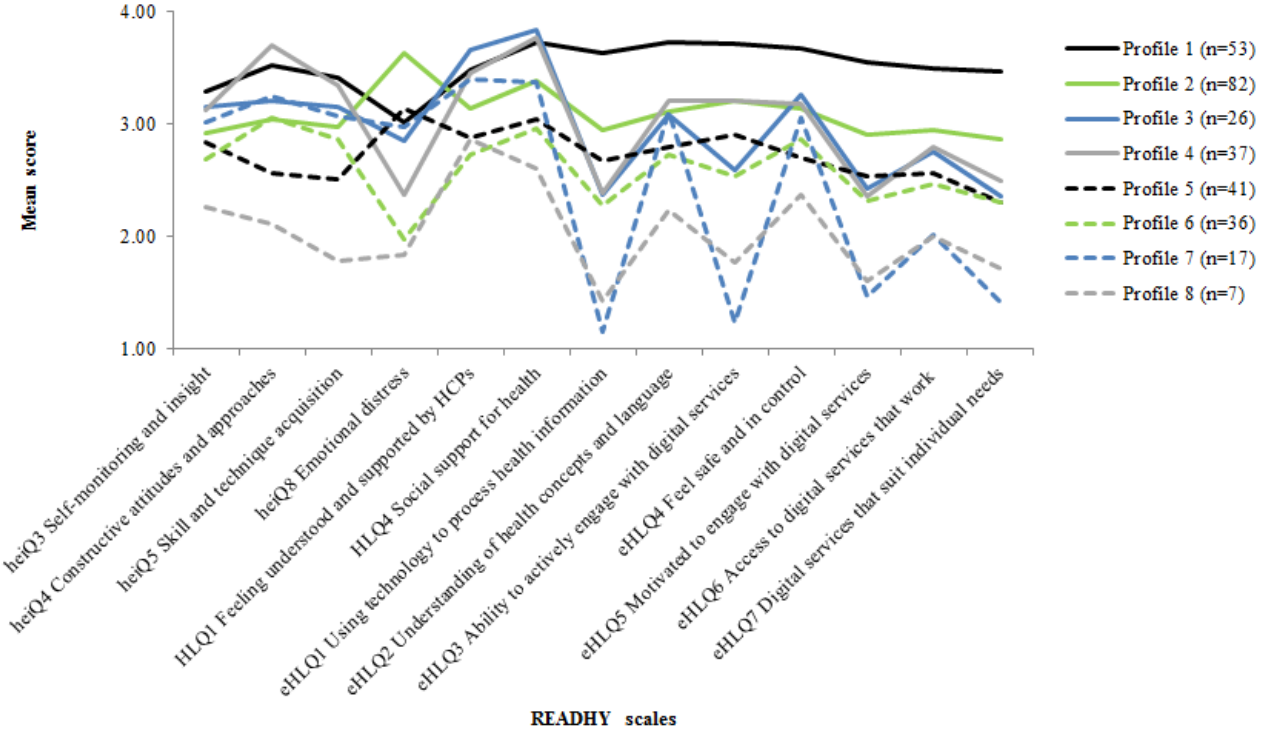


Table 1. Demographic characteristics of individuals across 4- and 8-cluster health technology readiness profiles.

Profiles	Age (years), mean (SD)	Own a smartphone, n (%)	Number of chronic conditions (other than cancer), mean (SD)
4-cluster			
Profile 1 (n=76)	53.9 (13.3)	70 (92)	0.59 (0.87)
Profile 2 (n=66)	58.1 (14.7)	52 (79)	0.67 (0.84)
Profile 3 (n=119)	58.1 (12.8)	106 (89)	0.73 (0.88)
Profile 4 (n=38)	67.8 (12.5)	16 (42)	1.34 (1.05)
8-cluster			
Profile 1 (n=53)	55.1 (13.7)	49 (92)	0.58 (0.93)
Profile 2 (n=37)	56.0 (15.5)	33 (89)	0.67 (0.93)
Profile 3 (n=82)	57.2 (13.2)	72 (88)	0.70 (0.80)
Profile 4 (n=26)	61.2 (12.7)	17 (65)	0.76 (0.97)
Profile 5 (n=36)	62.8 (12.4)	26 (72)	0.56 (0.65)
Profile 6 (n=41)	55.3 (13.9)	36 (88)	1.05 (1.02)
Profile 7 (n=17)	70.9 (10.1)	9 (53)	1.35 (1.17)
Profile 8 (n=7)	58.9 (15.1)	2 (29)	1.29 (0.95)

Figure 4. Eight health technology readiness profiles based on cluster analysis of questionnaires administered to 305 people with a recent diagnosis of cancer. Health Education Impact Questionnaire (heiQ) dimension 8 was reverse scored so that a high score means low level of distress. eHLQ: eHealth Literacy Questionnaire; HCPs: health care providers; HLQ: Health Literacy Questionnaire; READHY: Readiness and Enablement Index for Health Technology.



Discussion

Health Technology Readiness and Enablement

This study presents a new way to understand user health technology readiness by taking advantage of the recently developed eHLQ instrument, supplemented with scales from other psychometrically sound instruments, to create the READHY instrument. The READHY covers aspects of users' knowledge and skills, their self-management of disease, their perceptions and mindset, their experience with health technology systems, and an understanding of the extent to which users feel supported by relatives, peers, and health professionals. Understanding of technology readiness is usually associated with the maturity of a technology [36], including the organizational context [37]. READHY addresses the complementary situation: how ready and able the user is to engage with and take advantage of technologies. This adds to the understanding of the complex situation of a user's interaction with digital tools in a health context.

The eHLQ is a multidimensional instrument that measures eHealth literacy strengths and weaknesses. Each dimension is represented by an independent scale, and the dimensions collectively provide a comprehensive profile of the informant. The eHLQ can be used alone in digital or technology contexts [12]. By adding scales from 2 other psychometrically sound and well-tested instruments derived from the lived experience of users, we extend the utility of the eHLQ to generate a comprehensive description of health technology readiness, as well as ancillary elements that capture the degree to which they are enabled: that is, determinants of users' technological capability, their emotional distress, and their ability to manage their own condition over time. Technology-based interventions may also change the way the individual interacts with health professionals, peers, and relatives. Whether enablement is related to a higher or lower score on the HLQ scales (HLQ1, feeling understood and supported by health care providers; HLQ4, social support for health) remains to be explored.

The strength of the READHY instrument is that the eHLQ dimensions provide insight into users' knowledge and skills (eHLQ1, using technology to process health information; eHLQ2, understanding of health concepts and language; eHLQ3, ability to actively engage with digital services), their self-management of disease (heiQ3, self-monitoring and insight; heiQ5, skills and technique acquisition), their perceptions and mindset (eHLQ4, feel safe and in control; eHLQ5, motivated to engage with digital services; heiQ4, constructive attitudes and approaches; heiQ8, emotional distress), their experience with health technology systems (eHLQ6, access to digital services that work; eHLQ7, digital services that suit individual needs), and an understanding of the extent to which users feel supported by relatives, peers, and health professionals (HLQ1, feeling understood and supported by health care providers; HLQ4, social support for health).

Compared with the 2 other instruments we identified that measure readiness, PERQ [10] and SUTAQ [11], READHY is grounded in a modern concept of eHealth literacy (eHealth literacy framework [15]) and addresses users' knowledge and skills, their self-management of disease, their perceptions and mindset, their experience with health technology systems, and an understanding of the extent to which users feel supported by

relatives, peers, and health professionals. PERQ differs by not being conceptually based or psychometrically validated, but rather consists of specific questions in relation to internet use, social support, personal abilities, and economic burden [10]. SUTAQ is, like READHY, conceptually and psychometrically robust. In contrast to READHY, which conceptually builds on eHealth literacy, SUTAQ was developed to assess technology acceptance in the Whole System Demonstrator project [11]. SUTAQ predicts acceptance based on the user's experience with a given technology, where an experience of reduced privacy and comfort predicts rejection of the technology, whereas an experience of benefits by the users lowers the likelihood of rejection [10]. Compared with READHY, SUTAQ lacks information about the user's knowledge and skills, as well as the context of the burden of the informant's condition.

Evaluation of the Properties of READHY Using Confirmatory Factor Analysis

In our data from a large group of people with a cancer diagnosis engaging in rehabilitation services, we found that discriminant validity between some of the scales was not sufficient, suggesting that a more parsimonious model, with fewer items and scales, might be available. However, we were not able to find a better fit, although our results indicated clear grouping of some dimensions. One example is the high correlation between heiQ3 (self-monitoring and insight) and heiQ5 (skill and technique acquisition). This could be either because the 2 dimensions capture the same construct, or because one (heiQ 5) might *cause* the other (heiQ3); thus, we observed a high correlation between them, but they are different constructs and different interventions would affect changes in these dimensions. Likewise, the connection between eHLQ1, eHLQ3, eHLQ5, eHLQ6, and eHLQ7 may be explained by a causal pathway, that is, having the personal knowledge and skills to engage with digital services and information and having access to systems that work may motivate you to use the systems. While there appears to be some overlap between some of the 13 scales selected, and potential causal hypotheses can be drawn from the data, the 13-factor model works satisfactorily, notwithstanding some factor collapse. If such a pattern is replicated in other patient groups in other settings, particularly among the eHLQ scales, a reduced number of scales might be appropriate for future use, particularly in population studies where questionnaire length may be important. Careful testing of the READHY in other settings receiving technological interventions, such as people with complex conditions such as chronic obstructive pulmonary disease, diabetes, or cardiac heart failure, will provide important empirical information that will assist with reducing the total number of dimensions of the READHY. However, a premature reduction of dimensions may omit information critical in other contexts, such as when robust interventions are introduced, and computer media or system services are changed.

Cluster Analysis

The combination of hierarchical and k-means cluster analysis is a way of identifying clusters of individuals with respect to their attributes in relation to health technology. We further explored the clusters by adding disease-specific and sociodemographic data (age, smartphone ownership, and number of chronic conditions). These different profiles can be addressed in the form of tailored introductions to technology, educational programs, or variations in user interfaces. In addition, these groups can be used to understand the individual's attitude and experiences in relation to health technology, their emotional well-being, and the impact of their condition. For a health technology design or service design process, a lower number of clusters would be appropriate, while separating users into further clusters can help understand how different user types can be supported in their use of technology.

The profiles identified in the cluster analysis may be useful for identifying those who are at risk of being marginalized or need particular interventions. We identified such a group in the 8-cluster analysis. In this subgroup (see profile 8, Figure 4), individuals have high emotional distress and may need assistance to learn to cope with and manage their situation to reduce the burden of their condition before being ready to engage with digital services or technologies. Although only 7 individuals belonged to this group, they were likely to be underrepresented in the sample, given the recruitment process, and they represent at least 2% of the population and thus a considerable number of individuals in a national or regional perspective. This group may well be found to be higher users of health care services as their chronic conditions progress [38]. It is important to characterize high-risk groups, which is reflective in important programs such as Denmark's and other countries' focus on interventions for the 1% of the population that accounts for almost a third of health care expenses [39-45]. This data-driven approach may be a strong tool to identify particular subpopulations to which specific actions and interventions can be tailored. This approach has been effectively incorporated into health system improvement initiatives such as the OPHELIA process [31,32].

Strengths and Limitations of the Study

A strength of the study is the way we sought to minimize missing data for people with low literacy, as we ensured that respondents had the opportunity to have the instrument read aloud and filled out for them by a project member. A limitation to this study is the exclusion of ethnic minorities. Furthermore, it has been shown that referral to the Copenhagen Centre for Cancer and Health is not equally distributed by socioeconomic group. Higher educational level is associated with a higher rate of referral to rehabilitation services [46]. Data analysis used a combination of CFA, EFA, and ESEM. While we used confirmatory analyses at the commencement and completion of the sequence, the conclusion that a more parsimonious selection of scales might be possible also relied on information drawn from the exploratory analyses. As such, the analyses should be regarded as principally exploratory, and the final 8-factor solution should be replicated in further studies with different samples.

Perspectives

With the introduction of READHY, we now have an instrument that can help designers of services and health care providers to better understand an individual's readiness for using digital health services or technology, not only from the eHealth literacy

perspective, but also by including insight into the individual's social network, including health care providers, and their capability to manage their own health. This is particularly important when digital services are provided to people with 1 or more chronic conditions. For this study, we chose cancer survivors. This group covers a wider age range and includes both sexes. In cancer patients, as well as in other groups of people living with 1 or more chronic conditions, it is important to be aware of the burden of the disease and treatment, and the social support from health care professionals, peers, and relatives to be able to provide the digital services and technology in a way that raises the likelihood that these people will adopt the technology. The development of READHY is the first step. We have submitted data for publication that report on users' willingness to use technology and more detailed information about the group of cancer survivors (S Rossen MSc PhD, unpublished data, 2018), and a qualitative study exploring individuals' personal values, perceptions, and experiences in more detail is ongoing. Currently READHY is used in 150 Danish diabetes patients to widen our knowledge about groups living with chronic conditions. READHY is also being used in other studies in Denmark and Norway. In these studies, READHY is being used to understand the segments of users to better develop educative material and strategies, to better design digital solutions that can be tailored to the various segments' specific needs. We expect that future longitudinal studies will contribute to an understanding of how specific interventions will affect patients' knowledge and skills, self-management of disease, perceptions and mindset, experience with health technology systems, and social support.

This multidimensional, person-centered evaluation will help designers and providers to address the particular needs identified. Like in the OPHELIA process [31], identification of dimensions that score low in individuals or groups will help to design specific interventions. In this way, READHY can assist in providing more effective and efficient interventions, as they can be developed based on specific gaps or needs.

The READHY instrument, with further testing in a wide range of settings, may be a promising tool to provide technologists, researchers, health care providers, and policy makers with robust information to ensure that fit-for-purpose and inclusive digital health systems are developed and evaluated across health care systems.

Conclusion

While the READHY is relatively long (13 scales), further efforts to reduce the number of scales seems warranted. While the observed clustering and high correlation between some scales suggests redundancy, many scales are likely to be causally related or influenced by different interventions. Until empirical studies provide evidence on which scales predict future readiness outcomes or respond to specific interventions, we recommend using the current scales or selecting a smaller set based on well-defined a priori hypotheses of which scales are likely to predict target outcomes in new settings. The READHY provides an option for researchers and technology implementers to assess groups of individuals' readiness for health technology. The data derived from the tool will provide rich information in the form of a situational analysis (prior to implementation) and through the complex transitions that individuals, practitioners, and organizations as a whole go through as technological improvements are applied across the health and social care sectors.

Acknowledgments

The authors wish to thank Karen Trier, Carina Nees, Cecilie Fau Brinkløv, Mathilde Ingemann Källstrøm Damm, Sarah Leggett, and Jonas Ravn for being a part of the study team and actively engaging in discussing and planning the study, as well as for organizing and administering collection of questionnaires. This study was supported by TrygFonden. TrygFonden had no role in the study design, collection, analysis, or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication. The Centre for Physical Activity Research, Rigshospitalet, Copenhagen, Denmark, is supported by a grant from TrygFonden. RHO was funded in part through a National Health and Medical Research Council of Australia Senior Research Fellowship (#APP1059122). The eHLQ is available through license by either Deakin University, Geelong, Australia, or by University of Copenhagen. The HLQ and heiQ are available through license by Deakin University.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of heiQ dimensions affected during interventions.

[PDF File (Adobe PDF File), 23KB - jmir_v21i2e10377_app1.pdf]

Multimedia Appendix 2

Identification of the demarcation point.

[PDF File (Adobe PDF File), 16KB - jmir_v21i2e10377_app2.pdf]

Multimedia Appendix 3

Confirmatory factor analysis for the 13-factor model.

[XLSX File (Microsoft Excel File), 19KB - jmir_v21i2e10377_app3.xlsx]

Multimedia Appendix 4

Convergent and discriminant validity for the 13-factor model.

[XLSX File (Microsoft Excel File), 10KB - jmir_v21i2e10377_app4.xlsx]

Multimedia Appendix 5

Scree plot and parallel analysis for the exploratory factor analysis.

[PDF File (Adobe PDF File), 107KB - jmir_v21i2e10377_app5.pdf]

Multimedia Appendix 6

Confirmatory factor analysis for the 8-factor model.

[XLSX File (Microsoft Excel File), 14KB - jmir_v21i2e10377_app6.xlsx]

References

1. English: Oxford Living Dictionaries. Oxford, UK: Oxford University Press; 2018. Definition of readiness in English: readiness URL: https://en.oxforddictionaries.com/definition/readiness [accessed 2018-03-15] [WebCite Cache ID 6xvllHh4f]
2. English: Oxford Living Dictionaries. Oxford, UK: Oxford University Press; 2018. Definition of enable in English: enable URL: https://en.oxforddictionaries.com/definition/enable [accessed 2018-03-15] [WebCite Cache ID 6xvltErWR]
3. Expert Panel on Effective Ways of Investing in Health (EXPH). Disruptive innovation: considerations for health and health care in Europe. Luxembourg: European Commission, Directorate-General for Health and Food Safety; 2016. URL: https://ec.europa.eu/health/expert_panel/sites/expertpanel/files/012_disruptive_innovation_en.pdf [accessed 2018-12-03] [WebCite Cache ID 74OKtyOMA]
4. Kleijnen M, Lee N, Wetzels M. An exploration of consumer resistance to innovation and its antecedents. J Econ Psychol 2009 Jun;30(3):344-357. [doi: 10.1016/j.joep.2009.02.004] [Medline: 25904163]
5. Morison EE. Men, Machines, and Modern Times. Cambridge, MA: MIT Press; 2008.
6. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
7. Wakefield BJ, Turvey CL, Nazi KM, Holman JE, Hogan TP, Shimada SL, et al. Psychometric properties of patient-facing eHealth evaluation measures: systematic review and analysis. J Med Internet Res 2017 Oct 11;19(10):e346 [FREE Full text] [doi: 10.2196/jmir.7638] [Medline: 29021128]
8. Griebel L, Enwald H, Gilstad H, Pohl A, Moreland J, Sedlmayr M. eHealth literacy research--quo vadis? Inform Health Soc Care 2017 Oct 18:1-16. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
9. Mauco KL, Scott RE, Mars M. Critical analysis of e-health readiness assessment frameworks: suitability for application in developing countries. J Telemed Telecare 2018 Feb;24(2):110-117. [doi: 10.1177/1357633X16686548] [Medline: 28008790]
10. Jones R. Development of a Questionnaire and Cross-Sectional Survey of Patient eHealth Readiness and eHealth Inequalities. Med 2 0 2013;2(2):e9 [FREE Full text] [doi: 10.2196/med20.2559] [Medline: 25075244]
11. Hirani SP, Rixon L, Beynon M, Cartwright M, Cleanthous S, Selva A, et al. Quantifying beliefs regarding telehealth: development of the Whole Systems Demonstrator Service User Technology Acceptability Questionnaire. J Telemed Telecare 2017 May;23(4):460-469. [doi: 10.1177/1357633X16649531] [Medline: 27224997]
12. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
13. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
14. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]

15. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag Elearn 2015 Dec 26;7(4):522-540 [FREE Full text]
16. Gilstad H. Toward a comprehensive model of eHealth literacy. In: Jaatun EAA, Brooks E, Berntsen K, Gilstad H, Jaatun MG, editors. Proceedings of the 2nd European Workshop on Practical Aspects of Health Informatics. 2014 Presented at: 2nd European Workshop on Practical Aspects of Health Informatics; May 19-20, 2014; Trondheim, Norway p. 63-72 URL: http://ceur-ws.org/Vol-1251/paper7.pdf [doi: 10.13140/2.1.4569.0247]
17. May CR, Eton DT, Boehmer K, Gallacher K, Hunt K, MacDonald S, et al. Rethinking the patient: using burden of treatment theory to understand the changing dynamics of illness. BMC Health Serv Res 2014;14:281 [FREE Full text] [doi: 10.1186/1472-6963-14-281] [Medline: 24969758]
18. Buchbinder R, Batterham R, Elsworth G, Dionne CE, Irvin E, Osborne RH. A validity-driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. Arthritis Res Ther 2011;13(5):R152 [FREE Full text] [doi: 10.1186/ar3468] [Medline: 21933393]
19. Iacobucci D. Structural equations modeling: Fit Indices, sample size, and advanced topics. J Consum Psychol 2010 Jan;20(1):90-98. [doi: 10.1016/j.jcps.2009.09.003] [Medline: 25904163]
20. Bandalos DL. Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. Struct Equ Model 2014 Jan 31;21(1):102-116. [doi: 10.1080/10705511.2014.859510]
21. Forero CG, Maydeu-Olivares A, Gallardo-Pujol D. Factor analysis with ordinal indicators: a Monte Carlo study comparing DWLS and ULS estimation. Struct Equ Model 2009 Oct 06;16(4):625-641. [doi: 10.1080/10705510903203573]
22. McDonald RP. Semiconfirmatory factor analysis: the example of anxiety and depression. Struct Equ Model 2005 Jan;12(1):163-172. [doi: 10.1207/s15328007sem1201_9]
23. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39. [doi: 10.2307/3151312]
24. Farrell AM. Insufficient discriminant validity: a comment on Bove, Pervan, Beatty, and Shiu (2009). J Bus Res 2010 Mar;63(3):324-327. [doi: 10.1016/j.jbusres.2009.05.003]
25. Bollen KA. Structural Equations With Latent Variables. New York, NY: Wiley; 1989.
26. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990 Mar;107(2):238-246. [Medline: 2320703]
27. Browne M, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing Structural Equation Models. Newbury Park, CA: SAGE; 1993:136-162.
28. West A, Taylor A. Model fit and model selection in structural equation modeling. In: Hoyle RH, editor. Handbook of Structural Equation Modeling. New York, NY: Guilford Publications; 2012:209-231.
29. Yu CY. Evaluating Cut-Off Criteria of Model Fit Indices for Latent Variable Models With Binary and Continuous Outcomes [doctoral thesis]. Los Angeles, CA: University of California, Los Angeles; 2002. URL: https://www.statmodel.com/download/Yudissertation.pdf
30. Cooper A. Chapter 9. Designing for Pleasure. In: The Inmates are Running the Asylum. Indianapolis, IN: Sams Publishing; 1999.
31. Beauchamp A, Batterham RW, Dodson S, Astbury B, Elsworth GR, McPhee C, et al. Systematic development and implementation of interventions to OPtimise Health Literacy and Access (Ophelia). BMC Public Health 2017 Dec 03;17(1):230 [FREE Full text] [doi: 10.1186/s12889-017-4147-5] [Medline: 28253883]
32. Batterham RW, Buchbinder R, Beauchamp A, Dodson S, Elsworth GR, Osborne RH. The OPtimising HEalth LIterAcy (Ophelia) process: study protocol for using health literacy profiling and community engagement to create and implement health reform. BMC Public Health 2014;14:694 [FREE Full text] [doi: 10.1186/1471-2458-14-694] [Medline: 25002024]
33. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963 Mar;58(301):236-244. [doi: 10.1080/01621459.1963.10500845]
34. Mooi E, Sarstedt M. A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics. Berlin, Germany: Springer; 2011.
35. Norušis M. IBM SPSS Statistics 19 Advanced Statistical Procedures Companion. Upper Saddle River, NJ: Prentice Hall; 2012.
36. Mankins JC. Technology readiness assessments: a retrospective. Acta Astronautica 2009 Nov;65(9-10):1216-1223. [doi: 10.1016/j.actaastro.2009.03.058]
37. Flott K, Callahan R, Darzi A, Mayer E. A patient-centered framework for evaluating digital maturity of health services: a systematic review. J Med Internet Res 2016 Apr 14;18(4):e75 [FREE Full text] [doi: 10.2196/jmir.5047] [Medline: 27080852]
38. Runz-Jørgensen SM, Schiøtz ML, Christensen U. Perceived value of eHealth among people living with multimorbidity: a qualitative study. J Comorb 2017;7(1):96-111 [FREE Full text] [doi: 10.15256/joc.2017.7.98] [Medline: 29359124]
39. Aktiv patientstøtte. URL: https://www.regionh.dk/aktivpatientstoette/Sider/default.aspx [accessed 2018-03-15] [WebCite Cache ID 6xvmHRICu]
40. Produktion, Forskning og Innovation, Region Sjælland. Analyse af forudsætninger for øget patientkvalitet og ressourceeffektivitet gennem aktiv patientstøtte i Region Sjælland. 2014. URL: https://bit.ly/2WJ9W3z [accessed 2018-03-15] [WebCite Cache ID 6xvmX42CL]

41. Reinius P, Johansson M, Fjellner A, Werr J, Ohlén G, Edgren G. A telephone-based case-management intervention reduces healthcare utilization for frequent emergency department visitors. Eur J Emerg Med 2013 Oct;20(5):327-334. [doi: 10.1097/MEJ.0b013e328358bf5a] [Medline: 22960802]
42. LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Ann Emerg Med 2010 Jul;56(1):42-48. [doi: 10.1016/j.annemergmed.2010.01.032] [Medline: 20346540]
43. Althaus F, Paroz S, Hugli O, Ghali WA, Daeppen J, Peytremann-Bridevaux I, et al. Effectiveness of interventions targeting frequent users of emergency departments: a systematic review. Ann Emerg Med 2011 Jul;58(1):41-52.e42. [doi: 10.1016/j.annemergmed.2011.03.007] [Medline: 21689565]
44. Skinner J, Carter L, Haxton C. Case management of patients who frequently present to a Scottish emergency department. Emerg Med J 2009 Feb;26(2):103-105. [doi: 10.1136/emj.2008.063081] [Medline: 19164618]
45. Wennberg DE, Marr A, Lang L, O'Malley S, Bennett G. A randomized trial of a telephone care-management strategy. N Engl J Med 2010 Sep 23;363(13):1245-1255. [doi: 10.1056/NEJMsa0902321] [Medline: 20860506]
46. Moustsen IR, Larsen SB, Vibe-Petersen J, Trier K, Bidstrup PE, Andersen KK, et al. Social position and referral to rehabilitation among cancer patients. Acta Oncol 2015 May;54(5):720-726. [doi: 10.3109/0284186X.2014.997836] [Medline: 25761091]

Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
eHLQ: eHealth Literacy Questionnaire
ESEM: exploratory structural equation modeling
heiQ: Health Education Impact Questionnaire
HLQ: Health Literacy Questionnaire
OPHELIA: Optimising Health Literacy and Access
PERQ: Patient eHealth Readiness Questionnaire
READHY: Readiness and Enablement Index for Health Technology
RMSEA: root mean square error of approximation
SUTAQ: Service User Technology Acceptability Questionnaire
TLI: Tucker-Lewis fit index
WLSMV: weighted least squares mean and variance adjusted

Edited by G Eysenbach; submitted 15.03.18; peer-reviewed by D Ricard, T Kuhlmann; comments to author 18.08.18; revised version received 12.10.18; accepted 25.11.18; published 12.02.19.

Please cite as:
Kayser L, Rossen S, Karnoe A, Elsworth G, Vibe-Petersen J, Christensen JF, Ried-Larsen M, Osborne RH
Development of the Multidimensional Readiness and Enablement Index for Health Technology (READHY) Tool to Measure Individuals' Health Technology Readiness: Initial Testing in a Cancer Rehabilitation Setting
J Med Internet Res 2019;21(2):e10377
URL: http://www.jmir.org/2019/2/e10377/
doi:10.2196/10377
PMID:30747717

Review

Use of Learning Analytics Data in Health Care–Related Educational Disciplines: Systematic Review

Albert KM Chan[1], BS, MSc, CPE, LLB; Michael G Botelho[1], BDS, MSc, FRACDS, MRD, RCS, PhD; Otto LT Lam[1], BSc, BDS (Hons), MEd, PhD, ACRP-CP

Prosthodontics, Faculty of Dentistry, The University of Hong Kong, Sai Ying Pun, China (Hong Kong)

Corresponding Author:
Michael G Botelho, BDS, MSc, FRACDS, MRD, RCS, PhD
Prosthodontics
Faculty of Dentistry
The University of Hong Kong
Prince Philip Dental Hospital
34 Hospital Road
Sai Ying Pun,
China (Hong Kong)
Phone: 852 2859 0412
Fax: 852 2858 6114
Email: botelho@hku.hk

Abstract

Background: While the application of learning analytics in tertiary education has received increasing attention in recent years, a much smaller number have explored its use in health care-related educational studies.

Objective: This systematic review aims to examine the use of e-learning analytics data in health care studies with regards to how the analytics is reported and if there is a relationship between e-learning analytics and learning outcomes.

Methods: We performed comprehensive searches of papers from 4 electronic databases (MEDLINE, EBSCOhost, Web of Science, and ERIC) to identify relevant papers. Qualitative studies were excluded from this review. Papers were screened by 2 independent reviewers. We selected qualified studies for further investigation.

Results: A total of 537 papers were screened, and 19 papers were identified. With regards to analytics undertaken, 11 studies reported the number of connections and time spent on e-learning. Learning outcome measures were defined by summative final assessment marks or grades. In addition, significant statistical results of the relationships between e-learning usage and learning outcomes were reported in 12 of the identified papers. In general, students who engaged more in e-learning resources would get better academic attainments. However, 2 papers reported otherwise with better performing students consuming less e-learning videos. A total of 14 papers utilized satisfaction questionnaires for students, and all were positive in their attitude toward e-learning. Furthermore, 6 of 19 papers reported descriptive statistics only, with no statistical analysis.

Conclusions: The nature of e-learning activities reported in this review was varied and not detailed well. In addition, there appeared to be inadequate reporting of learning analytics data observed in over half of the selected papers with regards to definitions and lack of detailed information of what the analytic was recording. Although learning analytics data capture is popular, a lack of detail is apparent with regards to the capturing of meaningful and comparable data. In particular, most analytics record access to a management system or particular e-learning materials, which may not necessarily detail meaningful learning time or interaction. Hence, learning analytics data should be designed to record the time spent on learning and focus on key learning activities. Finally, recommendations are made for future studies.

(J Med Internet Res 2019;21(2):e11241) doi:10.2196/11241

KEYWORDS

education; e-learning; learning analytics; learning management systems; online learning; systematic review

Introduction

Learning analytics has been defined [1] as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs"; this broad definition allows the inclusion of virtually anything related to learning. From a more holistic perspective, Picciano [2] proposed that learning analytics is a process that can provide conclusions for decision making through the examination of data such as for helping colleges and universities to identify and evaluate strategies for improving the retention of students. In addition, it can help instructors to decide if an intervention is needed to assist students.

More recently, the 2016 Horizon Report [3] increasingly emphasized the usage of Web-based tools and platforms and described learning analytics as "an educational application of web analytics aimed at learner profiling, a process of gathering and analyzing details of individual student interactions in online learning activities."

While a universal definition of learning analytics has not yet reached a consensus, there is general agreement that it is a relatively new [4] and emerging [5,6] tool in academic research, which can be used to track and store students' Web-based learning activities [5,7]. Higher education institutions collect a vast amount of information from students regarding their use of e-learning resources in the form of activity logs and other digital footprints such as time and date, student demographics, course enrollments, survey questionnaires, library usage, and academic grades [8]. A wide range of e-learning resources ranging from administration, assessment, assignment, quiz, multimedia, to collaboration, to name a few, now can be found integrated into learning management systems (LMS). Looking at such learning analytics data can allow researchers to investigate and examine relations between students' e-learning use and academic performance [9].

Furthermore, learning analytics offers a convenient and potentially accurate method to capture students' interactions with the e-learning resources, which was not achievable in the past. Previously, student engagement was measured by class attendance [10] and self-reported questionnaires. For example, questions such as "How frequently do you use *X* " [11] would be administered to acquire information from students about their e-learning usage. However, self-reported answers have the disadvantage of being inaccurate (recall bias). With learning analytics, investigators can collect information such as the exact number of video watches, and time and date the videos were viewed.

One of the main practical applications for learning analytics data is the investigation of students' e-learning usage and the examination of its effect on their academic performance. Students' e-learning usage behaviors, such as the number of log-ins, time spent on e-learning platforms, and use of other resources, have been studied and found to be positively associated with academic performance outcomes such as summative multiple-choice question (MCQ) exam scores [12,13]. In addition, learning analytics can allow educators to examine an individual student's tracked Web-based activities and search for any at-risk students with one of its predictive functions, then intervene by providing feedback and instructional content [5,14,15].

Although learning analytics applications in higher education have received increasing attention in recent years [5], a limited number of studies have investigated its use in health care-related educational studies. With the acknowledgment of advantages in learning analytics data, this paper aims to review the use of learning analytics in e-learning in health care educational studies with regards to how it is reported and how this may be related to learning outcomes.

Methods

Textboxes 1 and 2 present details pertaining to study populations, interventions, comparisons, and outcomes are presented in accordance with the PICO (population, intervention, comparison, and outcome) model [16]. Textbox 3 details the keywords used in the systematic search of 4 electronic databases (MEDLINE, EBSCOhost, Web of Science, and ERIC). The end search date was August 25, 2017.

We limited the searches to papers that spanned from 2000 to 2017, were involved in tertiary-level education in health care-related disciplines, and were in English. We excluded gray literature. In addition, reviews and commentary columns were discarded, and multiple papers on the same research data were excluded.

After deleting duplicates, all papers retrieved from these initial search criteria were subjected to a screening process by reading titles and abstracts. The detailed information regarding the data type, e-learning content, learning outcome, and key findings were analyzed.

Textbox 1. The inclusion criteria based on the PICO (population, intervention, comparison, and outcome) model.

P: Studies that involved undergraduate or postgraduate students in health care-related disciplines (eg, dentistry, medicine, nursing, and pharmacy).

I: Studies that explored learning analytics were included. Owing to the fact that some studies reported learning analytics data but did not specifically utilize the term "learning analytics," this review also included studies that used learning management systems, educational technologies, or other tools that contained digital footprints of students' e-learning usage.

C: Regarding comparisons, Jin and Bridges [17] suggested that experimental designs should not be considered exclusively because a large proportion of educational research in these fields is case-based as well.

O: Studies that mentioned quantitative measurements of learning outcomes, such as student academic performance related to their knowledge or skill assessment were included.

Textbox 2. The exclusion criteria based on the PICO (population, intervention, comparison, and outcome) model.

P: Studies that reported only instructors, staff, or physicians were excluded.

I: Anything other than those included in the inclusion criteria.

C: Studies that involved only qualitative methods were not included.

O: Anything that did not mention the word “learning outcome” or “GPA/grade”

Textbox 3. The database search strategy.

(clinical OR dent* OR med* OR nursing OR pharmacy) AND

(undergraduate OR postgraduate) AND

(academic achievement OR academic attainment OR assessment OR GPA or grade or consumption) AND

(educational technologies OR learning management system OR content management system OR virtual learning environment OR technology enhanced learning OR learning analytics OR digital footprint OR e-learning OR logs) AND

(education OR learning OR training)

Results

Principal Results

The search of the 4 databases resulted in 537 papers (Figure 1), and a total of 337 papers were obtained after the removal of duplicate results. These were further screened by the same independent researchers, and 296 papers were excluded because of not satisfying the inclusion criteria. Based on the content in abstracts, full texts of 31 potentially effective papers were retrieved and screened, from which a total of 19 papers met all the inclusion criteria and were subjected to content evaluation.

Learning Analytics Data Types

A total of 19 studies reported analytic data based on the number of connections, time spent, or combinations of these and other analytic approaches such as the number of forum posts, MCQ exam scores, and responses of perception questionnaires (ie, satisfaction). Table 1 summarizes the types of learning analytics data among the studies.

Figure 1. Flowchart of the search process.

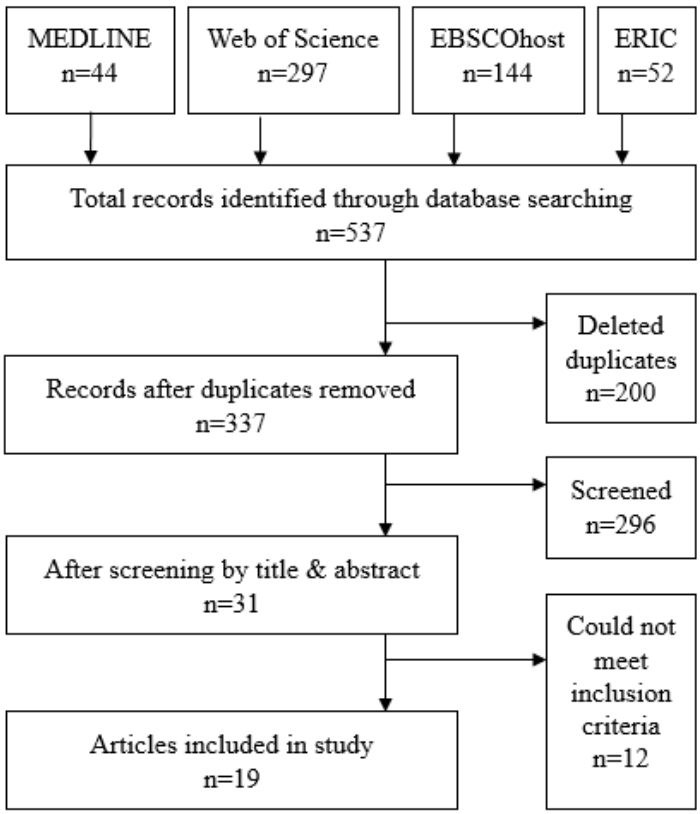


Table 1. The list of learning analytics data type.

Paper	Learning analytics data type				
	Number of connections	Time spent	Both	Number of posts	Multiple-choice question or Likert
Boye et al [18]		✓			
Catteau et al [19]			✓		✓
Chastonay et al [20]	✓			✓	
Colthorpe et al [21]	✓				
Costa-Santos et al [22]	✓				
Critchley et al [12]			✓		✓
Davidson and Candy [23]		✓			✓
DiLullo et al [24]	✓				
Franson et al [25]			✓		
Garrett et al [26]			✓		
Gurpinar et al [27]	✓				
Kleinsorgen et al [28]			✓		✓
Kukolja-Taradi et al [29]	✓			✓	✓
Lameris et al [30]			✓		
Mahnken et al [31]			✓		
Poot et al [13]	✓				
Reimer et al [32]			✓		
Romanov and Nevgi [33]			✓	✓	✓
Saqr et al [34]			✓	✓	✓

Of the 19 studies, 7 recorded the number of connections (ie, log-in or click or hit or visit or access, etc) to the learning platforms [13,20-22,24,27,29]. Of these, 3 studies only recorded connections to the platform or folder and not the specific learning items contained within [20,21,24].

Two studies reported only the length of time students spent on certain e-learning activities such as e-learning animations [18] and game-based platforms [23]. However, 10 studies used both the number of connections and length of time spent on the e-learning activity by students [12,19,25,26,28,30-34] (eg, the number of connections to the e-platform and the accumulative time spent completing the Web-based modules [19], or connections and time on pages of an e-portfolio system [26]).

The Nature of E-Learning Content

The types of e-learning interventions varied widely (Table 2), with the more commonly used e-learning formats including videos and animation clips [12,18,19,21,23,24,28,29,32,33], Web-based text documents (eg, handouts, book chapters, journal papers, and slides) [12,19,20,22,24-29], Web-based tests or quizzes [19,22,27-33], and URL links to external Web-based resources [22,27-29].

In particular, 10 papers reported the use of videos, which ranged from animations and lecture recordings, to laboratory precaution clips. In addition, 10 papers utilized e-learning content presented in the form of Web-based text, including but not limited to, slides, PDFs, text documents, and scanned book chapters. A total of 10 studies investigated the use of e-assessments (eg, tests and quizzes), and 4 studies used URL links to external websites. The majority (14/19) of the studies used >1 type of e-learning resource, and 5 studies used a single type [13,18,21,30,31].

Outcome Measures—Learning and Evaluation

The learning outcomes investigated in the studies were limited to examination or test results with outcomes reported as either a single exam score or a final course grade, which could be a combination of midsemester exams, written assignments, and the end-of-semester exam results. Overall, 14 papers documented learning outcomes (Table 3) as measured by either an end-of-course exam [18,25,27,29,30,32,33] or a combination with course assessments [12,13,21-23,31,34].

Table 2. The list of e-learning content.

Paper	e-Learning content				
	Video	Text doc	Test or quiz	URL	Others
Boye et al [18]	✓				
Catteau et al [19]	✓	✓	✓		✓
Chastonay et al [20]		✓			✓
Colthorpe et al [21]	✓				
Costa-Santos et al [22]		✓	✓	✓	✓
Critchley et al [12]	✓	✓			✓
Davidson and Candy [23]	✓				✓
DiLullo et al [24]	✓	✓			✓
Franson et al [25]		✓			✓
Garrett et al [26]		✓			✓
Gurpinar et al [27]		✓	✓	✓	✓
Kleinsorgen et al [28]	✓	✓	✓	✓	✓
Kukolja-Taradi et al [29]	✓	✓	✓	✓	✓
Lameris et al [30]			✓		
Mahnken et al [31]			✓		
Poot et al [13]					✓
Reimer et al [32]	✓		✓		
Romanov and Nevgi [33]	✓		✓		✓
Saqr et al [34]			✓		✓

Five studies did not include objective learning outcomes [19,20,24,26,28]; of these, 1 reported self-efficacy perception from students [19], while another paper mentioned it did monitor assessments and students completed exams [20] but did not subject these data to any statistical analysis nor descriptive statistics. The remaining 4 papers did not include any learning outcome variables in the study designs.

In this study, 14 papers reported student feedback or evaluation questionnaires (Table 3), which assessed their satisfaction toward the e-learning resources [12,13,18-20,22-24,26-30,32] and were all positive. For example, Boye et al [18] reported that the e-learning was well appreciated by students. Garrett et al [26] reported students valued the accessibility and convenience that came from an electronic portfolio. A total of 9 studies incorporated both objective learning outcomes and evaluation questionnaire data [12,13,18,22,23,27,29,30,32].

Statistical Analyses

Of 19 papers, 5 reported only descriptive statistics [20,23,24,26,32]. The remaining 14 studies [12,13,18,19,21,22,25,27-31,33,34] performed additional statistical analyses, including the Mann-Whitney *U*-test [18,21,27,33], Student *t* test [21,27,31,33], regression (linear, multiple, or logistic) analyses [12,18,25,30,34], analysis of variance (univariate analysis of variance and multivariate analysis of variance) [13,22,30,31], and correlation tests (Pearson and Spearman) [19,31,34].

Relating Learning Analytics Data to Learning Outcomes

With regards to the relationship of learning analytics data to objective learning outcomes, 12 of 13 studies demonstrated significant results [12,13,18,19,21,22,25,29-31,33,34], as presented below. One study [19] only analyzed students' learning analytics on their self-efficacy responses as a replacement of learning outcome.

Within the Cohort Relationship of Learning Outcomes and Learning Analytics Data

Perhaps, the most meaningful examination of learning analytics is how it relates to learning outcomes. In a study examining radiology students' consumption of "e-cases," a significant correlation was found between their "improvement in knowledge" (pre- and postcourse assessments) and the number of e-cases accessed for those who chose to access e-cases (r=0.50, P=.003) and those who were required to access (r=0.46, P=.008) [31].

In a blended learning course for "introductory medical education," the first-year medical students' frequency of log-ins to access the LMS was found to have the strongest correlation with their final grade (r=0.47, P<.01). The second strongest correlation to the final grade was the number of attempts that students took the formative quiz (r=0.46, P<.01), and the third was students' formative assessment grades (r=0.43, P<.01) [34].

Table 3. Reported learning outcomes, self-evaluations, and statistical tests.

Paper	Learning outcome (final exam or grade)	Evaluation (satisfaction)	Statistical test (eg, learning outcome versus e-learning)
Boye et al [18]	✓	✓	✓
Catteau et al [19]	Self-efficacy	✓	✓
Chastonay et al [20]		✓	Descriptive
Colthorpe et al [21]	✓		✓
Costa-Santos et al [22]	✓	✓	✓
Critchley et al [12]	✓	✓	✓
Davidson and Candy [23]	✓	✓	Descriptive
DiLullo et al [24]		✓	Descriptive
Franson et al [25]	✓		✓
Garrett et al [26]		✓	Descriptive
Gurpinar et al [27]	✓	✓	✓
Kleinsorgen et al [28]		✓	✓
Kukolja-Taradi et al [29]	✓	✓	✓
Lameris et al [30]	✓	✓	✓
Mahnken et al [31]	✓		✓
Poot et al [13]	✓	✓	✓
Reimer et al [32]	✓	✓	Descriptive
Romanov and Nevgi [33]	✓		✓
Saqr et al [34]	✓		✓

In a study in which case-based e-learning scenarios were used to assist medical students in learning anesthesia diagnostic decision making, learning analytics data, such as the number of students' completed e-cases, were recorded in an LMS. It was found that students' marks in the final MCQ test (r=0.21, P<.05), as well as 2 graded case reports (r=0.25, P<.01 and r=0.30, P<.01), were significantly correlated to their second attempt in completing the e-cases and the number of log-ins they made to the LMS (MCQ: r=0.18, P<.05; case report 1: r=0.24, P<.01; and case report 2: r=0.32, P<.01) [12].

In a study concerning a "Teaching Resource Centre (TRC) database" for clinical pharmacology, researchers performed a regression analysis and found that students in 2 separate academic years could increase their grades by 32% (P<.001) and 55% (P<.001) with more time spent on accessing the "database" [25].

Similarly, Boye et al [18] found that students with intermediate scores in their immunology tests improved their outcome by 3.6% for every hour of watching animated e-learning clips (P=.005) from regression analysis to their learning analytics data.

Within the Cohort Comparison of Learning Outcomes and Learning Analytics Data

Most studies in this review were interested in finding differences in learning outcomes from different groups of students according to their learning analytics records. Researchers would group students according to the e-learning usage intensity and looked for between-group differences. In a study [21] examining video consumption by physiotherapy or speech pathology students, a marked difference was observed between the groups of high and low "performers." The high or low performer categorization was defined by considering a range of students' access to lecture recordings, "meta-learning" tasks, student submission dates of assignments, and assignment and course grades through cluster analysis. Perhaps counterintuitively, students who "performed" better academically were found to have watched lesser Web-based lecture video recordings than students who performed poorer (P<.05). Furthermore, the paper reported that higher "performers" achieved a higher course grade than lower "performers" (P<.001).

Costa-Santos et al [22] stratified medical students' Web-based "mini-test" grade results as low, medium, and high and found significantly different outcomes between these on their final exam grades in the biostatistics (P<.001) and medical informatics (P<.001) modules. Furthermore, they observed that the average grades on the final exam were higher when the "mini-test" results were higher.

In a Web-based course about "acid-base balance in humans," it was reported that medical students' "knowledge gain" was significantly improved (P<.001) by the e-course [29]; this was determined by students' "knowledge gain" based on pre- and posttest scores. Furthermore, frequencies of "students' logs" and students' visited pages were recorded; however, no statistical analysis was performed relating to the learning outcomes.

A study examining completed modules on an internet-based app found that medical students classified as moderate (P=.04) and intensive users (P<.001) scored significantly higher than nonusers on the final exam [30]. Each module contained only MCQs, and the study aimed to determine whether a formative testing approach would be effective in stimulating students' study performance. The classification of nonuser or moderate or intensive was defined by counting the number of modules completed.

In a physiology course where students created and answered peer-generated questions, students who logged into the platform showed significantly higher scores in a summative test compared with those who did not log on to the platform (P=.001) [13]. While the number of questions created and answered was recorded, these were not tested for their association with summative test scores.

In another study, medical informatics students who watched Web-based videos had significantly higher final exam scores when compared with nonvideo-watchers (P=.007) [33]. Nevertheless, the frequencies of video watches were documented from students' self-administered questionnaires and were not recorded digitally.

While most papers had shown significant results, Gurpinar et al [27] and Mahnken et al [31] reported otherwise. Gurpinar et al [27] created a website with several pages that contained various e-learning resources. The frequencies of page visits were recorded and it was found that medical students who visited the pages during problem-based learning period exhibited slightly higher exam marks than those who did not visit any pages; however, the result was not statistically significant (P=.12).

Mahnken et al [31] did not find a significant difference (P=.54) in "improvement in knowledge" (pre- and postcourse assessments) between radiology students who viewed e-cases and those who did not; this is despite the finding that within the stratum of e-case consumers, marked correlations with "improvement in knowledge" were found for both groups of students who were required and not forced to consume the e-cases.

Discussion

Principal Findings

Prior educational studies utilizing learning analytics data in nonhealth care disciplines have reported improvements in learning outcomes with increased e-learning interactions [35-37]; these studies have supported the use of e-learning materials as effective in improving students' academic performance. The findings are congruent with the data from the health care disciplines.

The majority of papers reviewed in this study utilized >1 type of e-learning material, with videos, Web-based documents, and quizzes being the 3 most common; this reflects a diversity of needs and functionalities among course designers in health care curricula. For example, showing a video can provide easier understanding than a description in the text of a complex concept or clinical procedure. Conversely, presenting a written document may be more concise than other means of multimedia e-learning content. Woodham et al [38] found that students preferred a text-based format and believed that the use of video slowed their pace to review and appraise the learning materials in their problem-based learning curriculum. As a result, providing the required content appropriately and efficiently for students' use may increase their engagement in e-learning use. Based on the diverse formats of e-learning (ie, video, PDF, quiz, etc) and the combinations of these used, the most effective e-learning formats to support learning has not been determined. Hence, further research is needed.

While the majority of studies found that assessed learning outcomes were improved by using e-learning resources, 2 papers reported insignificant findings but did not explain why this occurred [27,31]. A possible reason could be that students in these 2 studies did not become more knowledgeable with the provided e-learning materials; this may be because the conventional in-class teaching was sufficient or the Web-based materials were not aligned to the assessment outcomes and, therefore, did not support the learning outcomes.

Although most papers reported a positive relationship between more e-learning usage and better academic attainment, there were 2 that reported otherwise. While Boye et al [18] observed a slight increase in students' grades through the watching of more animated e-learning videos, this was only for students with intermediate scores. Students with either good or weak academic grades did not appear to benefit. They reasoned that students with better academic performance might already be inclined to study more through conventional means, such as books, lectures, or tutorial sessions, limiting any potential additional benefit from supplementary e-learning materials. In contrast, weaker students may try to compensate for their lack of participation during regular teaching sessions by spending more time with the e-learning materials in "the last minute" prior to examinations. In addition, Colthorpe et al [21] showed that better performing students watched lesser videos than inferior students. They reasoned that students with greater understanding of the learning materials may not consume the video content as they have already mastered it. From this, we may infer that watching more videos may not lead to better learning outcomes for particular students, especially if students do not see the purpose or meaning for their learning.

Students' satisfaction with Web-based learning materials was attributed to its ease of access [22], as well as its usefulness as a "complement" to lectures [32] and a "good supplement" to regular teaching [18]. Students felt that e-learning could increase competence in the subject [13].

Aside from descriptive and diagnostic analytics, higher levels of learning analysis, such as predictive analysis (ie, regression and modeling), could be used to apply learning analytics data within the scope of students' need. For example, Purdue University's Signals project is, perhaps, the most famous example of the successful application of learning analytics' predictive modeling to identify at-risk students. By providing a real-time "red or amber or green traffic light" to students and teachers based on the data collected from the LMS, learning

analytics can help in identifying students who are at risk so that help and support could be provided to them [39-42]. Teachers can then target interventions ranging from emailing those who are at risk to referring them to academic advisors or meeting them face-to-face [39]. In another case, the University of Alabama developed a model to predict student retention rate by using freshmen's Web-based data records with various parameters such as students' English course grades, total hours earned, and their demographics information [5,43]. These are just a few examples of learning analytics applications in higher education. It is predicted that in the coming years, learning analytics will be widely implemented in Web-based education to identify the patterns of student behaviors for improving students' learning and their retention rates [5].

Variations of Learning Analytics Variables

In this review, a diverse range of learning analytics data was reported, making comparisons between the studies at the least difficult. In addition, there was ambiguity with regards to the interpretation and definitions of the learning analytics data. For example, "log-ins" was not clearly defined if it related to the duration of log-ins or the number of successful log-ins [12]. It is important for future studies to standardize and include clear definitions of such key variables to facilitate comparisons.

The most common learning analytics parameter documented was the number of connections to a specific e-learning material. However, some only recorded connections to folders of resources such as webpages containing multiple e-learning resources. For example, Colthorpe et al [21] deposited all of their teaching videos into a single folder for which they registered the frequency of access, and Franson et al [25] tracked students' access to "TRC database material," which included schematic graphics, explanation texts, and feedback questions, rather than each type of resource individually. Therefore, we do not have a clear picture of what learning materials students are actually engaging with. It would be preferable to record access to individual e-learning resources, as this would provide more precise data regarding the effectiveness and popularity of each document. Furthermore, this would allow differentiation between accessing e-learning materials as opposed to log-ins for viewing other included documents, such as timetables or course announcements, which would not be considered learning materials.

Another way used to analyze students' engagement of e-learning was the time they spent on tasks. Nevertheless, the majority of studies, except Boye et al [18] and Franson et al [25], did not provide sufficient details on how this parameter was recorded. Boye et al [18] reported that each student's individual access to an e-learning animation was recorded at 10-second intervals, and when students had stayed idle for >3 minutes, the tracking mechanism would stop until another action was taken. Franson et al [25] applied filters to time spent on their "TRC database material" such that a too short engagement (3 seconds) was considered not to be meaningful. Likewise, greater than the anticipated duration (6 minutes) suggested nonstudy activities. However, the total time spent on a task may provide limited information, as this may simply be an open webpage not being read or a playing video not being watched. Furthermore, this would neither necessarily reflect the actual physical presence of students nor cognitive engagement.

A total of 8 studies documented the frequency of messages posted in forums. However, the value of such learning analytics data was questionable, as it does not reflect the quality of the discussion students created. As such, unless the quality of forum posts is monitored and assessed, the benefit of simply registering the quantity of posts is in doubt. Therefore, an assessment rubric is required that can evaluate the quality of the discussion posts.

From the above, we can assert that detailing of learning analytics data needs to be improved so that research studies can allow meaningful outcomes and interpretations. Furthermore, diverse analytics need to be recorded for the individual types of e-learning resources, as well as a way of capturing active engagement with the content.

Recommendations

This review observed diverse approaches in recording and defining learning analytics data of students' e-learning use. The varied and imprecise nature of learning analytic data that has been used in the current studies does not help further our understanding of how Web-based learning helps students learn and how we can use the analytics to help students at risk. Hence, a more detailed and precise approach to analytics and e-learning research is required to answer these questions.

From an examination of this literature and an analysis of its shortcomings, we propose the following recommendations:

- Learning analytics data should include well-defined terms and conditions used to describe data collected.
- A detailed collection of individual e-learning items should be performed, as opposed to merely platform log-ins or folder connection frequencies to gain greater knowledge regarding learner engagement.
- The length of time spent by students on e-learning materials, should be recorded when appropriate for certain e-learning resources. For instance, the collection of data on the amount of a video watched is now possible with many LMS and even on YouTube. Conversely, time analysis of access to a PDF document would not be appropriate.
- A mechanism to identify idle time in an e-learning activity may be appropriate to identify when students are not engaged with learning by way of analyzing keystrokes, mouse use, or video playback and pause buttons.
- Course designers and researchers need to plan their learning objectives and how these map to in-class and Web-based learning activities. To truly identify the benefits of e-learning, some materials should be exclusively focused on one particular course learning objective, and students should be informed this will not be covered in-class and that this learning objective will be assessed. This will drive consumption of the learning materials and will also allow more meaningful outcome analysis of e-learning to assessment outcomes, as it is not known if in-class learning may be merely duplicating Web-based learning. This requires careful design and planning as to how e-learning is effective and how it helps the course objectives.

- Course designers and researchers also need to consider the best delivery mode of the Web-based learning material, and as of yet research has not answered this question. Ideally, a learning taxonomy needs to be used to classify the nature of the learning objective (ie, understanding, analysis, critical thinking, etc) and this in turn also needs to be mapped to the e-learning material. With such mapping, researchers will be able to build understanding about which knowledge domain is best suited for which e-learning presentation type.
- Indicators as to how cognitively engaged students are with the Web-based learning materials are desirable, as the use of e-learning does not necessarily ensure cognitive engagement. For example, there is a need to test understanding that can be achieved by questions embedded with videos or as standalone assessments to Web-based learning resource; this will help assess understanding of the material as a proxy for engagement.
- Furthermore, course designers may want to look for ways to motivate students to consume the e-learning materials in a more spaced manner rather than at the last moment. Studies have reported that students' use of e-learning materials were crammed days before exams. Therefore, there may be a need to space smaller assessments over time during the course, which may facilitate consumption and learning.
- Another possible motivator to engage students' learning with electronic resources may be a personal analytics dashboard to let students see their current activity in e-learning usage and how this compares to others; this may motivate lower consumers to consume more to help keep them on track.
- A particular and significant point on the relation of e-learning analytics to learning outcomes may actually be that all we are doing is measuring the motivation of students and not the benefit of e-learning. Well-motivated students will succeed in virtually any learning environment and, therefore, we may well not be measuring the impact of the e-learning experience. Our efforts should, therefore, be targeted at identifying students who are underperforming in consumption of the e-learning materials and identifying how to motivate or support them. This is the next challenge for researchers.

Table 4 provides a concise table suggesting several points one might consider for example in studying video e-learning through learning analytics.

Conclusions

This systematic review of e-learning health care education supports the general literature that greater consumption of e-learning as recorded by learning analytics generally supports learning outcomes. However, the detail and nature of the studies were heterogeneous both in learning analytics data and in e-learning content. More detailed and more focused research is required to help understand how e-learning, learning analytics, and learning outcomes can be more effective and in how they help students learning. Recommendations have been proposed for future course designers and researchers to create content and provide evidence for meaningful e-learning and support of all learners as this pedagogical approach grows further.

Table 4. An example of learning analytics of a video e-learning study.

Suggestion	Example
Well-defined variables and conditions	• Reporting *individual* video access: (a) Number of times video A is watched; (b) Number of times video B is watched; (c) Duration of video A being played; (d) Duration of video B being played • Precise *conditions* in recording: (a) <3 minutes between consecutive clicks of videos would not be counted; (b) <5 seconds of duration in playback would not be counted
Exclusive materials	• Exclusive learning materials via e-learning (not overlapping with in-class materials) • Informing students the e-learning materials would be tested
Mapping learning taxonomy	• When creating videos, one may consider classifying into: (a) Video A is for understanding; (b) Video B is for critical thinking
Engaging cognitively	• Multiple-choice questions to be embedded within the videos, such that students need to answer them to continue watching
Facilitate consumption and learning	• Space smaller assessments over time during the course • Personal analytics dashboard showing consumption status and allowing comparisons with the rest of the class

Conflicts of Interest

None declared.

References

1. Siemens G. Learning Analytics & Knowledge. New York, USA: ACM; 2011 Presented at: 1st International Conference on Learning Analytics and Knowledge; February 27-March 1, 2011; Banff, Alberta URL: https://tekri.athabascau.ca/analytics/

2. Picciano AG. The Evolution of Big Data and Learning Analytics in American Higher Education. OLJ 2012 Jun 18;16(3):9-20. [doi: 10.24059/olj.v16i3.267]
3. Johnson L, Adams BS, Cummins M, Estrada V, Freeman A, Hall C. NMC Horizon Report: 2016 Higher Education Edition. Austin, Texas: The New Media Consortium; 2016.
4. Goldie JGS. Connectivism: A knowledge learning theory for the digital age? Med Teach 2016 Oct;38(10):1064-1069. [doi: 10.3109/0142159X.2016.1173661] [Medline: 27128290]
5. Nunn S, Avella JT, Kanai T, Kebritchi M. Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. OLJ 2016 Jan 10;20(2):13-29. [doi: 10.24059/olj.v20i2.790]
6. Topps D, Helmer J, Ellaway R. YouTube as a platform for publishing clinical skills training videos. Acad Med 2013 Feb;88(2):192-197. [doi: 10.1097/ACM.0b013e31827c5352] [Medline: 23269305]
7. Alturki UT, Aldraiweesh A, Kinshuck D. Evaluating The Usability And Accessibility Of LMS "Blackboard" At King Saud University. CIER 2016 Jan 15;9(1):33-44. [doi: 10.19030/cier.v9i1.9548]
8. de Freitas S, Gibson D, Du Plessis C, Halloran P, Williams E, Ambrose M, et al. Foundations of dynamic learning analytics: Using university student data to increase retention. Br J Educ Technol 2014 Oct 15;46(6):1175-1188. [doi: 10.1111/bjet.12212]
9. Caruso J. EDUCAUSE Center for Applied Research. 2006 Sep 12. Measuring student experiences with course management systems URL: https://cse.sc.edu/~buell/References/ComputingHigherEdMisc/ERB0619.pdf [accessed 2018-05-08] [WebCite Cache ID 6zFp6nAuZ]
10. Broadbent J. Academic success is about self-efficacy rather than frequency of use of the learning management system. AJET 2016 Jul 04;32(4):38-49. [doi: 10.14742/ajet.2634]
11. Kalmus V, Realo A, Siibak A. MOTIVES FOR INTERNET USE AND THEIR RELATIONSHIPS WITH PERSONALITY TRAITS AND SOCIO-DEMOGRAPHIC FACTORS. Trames 2011;15(4):385-403. [doi: 10.3176/tr.2011.4.04]
12. Critchley LAH, Kumta SM, Ware J, Wong JW. Web-based formative assessment case studies: role in a final year medicine two-week anaesthesia course. Anaesth Intensive Care 2009 Jul;37(4):637-645 [FREE Full text] [Medline: 19681426]
13. Poot R, de Kleijn RA, van Rijen VM, van Tartwijk J. Students generate items for an online formative assessment: Is it motivating? Med Teach 2017 Mar;39(3):315-320. [doi: 10.1080/0142159X.2017.1270428] [Medline: 28024432]
14. Next Generation Learning Challenges. 2010. Underlying Premises: Learner Analytics URL: https://library.educause.edu/~/media/files/library/2010/10/nglc003-pdf.pdf [accessed 2018-05-08] [WebCite Cache ID 6zFqrC4mU]
15. Johnson L, Smith R, Willis H, Levine A, Haywood K. The 2011 Horizon Report. Austin, Texas: The New Media Consortium; 2011.
16. Cook DA, West CP. Conducting systematic reviews in medical education: a stepwise approach. Med Educ 2012 Oct;46(10):943-952. [doi: 10.1111/j.1365-2923.2012.04328.x] [Medline: 22989128]
17. Jin J, Bridges SM. Educational technologies in problem-based learning in health sciences education: a systematic review. J Med Internet Res 2014 Dec 10;16(12):e251 [FREE Full text] [doi: 10.2196/jmir.3240] [Medline: 25498126]
18. Boye S, Moen T, Vik T. An e-learning course in medical immunology: does it improve learning outcome? Med Teach 2012;34(9):e649-e653. [doi: 10.3109/0142159X.2012.675456] [Medline: 22497322]
19. Catteau C, Faulks D, Mishellany-Dutour A, Collado V, Tubert-Jeannin S, Tardieu C, et al. Using e-learning to train dentists in the development of standardised oral health promotion interventions for persons with disability. Eur J Dent Educ 2013 Aug;17(3):143-153. [doi: 10.1111/eje.12024] [Medline: 23815691]
20. Chastonay P, Zesiger V, Moretti R, Cremaschini M, Bailey R, Wheeler E, et al. A public health e-learning master's programme with a focus on health workforce development targeting francophone Africa: the University of Geneva experience. Hum Resour Health 2015 Aug 13;13:68 [FREE Full text] [doi: 10.1186/s12960-015-0065-8] [Medline: 26268723]
21. Colthorpe K, Zimbardi K, Ainscough L, Anderson S. Know thy student! Combining learning analytics and critical reflections to develop a targeted intervention for promoting self-regulated. JLA 2015;2(1):134-155. [doi: 10.18608/jla.2015.21.7]
22. Costa-Santos C, Coutinho A, Cruz-Correia R, Ferreira A, Costa-Pereira A. E-learning at Porto Faculty of Medicine. A case study for the subject 'Introduction to Medicine'. Stud Health Technol Inform 2007;129(Pt 2):1366-1371. [Medline: 17911937]
23. Davidson SJ, Candy L. Teaching EBP Using Game-Based Learning: Improving the Student Experience. Worldviews Evid Based Nurs 2016 Aug;13(4):285-293. [doi: 10.1111/wvn.12152] [Medline: 27028987]
24. DiLullo C, Morris HJ, Kriebel RM. Clinical competencies and the basic sciences: an online case tutorial paradigm for delivery of integrated clinical and basic science content. Anat Sci Educ 2009 Oct;2(5):238-243. [doi: 10.1002/ase.97] [Medline: 19670214]
25. Franson KL, Dubois EA, de Kam ML, Cohen AF. Measuring learning from the TRC pharmacology E-Learning program. Br J Clin Pharmacol 2008 Jul;66(1):135-141 [FREE Full text] [doi: 10.1111/j.1365-2125.2008.03167.x] [Medline: 18507664]
26. Garrett BM, MacPhee M, Jackson C. Evaluation of an eportfolio for the assessment of clinical competence in a baccalaureate nursing program. Nurse Educ Today 2013 Oct;33(10):1207-1213. [doi: 10.1016/j.nedt.2012.06.015] [Medline: 22789875]
27. Gurpinar E, Zayim N, Ozenci C, Kemal M. First report about an e-learning application supporting PBL: Students' usages, satisfactions, and achievements. The Turkish Online Journal of Educational Technology 2009;8(2):55-62 [FREE Full text]
28. Kleinsorgen C, Kankofer M, Gradzki Z, Mandoki M, Bartha T, von Köckritz-Blickwede M, et al. Utilization and acceptance of virtual patients in veterinary basic sciences - the vetVIP-project. GMS J Med Educ 2017;34(2):Doc19 [FREE Full text] [doi: 10.3205/zma001096] [Medline: 28584867]

29. Kukolja-Taradi S, Dogas Z, Dabić M, Drenjancevic Perić I. Scaling-up undergraduate medical education: enabling virtual mobility by online elective courses. Croat Med J 2008 Jun;49(3):344-351 [FREE Full text] [Medline: 18581612]
30. Lameris AL, Hoenderop JGJ, Bindels RJM, Eijsvogels TMH. The impact of formative testing on study behaviour and study performance of (bio)medical students: a smartphone application intervention study. BMC Med Educ 2015 Apr 10;15:72 [FREE Full text] [doi: 10.1186/s12909-015-0351-0] [Medline: 25889923]
31. Mahnken AH, Baumann M, Meister M, Schmitt V, Fischer MR. Blended learning in radiology: is self-determined learning really more effective? Eur J Radiol 2011 Jun;78(3):384-387. [doi: 10.1016/j.ejrad.2010.12.059] [Medline: 21288674]
32. Reimer S, Hörnlein A, Tony H, Kraemer D, Oberück S, Betz C, et al. Assessment of a case-based training system (d3web.Train) in rheumatology. Rheumatol Int 2006 Aug;26(10):942-948. [doi: 10.1007/s00296-006-0111-x] [Medline: 16432685]
33. Romanov K, Nevgi A. Do medical students watch video clips in eLearning and do these facilitate learning? Med Teach 2007 Jun;29(5):484-488. [doi: 10.1080/01421590701542119] [Medline: 17885978]
34. Saqr M, Fors U, Tedre M. How learning analytics can early predict under-achieving students in a blended medical education course. Med Teach 2017 Jul;39(7):757-767. [doi: 10.1080/0142159X.2017.1309376] [Medline: 28421894]
35. Zhang D, Zhou L, Briggs RO, Nunamaker JF. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. Information & Management 2006 Jan;43(1):15-27. [doi: 10.1016/j.im.2005.01.004]
36. Rodgers T. Student engagement in the e-learning process and the impact on their grades. International Journal of Cyber Society and Education 2008;1(2):143-156 [FREE Full text]
37. Donkin R, Askew E. An Evaluation of Formative "In-Class" versus "E-Learning" Activities to Benefit Student Learning Outcomes in Biomedical Sciences. Journal of Biomedical Education 2017 Sep 13;2017:1-7. [doi: 10.1155/2017/9127978]
38. Woodham LA, Ellaway RH, Round J, Vaughan S, Poulton T, Zary N. Medical Student and Tutor Perceptions of Video Versus Text in an Interactive Online Virtual Patient for Problem-Based Learning: A Pilot Study. J Med Internet Res 2015 Jun 18;17(6):e151 [FREE Full text] [doi: 10.2196/jmir.3922] [Medline: 26088435]
39. Arnold KE, Pistilli MD. Course signals at Purdue: using learning analytics to increase student success. In: LAK '12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. New York: ACM; 2012 Presented at: Second International Conference on Learning Analytics and Knowledge; 2012 April 29 - May 02; Vancouver, BC, Canada p. 267-270. [doi: 10.1145/2330601.2330666]
40. Buckingham-Shum S, Ferguson R. Social learning analytics: five approaches. Journal of Educational Technology & Society 2012;15(3):3-26 [FREE Full text] [doi: 10.1145/2330601.2330616]
41. Dietz-Uhler B, Hurn JE. Using learning analytics to predict (and improve) student success: A faculty perspective. Journal of Interactive Online Learning 2013;12(1):17-26.
42. Lonn S, Aguilar SJ, Teasley SD. Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. Computers in Human Behavior 2015 Jun;47:90-97. [doi: 10.1016/j.chb.2014.07.013]
43. Campbell JP, DeBlois PB, Oblinger D. Academic analytics: A new tool for a new era. EDUCAUSE Review 2007;42(4):40-57 [FREE Full text]

Abbreviations

LMS: learning management system
MCQ: multiple-choice question
TRC: Teaching Resource Centre

Edited by G Eysenbach; submitted 07.06.18; peer-reviewed by B Arnoldussen, C Perrin; comments to author 22.08.18; revised version received 28.08.18; accepted 28.08.18; published 13.02.19.

Please cite as:
Chan AKM, Botelho MG, Lam OLT
Use of Learning Analytics Data in Health Care–Related Educational Disciplines: Systematic Review
J Med Internet Res 2019;21(2):e11241
URL: http://www.jmir.org/2019/2/e11241/
doi:10.2196/11241
PMID:30758291

Original Paper

Developing a Web-Based Comic for Newly Diagnosed Women With Breast Cancer: An Action Research Approach

Tzu-I Lee[1], RN, MS; Shuh-Jen Sheu[2], RN, PhD; Hsueh-Chin Chang[1], RN, MSN; Yu-Ting Hung[1,3], RN, MSN; Ling-Ming Tseng[4,5], MD; Shin-Shang Chou[6], RN, PhD; Te-Hsin Liang[7], PhD; Hui-Ju Liu[8], BA; Hui-Ling Lu[9], BA; Mei-Chun Chen[10], BA, RN; Ying-Chun Liu[11], RN, PhD; Chi-Shan Tsai[12], RN, MSN; Jui-Chiung Sun[13,14], RN, PhD

[1]School of Nursing, National Yang-Ming University, Taipei, Taiwan

[2]Institute of Community Health Care, School of Nursing, National Yang-Ming University, Taipei, Taiwan

[3]Public Health Center of Taoyuan District, Department of Public Health, Taoyuan City Government, Taoyuan, Taiwan

[4]Comprehensive Breast Health Center & Division of General Surgery, Department of Surgery, Taipei Veterans General Hospital, Taipei, Taiwan

[5]Department of Surgery, School of Medicine, National Yang-Ming University, Taipei, Taiwan

[6]Department of Nursing, Taipei Veterans General Hospital, Taipei, Taiwan

[7]Department of Statistis and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan

[8]New York University, New York, NY, United States

[9]Health Education, Taiwan Tanabe Seiyaku Co, Ltd, Taipei, Taiwan

[10]Department of Nursing, National Taiwan University Hospital, Taipei, Taiwan

[11]School of Nursing, National Defense Medical Center, Taipei, Taiwan

[12]Wan Fang Hospital, Taipei, Taiwan

[13]Department of Nursing, Chang Gung University of Science and Technology, Taoyuan, Taiwan

[14]Department of Obstetrics and Gynecology, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

Corresponding Author:
Shuh-Jen Sheu, RN, PhD
Institute of Community Health Care
School of Nursing
National Yang-Ming University
Nursing Building Room 404
Number155, Section 2, Linong Street
Taipei, 112
Taiwan
Phone: 886 2 28267271
Fax: 886 2 28229973
Email: sjsheu@ym.edu.tw

Abstract

Background: Personal narratives have been seen as a useful way of communicating about cancer treatment options and providing recovery information. Many printed versions of such material are available, including comics that explore the individual memories of patients who have gone through cancer treatment. These studies have been used to orientate patients, patients' relatives, and physicians. However, only a few Web-based comics have been specifically designed for patients with breast cancer and used as aids to decision making.

Objective: We aimed to describe the developmental process of creating an animated comic as a Web-based surgery decision-making tool; the comic was aimed at illustrating the feelings, thoughts, and meanings when a patient suffers from breast cancer. This was done by recounting the symptoms, diagnostic process, treatments, and treatment effects of such women from the diagnosis stage onward.

Methods: Using cycles of planning, action, evaluation, and reflection, which involved collaborative work, action research was conducted to develop a Web-based animated comic. The stages of action research consisted of (1) semistructured and in-depth interviews to collect experiences of women with breast cancer; (2) construction of an animated comic by editors, graphics designers, dubbers, and information technology engineers; (3) redrawing of pictures of the comic after gathering feedback from a breast surgeon; and (4) evaluation of the Web-based animated comic using 6 patient focus groups.

Results: The comic was produced and showcased on the website "The Network of Making-decision Aids for Breast Cancer Surgery"; the comic was accompanied by soft music and audio explanations. The comic functions as a personal statement that describes experiencing breast cancer. The animated comic consists of 8 chapters, based on the 8 themes deducted from the findings obtained during the analysis of relevant interviews. The 8 chapters include (1) the appearance of a lump; (2) confirmation by medical diagnosis; (3) the uncertainty of waiting (4) fear of life-threatening disease; (5) choosing life over despair; (6) being brave and deciding to undergo treatment; (7) choosing the type of surgery; and (8) being reborn.

Conclusions: Using action research, this study illustrated that the comic that sheds light on issues of feelings, emotions, and thoughts that are present when a woman is diagnosed with breast cancer and provides a communication medium to explain the steps in the process. Meanwhile, it implies that hope will be able to overcome the challenges that will be faced. Within the Web-based decision aid for patients with breast cancer, the animated comic acts as an information resource and is aimed at patients' understanding of impacts of emotions arising when suffering from breast cancer. It is potentially applicable as a therapeutic tool that facilitates self-reflection and self-healing among newly diagnosed patients with breast cancer.

(J Med Internet Res 2019;21(2):e10716) doi:10.2196/10716

KEYWORDS

action research; breast cancer; comic; narrative

Introduction

Over the past few years, the number of breast cancer cases has been consistently increasing worldwide [1]. Over the same period, breast cancer has been ranked as the primary cause of cancer-related death among women in Taiwan [2]. Most patients with early-stage breast cancer are presented with lumpectomy or mastectomy as the recommended treatment [3]. Generally, after being diagnosed with breast cancer, patients must make a decision regarding the type of surgery they would prefer to undergo. However, even without having decided on the type of surgery, most hospitals in Taiwan offer their patients a surgery date, which normally ranges from 1-2 weeks in the future. With such a constrained time window owing to the treatment process, patients are usually asked to make a quick decision regarding the type of surgery they prefer. In Taiwan, most breast surgeons and oncologists still focus mainly on the biomedical aspects of treatment, paying much less attention (perhaps, unwillingly) to the emotional impact of this condition on patients and their relatives [4].

Psychologists have shown that emotions do, indeed, play an important role in the patient decision-making process regarding the type of treatment they want to undergo when treating their cancer [5]. From the moment they are diagnosed with the condition, patients begin to experience different levels of emotional disturbance [5] such as shock, anxiety, uncertainty, ambiguity, sadness, anger, guilt, fear of death, fear of suffering, and depression [5-8]. For most women, breast cancer becomes an overwhelming emotional experience [7]. The breast "cancer diagnosis" can provoke reactions of the negative emotions mentioned above and inflate a range of emotional experiences [5-8]. This wide range of emotions and psychological episodes may end up submerging a patient and her relatives in a state of intense uncertainty [8]. Briefly, a woman who gets diagnosed with breast cancer might be daunted by the uncertainties associated with the prognosis and treatment [8,9] and by considerations of herself, her relationships, and her life [10]. Research has shown that breast cancer is often very stressful for women from the very moment of diagnosis and all the way through surgery, including the period when they are waiting for diagnostic results [7,11,12]. In addition, the way these women are diagnosed can directly cause an increase in their stress level, which strengthens negative emotions [13]. Moreover, time constraints in the clinical setting in addition to certain aspects of the Chinese social culture may contribute to aggravating the situation. For instance, the fact those patients are not too willing to openly discuss cancer-related feelings [14]. Therefore, this makes them feel alone and not understood [15]. In this context, the specific needs of given individual's psychological issues might not be properly identified and addressed [15,16]. Studies have found that being diagnosed with cancer may affect a patient's cognitive capability by temporarily reducing comprehension and the ability to process information [17,18]. The studies cited above provide evidence of the impact that negative emotions can have during the diagnosis period. This indicates that assisting women during this stage should be able to contribute to more informed and conscious decision making about the type of surgery and type of treatment better suited for them.

There are several alternatives ways to provide emotional support to patients with breast cancer in a clinical setting; these include support groups, individual counseling, and others. These kinds of arrangements often seek to manifest narratives naturally. Personal narratives are often used in the realms of health and medicine to make up for lack of experience, to help the patients' interpretation of their health status, and to embody such status [19]. Through narratives, a person undergoing a physical illness may be able to make sense of or revise his or her experiences during the illness [19]. This is because, when narratives are used as a medium, it helps the environment surrounding a person to be understood, which, in turn, informs a person about his or her concept of himself or herself [20]. This allows patients to make better sense of events and attempt to act in accordance with the meanings they have gathered [21]. Once it is understood that narratives construct a story, research has been able to show that cancer survivors' stories not only have a positive influence on both the listeners' and storytellers' health [22-24] but it is also clear that they have a positive effect by reducing a reader's or a listener's feelings of fear and isolation during his or her illness [25]. This shows that patient stories seem to be more authentic,

credible, and appealing compared with the mere documentation of facts [26].

Comics are a type of narrative that integrates sequential visual imagery and text [27]. Comics can convey affective and aesthetic sensibilities that have an impact on the affective processes and responses of readers [28,29]. They can also make use of a set of visual narrative interpretation structures that allow them to tell complex and emotionally rich stories, while conveying, at the same time, information without an individual having to read any complex text [30]. Comics are uniquely suited to represent vivid and valid embodied illness experiences [30]. Several prominent cancer-related comic narratives from the patients' and family perspectives have been published such as Marisa Acocella's "Marchetto's autobiographical Cancer Vixen" (breast cancer) [31], Brain Fies's "Mom's Cancer" (lung cancer) [32], and Stan Mack's "Janet and Me" (breast cancer) [33]. These comics represent individual experiences that are relevant to a specific illness or type of cancer and include the medical and diagnostic approaches to the disease in question [34]. In addition, they include the dialogical use of images by purposely adjusting hue, tone, saturation, and colors conjointly in parallel with the written language. These processes are used to deploy a given narrative and, as a result, the comics can catch and juxtapose multiple different layers of cancer-embodied experiences, especially those that are unable to be adequately and fully captured by narrative alone [31-33,35]; this can be done in the comics from various perspectives, namely the patient, the patient's family members, and the patient's caregivers [31-33,35].

Emotions can influence humans' decision making [36,37]. A few stories of the breast cancer journey have focused on women with breast cancer before undergoing surgery and decision making. Where they exist, these stories do not necessarily include comprehensive descriptions or analysis of the psychological impacts of a breast cancer diagnosis. Furthermore, and obviously, illustrated comics with this discourse are even rarer. We believe that alleviating the potentially pessimist mindset of women with breast cancer is an important task during the decision-making process. Hence, the purpose of our creation of a comic story is to serve as a psychosocial support tool, which is aimed at helping newly diagnosed patients with breast cancer so that the possibility of emotional distress is reduced during the decision-making period for treatment.

Methods

Design

An action research (AR) design was utilized to generate the comic through engagement with women who have had real experiences with breast cancer. AR implies "research with people" instead of "research on people" [38]; it works in a mutually cooperative way linking researchers and participants to develop a project [38,39]. AR has various forms, different emphases, and several key principles, all of which have been presented in the literature [40,41]. AR is a diversified, reflexive, and cyclical research methodology that typically uses several different methods to collect and analyze data that are regardless of the data's various forms [38,40,41]. AR is characterized by a cyclical and continuous interleaving of planning, action, evaluation, and reflection [39]. AR does not seem to have been applied to the design of Web-based decision-making aids for patients with breast cancer before. With thorough analysis and team discussions, we concluded that applying AR, in combination with several different data collection methods, was the most suitable and flexible way of conducting this study. It allows more brainstorming and creative (innovative) thinking and performance emergence. This mixed-method gives us deeper understanding, which, in turn, allows us to build a fuller picture; this then could be translated into the creation of our proposed Web-based animated comic.

The cyclical process of our AR included the sequential development of our outline, drafting of a script, and related graphics (this needs to be easily understandable by any audience with a middle school level of education), recording of the script, and developing of animated graphics to produce an initial prototype of the Web-based comic (see Figure 1). Progress and the accuracy of the project were assessed using biweekly or weekly team meetings, during which adjustments were made where needed after reaching a consensus among all the team members. Our research team included a Principal Investigator (SJS), nursing students of different levels, information technology engineers, and computer graphics designers. This research was approved by the hospital's Institutional Review Board, and informed consent was obtained from all research participants.

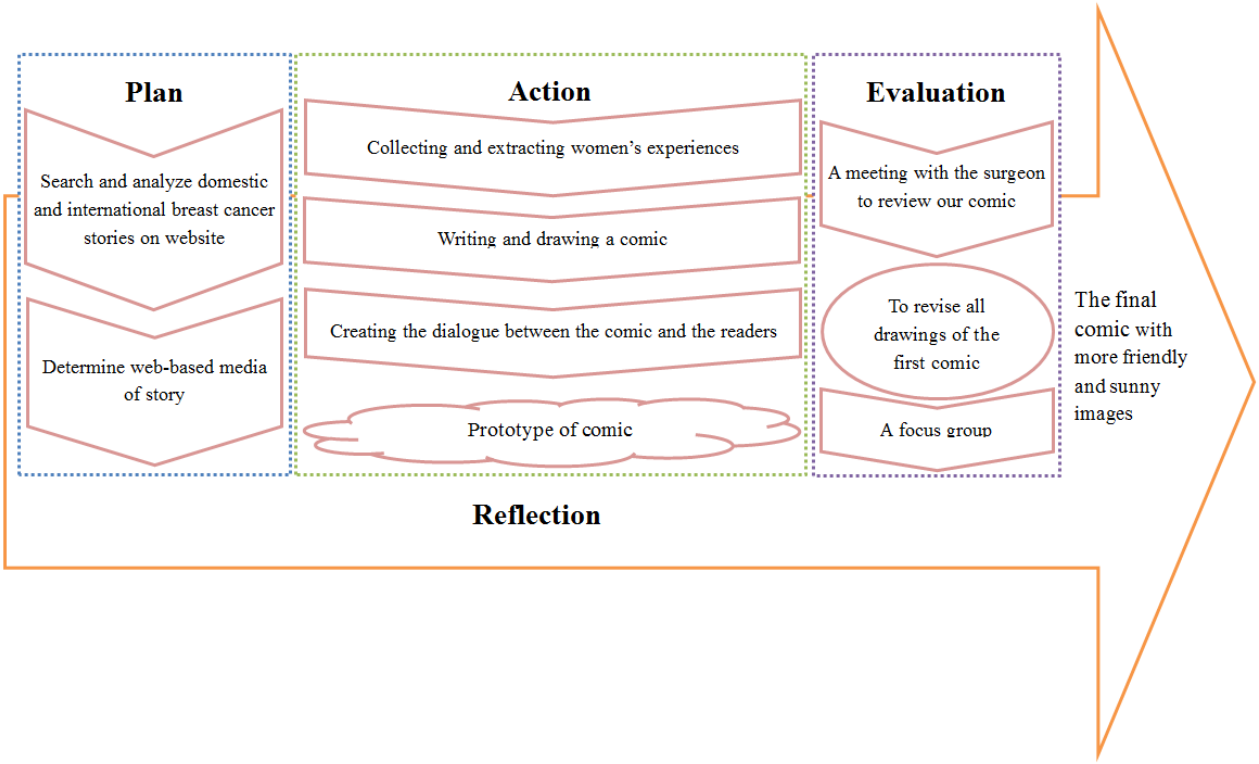
Planning Phase

During the Planning Stage, the team defined and planed a Web-based multimedia module. This was designed by searching and analyzing Web-based personal stories of patients with breast cancer. We moved forward with the designing of a Web narrative-comic structure by applying visual aesthetics and the healing concept. Working on the premise that an animated comic as a breast cancer decision-making aid needed most importantly to provide accurate, empathetic, and easy-to-understand emotional information for patients with newly diagnosed breast cancer, we emphasized both the visual and psychological effects of the comic.

Action Phase

During this stage, we proceeded with the design of the animated comic, which was constructed on the basis of the results obtained from the detailed analysis of the interviews with patients with breast cancer. Particular attention was paid to the media aspects and content aspects of the comic, with the leading investigator and the rest of the team focusing on these from the outset of the project. To allow the interviews with patients guide the development of the final comic, this phase followed the following 3 steps.

Figure 1. Research Process.



Step 1: Collecting and Extracting Women's Experiences

We used purposive sampling to recruit 11 patients with breast cancer from a hospital in Taipei, Taiwan, between September 2010 and July 2011 (average age, 49.5 years; range 41-60 years). Among these patients, 5 had undergone mastectomy, 5 had undergone lumpectomy, and 1 had undergone a mastectomy followed by immediate breast reconstruction. We conducted semistructured individual interviews, using prompt, reflective, and open-ended questions in a conversational style to facilitate the participants' conversions about a range of topics associated with their experiences with breast cancer. All interviews were transcribed verbatim. The data collection was based on the principles of saturation. We used the Objective and Systematic Approach of Miles and Huberman for the content analysis of the transcripts [42]. The text of each interview was cataloged and analyzed. Furthermore, the frequency, order, and intensity of the occurrence of words, phrases, or sentences related to themes that represented specific concepts were enumerated for each participant [42].

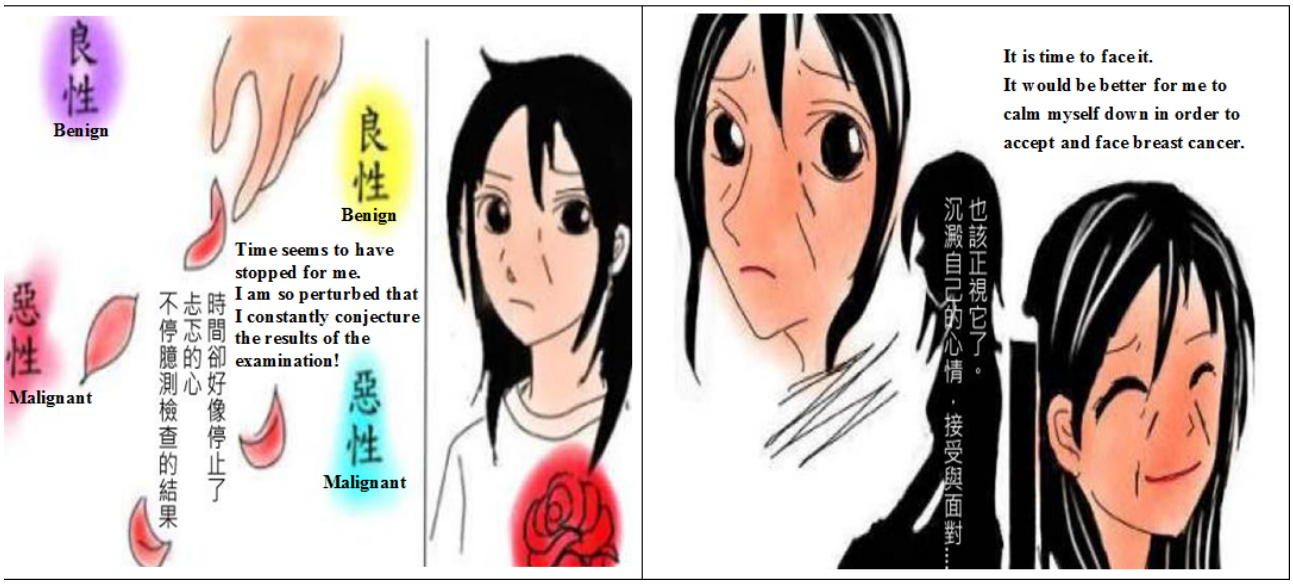
Findings of our analysis of the interviews indicated that women with breast cancer were dealing with conflicting emotions and thoughts that resulted from being diagnosed with breast cancer. A simple thematic analysis of the interviews' data revealed 8 themes. These thematic perspectives were closely associated and appeared to affect the ways in which these women described their experiences of being diagnosed with breast cancer. In addition, these themes were related to the physical, emotional, and social perspectives of patients, especially since the diagnostic period through presurgery. The 8 emerging themes included (1) the appearance of a lump; (2) confirmation by medical diagnosis; (3) the uncertainty of waiting; (4) fear of a life-threatening disease; (5) choosing life over despair; (6) being brave and deciding to undergo treatment; (7) choosing the type of surgery; and (8) being reborn.

Step 2: Writing and Drawing the Comic

Writing the Comic

The collection of individual interviews provided us with actual dialogue and character descriptions that were then incorporated into one storyline of the comic to ensure that it would be relevant to our targeted audience and relevant situations. Upon comparing it with the narratives of all participants, 1 participant's story resonated especially with us, and it became the main plotline for our comic. This woman had undertaken many roles, tasks, and responsibilities during her life, including maintaining her marriage, raising a child, and so on; her story and life experiences conveyed a wide array of feelings worth incorporating into our comic. Based on the purpose of our research and our targeted audience, we thought one storyline was enough in this study because aspects of multiculture and different races are not significantly present in the Taiwanese population. We employed the 8 themes resulting from the interview data as individual chapters of the comic's plot. The narrative of each chapter is the content extracted from these interviews after they were correctly modified to reflect typical scenarios that result from the experience of being diagnosed with breast cancer.

Figure 2. Sample drawings from the initial comic.



Drawing the Comic

After a thorough discussion and consultations among team members, we agreed on the inclusion of 32 slides in the form of webpages, 4 slides per theme to make up our proposed animated comic strip. While the team discussed the context of these interviews and the comic plot, 1 comic illustrator was asked to sit in during meetings to listen to the discussion or even take part where appropriate. The illustrator could draw during meetings or produce the images after reflection at a later time. The pictures of the comic are simple hand drawings influenced by manga and stresses on the facial expressions of the character within which the emotions are expressed and perceived. From descriptions of the team members, the illustrator captured some facial expressions and actions of the character in this comic from patients whom the team members had met and cared for previously or further researched patient blogs and self-help sites, as well as academic papers. The team then held a series of iterative discussions to determine the final list of comic images, which best exemplified the comic scenario and were sketched and redrawn several times before they were agreed on. Representative quotes are shown in the comics to point out participants' cancer-related emotions and thoughts and, thus, provide a more holistic context.

To enhance a better user's experiences and more comprehensive involvement in short viewing time, audio explanations and soft music tracks were embedded within the comic. Overall, all pictures have been drawn so that they purposefully follow a simple color-matching structure, which, combined with the oral and textual narrative, embodies the healing and calming purpose of the comic (see Figure 2).

Step 3: Creating a Dialogue Between the Comic and its Readers

We added 6 questions from the Chinese version of the *Mishel Uncertainty in Illness Scale* [43] and 8 questions from the *Mini-Mental Adjustment to Cancer Scale—Chinese Version* [44] to the slide at the end of each theme to provide the readers with an opportunity to confront and understand their cancer-related feelings and thoughts. Following these questions, an additional blank column was provided on the same slide to allow readers writing their own feelings and thoughts.

Evaluation Phase

The evaluation of developing comic was conducted by interviews with a breast cancer surgeon, focus groups meetings with former patients with breast cancer, and an internal assessment by regular team meetings.

The Surgeon's Comment

Once the animated comic was developed, we convened a meeting with a renowned breast surgeon in Taipei (LMT) to have him assess our work and provide some feedback. The surgeon gave our work an overall positive review; however, he expressed some concerns regarding the drawings, specifically regarding the appearance of the character in the comic. He suggested that she appeared to be too sad and transmit sorrow to the comic's audience, especially to patients with newly diagnosed breast cancer. He pointed out that, based on his experience dealing with these kinds of patients, they might have a far more pessimist mindset than health providers can see and expect in such a situation. Moreover, he indicated that such women might not be able to afford any extra negative emotional suffering; therefore, he deemed it unnecessary to expose this group to a sad and depressive visual stimulation through the comic-related images.

Response to the Surgeon's Comments

Considering the surgeon's comments, we invited and tasked another different graphics designer to review and redraw of our first comic. As this graphics designer reflected on the explanations of the comic storyline from the initial design, the pictures of the initial comic, and his drawing experience, he was simultaneously acting as a viewer and (re)positioning himself as an illustrator. Furthermore, he was deeply considering viewers in the shoes of newly diagnosed breast cancer women. Upon agreement of the team, he used the same comic storyline but expertly laid down a different drawing style from the first comic through using images, color, symbolism, and spatial

arrangement of objects. Still, facial expressions and actions of the character in this new-version comic kept the rigorous drawing process of the first version comic. Therefore, the second-version comic was made more visually friendly, thus conveying a more optimistic feeling.

Feedback From the Focus Groups

Upon completion of the second-version comic, we formed a focus group with the aim of helping us to evaluate the experience and impact of our work on real users once it had been launched to the general public [42]. We believed that a focus group was the best way to conduct such an evaluation, especially because it offered participants an opportunity to openly and freely express their personal thoughts and emotions [42]. This, in turn, would allow us to generate and collect a rich dataset [42]. The goals of the focus groups were to obtain a broad spectrum of feedback regarding the framework of the comic strips, as well as to gather ideas for promoting the comic, and receive suggestions on ways to improving it. We conducted 6 focus group meetings, which comprised 7 breast cancer survivors (average age, 60.4 years; range 51-72 years), from January 2015 to March 2015. Overall, 6 of these participants had been diagnosed with stage I cancer, while 1 had been diagnosed with stage II. They all had undergone surgery between 1993 and 2014. Regarding this new comic, we did not receive any suggestions of a modified form from these focus group members but got some positive appraisal of this comic from them.

The focus group members were enthusiastic after viewing the 2 versions of the comic during focus group meetings. Their opinions of both versions of the comic were generally positive. Although they preferred the images of the initial comic, they were satisfied with the second version as well and understood why we had adopted the second-version comic based on the recommendation of the breast surgeon. In addition, they positively valued the plot and emotion descriptions within the comic, which they found easy to follow and understand. Furthermore, they indicated that they felt very touched and empathetic after watching the animated comic because the images from the comic were able to truly represent their very own feelings and experiences when suffering from breast cancer.

> *...This was a road I have walked through... The emotional information was complete... It was like a very detailed description of my journey.* [C2, 164]
>
> *I thought this comic was very vivid and realistic to me describing how I felt. And, it was my spiritual journey...* [E2, 186]
>
> *It felt real, it was very empathetic. As if I had just breast cancer at that time. Really, I did not know exactly what to do, and then it seemed like me, when I was afraid that others knew about my disease and I almost did not dare to ask what was going on with my breast. So, I could only discuss about my breast with my husband. In really, it would have been better to watch this kind of comic story on Internet. If I could have seen it, then I naturally would have felt calmer, right, and I also felt I was ok.* [G6, 6]

Moreover, they agreed that they would not have felt as alone during the course of their breast cancer after viewing this comic; this was because this comic would have made them feel less anxious, less depressed, and more cared about by others.

> *...she (the patient) had to face various ups and downs in her mind before making a decision about surgery. The effect of reading the comic would have made her feel that this is a general phenomenon and it is normal to experience emotional changes after receiving a breast cancer diagnosis. It can be said that the comic seems to provide empathy and mental support for her, and she could counsel herself with this because she was not alone.* [A2, 193]
>
> *It is actually quite healing due to the caring effect of this comic story.* [A2, 198]

Furthermore, the breast cancer survivors thought that character modeling was more real because it addressed their embodiment of the breast cancer journey. Finally, they suggested that the initial version of the comic ought to become a comic book and acts as another information resource for newly diagnosed patients with breast cancer in hospitals.

Results

Based on the results obtained from analyzing patients' interviews and, with the focus on "patient's mindset while receiving breast cancer diagnosis," the final product of our work was drawn by hand, digitally modified, and colored in Photoshop CC to be hosted on the website (Multimedia Appendix 1), where it was run for a relatively short time (8 minutes). The comic is accompanied by a female narration and soft background music. It consists of uncertainty, fear, and anxiety that are of intense emotional significance and which are conveyed through the flash-animated images. Named "Women's Voice," the story attempts to convey echoes of the personal words of patients with breast cancer. The comic combines narrative with strong visuals to present and communicate emotional information through the protagonist named Shu-Jun (a common Chinese women's name). She autobiographically articulates her own life experiences with breast cancer, while simultaneously acknowledging the sufferings she had to endure along the way. Each theme's full context is summarized in a screenshot (Figure 3).

Figure 3. Examples of screenshots from the comic.

The appearance of a lump:

Shu-Jun wonders who she should consult about the abnormal mass that has appeared on her breast.

Confirmation by medical diagnosis:

Shu-Jun automatically thinks that she has breast cancer. She considers whether she should see a doctor and get a second medical opinion.

The uncertainty of waiting:

Shu-Jun is feeling depressed and worried while she waits for the results of her biopsy. She is wondering if the abnormal mass is malignant or not.

Fear of life-threatening disease:

Shu-Jun has been diagnosed with breast cancer. She jumps to the conclusion that she is about to die. She is upset because there are many things she still wants to do in her life!

Choosing life over despair:

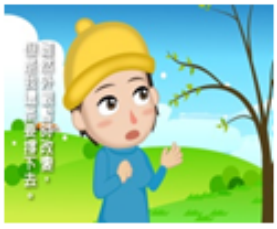
Shu-Jun struggles with her diagnosis; she wants to live, even though she often thinks of dying.

Being brave and deciding to undergo treatment:

Shu-Jun has told herself that she must adopt a positive attitude in order to make the most of her life, for both her and her family.

Choosing the type of surgery:

Shu-Jun wonders how others will react to and how she will feel about her physical appearance after the surgery to remove the breast cancer.



Being reborn:

Cancer treatment was difficult for Shu-Jun because her physical appearance changed and she felt discomfort. Nevertheless, she finally completed the treatment and overcame the disease. Her life now seems to be like new leaves growing on a tree.

Discussion

Principal Findings

"Cancer" can elicit negative images and stigmatizing responses from individuals associated with patients [45]. Furthermore, cancer-related emotional distress can affect both their thinking and information-processing capability of dealing with breast cancer [46] and their coping skills [47]. Even though there are many forms of mental support for women with breast cancer, mental support and emotional support seem to be difficult to define attending to the particular characteristics and culture of an individual. Comic serves as a medium for making trauma visible, and it does this by not only illustrating the fragmentation of such experienced memory [48] but also adds articulation to, otherwise incomplete, narratives, which often may be left unrepresented because they are hard to express or are too personal [48,49]. Compared with the form of plain text and printed material, which is helpful to certain extent, recent advances in digital content and information technologies inspired and allowed us to bravely create the available information in the form of Web-based animated comic story, which would be expected to be more accessible to wider groups of patients, caregivers, and relatives; this medium would also be appealing to the recipients [50]. Hence, the focus group members in this study highly recommended and appreciated viewing our Web-based animated comic to be as a consoling tool in future while they volunteer newly diagnosed patients with breast cancer in a hospital.

It is essential that care providers for patients with breast cancer let them feel that they are "not alone" and that they are "being supported" [51]. Compassion can make people become less self-absorbed and helps them to reach a better perspective regarding a problem [52]. Although one focus group member complained not to get used to the comic instead of a friend, the team hopes our comic, applying the drawing style, detailed expressions, and color aesthetics, would have relatable implications and allow reader-text intimacy; furthermore, it would give viewers suffering from breast cancer a sense of "we were in the same boat, we have also gone through the experience of having breast cancer, and we are not alone in this journey," to induce feelings of compassion when a person faces breast cancer. From viewing the appraisal of focus group members, these women generally felt that the storyline was very realistic and then recalled that they experienced very touching feelings and deep empathy immediately after viewing our animated comic because our comic represents a character in a plot that is quite similar to their journey after receiving their cancer diagnosis.

Overall, focus group members appreciated the pictures and colors used in the comic and agreed that the use of bright colors added a description. We hope that our animated comic will become a healing story that exerts a positive influence on the wounds inflicted by breast cancer and that patients would be able to connect to other individuals through their experiences and thoughts after they have viewed this comic. However, this study is at the initial stage of developing the comic, therefore we cannot be sure whether this approach of reading the comic, the content of the comic, or both of them might be more meaningful for the field of mental support for Taiwanese women with newly diagnosed breast cancer. Furthermore, one single type or style of the comic content of this study is enough for the target of this study—the Taiwanese women population. Given that there are no multicultural and multiracial aspects to consider in Taiwan in the implementation of more than one type or style as would be the case in other multicultural or multiracial societies, we consider unnecessary to embark ourselves with the creation of more complex versions of this comic. Future evaluation studies of this comic approach will be used to assess the impact of comics as part of the wider health information and mental experience of women with breast cancer, as well as to evaluate this individual comic, including the narrative element and its visual component. Ideally, the results would benefit by repeating the project using empirical studies.

Interestingly, a preference difference existed between the breast cancer surgeon and focus group members with breast cancer regarding 2 different drawing styles of the same comic storyline. For the initial comic images, the surgeon suggested that the sad and depressed drawing style might contribute to making patients with breast cancer more miserable, more anxious, and more hopeless after reading this comic. However, focus group members thought these images were closer to their deeper insights of suffering breast cancer. They suggested that because facial expressions in this comic looked more like photographs and were more realistic than those in the second version and that they better represented the emotional conflicts and thoughts that they could not find the right words to express. Some researchers have pointed out that age, gender, cultural beliefs, and values have various influences on responses to the diagnosis and treatment of breast cancer [45,50] and, thus, might affect what is perceived when watching a comic of this nature [29]. This may explain the different senses of the 2 versions of the comic by the surgeon and patients. This kind issue of having different perceptions about organizing the Web-based comic construction also made the team to think more and overcome some challenges, such as aspects of different professional terminology, communication, information engineering, and art design, as well as how to comfort patients dealing with the effects of being diagnosed with breast cancer. Compared with conventional research designs, the AR cycles provided an opportunity to multiprofessionals working collaboratively to modify and confirm through each process repeatedly. This study may be regarded as a reference for describing how to develop a psychological and decision support tool for women newly diagnosed with breast cancer by considering their own corresponding culture and the region where they live.

Conclusion

In this study, AR was used as the most appropriate methodological approach to understand the emotion-related suffering of newly diagnosed breast cancer women, leading to the development of our Web-based animated comic. We attempted to generalize the content of this animated comic about breast cancer stories thematically and simply to facilitate the readers' comprehension of the emotional information. We believe this animated comic has the potential to serve as a positive influence on those women who need further orientation

about the decision process regarding their choice of cancer-related treatment; this is because it seems better suited to their situation and provides these women with a channel that should help alleviate possible negative cancer-related emotions. This research is valuable because it presents and shares a different process of creation for a relevant media tool and because it uses a new creative way to bring a sense of hope to patients newly diagnosed with breast cancer.

Acknowledgments

We appreciate the funding from the Ministry of Science and Technology of Republic of China (2000-2012: The exploration and promotion of suspected breast cancer women's medical autonomy and decision making for receiving a surgery, NSC 99-2629-H-010-001MY2; 2013-2015: Construction and exploration of breast surgery decisions website interactive content—comic animation and decisions games design, NSC 102-2629-H-010-001MY2) and all team members in this research. Special thanks to Associate Professor Chiu-Yueh Yang, Professor Jui-Ch'i Liu, Ching-Fang Wu, Pei-Ni Wu, Guan-Rong Su, and Yu-Sheng Lin. We also thank the NYMU-NTUST Joint Research Program (107TY02, The Exploration of Constructing Breast Surgery Decision Aide Website on Public Network Platform and Using Experience) for providing funds and Professor Kung-Jeng Wang for his support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The screenshot of the animation on the website.

[PNG File, 317KB - jmir_v21i2e10716_app1.png]

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015 Mar 01;136(5):E359-E386 [FREE Full text] [doi: 10.1002/ijc.29210] [Medline: 25220842]
2. Health Promotion Administration Ministry of Health and Welfare Taiwan. Cancer Registry Annual Report, 2015 Taiwan URL: https://www.hpa.gov.tw/Pages/ashx/File.ashx?FilePath=~/File/Attach/8084/File_7635.pdf [WebCite Cache ID 6yQUkqnRr]
3. Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, et al. NCCN Guidelines Insights: Breast Cancer, Version 1.2017. J Natl Compr Canc Netw 2017 Dec;15(4):433-451. [Medline: 28404755]
4. Lam WWT, Bonanno GA, Mancini AD, Ho S, Chan M, Hung WK, et al. Trajectories of psychological distress among Chinese women diagnosed with breast cancer. Psychooncology 2010 Oct;19(10):1044-1051. [doi: 10.1002/pon.1658] [Medline: 20014074]
5. Reyna VF, Nelson WL, Han PK, Pignone MP. Decision making and cancer. Am Psychol 2015;70(2):105-118 [FREE Full text] [doi: 10.1037/a0036834] [Medline: 25730718]
6. Lam WW, Fielding R, Chan M, Chow L, Or A. Gambling with your life: the process of breast cancer treatment decision making in Chinese women. Psychooncology 2005 Jan;14(1):1-15. [doi: 10.1002/pon.803] [Medline: 15386793]
7. Shapiro SL, Lopez AM, Schwartz GE, Bootzin R, Figueredo AJ, Braden CJ, et al. Quality of life and breast cancer: relationship to psychosocial variables. J Clin Psychol 2001 Apr;57(4):501-519. [Medline: 11255204]
8. Liu L, Li C, Tang ST, Huang C, Chiou A. Role of continuing supportive cares in increasing social support and reducing perceived uncertainty among women with newly diagnosed breast cancer in Taiwan. Cancer Nurs 2006;29(4):273-282. [Medline: 16871094]
9. Yusuf A, Ab Hadi IS, Mahamood Z, Ahmad Z, Keng S. Understanding the breast cancer experience: a qualitative study of Malaysian women. Asian Pac J Cancer Prev 2013;14(6):3689-3698 [FREE Full text] [Medline: 23886167]
10. Halkett GKB, Arbon P, Scutter SD, Borg M. The phenomenon of making decisions during the experience of early breast cancer. Eur J Cancer Care (Engl) 2007 Jul;16(4):322-330. [doi: 10.1111/j.1365-2354.2007.00778.x] [Medline: 17587355]
11. Liao M, Chen M, Chen S, Chen P. Uncertainty and anxiety during the diagnostic period for women with suspected breast cancer. Cancer Nurs 2008;31(4):274-283. [doi: 10.1097/01.NCC.0000305744.64452.fe] [Medline: 18600114]
12. Drageset S, Lindstrøm TC, Giske T, Underlid K. Being in suspense: women's experiences awaiting breast cancer surgery. J Adv Nurs 2011 Sep;67(9):1941-1951. [doi: 10.1111/j.1365-2648.2011.05638.x] [Medline: 21466581]
13. Gibbons A, Groarke A, Curtis R, Groarke J. The effect of mode of detection of breast cancer on stress and distress. Psychooncology 2017 Dec;26(6):787-792. [doi: 10.1002/pon.4227] [Medline: 27449013]
14. Lee S, Chen L, Ma GX, Fang CY, Oh Y, Scully L. Challenges and Needs of Chinese and Korean American Breast Cancer Survivors: In-Depth Interviews. N Am J Med Sci (Boston) 2013 Feb 02;6(1):1-8 [FREE Full text] [Medline: 24019995]

15. Wang K, Liu C, Leu J. Anxiety and Concern in Breast Cancer Patients. Taipei City Medical Journal 2005;2(9):844-850. [doi: 10.6200/tcmj.2005.2.9.10]
16. Howie JG, Hopton JL, Heaney DJ, Porter AM. Attitudes to medical care, the organization of work, and stress among general practitioners. Br J Gen Pract 1992 May;42(358):181-185 [FREE Full text] [Medline: 1389427]
17. Kuang K, Wilson SR. A Meta-Analysis of Uncertainty and Information Management in Illness Contexts. J Commun 2017 May 11;67(3):378-401. [doi: 10.1111/jcom.12299]
18. Mishel M. Reconceptualization of the uncertainty in illness theory. Image J Nurs Sch 1990;22(4):256-262. [Medline: 2292449]
19. Harter L, Japp P, Beck C. Vital problematics of narrative theorizing about healthhealing. In: Harter LM, Japp PM, Beck CS, editors. Narratives, health, and healing: Communication theory, research, and practice. New Jersey: Lawrence Erlbaum Association; 2005:7-29.
20. Johnstone B. 2010. Language and place URL: https://bit.ly/2FXJD4r
21. Cohn N. The limits of time and transitions: challenges to theories of sequential image comprehension. studies in comics 2010 Apr 01;1(1):127-147. [doi: 10.1386/stic.1.1.127/1]
22. Cepeda MS, Chapman CR, Miranda N, Sanchez R, Rodriguez CH, Restrepo AE, et al. Emotional disclosure through patient narrative may improve pain and well-being: results of a randomized controlled trial in patients with cancer pain. J Pain Symptom Manage 2008 Jun;35(6):623-631. [doi: 10.1016/j.jpainsymman.2007.08.011] [Medline: 18359604]
23. Charon R. The Novelization of the Body, or, How Medicine and Stories Need One Another. Narrative 2011;19(1):33-50. [doi: 10.1353/nar.2011.0004]
24. Chou WS, Hunt Y, Folkers A, Augustson E. Cancer survivorship in the age of YouTube and social media: a narrative analysis. J Med Internet Res 2011 Jan 17;13(1):e7 [FREE Full text] [doi: 10.2196/jmir.1569] [Medline: 21247864]
25. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Education and Counseling 2004 Apr;53(1):57-64. [doi: 10.1016/S0738-3991(03)00116-2]
26. Aaker D, Aaker JL. What are Your Signature Stories? California Management Review 2016 May;58(3):49-65. [doi: 10.1525/cmr.2016.58.3.49]
27. Mayer RE, Sims VK. For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. Journal of Educational Psychology 1994;86(3):389-401. [doi: 10.1037//0022-0663.86.3.389]
28. Adam MT, Astor PJ, Krämer J. Affective Images, Emotion Regulation and Bidding Behavior: An Experiment on the Influence of Competition and Community Emotions in Internet Auctions. Journal of Interactive Marketing 2016 Aug;35:56-69. [doi: 10.1016/j.intmar.2015.12.002]
29. Mccloud S. Understanding Comics. New York, USA: William Morrow Paperbacks; 1994.
30. Green MJ, Myers KR. Graphic medicine: use of comics in medical education and patient care. BMJ 2010 Mar 03;340(mar03 2):c863-c863. [doi: 10.1136/bmj.c863] [Medline: 20200064]
31. Marchetto M. Cancer Vixen: A True Story. New York, USA: Pantheon; 2009.
32. Fies B. Mom's Cancer. New York, USA: Abrams, Harry N, Inc; 2009.
33. Mack S, Mack S. Janet & Me: An Illustrated Story of Love and Loss. New York, USA: Simon & Schuster Adult Publishing Group; 2004.
34. Lo-Fo-Wong DNN, Beijaerts A, de Haes HCJM, Sprangers MAG. Cancer in full-colour: use of a graphic novel to identify distress in women with breast cancer. J Health Psychol 2014 Dec;19(12):1554-1563. [doi: 10.1177/1359105313495905] [Medline: 23928986]
35. Czerwiec MK, Huang MN. Hospice Comics: Representations of Patient and Family Experience of Illness and Death in Graphic Novels. J Med Humanit 2017 Jun;38(2):95-113. [doi: 10.1007/s10912-014-9303-7] [Medline: 25138207]
36. Green L, Myerson J. A discounting framework for choice with delayed and probabilistic rewards. Psychol Bull 2004 Sep;130(5):769-792 [FREE Full text] [doi: 10.1037/0033-2909.130.5.769] [Medline: 15367080]
37. Seymour B, Dolan R. Emotion, decision making, and the amygdala. Neuron 2008 Jun 12;58(5):662-671 [FREE Full text] [doi: 10.1016/j.neuron.2008.05.020] [Medline: 18549779]
38. McNiff J. Action Research: All You Need to Know. London: SAGE Publications Ltd; 2017.
39. Koshy, Valsa. Action Research for Improving Educational Practice: A Step-By-Step Guide. London: SAGE Publications Ltd; 2018.
40. Carr W. Philosophy, Methodology and Action Research. J Philosophy of Education 2006 Nov;40(4):421-435. [doi: 10.1111/j.1467-9752.2006.00517.x]
41. Titchen A. International Practice Development Journal;5(1). 2015 May. Action research: genesis, evolution and orientations URL: https://www.fons.org/Resources/Documents/Journal/Vol5No1/IPDJ_0501_01.pdf [accessed 2018-12-04] [WebCite Cache ID 74PjIaNDi]
42. Miles M, Huberman A. An Expanded Sourcebook. 2nd Edition. California. Qualitative Data Analysis: SAGE Publications, Inc; 1994.
43. Taylor-Piliae R, Molassiotis A. Psychometric evaluation of the Chinese version of Mishel's Uncertainty in Illness Scale. Hong Kong Nursing Journal 2000 Jan;36(4):14-22.

44. Ho SMY, Fung WK, Chan CLW, Watson M, Tsui YKY. Psychometric properties of the Chinese version of the Mini-Mental Adjustment to Cancer (MINI-MAC) scale. Psychooncology 2003 Sep;12(6):547-556. [doi: 10.1002/pon.672] [Medline: 12923795]
45. Daher M. Cultural beliefs and values in cancer patients. Ann Oncol 2012 Apr;23 Suppl 3:66-69. [doi: 10.1093/annonc/mds091] [Medline: 22628419]
46. Bowen D, McTiernan A, Burke W, Powers D, Pruski J, Durfy S, et al. Participation in breast cancer risk counseling among women with a family history. Cancer Epidemiol Biomarkers Prev 1999 Jul;8(7):581-585 [FREE Full text] [Medline: 10428194]
47. Brandão T, Tavares R, Schulz MS, Matos PM. Measuring emotion regulation and emotional expression in breast cancer patients: A systematic review. Clin Psychol Rev 2016 Feb;43:114-127. [doi: 10.1016/j.cpr.2015.10.002] [Medline: 26520599]
48. Brunner E. Making the unseenunspoken visibleaudible: traumathe graphic novel. In: Dong L, editor. Teaching comics and graphic narrativesays on theory, strategy and practice. North Carolina: McFarland; 2012.
49. Ariela F. Sorting through My GriefPutting It into Boxes: ComicsPain. In: Cohen E, editor. Knowledge and Pain. New York: Rodopi; 2012:84-381.
50. Scott DW. Anxiety, critical thinking and information processing during and after breast biopsy. Nurs Res 1983;32(1):24-28. [Medline: 6549839]
51. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249. [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
52. Radley A. The aesthetics of illness: narrative, horror and the sublime. Sociol Health & Illness 1999 Nov;21(6):778-796. [doi: 10.1111/1467-9566.00183]

Abbreviations

AR: action research

Edited by G Eysenbach; submitted 01.06.18; peer-reviewed by R Yamanishi, Z Zhou; comments to author 26.07.18; revised version received 20.09.18; accepted 25.09.18; published 04.02.19.

Please cite as:
Lee TI, Sheu SJ, Chang HC, Hung YT, Tseng LM, Chou SS, Liang TH, Liu HJ, Lu HL, Chen MC, Liu YC, Tsai CS, Sun JC
Developing a Web-Based Comic for Newly Diagnosed Women With Breast Cancer: An Action Research Approach
J Med Internet Res 2019;21(2):e10716
URL: https://www.jmir.org/2019/2/e10716/
doi:10.2196/10716
PMID:30714947

Original Paper

Step Counts of Middle-Aged and Elderly Adults for 10 Months Before and After the Release of Pokémon GO in Yokohama, Japan

Kimihiro Hino[1], PhD; Yasushi Asami[1], PhD; Jung Su Lee[2], PhD

[1]Department of Urban Engineering, Graduate School of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

[2]Department of Public Health, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

Corresponding Author:
Kimihiro Hino, PhD
Department of Urban Engineering
Graduate School of Engineering
The University of Tokyo
7-3-1 Hongo
Bunkyo-ku, Tokyo, 113-8656
Japan
Phone: 81 3 5841 6225
Email: hino@ua.t.u-tokyo.ac.jp

Abstract

Background: Smartphones have been integrated into our society and are expected to serve as tools to improve health outcomes. In the summer of 2016, Pokémon GO, a location-based augmented reality game for smartphones was released; it attracted attention from the perspective of health, especially with its potential to increase physical activity (PA). A few studies have compared objectively measured step counts before and after the release of the game; however, they were conducted over a short study period and evaluated only young people.

Objective: The objective of this study was to confirm whether there was a difference in step counts between middle-aged and elderly players and nonplayers before and after the release of Pokémon GO.

Methods: A total of 46 players and 184 nonplayers aged ≥40 years were matched for sex, age group, and PA level; they were respondents to a questionnaire randomly sent to citizens who were given free pedometers by Yokohama city. Their play status was identified through the questionnaire. To investigate the change in step counts before and after the release of Pokémon GO according to play status, a 2-way repeated-measures analysis of variance was performed. Step counts 1 month before the release of the game were compared with those 8 months after the release. In addition, subgroup analyses according to sex, age group, PA level, and subjective health status were performed.

Results: The mean ages of players and nonplayers were 56.5 (SD 9.9) years and 57.3 (SD 9.6) years, respectively, and the mean baseline step counts of players and nonplayers were 7641.8 (SD 2754.5) and 7903.3 (SD 2674.7), respectively. There was no significant difference in the age and baseline step counts according to a t test (2-tailed). In the analysis of all samples, the interaction between play status and time effect was significant for 3 of 8 months after release. In the subgroup analyses, the interaction was significant for 3 months in men, 7 months in the 55-64-year-old group, 2 months in workers, 4 months in the active group in PA level, and 2 months in participants with subjectively good health. The interaction was significant for only 1 month, at most, in other subgroups.

Conclusions: The present study confirmed a difference in step counts between players and nonplayers before and after the release of Pokémon GO. According to our analysis, step counts were higher until 7 months after the release. The player group maintained their step counts in winter, despite the decrease in step counts of nonplayers. In subgroup analyses, players were more likely to be men, aged <55 years, workers, active, and subjectively in good health.

(J Med Internet Res 2019;21(2):e10724) doi:10.2196/10724

KEYWORDS

augmented reality; location-based games; mobile phone; pedometer; physical activity; Pokémon GO; smartphone

Introduction

Background

Smartphones have become integrated into our society and can serve as tools to improve health outcomes [1], especially in increasing physical activity (PA), which, in turn, is associated with physical and mental health [2]. However, until a few years ago, most apps incorporated few behavioral change techniques, such as feedback on behavior [3,4], and only a few studies have assessed the effects on PA of interventions using smartphone technology [5-7].

Pokémon GO is a location-based augmented reality game for smartphones released in the summer of 2016. According to the Guinness World Records, it was the "most downloaded mobile game in its first month" [8] and was the number 1 free app in 36 countries [9]. Although the combination of location-based service and augmented reality is not novel, Pokémon GO attracted attention because of the perfect incorporation of technology into the story within the game [10]. Players are encouraged to walk outside to catch Pokémon characters, hatch their eggs, and acquire game items by visiting Pokéstops and Gyms.

The game has attracted attention from the perspective of health [11-16] as a "game changer" for the physical inactivity crisis [17]. Pokémon GO players collectively walked 4.6 billion kilometers in only 8 weeks, surpassing 500 million downloads worldwide [18]. Twitter postings demonstrated that Pokémon GO players walked >30 minutes a day [19], and the game even changed peoples' hot-spots [20] and the general mobility patterns [21] in cities.

Previous Studies

For Pokémon GO players, health or PA is an important and unique motivation to play, which is not usually the case in other Web-based gaming [22-24]. In addition, positive outcomes have been observed in not only PA seekers but also players with other motivations for playing [25].

From the viewpoint of psychological health, Pokémon GO moderated psychological distress [26] and social anxiety [27], as well as increased place attachment [28]. These psychological factors relate to behavioral consequences; enjoyment of Pokémon GO remarkably resulted in perceived increases in outdoor PA [29].

Some studies used the International Physical Activity Questionnaire to report an increase in PA in players after playing Pokémon GO [22,30,31]. Several other studies used different questionnaires and reported an increase in PA frequency and duration [32], increase in PA and decrease in sedentary behaviors [33], and increase in time spent walking dogs [34]. However, self-serving bias cannot be avoided when using self-reported measurements [35-37].

Some studies have compared objectively measured step counts using a pedometer or an accelerometer, which can explain most of the variation in PA [38] before and after playing Pokémon GO. One study compared the number of daily steps taken by players and nonplayers for 4 weeks before and 6 weeks after installation of Pokémon GO. That study concluded that step counts moderately increased after installation of the game, but this result was no longer observed 6 weeks after installation [39]. Another study compared players' step counts 3 weeks before and 3 weeks after the Pokémon GO release date and observed the largest increases in participants who spent more time playing, who were overweight or obese, or who had a low baseline PA level [40]. However, both of the abovementioned studies measured the number of step counts using the iPhone Health app, which has considerable difference from the accelerometer under free-living conditions [41]. Another study used Microsoft Band, which includes a 3-axis accelerometer and gyrometer, and observed that players identified through Microsoft's search engine queries had notably increased PA over an observation period of approximately 4 weeks [42].

Study Aim

This study aimed to confirm whether there was a difference in step counts between middle-aged and elderly players and nonplayers before and after the release of Pokémon GO in Japan. The long-term impact of the game on PA should be examined; however, previous studies have evaluated its impact for only several weeks [42,43]. The uniqueness of this study lies in its long survey period (10 months), use of a pedometer, and focus on middle-aged to elderly people, while most previous studies focused on younger people [30,31,33,44]. As PA patterns change by age [45], the impact of the game for middle-aged people, who tend to be more occupied with social responsibilities, and for elderly people after retirement is expected to be different than the impact on younger people. Pokémon GO was released in Japan on July 22, 2016.

Methods

Yokohama Walking Point Program

Yokohama city is the second largest city in Japan, with a population of approximately 3.7 million and a humid subtropical climate with 4 distinct seasons. The city launched the Yokohama Walking Point Program (YWPP) in November 2014 to encourage citizens to improve their health and enjoy a healthy life expectancy, as the population ages and the disease structure changes in Japan.

A free pedometer (Omron HJ-326F, Japan) was given to citizens aged ≥18 years who volunteered to participate in the program. Participants were awarded points by scanning their pedometers via special readers installed at approximately 1000 stores and other facilities in the city. Accumulation of a certain number of points made participants eligible to win prizes. The scanned data were sent to a data server through the internet, and participants could monitor step counts and rank among all participants using a computer or a smartphone [46].

Pedometer Data

Pedometer data, which were collected from 231,606 YWPP participants who registered before March 31, 2017, were used in this study. The daily mean step count for each month was calculated for each participant using the monthly step count and the number of days recorded for that month. Data on sex and

date of birth were obtained from registered information, and age as of May 31, 2016, was considered.

Questionnaire

In May 2017, Yokohama city conducted a questionnaire survey among 2580 participants, who were randomly selected among 99,462 individuals who sent their recorded data of >80% days after registration; among the selected participants, 2055 replied to the survey (response rate 2055/2580, 79.65%). The questionnaire data and the pedometer data were combined in the analysis.

The questionnaire mainly asked about the change in participants' walking habit and health attitude as well as their use of "smartphone game apps in which players catch characters and occupy bases using location services." Participants who selected the first of 4 options ranging from "frequently" to "not at all" in a question on the frequency of use of such smartphone game apps were regarded as "players" and others as "nonplayers" in this study. Regarding occupation, participants were classified into workers (including part-time employees) and nonworkers. While health status was graded according to 4 points ranging from good to poor, it was also classified into "good" and "others" because the latter two scales (rather poor or poor) were not often selected.

Study Participants

Among the 2055 participants who replied to the questionnaire, those who had records of limited days or had abnormal records were excluded. First, 1079 participants with pedometer data for <10 days per month in any month from June 2016 to March 2017, except July 2016 (which was when Pokémon GO was released in Japan), were excluded. In addition, 62 participants whose step counts in any of those months were out of the range of the mean (SD 2) in all samples were excluded as outliers.

Among the remaining 914 participants, 46 (5.0%) were game players. Among nonplayers, 184 participants, which was equivalent to 4 times the number of players, were randomly selected so that the ratio of players to nonplayers was 1:4 after stratification by sex, age group (<55, 55-64, and ≥65 years), and PA level. We considered the age of 65 years as the threshold for elderly persons in Japan, with the employment rate decreasing the most at this age. Participants aged <65 years were divided into 2 groups so that the numbers of participants in the 2 groups were similar. The participants' PA level was based on whether their baseline step counts were higher than the target in the Japanese national health promotion plan (men aged <65 years, 9000; women aged <65 years, 8500; men aged ≥65 years, 7000; women aged ≥65 years, 6000) [47]. Finally, a total of 230 participants were included in the analysis.

Statistical Analysis

To investigate the change in step counts before and after the release of Pokémon GO according to play statuses, a 2-way repeated-measures analysis of variance was performed. Step counts in June 2016 (baseline) were compared with those for each month from August 2016 to March 2017. After analysis of all samples, subgroup analyses according to sex, age group, PA level, and subjective health status were performed. The effect of Pokémon GO could be determined based on the interaction between the play status and the time effect. The significance level was set at $P<.05$. Shapiro-Wilk test was used to assess the normality of the distribution of samples' step counts in each group in each month. Levene test was used to assess the homogeneity of variances between groups in each month. All statistical analyses were conducted using IBM SPSS Statistics 23 (IBM Corp).

Results

Sample Statistics

Participants' characteristics are presented in Table 1. The youngest participant was 40 years old because YWPP recruited citizens aged 18-39 years since June 2016. The mean ages of players and nonplayers were 56.5 (SD 9.9) years and 57.3 (SD 9.6) years, respectively. The mean baseline step counts of players and nonplayers were 7641.8 (SD 2754.5) and 7903.3 (SD 2674.7), respectively. No significant difference was observed between players and nonpalyers regarding age and baseline step counts using a *t* test (2-tailed; $P=.56$ and.61, respectively). Although the percentage of players among nonworkers was lower than that among workers, the difference was not significant according to chi-square test ($P=.22$). Regarding subjective health status, the proportion of participants in good health was almost the same in players and nonplayers.

Analysis of All Samples

The results of the analyses of all samples are presented in Table 2. Although nonplayers walked more than players by 261 steps per day in June 2016 (baseline), the difference in step counts between the 2 groups got smaller between August and October, and players walked more than nonplayers later on. Time effect was significant in August, September, January, and March, when the step counts of participants decreased compared with that at baseline. The interaction between the play status and time effect was significant in November, December, and February. Although a decrease in step counts was observed in the nonplayer group in winter, the player group maintained their step counts (Figure 1). The largest difference in step counts between players and nonplayers was 583 steps per day in December.

Table 1. Characteristics of the study participants.

Characteristic	Pokémon GO Players, n (%)	Nonplayers, n (%)	Total, n
Sex			
Male	23 (20.0)	92 (80.0)	115
Female	23 (20.0)	92 (80.0)	115
Age in years			
<55	20 (20.0)	80 (80.0)	100
55-64	16 (20.0)	64 (80.0)	80
≥65	10 (20.0)	40 (80.0)	50
Occupation			
Worker[a]	33 (22.4)	114 (77.6)	147
Nonworker	13 (15.7)	70 (84.3)	83
Physical activity level[b]			
Active	17 (20.0)	68 (80.0)	85
Nonactive	29 (20.0)	116 (80.0)	145
Health status			
Good	17 (20.7)	65 (79.3)	82
Others[c]	29 (19.6)	119 (80.4)	148
Total	46 (20.0)	184 (80.0)	230

[a]Worker included part-time employees.

[b]Participants' physical activity level was based on whether their baseline step count was higher than the target in the Japanese national health promotion plan (men aged <65 years, 9000; women aged <65 years, 8500; men aged ≥65 years, 7000; women aged ≥65 years, 6000).

[c]Others health status included rather good, rather poor, and poor.

Table 2. Results of a 2-way repeated-measures analysis of variance investigating changes in step counts before and after the release of Pokémon GO according to play statuses (N=230).

Time point[a]	Players (n=46), mean (SD)	Nonplayers (n=184), mean (SD)	Time effect		Interaction term (time×player)	
			F test (df_1, df_2)	*P* value	*F* test (df_1, df_2)	*P* value
Baseline						
June '16	7642 (2754)	7903 (2675)	N/A[b]	N/A	N/A	N/A
Follow-up						
August '16	7207 (2888)	7063[c] (2565)	23.20 (1, 228)	<.001	2.35 (1, 228)	.13
September '16	7331 (2546)	7383 (2479)	15.81 (1, 228)	<.001	1.00 (1, 228)	.32
October '16	7789 (2754)	7712 (2493)	0.04 (1, 228)	.84	2.46 (1, 228)	.12
November '16	7924 (3012)	7654 (2500)	0.02 (1, 228)	.90	4.12 (1, 228)	.04[d]
December '16	8010 (2946)	7427 (2434)	0.19 (1, 228)	.67	11.36 (1, 228)	.001[d]
January '17	7613 (3000)	7343 (2512)	4.16 (1, 228)	.04[d]	3.39 (1, 228)	.07
February '17	7870 (3086)	7450 (2497)	0.71 (1, 228)	.40	6.55 (1, 228)	.01[d]
March '17	7547 (2857)	7398 (2493)	5.39 (1, 228)	.02[d]	2.52 (1, 228)	.11

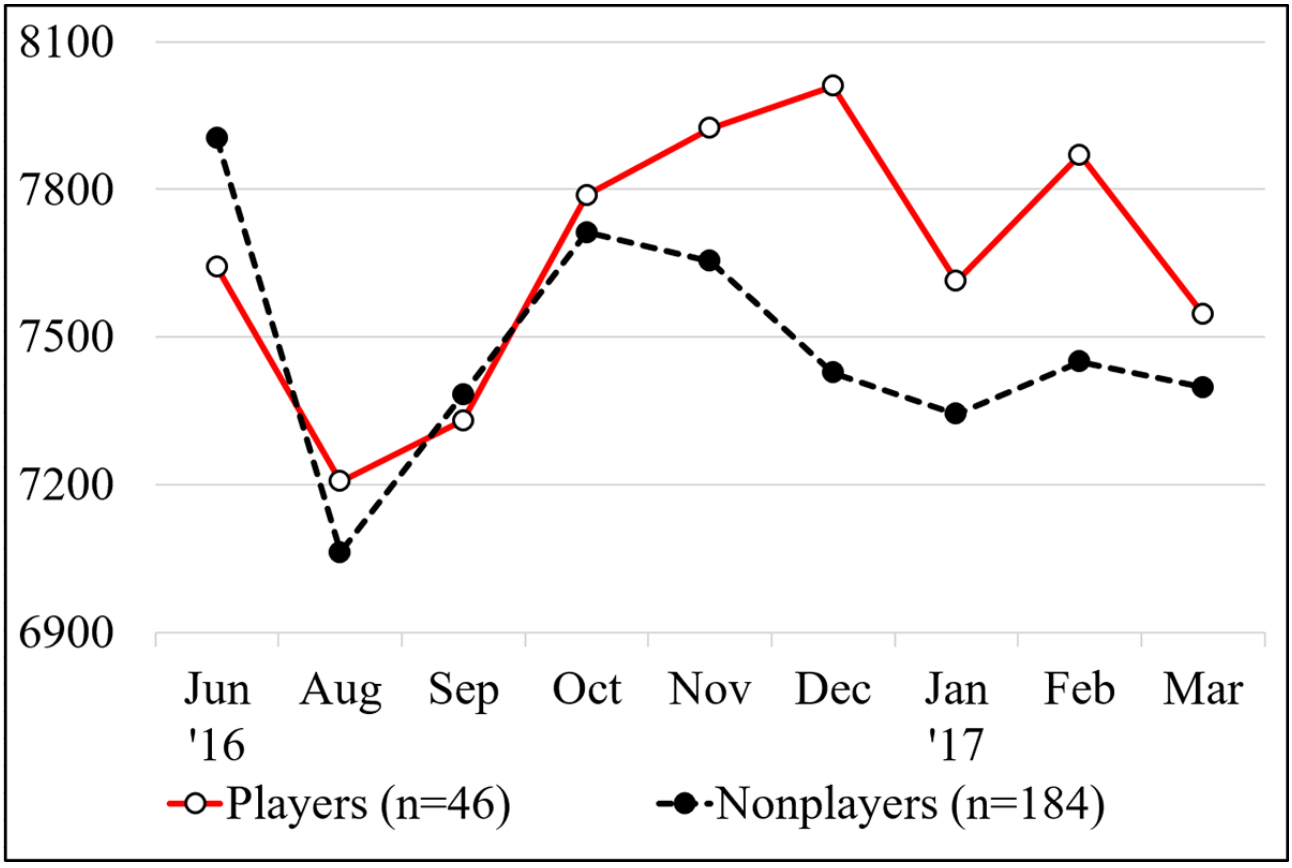
[a]No violation of homogeneity of variance assumption between groups in any month occurred.

[b]Not applicable.

[c]Significantly deviated from normality (Shapiro-Wilk test: $P<.05$).

[d]$P<.05$.

Figure 1. Mean daily step counts at baseline (June 2016) and follow-up (August 2016 to March 2017).



Subgroup Analysis

The results of the subgroup analyses are summarized in Table 3 (the complete results are presented in Multimedia Appendix 1). The interaction was significant for 3 months in men, but only 1 month in women. Regarding age, the interaction was significant for 7 of 8 months in the 55-64-year-old group but no month in the other two groups. In the 55-64-year-old group, the difference between players and nonplayers amounted to 1891 steps per day in December, whereas the difference was 320 steps per day at baseline. Regarding occupation, the interaction was significant for 2 months in workers and 1 month in nonworkers. Regarding PA level, the interaction was significant for 4 months in the active group, but only 1 month in the nonactive group. Finally, regarding subjective health status, the interaction was significant for 2 months in participants with good health and 1 month in others.

Table 3. *P* value of interaction by subgroup analyses according to sex, age group, physical activity level, and subjective health status.

Time point	Sex		Age group (years)			Occupation		Physical activity level		Health status	
	Male (n=115)	Female (n=115)	<55 (n=100)	55-64 (n=80)	≥65 (n=50)	Worker (n=147)	Nonworker (n=83)	Active (n=85)	Nonactive (n=145)	Good (n=82)	Others (n=148)
August ’16	.28	.29	.35	.03[a]	.24	.08	.96	.02[a]	.80	.61	.13
September ’16	.58	.37	.43	.31	.71	.24	.87	.11	.96	.64	.35
October ’16	.36	.18	.60	.007[a]	.73	.19	.27	.02[a]	.81	.22	.31
November ’16	.03[a]	.64	.77	.005[a]	.77	.02[a]	.86	.11	.17	.02[a]	.54
December ’16	.02[a]	.01[a]	.31	.001[a]	.23	.02[a]	.01[a]	.01[a]	.02[a]	.01[a]	.03[a]
January ’17	.06	.67	.61	.02[a]	.94	.08	.39	.30	.13	.10	.31
February ’17	.006[a]	.48	.71	.008[a]	.24	.07	.08	.07	.07	.06	.08
March ’17	.09	.70	.93	.02[a]	.99	.12	.66	.02[a]	.84	.36	.20

[a]*P*<.05.

Discussion

Principal Findings

This study confirmed the difference in step counts between players and nonplayers before and after the release of Pokémon GO. On analyzing all samples, we observed that step counts were significantly high even in February 2017, 7 months after the release of the game, which was not observed in previous short-term studies. Players did not decrease their step counts even in cold months, unlike nonplayers and other YWPP participants [46].

In subgroup analyses, step counts of the 55-64-year-old group were significantly high until the last surveyed month. This finding, along with the result that significantly higher step counts were noted in men and workers than in women and nonworkers, indicates that older, middle-aged male workers may take long, indirect routes to visit Pokéstops and Gyms for catching rare Pokémon on their way to the work place, visiting other places, and while going home. This result is in accordance with a previous finding, according to which people did not change their daily routines but slightly modified them to play the game [21]. Although the higher step counts of players aged ≥65 years was not significant in any month in this study, some senior citizen's clubs in Japan have promoted Pokémon GO to increase outdoor activities, as older people continue to play the game despite its diminishing popularity among younger players [48]. Such trials are recommended along with the development of additional location-based augmented reality games following Pokémon GO [49] for young people who lose interest quickly.

Comparison With Previous Studies

Previous studies have observed short-term effects on the PA of young people [30,31,33,44]. For example, a US study conducted in survey participants aged 18-35 years reported that the difference in daily step counts between players and nonplayers was approximately 1000 in the first week after installation of Pokémon GO [39]. The difference in this study was smaller (583 steps in December, at most), which may be because the players were older and did not play the game as intensely as younger players or because the analysis was conducted for each month and a short-term difference may have been overlooked. We observed significantly higher step counts in more months in the active group than in the nonactive group, unlike a previous study in which it was reported that players with a lower baseline PA level had increased PA [40,42]. In addition, it may be due to the difference in sample characteristics and length of the study period.

Limitations

Our study had some limitations. First, all characteristics of the study sample could not be simultaneously considered because of the limited number of participants who played the game. If there were more participants, we could have specified in more detail those who increased their step counts and continued to play the game. Their motivations for taking part in the YWPP, such as family structure and living area, should be considered as well. Second, the distribution of step counts in some subgroups did not meet the assumptions of normality and homogeneity of variance. The results of these analyses might not be as valid as those of others, although the F test is considered to be robust for nonnormal data [50]. Third, it is unclear whether the results of this study depend on the characteristics of the country, city, or participants. Further studies under different conditions are required to further examine the results. Fourth, there was a lack of variables regarding when, where, and why the participants played Pokémon GO, besides their intensity for the game, such as playing time or experience points in the game [40]. Finally, the effect on physical and psychological health, other than step counts, was not confirmed. Its negative impact, especially accidents [51-53] and injury [54,55], should be considered as well. However, our study demonstrated the potential of location-based augmented reality apps in promoting PA for a longer period than expected in the literature, at least for particular groups.

Conclusions

This study confirmed the difference in step counts between players and nonplayers before and after the release of Pokémon GO. According to our analysis, step counts were higher until 7 months after the release. The player group maintained their step counts in winter, despite the decrease in step counts of nonplayers. In subgroup analyses, players were more likely to be men, aged <55 years, workers, active, and subjectively in good health.

Acknowledgments

This study was funded by the Japan Society for the Promotion of Science KAKENHI, grant numbers 25242063, 26249073, and 18H01602.

Authors' Contributions

All authors contributed to the study design and data interpretation. KH contributed to data collection and data analysis. KH wrote the first draft, and YA and JSL critically reviewed the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The full results of subgroup analysis by sex, age group, physical activity level, and subjective health status.

[PDF File (Adobe PDF File), 98KB - jmir_v21i2e10724_app1.pdf]

References

1. Kamboj AK, Krishna SG. Pokémon GO: An innovative smartphone gaming application with health benefits. Prim Care Diabetes 2017 Dec;11(4):397-399. [doi: 10.1016/j.pcd.2017.03.008] [Medline: 28457897]
2. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry 2005 Mar;18(2):189-193. [Medline: 16639173]
3. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
4. Yang C, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. Am J Prev Med 2015 Apr;48(4):452-455. [doi: 10.1016/j.amepre.2014.10.010] [Medline: 25576494]
5. 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]
6. Fukuoka Y, Vittinghoff E, Jong SS, Haskell W. Innovation to motivation--pilot study of a mobile phone intervention to increase physical activity among sedentary women. Prev Med 2010;51(3-4):287-289 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.006] [Medline: 20600263]
7. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Using smartphone technology to monitor physical activity in the 10,000 Steps program: a matched case-control trial. J Med Internet Res 2012 Apr 20;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]
8. Swatman R. Guinness World Records. 2016 Aug 26. Pokémon Go catches five new world records Internet URL: http://www.guinnessworldrecords.com/news/2016/8/pokemon-go-catches-five-world-records-439327 [accessed 2018-03-21] [WebCite Cache ID 6yUfaCUN7]
9. Leswing K. Business Insider. 2016 Jul 26. Pokemon Go has been dethroned as the top free iPhone app-by Snapchat's $100 million stickers URL: http://www.businessinsider.com/pokemon-go-has-been-dethroned-as-the-top-free-app-by-snapchats-100-million-stickers-2016-7 [accessed 2018-03-21] [WebCite Cache ID 6yUfo35eT]
10. Tang AK. Key factors in the triumph of Pokémon GO. Business Horizons 2017 Sep;60(5):725-728. [doi: 10.1016/j.bushor.2017.05.016]
11. American Heart Association News. 2016 Jul 19. Pokémon Go brings video games outside URL: https://news.heart.org/pokemon-go-brings-video-games-outside/ [accessed 2018-04-07] [WebCite Cache ID 6yUfpzSPG]
12. Ayers JW, Leas EC, Dredze M, Allem J, Grabowski JG, Hill L. Pokémon GO-A New Distraction for Drivers and Pedestrians. JAMA Intern Med 2016 Dec 01;176(12):1865-1866. [doi: 10.1001/jamainternmed.2016.6274] [Medline: 27635638]
13. Clark AM, Clark MTG. Pokémon Go and Research. International Journal of Qualitative Methods 2016 Aug 29;15(1):160940691666776. [doi: 10.1177/1609406916667765]
14. McCartney M. Margaret McCartney: Game on for Pokémon Go. BMJ 2016 Aug 09;354:i4306. [doi: 10.1136/bmj.i4306] [Medline: 27510374]
15. Williamson JW. Will the 'Pokémon' be Heroes in the Battle Against Physical Inactivity? Sport Exerc Med Open J 2016 Sep 14;2(1):13-14. [doi: 10.17140/SEMOJ-2-133]
16. Ghosh A, Misra A. Pokémon Go, Obesity and Diabetes: A Perspective from India. Diabetes Technol Ther 2016 Dec;18(11):725-726. [doi: 10.1089/dia.2016.0303] [Medline: 27753510]
17. LeBlanc AG, Chaput J. Pokémon Go: A game changer for the physical inactivity crisis? Prev Med 2017 Aug;101:235-237. [doi: 10.1016/j.ypmed.2016.11.012] [Medline: 27856340]
18. Niantic, Inc. 2016 Sep 07. Heads Up! URL: https://www.nianticlabs.com/blog/headsup/ [accessed 2018-03-21] [WebCite Cache ID 6yUfwC2qC]
19. Krittanawong C, Aydar M, Kitai T. Pokémon Go: digital health interventions to reduce cardiovascular risk. Cardiol Young 2017 Oct;27(8):1625-1626. [doi: 10.1017/S1047951117000749] [Medline: 28414007]
20. Wang D, Wu T, Wen S, Liu D, Xiang Y, Zhou W, et al. Pokémon GO in Melbourne CBD: A case study of the cyber-physical symbiotic social networks. Journal of Computational Science 2018 May;26:456-467. [doi: 10.1016/j.jocs.2017.06.009]
21. Graells-Garrido E, Ferres L, Caro D, Bravo L. The effect of Pokémon Go on the pulse of the city: a natural experiment. EPJ Data Sci 2017 Sep 15;6(1):a. [doi: 10.1140/epjds/s13688-017-0119-3]
22. Wasmuth EV, Lima CD. The Rrp6 C-terminal domain binds RNA and activates the nuclear RNA exosome. Nucleic Acids Res 2017 Dec 25;45(2):846-860 [FREE Full text] [doi: 10.1093/nar/gkw1152] [Medline: 27899565]
23. Zsila Á, Orosz G, BÅ'the B, Tóth-Király I, Király O, Griffiths M, et al. An empirical study on the motivations underlying augmented reality games: The case of Pokémon Go during and after Pokémon fever. Personality and Individual Differences 2017 Jun:-. [doi: 10.1016/j.paid.2017.06.024]

24. Rauschnabel P, Rossmann A, tom Dieck MC. An adoption framework for mobile augmented reality games: The case of Pokémon Go. Computers in Human Behavior 2017 Nov;76:276-286. [doi: 10.1016/j.chb.2017.07.030]
25. Marquet O, Alberico C, Adlakha D, Hipp JA. Examining Motivations to Play Pokémon GO and Their Influence on Perceived Outcomes and Physical Activity. JMIR Serious Games 2017 Oct 24;5(4):e21 [FREE Full text] [doi: 10.2196/games.8048] [Medline: 29066423]
26. Watanabe K, Kawakami N, Imamura K, Inoue A, Shimazu A, Yoshikawa T, et al. Pokémon GO and psychological distress, physical complaints, and work performance among adult workers: a retrospective cohort study. Sci Rep 2017 Sep 07;7(1):10758 [FREE Full text] [doi: 10.1038/s41598-017-11176-2] [Medline: 28883633]
27. Bonus JA, Peebles A, Mares M, Sarmiento IG. Look on the Bright Side (of Media Effects): Pokémon Go as a Catalyst for Positive Life Experiences. Media Psychology 2017 Apr 07;21(2):263-287. [doi: 10.1080/15213269.2017.1305280]
28. Oleksy T, Wnuk A. Catch them all and increase your place attachment! The role of location-based augmented reality games in changing people - place relations. Computers in Human Behavior 2017 Nov;76:3-8. [doi: 10.1016/j.chb.2017.06.008]
29. Zach FJ, Tussyadiah IP. To Catch Them All—The (Un)intended Consequences of Pokémon GO on Mobility, Consumption, and Wellbeing. In: Schegg R, Stangl B, editors. Information and Communication Technologies in Tourism 2017. Cham, Switzerland: Springer; 2017:217-227.
30. Barkley JE, Lepp A, Glickman EL. "Pokémon Go!" May Promote Walking, Discourage Sedentary Behavior in College Students. Games Health J 2017 Jun;6(3):165-170. [doi: 10.1089/g4h.2017.0009] [Medline: 28628384]
31. Wong FY. Influence of Pokémon Go on physical activity levels of university players: a cross-sectional study. Int J Health Geogr 2017 Dec 22;16(1):8 [FREE Full text] [doi: 10.1186/s12942-017-0080-1] [Medline: 28228102]
32. Liu W, Ligmann-Zielinska A. A Pilot Study of Pokémon Go and Players' Physical Activity. Games Health J 2017 Dec;6(6):343-350. [doi: 10.1089/g4h.2017.0036] [Medline: 28853912]
33. Nigg CR, Mateo DJ, An J. Pokémon GO May Increase Physical Activity and Decrease Sedentary Behaviors. Am J Public Health 2017 Jan;107(1):37-38. [doi: 10.2105/AJPH.2016.303532] [Medline: 27854536]
34. Kogan L, Hellyer P, Duncan C, Schoenfeld-Tacher R. A pilot investigation of the physical and psychological benefits of playing Pokémon GO for dog owners. Computers in Human Behavior 2017 Nov;76:431-437. [doi: 10.1016/j.chb.2017.07.043]
35. Colpani V, Spritzer PM, Lodi AP, Dorigo GG, Miranda IASD, Hahn LB, et al. Physical activity in climacteric women: comparison between self-reporting and pedometer. Rev Saude Publica 2014 Apr;48(2):258-265 [FREE Full text] [Medline: 24897047]
36. Sabia S, van Hees VT, Shipley MJ, Trenell MI, Hagger-Johnson G, Elbaz A, et al. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol 2014 Mar 15;179(6):781-790 [FREE Full text] [doi: 10.1093/aje/kwt330] [Medline: 24500862]
37. Senso MM, Anderson CP, Crain AL, Sherwood NE, Martinson BC. Self-reported activity and accelerometry in 2 behavior-maintenance trials. Am J Health Behav 2014 Mar;38(2):254-264 [FREE Full text] [doi: 10.5993/AJHB.38.2.11] [Medline: 24629554]
38. Tudor-Locke C, Johnson WD, Katzmarzyk PT. Relationship between accelerometer-determined steps/day and other accelerometer outputs in US adults. J Phys Act Health 2011 Mar;8(3):410-419. [Medline: 21487141]
39. Howe KB, Suharlim C, Ueda P, Howe D, Kawachi I, Rimm EB. Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study. BMJ 2016 Dec 13;355:i6270 [FREE Full text] [doi: 10.1136/bmj.i6270] [Medline: 27965211]
40. Xian Y, Xu H, Xu H, Liang L, Hernandez AF, Wang TY, et al. An Initial Evaluation of the Impact of Pokémon GO on Physical Activity. J Am Heart Assoc 2017 May 16;6(5):pii: e005341 [FREE Full text] [doi: 10.1161/JAHA.116.005341] [Medline: 28512111]
41. Duncan MJ, Wunderlich K, Zhao Y, Faulkner G. Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. J Sports Sci 2018 Aug;36(15):1695-1704. [doi: 10.1080/02640414.2017.1409855] [Medline: 29179653]
42. Althoff T, White RW, Horvitz E. Influence of Pokémon Go on Physical Activity: Study and Implications. J Med Internet Res 2016 Dec 06;18(12):e315 [FREE Full text] [doi: 10.2196/jmir.6759] [Medline: 27923778]
43. Freeman B, Chau J, Mihrshahi S. Why the public health sector couldn't create Pokémon Go. Public Health Res Pract 2017 Jul 26;27(3):pii: 2731724 [FREE Full text] [doi: 10.17061/phrp2731724] [Medline: 28765857]
44. Marquet O, Alberico C, Hipp AJ. Pokémon GO and physical activity among college students. A study using Ecological Momentary Assessment. Computers in Human Behavior 2018 Apr;81:215-222. [doi: 10.1016/j.chb.2017.12.028]
45. Caspersen CJ, Pereira MA, Curran KM. Changes in physical activity patterns in the United States, by sex and cross-sectional age. Med Sci Sports Exerc 2000 Sep;32(9):1601-1609. [Medline: 10994912]
46. Hino K, Lee JS, Asami Y. Associations between seasonal meteorological conditions and the daily step count of adults in Yokohama, Japan: Results of year-round pedometer measurements in a large population. Prev Med Rep 2017 Dec;8:15-17 [FREE Full text] [doi: 10.1016/j.pmedr.2017.07.009] [Medline: 28831368]
47. Ministry of Health, Labour and Welfare. 2012. Health Japan 21 (second) URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/kenkounippon21.html [WebCite Cache ID 6yUg2aiGt]

48. Kyodo News. 2017 Jul 15. One year after release, Pokemon Go is fitness tool for older people URL: https://english.kyodonews.net/news/2017/07/2260c4fc6b84-feature-one-year-after-release-pokemon-go-is-fitness-tool-for-older-people.html [WebCite Cache ID 6yUh6tOoO]
49. Kamel Boulos MN, Lu Z, Guerrero P, Jennett C, Steed A. From urban planning and emergency training to Pokémon Go: applications of virtual reality GIS (VRGIS) and augmented reality GIS (ARGIS) in personal, public and environmental health. Int J Health Geogr 2017 Dec 20;16(1):7 [FREE Full text] [doi: 10.1186/s12942-017-0081-0] [Medline: 28219378]
50. Blanca MJ, Alarcón R, Arnau J, Bono R, Bendayan R. Non-normal data: Is ANOVA still a valid option? Psicothema 2017 Nov;29(4):552-557. [doi: 10.7334/psicothema2016.383] [Medline: 29048317]
51. Barbieri S, Vettore G, Pietrantonio V, Snenghi R, Tredese A, Bergamini M, et al. Pedestrian Inattention Blindness While Playing Pokémon Go as an Emerging Health-Risk Behavior: A Case Report. J Med Internet Res 2017 Dec 01;19(4):e86 [FREE Full text] [doi: 10.2196/jmir.6596] [Medline: 28365563]
52. Joseph B, Armstrong DG. Potential perils of peri-Pokémon perambulation: the dark reality of augmented reality? Oxf Med Case Reports 2016 Oct;2016(10):omw080 [FREE Full text] [doi: 10.1093/omcr/omw080] [Medline: 27713831]
53. Wagner-Greene VR, Wotring AJ, Castor T, Kruger J, Mortemore S, Dake JA. Pokémon GO: Healthy or Harmful? Am J Public Health 2017 Jan;107(1):35-36. [doi: 10.2105/AJPH.2016.303548] [Medline: 27925813]
54. Anderson N, Steele J, O'Neill L, Harden LA. Pokémon Go: mobile app user guides. Br J Sports Med 2016 Aug 18;51(20):1505-1506. [doi: 10.1136/bjsports-2016-096762]
55. Serino M, Cordrey K, McLaughlin L, Milanaik RL. Pokémon Go and augmented virtual reality games: a cautionary commentary for parents and pediatricians. Curr Opin Pediatr 2016 Dec;28(5):673-677. [doi: 10.1097/MOP.0000000000000409] [Medline: 27479151]

Abbreviations

PA: physical activity
YWPP: Yokohama Walking Point Program

Edited by G Eysenbach; submitted 30.04.18; peer-reviewed by Y Yang, A Centi; comments to author 24.09.18; revised version received 18.10.18; accepted 05.11.18; published 05.02.19.

Please cite as:
Hino K, Asami Y, Lee JS
Step Counts of Middle-Aged and Elderly Adults for 10 Months Before and After the Release of Pokémon GO in Yokohama, Japan
J Med Internet Res 2019;21(2):e10724
URL: https://www.jmir.org/2019/2/e10724/
doi:10.2196/10724
PMID:30720438

Original Paper

An Integrated Influenza Surveillance Framework Based on National Influenza-Like Illness Incidence and Multiple Hospital Electronic Medical Records for Early Prediction of Influenza Epidemics: Design and Evaluation

Cheng-Yi Yang[1*], PhD; Ray-Jade Chen[2,3*], MSc, MD; Wan-Lin Chou[3], MS; Yuarn-Jang Lee[4], MD; Yu-Sheng Lo[1], PhD

[1]Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan

[2]Department of Surgery, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[3]Taipei Medical University Hospital, Taipei, Taiwan

[4]Division of Infectious Disease, Department of Internal Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yu-Sheng Lo, PhD
Graduate Institute of Biomedical Informatics
Taipei Medical University
250 Wuxing Street
Taipei, 11031
Taiwan
Phone: 886 939588576
Email: Loyusen@tmu.edu.tw

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2019/3/e13699/

Abstract

Background: Influenza is a leading cause of death worldwide and contributes to heavy economic losses to individuals and communities. Therefore, the early prediction of and interventions against influenza epidemics are crucial to reduce mortality and morbidity because of this disease. Similar to other countries, the Taiwan Centers for Disease Control and Prevention (TWCDC) has implemented influenza surveillance and reporting systems, which primarily rely on influenza-like illness (ILI) data reported by health care providers, for the early prediction of influenza epidemics. However, these surveillance and reporting systems show at least a 2-week delay in prediction, indicating the need for improvement.

Objective: We aimed to integrate the TWCDC ILI data with electronic medical records (EMRs) of multiple hospitals in Taiwan. Our ultimate goal was to develop a national influenza trend prediction and reporting tool more accurate and efficient than the current influenza surveillance and reporting systems.

Methods: First, the influenza expertise team at Taipei Medical University Health Care System (TMUHcS) identified surveillance variables relevant to the prediction of influenza epidemics. Second, we developed a framework for integrating the EMRs of multiple hospitals with the ILI data from the TWCDC website to proactively provide results of influenza epidemic monitoring to hospital infection control practitioners. Third, using the TWCDC ILI data as the gold standard for influenza reporting, we calculated Pearson correlation coefficients to measure the strength of the linear relationship between TMUHcS EMRs and regional and national TWCDC ILI data for 2 weekly time series datasets. Finally, we used the Moving Epidemic Method analyses to evaluate each surveillance variable for its predictive power for influenza epidemics.

Results: Using this framework, we collected the EMRs and TWCDC ILI data of the past 3 influenza seasons (October 2014 to September 2017). On the basis of the EMRs of multiple hospitals, 3 surveillance variables, TMUHcS-ILI, TMUHcS-rapid influenza laboratory tests with positive results (RITP), and TMUHcS-influenza medication use (IMU), which reflected patients with ILI, those with positive results from rapid influenza diagnostic tests, and those treated with antiviral drugs, respectively,

showed strong correlations with the TWCDC regional and national ILI data (r=.86-.98). The 2 surveillance variables—TMUHcS-RITP and TMUHcS-IMU—showed predictive power for influenza epidemics 3 to 4 weeks before the increase noted in the TWCDC ILI reports.

Conclusions: Our framework periodically integrated and compared surveillance data from multiple hospitals and the TWCDC website to maintain a certain prediction quality and proactively provide monitored results. Our results can be extended to other infectious diseases, mitigating the time and effort required for data collection and analysis. Furthermore, this approach may be developed as a cost-effective electronic surveillance tool for the early and accurate prediction of epidemics of influenza and other infectious diseases in densely populated regions and nations.

(J Med Internet Res 2019;21(2):e12341) doi:10.2196/12341

KEYWORDS

influenza; epidemics; influenza surveillance; electronic disease surveillance; electronic medical records; electronic health records; public health

Introduction

Influenza is a contagious respiratory illness caused by influenza viruses, which are easily transmitted from person to person during seasonal epidemics. Infection can result in severe illness, hospitalization, and even death, particularly among the elderly, very young individuals, and individuals with preexisting medical conditions [1,2]. According to the World Health Organization (WHO), an estimated 3 to 5 million cases of seasonal influenza epidemics develop severe illness, resulting in 250,000 to 500,000 annual deaths worldwide [3]. Moreover, influenza contributes to heavy economic losses to individuals and communities [4,5]. Therefore, early detection and treatment are the top priorities of health care professionals to reduce mortality and morbidity because of influenza.

Several surveillance methods for the real-time detection and routine monitoring of influenza activity have been proposed. Traditional influenza surveillance systems primarily rely on influenza-like illness (ILI) and virology data reported by health care providers, including hospitals, clinics, and contract laboratories [6,7]. The most notable are the official Centers for Disease Control and Prevention (CDC) reports on influenza activity in the United States and other countries, which despite their high detection accuracy, typically have a time lag of 1 to 2 weeks from incidence to report completion [8-10]. Therefore, simple, accurate, and timely surveillance approaches may be beneficial for the society.

Nontraditional influenza surveillance systems have attempted to provide more timely estimates of influenza activity using alternative or novel data sources. Compared with traditional approaches, nontraditional surveillance systems are more often aimed at the real-time detection of influenza epidemics. Some use sentinel-based surveillance systems with alternative data sources, such as flu prescription drug sales and school absenteeism data, to estimate the total number of patients with influenza [11-13]. These sentinel-based surveillance systems often have limitations, such as participating rates in the official sentinel surveillance, lack of influenza-specific data, and potential overestimation of the number of patients with influenza [13-15]. Some systems attempt to use novel data sources from the internet, such as Google user search logs and social media stream data, to capture influenza trends in near real time [16-19]; however, such strategies capture user behaviors, thoughts, and subjective judgments, as reflected through keywords, without physician confirmation. Unfortunately, these surveillance systems remain plagued by limitations in the accuracy, difficulties in recognizing influenza outbreaks at local and regional scales, and lack of support by clinical evidence and are often viewed as an extension of traditional surveillance systems [20-25]. Other approaches might use hospital-based data—electronic health records, emergency department (ED) patients' diagnosis codes or chief complaints, or rapid influenza diagnostic tests (RIDTs) results—to conduct near-real-time influenza surveillance and detect local and regional outbreaks. The results of such approaches are considered to be complementary to the regular reports of the national CDC [26-29]. Therefore, real-time and accurate monitoring, early prediction, and control of influenza epidemics remain major global public health challenges [16,30,31].

In Taiwan, seasonal influenza epidemics occur in winter, from late November to March, although smaller epidemics may occur during summer. The influenza surveillance and reporting system of Taiwan CDC (TWCDC) is centrally organized and aided by the regional departments of health, contracted laboratories, and all health care institutes. The Taiwan National Infectious Disease Statistics System [32] reports information regarding influenza activity, such as (1) viral surveillance data, including the percentage of positive specimens, viral types, trends, antigenicity, and antiviral resistance; (2) ILI surveillance data, including proportions of outpatient department (OPD) and emergency department visits attributable to ILI; and (3) reports of severe complicated influenza and mortality surveillance data. The TWCDC ILI reports, including national and regional ILI data, are freely available on the website but show a similar delay of at least one week for studies of OPD and ED visits for ILI and 2 weeks for studies of virology [32]. This information represents the gold standard for clinical and epidemiological studies of influenza epidemics in Taiwan. Over the past 3 years (October 2014 to September 2017), TWCDC has reported 2 influenza epidemics in Taiwan, 1 from the winter of 2015 until the spring of 2016 and another during the summer of 2017.

In 1995, Taiwan introduced a government-administered, insurance-based National Health Insurance (NHI) system. This single-payer NHI system of Taiwan comprehensively covers the population and has good accessibility, short waiting times, and relatively low costs. However, the weaknesses of this type

of health care system include the variable quality of care, weak gatekeeper role, and increasing financial pressures [33,34]. Therefore, patients are free to seek clinical services without referral when they feel uncomfortable. Consequently, the first-line physicians in OPDs or EDs may be in the best position to detect influenza activity early. Most patients with positive RIDT results are treated with antiviral drugs. RIDTs take only 15 to 20 min to obtain results, but they vary in sensitivity and specificity when compared with the other methods, such as viral culture or reverse transcription polymerase chain reaction (RT-PCR). In particular, RIDTs have high specificities (90%-95%), but their sensitivities are approximately 50% to 70% [35]. However, because of the long examination time (ie, several hours to days to obtain test results), specific testing equipment requirements, and higher costs of viral cultures or RT-PCR, most health care facilities tend to consider these as second-line diagnostic tools while mainly relying on RIDTs for influenza. Furthermore, because of low sensitivities of RIDTs, administration of antiviral drugs is suggested for individuals with ILI as well as for those presenting with other risk factors, such as underlying diseases, immunosuppression, or family environments with the potential for cluster infections in the absence of positive RIDT results.

When a patient is suspected or confirmed to have a diagnosis of influenza, most of the data on clinical history, laboratory results, received medications, and diagnostic codes are electronically documented in the hospital's electronic medical records (EMRs). Therefore, using EMR systems, patients with ILI, patients with positive RIDTs, and patients prescribed antiviral drugs can be easily identified, thereby allowing timely monitoring of influenza trends. Furthermore, by scaling up EMRs of a single hospital to a regional health care system, influenza epidemics may be predicted more accurately. Therefore, the use of selected surveillance data from hospital-based EMRs may be considered to aid the early prediction of regional and national influenza epidemics [36].

In this study, we aimed to develop a framework for integrating EMRs of multiple hospitals and the TWCDC ILI data, with the goal of reducing the time and effort involved in data collection and analysis for influenza surveillance. To improve regional and national influenza surveillance and control, we identified surveillance variables on the basis of physician behaviors for the early prediction of influenza epidemics and evaluated the predictive ability of each variable over the study period.

Methods

Settings

This study was conducted in the Taipei Medical University Health Care System (TMUHcS), which operates 3 hospitals: Taipei Medical University Hospital, Wan Fang Hospital, and Shuang Ho Hospital. All 3 hospitals are located in the Taipei metropolitan area (Taipei City and New Taipei City), within an 11-km driving distance from each other, and are Joint Commission International-certified academic research centers providing tertiary care. The EMRs of the 3 hospitals are hosted on a secure and private network for easy accessibility and interoperability. The total capacity of the 3 hospitals is 2800 beds, and they receive over 300,000 OPD visits each month. On an average, clinical reimbursement of TMUHcS accounts for 3.2% of the total NHI budget.

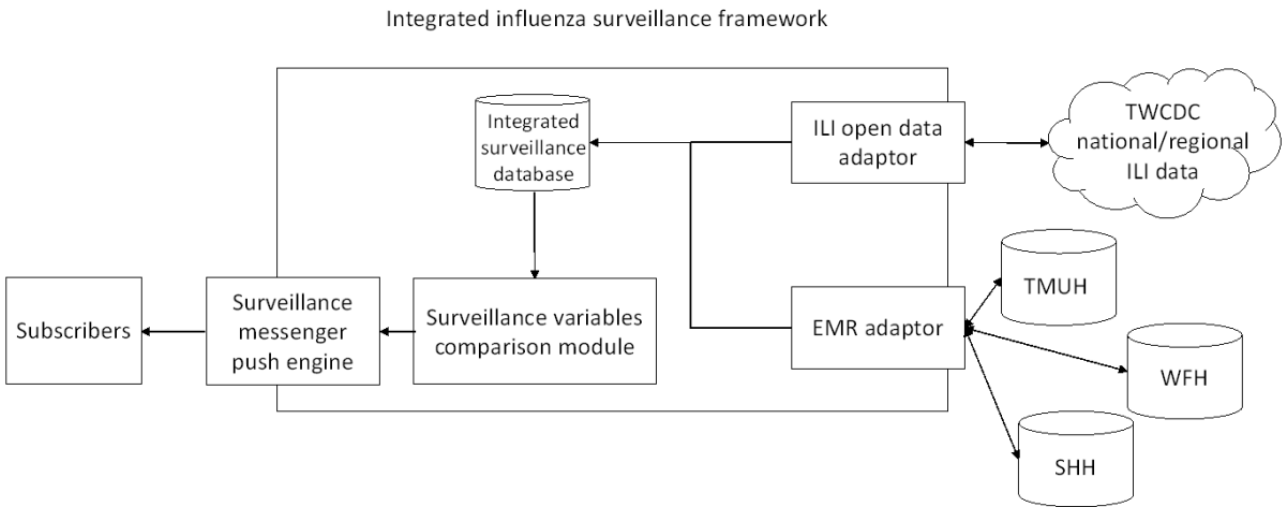
Integrated Influenza Surveillance Framework

We developed an Integrated Influenza Surveillance Framework that included 5 components (Figure 1): ILI Open Data Adapter, EMR Adapter, Integrated Surveillance Database, Surveillance Variables Comparison Module, and Surveillance Messenger Push Engine.

Influenza-Like Illness Open Data Adapter

The ILI Open Data Adapter included a Web crawler that can periodically initiate access requests to the TWCDC ILI open data website [32]. The TWCDC ILI data included the year, week number, ILI case counts, and the corresponding areas. Once the ILI Open Data Adapter retrieved the ILI data, these were classified into national (Taiwan)- and regional (Taipei City and New Taipei City)-level data and then deposited into the Integrated Surveillance Database.

Figure 1. Integrated Influenza Surveillance Framework for influenza epidemic surveillance. ILI: influenza-like illness; EMR: electronic medical record; TWCDC: Taiwan Centers for Disease Control and Prevention; TMUH: Taipei Medical University Hospital; WFH: Wan Fang Hospital; SHH: Shuang Ho Hospital.



Electronic Medical Records Adapter

The EMR Adapter created a connection with TMUHcS hospital EMRs and extracted EMRs according to the definition of each surveillance variable. Once the specific surveillance variables were retrieved from the EMRs, these data were stored in the Integrated Surveillance Database.

Integrated Surveillance Database

The Integrated Surveillance Database stored national- and regional-level ILI data from TWCDC and specific surveillance variables data from TMUHcS hospital EMRs. These data were then input into the Surveillance Variables Comparison Module for further processing.

Surveillance Variables Comparison Module

This module included an automated procedure developed using the Java programming language. It could access surveillance variable data from the Integrated Surveillance Database and conduct Pearson correlation analyses as well as Moving Epidemic Method (MEM) analyses. In this module, the Java programming language invoked the SAS Integrated Object Model [37] to conduct Pearson correlation analyses and rJava [38] to access the R server to perform MEM analysis.

In the initial stage of Surveillance Variables Comparison Module development, Pearson correlation analysis was performed to assess the association between each TMUHcS surveillance variable from the EMRs and the national and regional TWCDC ILI data. After the initial stage, Pearson correlation analysis was not performed at the time of data collection but rather after accumulating the annual data to ascertain that TMUHcS EMRs and TWCDC ILI data reached a significant correlation. For example, once a surveillance variable from TMUHcS EMR data was identified to have a strong correlation (correlation coefficient >.7 and $P \leq .05$) with TWCDC regional and national ILI data, MEM was used to assess whether the daily number of accumulated cases of the surveillance variable surpassed its influenza epidemic threshold. If the value was over the threshold, the Surveillance Variable Comparison Module was made to activate the Surveillance Messenger Push Engine to conduct subsequent tasks.

Surveillance Messenger Push Engine

To provide immediate monitoring results to hospital infection control practitioners, the publish-subscribe pattern was adopted in the Surveillance Messenger Push Engine. The publish-subscribe pattern was developed using Amazon Simple Notification Service [39,40] to manage the subscribes and push messages, such as influenza epidemic alert information. Subscribers (eg, hospital infection control practitioners) could subscribe to this messaging service using the hospital's mobile app or the computerized physician order entry (CPOE) system. When the Surveillance Messenger Push Engine was activated by the Surveillance Variables Comparison Module, it sent the alert messages to each subscriber; therefore, the subscribers could receive and check the alert messages via their mobile devices or the CPOE system.

Variables for Influenza Surveillance

A team of infection control specialists, clinicians, and information technology professionals worked together to identify relevant surveillance variables from EMRs. Unstructured data (ie, data in free-text format) were excluded, and only the structured data including diagnosis codes, medications prescribed, and laboratory tests and results were retained. All of the selected surveillance variables can be automatically extracted from TMUHcS EMRs. In this study, we identified 3 relevant surveillance variables: (1) TMUHcS-ILI, which referred to patients whose diagnosis codes were related to ILI symptoms on the basis of the TWCDC ILI case definition [32], (2) TMUHcS-RITP, which referred to patients whose RIDTs were positive on the basis of ILI diagnosis codes, and (3) TMUHcS-IMU, which referred to patients who were prescribed antiviral drugs for influenza on the basis of ILI diagnosis codes.

Each surveillance variable was independently evaluated for the EMR of each outpatient. For example, if an outpatient's EMR contained ILI diagnosis codes and a positive RIDT result, both TMUHcS-ILI and TMUHcS-RITP were considered positive. Descriptions of the surveillance variables and their corresponding definitions are summarized in Table 1.

Moving Epidemic Method

The MEM can be used to define baseline influenza activity from historical data and establish an absolute threshold, above which the weekly infection rates are considered to represent an epidemic period [41-43]. The absolute threshold can be used to define the start and end of the epidemic period. We used MEM to evaluate the predictive power of each surveillance variable for influenza epidemics. In addition, MEM was used to determine the absolute threshold for each surveillance variable and to determine the surveillance variables that reached the epidemic peak the soonest. Moreover, we estimated and compared the absolute threshold and epidemic period for each surveillance variable to determine the most useful variables for early prediction of influenza epidemics.

Data Collection

We retrospectively collected data from TMUHcS EMRs and TWCDC ILI data (both regional and national) over a 3-year period from October 2014 to September 2017. For TMUHcS EMRs, all outpatients who visited TMUHcS hospitals between October 1, 2014 and September 30, 2017 were included in this study. However, TMUHcS ED patients were excluded in this study because the case definition of the TWCDC ED ILI was different and had a wider range than that of the TWCDC OPD ILI [44]. Moreover, official TWCDC ILI reports to the public were mainly on the basis of the OPD ILI data. Therefore, in this study, we only used TMUHcS OPD EMRs for comparison with TWCDC ILI data. TWCDC ILI data are freely distributed and available from the TWCDC website [32]. Generally, TWCDC announces the latest ILI reports to the public with a time lag of approximately 1 week.

Table 1. Influenza surveillance variables and the corresponding data definitions.

Surveillance variable (abbreviation)	Description	Definition
ILI[a] diagnosis codes [32] (TMUHcS[b]-ILI)	Patients who were diagnosed with ILI-related symptoms	ILI diagnosis codes: (1) ICD-9-CM[c] codes: 480.xx~487.xx.; (2) ICD-10-CM codes[d]: J09-~J18-, A22-, A37-, B25-, and B44-
Rapid influenza laboratory tests with positive results (TMUHcS-RITP[e])	Patients who had positive results on rapid influenza tests based on ILI diagnosis codes	Rapid Influenza laboratory tests: positive, + And ILI diagnosis codes: (1) ICD-9-CM codes: 480.xx~487.xx (2) ICD-10-CM codes[d]: J09-~J18-, A22-, A37-, B25-, and B44-
Influenza medication use (TMUHcS-IMU[f])	Patients who were prescribed with antiviral drugs against influenza based on ILI diagnosis codes	Medications: Tamiflu, Relenza, Rapiacta, and Avigan And ILI diagnosis codes: (1) ICD-9-CM codes: 480.xx-487.xx (2) ICD-10-CM codes[d]: J09- to J18-, A22-, A37-, B25-, and B44-

[a]ILI: influenza-like illness.

[b]TMUHcS: Taipei Medical University Health Care System.

[c]ICD-9-CM: International Classification of Disease, Ninth Revision, Clinical Modification.

[d]Implemented in Taiwan in January 2016.

[e]RITP: rapid influenza laboratory tests with positive results.

[f]IMU: influenza medication use.

Data Analysis

Data analysis was performed weekly over a 12-month period to determine the week with the highest influenza incidence, according to different surveillance variables. The entire study duration was categorized into 3 periods: Period 1 (October 2014 to September 2015), Period 2 (October 2015 to September 2016), and Period 3 (October 2016 to September 2017).

According to the definitions of each surveillance variable, the number of influenza cases was analyzed to determine the proportion of the corresponding population affected on a weekly basis. To investigate in more detail the comparability of the weekly patterns, we used the locally estimated scatterplot smoothing for weekly data of each surveillance variable [45]. Using the statistical software SAS 9.4 (SAS Institute), Pearson correlation coefficients were calculated to assess the association of each surveillance variable from the EMRs with that from the national and regional TWCDC ILI data. A strong correlation was defined as correlation coefficient of >.7 and statistical significance at a $P \leq .05$. We used MEM to estimate the early predictive power of each surveillance variable. Data analysis was performed in R 3.4.2 (CRAN) [46] using the MEM library version 2.11 [47]. The Joint Institutional Review Board of Taipei Medical University and TMUHcS approved this study.

Results

Study Population

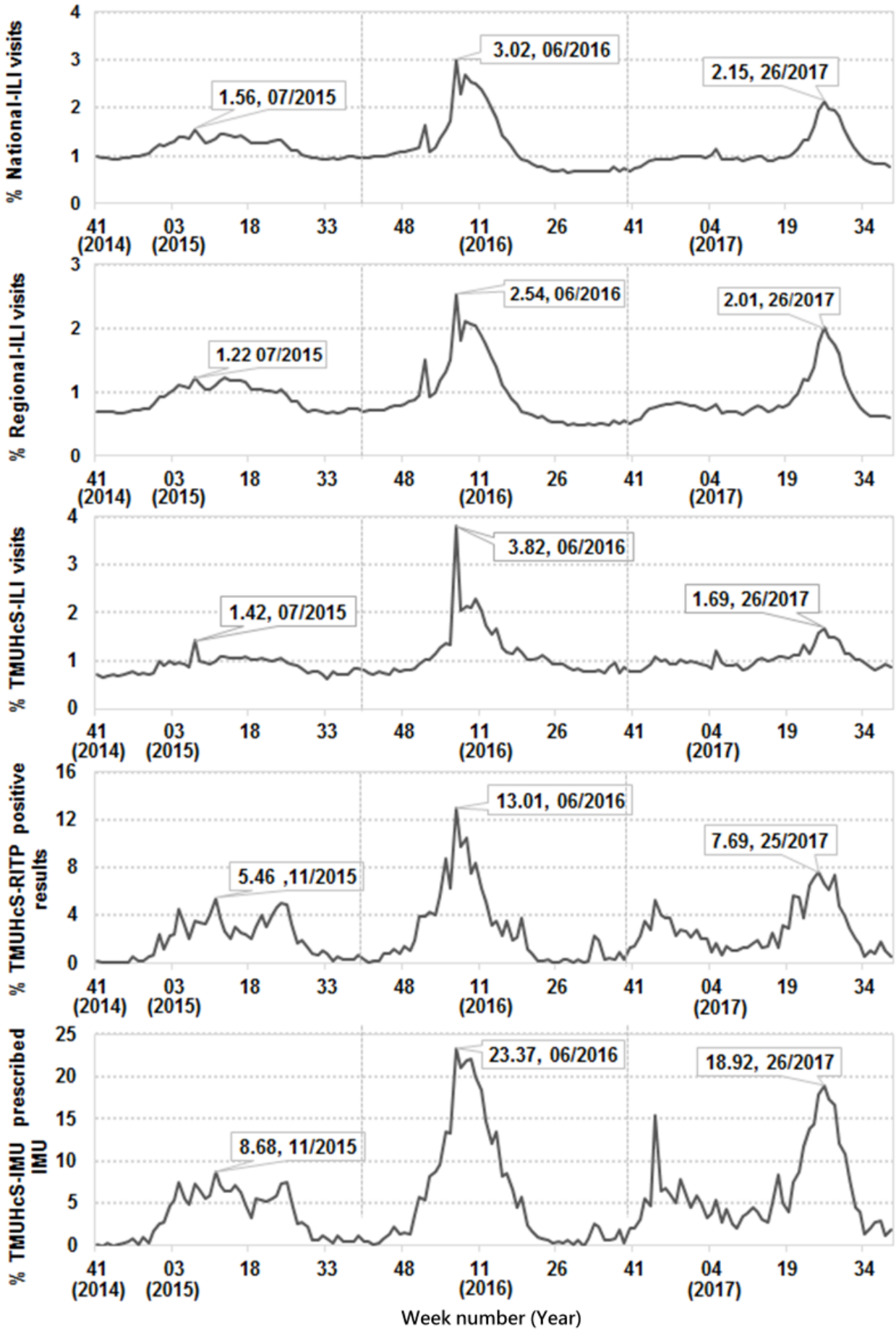
From October 1, 2014 to September 30, 2017 we collected a total of 156 weeks of the TWCDC ILI (both regional- and national-level) and TMUHcS EMRs data through the Integrated Influenza Surveillance Framework. Over this 3-year period, the total number of outpatient visits for the national (Taiwan), regional (Taipei Metropolitan), and TMUHcS hospitals were 759,306,494, 214,161,277, and 12,679,529, respectively. Overall, 8,772,125 (1.16%), 1,950,577 (0.91%), 128,378 (1.01%), 3888 (3.03%), and 8463 (6.59%) OPD cases presented with ILI in the national ILI, regional ILI, TMUHcS-ILI, TMUHcS-RITP, and TMUHcS-IMU datasets, respectively. Moreover, of all the TMUHcS-ILI OPD cases, 8094 (6.30%) were advised to undergo RIDTs; of these RIDTs cases, 3888 (48.04%) had positive results. The descriptive statistics for the study population are presented in Multimedia Appendix 1.

Time Series Trends for Influenza Surveillance Variables

The numbers and proportions of cases occurring over a week varied according to each surveillance variable (ie, national ILI, regional ILI, TMUHcS-ILI, TMUHcS-RITP, and TMUHcS-IMU; Figure 2). The highest peak for all surveillance variables occurred during the sixth week of 2016, during which the proportions of influenza-related OPD visits were 3.02%, 2.54%, 3.82%, 13.01%, and 23.37% of the national ILI, regional ILI, TMUHcS-ILI, TMUHcS-RITP, and TMUHcS-IMU data, respectively.

In 2017, the peak number of cases for all surveillance variables was lower than that in 2016. However, all variables reached the peak between the twenty-fifth and twenty-sixth week of 2017, during which the proportions of influenza-related OPD visits were 2.15%, 2.01%, 1.69%, 7.69%, and 18.92% of the national ILI, regional ILI, TMUHcS-ILI, TMUHcS-RITP, and TMUHcS-IMU data, respectively. However, the epidemic peaks identified using each surveillance variable in 2015 involved fewer cases and differed from those reported in the subsequent 2 years. In 2015, the peaks identified using the national ILI, regional ILI, and TMUHcS-ILI data occurred during the seventh week, with influenza-related case proportions of 1.56%, 1.22%, and 1.42%, respectively. In contrast, the epidemic peaks identified using TMUHcS-RITP and TMUHcS-IMU data occurred during the eleventh week, with influenza-related case proportions of 5.46% and 8.68%, respectively.

Figure 2. Timeline of influenza-related case proportions based on national influenza-like illness (ILI) data, regional ILI data, and surveillance variables from the electronic medical records of Taipei Medical University Health Care System (TMUHcS; October 2014 to September 2017). IMU: influenza medication use; RITP: rapid influenza laboratory tests with positive results.



Correlation Analysis for Surveillance Variables

Tables 2 and 3 show the Pearson correlation coefficients assessing the relationships between each surveillance variable of TMUHcS EMRs data and those of TWCDC regional and national ILI data. Over the entire study period, the correlation coefficient between the national and regional ILI data was .99 (P<.05) and remained constant over the 3-year study period.

Overall, Pearson correlation coefficients between the regional ILI data and each TMUHcS variable ranged from 0.86 to 0.90 (Table 2). All of the surveillance variables showed statistically significant correlation coefficients (P<.05) with regional ILI, with TMUHcS-IMU showing the strongest overall correlation (r=.90, P<.05). The strongest correlation coefficient of .97 (P<.05) was observed between TMUHcS-ILI and regional ILI during period 1 and between TMUHcS-IMU and regional ILI during period 2. Moreover, TMUHcS-RITP was strongly correlated with regional ILI during period 2 (r=.92, P<.05).

Similarly, over the entire study period, Pearson correlation coefficients between national ILI and each of the TMUHcS surveillance variables ranged from 0.88 to 0.93 (P<.05; Table 3). All variables showed statistically significant correlation coefficients (P<.05), whereas TMUHcS-IMU showed the strongest correlation with national ILI (r=.93, P<.05). The strongest correlation coefficient of .98 (P<.05) was observed between TMUHcS-ILI and national ILI during period 1. TMUHcS-RITP during period 2 (r=0.95, P<.05) and TMUHcS-IMU during period 2 (r=0.97, P<.05) were strongly correlated with national ILI.

Early Prediction of Influenza Epidemics

In Figure 3, the absolute thresholds for the 3-year study period were 1.56% to 1.52% for national ILI, 1.25% to 1.26% for regional ILI, 1.4% to 1.4% for TMUHcS-ILI, 4.6% to 3.25% for TMUHcS-RITP, and 8.69% to 6.39% for TMUHcS-IMU. The influenza epidemic threshold was not reached by any of the surveillance variables during period 1. During period 2, all of the surveillance variables surpassed the influenza epidemic threshold. TMUHcS-IMU (E2) and TMUHcS-RITP (D2) reached peak epidemic activity during the second week of 2016, followed by TMUHcS-ILI (C2) during the third week of 2016, then by national ILI (A2) and regional ILI (B2) during the fifth week of 2016 (1-week lag). Therefore, the peak epidemic activity according to the TMUHcS-RITP and TMUHcS-IMU was reached 3 weeks before that according to the national ILI and regional ILI during period 2.

During period 3, the influenza epidemic threshold was surpassed by each surveillance variable. TMUHcS-IMU (E3) and TMUHcS-RITP (D3) reached the peak epidemic activity during the twenty-first week of 2017, followed by regional ILI (B3) during the twenty-fourth week of 2017 (1-week lag) and subsequently by TMUHcS-ILI (C3) during the twenty-fifth week of 2017 and national ILI (A3; 1-week lag). Again, the peak influenza activity according to TMUHcS-RITP and TMUHcS-IMU was observed 3 to 4 weeks before that according to the national ILI and regional ILI during period 3.

The TWCDC ILI reports were in complete agreement with our study results (Figure 2). Moreover, the peak influenza epidemic activity predicted by MEM using TMUHcS-IMU and TMUHcS-RITP occurred 3 to 4 weeks earlier than that using national ILI and regional ILI, both in 2016 and 2017 (Figure 3).

Table 2. Pearson correlation coefficients between the Taiwan Center for Disease Control Regional influenza-like illness data and each surveillance variable from the electronic medical records of Taipei Medical University Health Care System (October 2014 to September 2017).

Time period	National ILI[a]	TMUHcS[b]-ILI	TMUHcS-RITP[c]	TMUHcS-IMU[d]
Regional ILI of the entire period (October 2014 to September 2017)	.99	.87	.86	.90
Regional ILI of period 1 (October 2014 to September 2015)	.99	.97	.88	.96
Regional ILI of period 2 (October 2015 to September 2016)	.99	.93	.92	.97
Regional ILI of period 3 (October 2016 to September 2017)	.99	.96	.82	.91

[a]ILI: influenza-like illness.

[b]TMUHcS: Taipei Medical University Health Care System.

[c]RITP: rapid influenza laboratory tests with positive results.

[d]IMU: influenza medication use.

Table 3. Pearson correlation coefficients between the Taiwan Center for Disease Control National influenza-like illness data and each surveillance variable from the electronic medical records of Taipei Medical University Health Care System (October 2014 to September 2017).

Time period	Regional ILI[a]	TMUHcS[b]-ILI	TMUHcS-RITP[c]	TMUHcS-IMU[d]
National ILI of the entire period (October 2014 to September 2017)	.99	.88	.90	.93
National ILI of period 1 (October 2014 to September 2015)	.99	.98	.87	.96
National ILI of period 2 (October 2015 to September 2016)	.99	.94	.95	.97
National ILI of period 3 (October 2016 to September 2017)	.99	.97	.88	.96

[a]ILI: influenza-like illness.

[b]TMUHcS: Taipei Medical University Health Care System.

[c]RITP: rapid influenza laboratory tests with positive results.

[d]IMU: influenza medication use.

Figure 3. Epidemic thresholds for each surveillance variable (October 2014 to September 2017; locally estimated scatterplot smoothing adjusted). TMUHcS: Taipei Medical University Health Care System; ILI: influenza-like illness; IMU: influenza medication use; RITP: rapid influenza laboratory tests with positive results.

Discussion

Principal Findings

The Integrated Influenza Surveillance Framework described in this study had the scalable ability to periodically integrate the TWCDC ILI open data with EMRs of multiple hospitals. The Surveillance Comparison Module could perform daily comparisons with the value of surveillance variables in this study. When the value of each surveillance variable reached or exceeded the epidemic thresholds, the Surveillance Messenger Push Engine proactively sent the emerging infectious disease surveillance alert to the subscribers (eg, health care facilities or epidemic prevention personnel). As our Surveillance Messenger Push Engine adopted the publish-subscribe pattern, the subscribers could receive the emerging infectious disease surveillance alert immediately through the mobile apps or CPOE systems; this novel approach can provide more efficient and instant infection warning compared with the Web application programming interface call or website of health care facilities or passive checking by the epidemic prevention personnel.

In this study, we identified 3 surveillance variables (TMUHcS-ILI, TMUHcS-RITP, and TMUHcS-IMU) from TMUHcS EMRs that showed a predictive value for influenza epidemics in Taiwan. Each TMUHcS surveillance variable was strongly correlated with regional and national ILI data. Compared with the TWCDC national and regional ILI populations, the TMUHcS population showed ILI case proportions of only 1.46% and 6.58%; however, the ILI case proportions for the national ILI, regional ILI, and TMUHcS-ILI were similar at 1.16%, 0.91%, and 1.01%, respectively. On the basis of the strong correlations and similar ILI case proportions between the 2 study datasets, the 10- to 70-fold smaller sample size of TMUHcS EMR data appeared to provide the same power as did the regional and national ILI data in supporting near-real-time influenza surveillance. These findings represent significant savings in the time and effort required for data collection and analysis.

There is a growing interest in exploring more timely and cost-effective surveillance approaches to enhance the early prediction of infectious disease epidemics [19,23,24,28,48,49]. Unlike previous studies that used only a single surveillance variable or a combination of algorithms to predict seasonal influenza epidemics, our approach was simple, easy to implement, and logically based on physician behaviors. Most importantly, the predictive power of the selected surveillance variables was evaluated using MEM. Of the 3 selected surveillance variables, TMUHcS-RITP and TMUHcS-IMU showed the best predictive power for influenza epidemics and predicted seasonal influenza epidemics 3 to 4 weeks ahead of the TWCDC ILI reports. Furthermore, the types of influenza vaccines recommended by WHO in 2018 might imperfectly match the circulating strains; they were approximately 30% effective against H3N2 viruses [50] but failed to prevent some epidemics. Therefore, the importance of early and accurate predictive ability of our approach is further highlighted for early epidemic control because it provided insights into the mismatches of the influenza virus types for national flu vaccination programs.

Taiwan is located in the subtropical regions of Southeast Asia and has several densely populated urban areas with well-established public transportation networks. There are frequent travelers from Taiwan to regions in which influenza pandemics have originated in the past, such as Hong Kong and mainland China [51]. Therefore, emerging infectious diseases can easily be transmitted from person to person in Taiwan, resulting in the rapid spread of an epidemic. Efficient and real-time monitoring tools for such diseases have become increasingly important. Our approach can be used for predicting influenza epidemics and can be easily extended to other infectious diseases by simply selecting different surveillance variables from EMRs. Accordingly, our approach can play a key role in public health and hospital preparedness, which is the top priority now more than ever.

The peaks of influenza epidemics appear to recur annually, as shown in our study. The influenza season in Taiwan commonly begins around October, peaks after the end of December, and ends in February. We closely examined the different time periods of our study datasets and noticed the following trends. During the 3-year study period, no influenza epidemic was observed during the first year; however, high-intensity influenza activity and an epidemic were observed during the second year, but no obvious influenza activity was noted during the third year (ie, October 2016 to February 2017). However, during the summer of 2017 (June-July), unprecedented levels of influenza activity were observed. Our medical team hypothesized that this delay in the occurrence of influenza epidemic resulted from the unusually warm winter from December 2016 to January 2017. The winter of 2016 to 2017 was the hottest on record, with an average monthly temperature that was 2°C higher than the historical average. This might indicate an association between warm winters and delayed occurrence of influenza epidemics.

Accumulating evidence has shown that EMR systems can facilitate better documentation, communication, and integration of health care data [52]. Accordingly, our results demonstrated that an integrated approach generated a simple, accurate, and timely electronic surveillance tool for predicting influenza epidemics, particularly in densely populated areas and smaller countries. With increased adoption of EMR systems, more meaningful surveillance variables and their corresponding EMRs can be collected and used to support various forms of emerging infectious disease surveillance in the future.

Conclusions

In this study, we developed an Integrated Influenza Surveillance Framework that had a scalable ability to periodically integrate TWCDC ILI open data with EMRs of multiple hospitals. The system allowed for the automatic monitoring of influenza activity, supporting the early surveillance of influenza epidemics. Using EMRs of multiple hospitals, we identified 3 selected surveillance variables for the early prediction of influenza epidemics. Each surveillance variable was strongly correlated with TWCDC regional and national ILI data. Most importantly, we observed that the predictive power of the 2

selected surveillance variables was 3 to 4 weeks advanced compared with that of the TWCDC ILI reports. Future research should involve further development of this accurate, timely, and cost-effective electronic influenza surveillance tool for the prediction of influenza epidemics as well as the extension of this approach to other infectious diseases in ways that can be particularly relevant to densely populated regions and nations.

Acknowledgments

The authors would like to thank all infection control team members of Taipei Medical University Hospital for their support. This research was supported in part by the Ministry of Science and Technology (MOST) of Taiwan (grant numbers MOST 105-2218-E-038 -004 and MOST 106-2221-E-038 -021).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Descriptive statistics for the outpatient visits associated with influenza-like illness from October 2014 to September 2017.

[DOCX File, 25KB - jmir_v21i2e12341_app1.docx]

References

1. World Health Organization. Battle against Respiratory Viruses (BRaVe) initiative. 2017. URL: http://www.who.int/influenza/patient_care/clinical/brave/en/ [WebCite Cache ID 74fvoYBix]
2. Ortiz JR, Neuzil KM, Shay DK, Rue TC, Neradilek MB, Zhou H, et al. The burden of influenza-associated critical illness hospitalizations. Crit Care Med 2014 Nov;42(11):2325-2332 [FREE Full text] [doi: 10.1097/CCM.0000000000000545] [Medline: 25148596]
3. World Health Organization. Influenza (Seasonal). 2018. URL: https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal) [accessed 2018-12-15] [WebCite Cache ID 74fw6Gv3Z]
4. Kochanek K, Murphy S, Xu J, Arias E. Mortality in the United States, 2013. NCHS Data Brief 2014 Dec(178):1-8 [FREE Full text] [Medline: 25549183]
5. Molinari NA, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, et al. The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine 2007 Jun 28;25(27):5086-5096. [doi: 10.1016/j.vaccine.2007.03.046] [Medline: 17544181]
6. Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, Pavlin JA. Code-based syndromic surveillance for influenzalike illness by International Classification of Diseases, Ninth Revision. Emerg Infect Dis 2007 Feb;13(2):207-216 [FREE Full text] [doi: 10.3201/eid1302.060557] [Medline: 17479881]
7. Brammer L, Budd A, Cox N. Seasonal and pandemic influenza surveillance considerations for constructing multicomponent systems. Influenza Other Respir Viruses 2009;3(2):51-58. [doi: 10.1111/j.1750-2659.2009.00077.x] [Medline: 19496841]
8. Centers for Disease Control and Prevention (CDC). Assessment of ESSENCE performance for influenza-like illness surveillance after an influenza outbreak--US Air Force Academy, Colorado, 2009. MMWR Morb Mortal Wkly Rep 2011 Apr 8;60(13):406-409 [FREE Full text] [Medline: 21471947]
9. Daughton AR, Velappan N, Abeyta E, Priedhorsky R, Deshpande A. Novel use of flu surveillance data: evaluating potential of sentinel populations for early detection of influenza outbreaks. PLoS One 2016;11(7):e0158330 [FREE Full text] [doi: 10.1371/journal.pone.0158330] [Medline: 27391232]
10. Baltrusaitis K, Brownstein J, Scarpino S, Bakota E, Crawley A, Conidi G, et al. Comparison of crowd-sourced, electronic health records based, and traditional health-care based influenza-tracking systems at multiple spatial resolutions in the United States of America. BMC Infect Dis 2018 Aug 15;18(1):403 [FREE Full text] [doi: 10.1186/s12879-018-3322-3] [Medline: 30111305]
11. Nakamura Y, Sugawara T, Kawanohara H, Ohkusa Y, Kamei M, Oishi K. Evaluation of estimated number of influenza patients from national sentinel surveillance using the national database of electronic medical claims. Jpn J Infect Dis 2015;68(1):27-29 [FREE Full text] [doi: 10.7883/yoken.JJID.2014.092] [Medline: 25420664]
12. Patwardhan A, Bilkovski R. Comparison: Flu prescription sales data from a retail pharmacy in the US with Google Flu trends and US ILINet (CDC) data as flu activity indicator. PLoS One 2012;7(8):e43611 [FREE Full text] [doi: 10.1371/journal.pone.0043611] [Medline: 22952719]
13. Egger JR, Hoen AG, Brownstein JS, Buckeridge DL, Olson DR, Konty KJ. Usefulness of school absenteeism data for predicting influenza outbreaks, United States. Emerg Infect Dis 2012 Aug;18(8):1375-1377 [FREE Full text] [doi: 10.3201/eid1808.111538] [Medline: 22840354]

14. Sugawara T, Ohkusa Y, Ibuka Y, Kawanohara H, Taniguchi K, Okabe N. Real-time prescription surveillance and its application to monitoring seasonal influenza activity in Japan. J Med Internet Res 2012 Jan 16;14(1):e14 [FREE Full text] [doi: 10.2196/jmir.1881] [Medline: 22249906]
15. Patterson-Lomba O, van Noort S, Cowling B, Wallinga J, Gomes MG, Lipsitch M, et al. Utilizing syndromic surveillance data for estimating levels of influenza circulation. Am J Epidemiol 2014 Jun 1;179(11):1394-1401 [FREE Full text] [doi: 10.1093/aje/kwu061] [Medline: 24748609]
16. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
17. Simonsen L, Gog J, Olson D, Viboud C. Infectious disease surveillance in the big data era: towards faster and locally relevant systems. J Infect Dis 2016 Dec 1;214(suppl_4):S380-S385 [FREE Full text] [doi: 10.1093/infdis/jiw376] [Medline: 28830112]
18. Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: a comparative analysis. JMIR Public Health Surveill 2016 Oct 20;2(2):e161 [FREE Full text] [doi: 10.2196/publichealth.5901] [Medline: 27765731]
19. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. PLoS One 2013;8(5):e64323 [FREE Full text] [doi: 10.1371/journal.pone.0064323] [Medline: 23750192]
20. 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]
21. Cho S, Sohn CH, Jo MW, Shin SY, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. PLoS One 2013;8(12):e81422 [FREE Full text] [doi: 10.1371/journal.pone.0081422] [Medline: 24339927]
22. Ortiz JR, Zhou H, Shay DK, Neuzil KM, Fowlkes AL, Goss CH. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. PLoS One 2011 Apr 27;6(4):e18687 [FREE Full text] [doi: 10.1371/journal.pone.0018687] [Medline: 21556151]
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. Martin LJ, Lee B, Yasui YJ. Google Flu Trends in Canada: a comparison of digital disease surveillance data with physician consultations and respiratory virus surveillance data, 2010-2014. Epidemiol Infect 2016 Jan;144(2):325-332. [doi: 10.1017/S0950268815001478] [Medline: 26135239]
25. Milinovich GJ, Williams GM, Clements AC, Hu WJ. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168. [doi: 10.1016/S1473-3099(13)70244-5] [Medline: 24290841]
26. Ghosh TS, Vogt RJ. Active influenza surveillance at the local level: a model for local health agencies. Am J Public Health 2008 Feb;98(2):213-215. [doi: 10.2105/AJPH.2006.106138] [Medline: 18172149]
27. Shimoni Z, Rodrig J, Dusseldorp N, Niven M, Froom PJ. Increased emergency department chief complaints of fever identified the influenza (H1N1) pandemic before outpatient symptom surveillance. Environ Health Prev Med 2012 Jan;17(1):69-72 [FREE Full text] [doi: 10.1007/s12199-011-0213-2] [Medline: 21448581]
28. Santillana M, Nguyen A, Louie T, Zink A, Gray J, Sung I, et al. Cloud-based electronic health records for real-time, region-specific influenza surveillance. Sci Rep 2016 Dec 11;6:25732 [FREE Full text] [doi: 10.1038/srep25732] [Medline: 27165494]
29. Escuyer KL, Waters CL, Gowie DL, Maxted AM, Farrell GM, Fuschino ME, et al. The assessment of data sources for influenza virologic surveillance in New York State. Influenza Other Respir Viruses 2017 Dec;11(2):138-147 [FREE Full text] [doi: 10.1111/irv.12433] [Medline: 27718314]
30. Grohskopf LA, Olsen SJ, Sokolow LZ, Bresee JS, Cox NJ, Broder KR, Centers for Disease Control and Prevention. Prevention and control of seasonal influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices (ACIP)--United States, 2014-15 influenza season. MMWR Morb Mortal Wkly Rep 2014 Aug 15;63(32):691-697 [FREE Full text] [doi: 10.15585/mmwr.rr6703a1] [Medline: 25121712]
31. Caini S, Alonso WJ, Balmaseda A, Bruno A, Bustos P, Castillo L, Global Influenza B Study group–Latin America. Characteristics of seasonal influenza A and B in Latin America: influenza surveillance data from ten countries. PLoS One 2017;12(3):e0174592 [FREE Full text] [doi: 10.1371/journal.pone.0174592] [Medline: 28346498]
32. Taiwan Centers for Disease Control. Taiwan Centers for Disease Control. 2018. Taiwan National Infectious Disease Statistic System URL: http://nidss.cdc.gov.tw/en/ [WebCite Cache ID 74fwDxqmR]
33. Wu T, Majeed A, Kuo KN. An overview of the healthcare system in Taiwan. London J Prim Care (Abingdon) 2010 Dec;3(2):115-119 [FREE Full text] [doi: 10.1080/17571472.2010.11493315] [Medline: 25949636]
34. Yang CY, Lo YS, Chen RJ, Liu CT. A clinical decision support engine based on a national medication repository for the detection of potential duplicate medications: design and evaluation. JMIR Med Inform 2018 Jan 19;6(1):e6 [FREE Full text] [doi: 10.2196/medinform.9064] [Medline: 29351893]
35. Centers for Disease Control and Prevention. Rapid Diagnostic Testing for Influenza: Information for Clinical Laboratory Directors. 2018. URL: https://www.cdc.gov/flu/professionals/diagnosis/rapidlab.htm [WebCite Cache ID 74fwNlG5N]

36. Lo YS, Lee WS, Liu CT. Utilization of electronic medical records to build a detection model for surveillance of healthcare-associated urinary tract infections. J Med Syst 2013 Apr;37(2):9923. [doi: 10.1007/s10916-012-9923-2] [Medline: 23321977]
37. SAS Institute. Developing Java Clients. 2018. URL: http://support.sas.com/rnd/itech/doc/dist-obj/javaclnt/index.html, [accessed 2019-01-28] [WebCite Cache ID 74fwVSV6U]
38. Urbanek S. The Comprehensive R Archive Network. 2018. rJava: Low-Level R to Java Interface URL: https://cran.r-project.org/web/packages/rJava/index.html [WebCite Cache ID 74fwmzWm4]
39. Amazon. Amazon CloudWatch User Guide. 2017. URL: https://www.amazon.com/Amazon-CloudWatch-User-Guide-Services-ebook/dp/B007S1I894 [accessed 2019-01-29] [WebCite Cache ID 75mwX5TtS]
40. Amazon. Amazon Simple Notification Service. URL: https://aws.amazon.com/sns/ [accessed 2018-12-15] [WebCite Cache ID 74fx1rSI1]
41. Vega T, Lozano JE, Meerhoff T, Snacken R, Mott J, de Lejarazu RO, et al. Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method. Influenza Other Respir Viruses 2013 Jul;7(4):546-558 [FREE Full text] [doi: 10.1111/j.1750-2659.2012.00422.x] [Medline: 22897919]
42. Vega T, Lozano JE, Meerhoff T, Snacken R, Beauté J, Jorgensen P, et al. Influenza surveillance in Europe: comparing intensity levels calculated using the moving epidemic method. Influenza Other Respir Viruses 2015 Sep;9(5):234-246 [FREE Full text] [doi: 10.1111/irv.12330] [Medline: 26031655]
43. Drumright L, Frost SD, Elliot AJ, Catchpole M, Pebody R, Atkins M, et al. Assessing the use of hospital staff influenza-like absence (ILA) for enhancing hospital preparedness and national surveillance. BMC Infect Dis 2015 Mar 1;15:110 [FREE Full text] [doi: 10.1186/s12879-015-0789-z] [Medline: 25886745]
44. Chapman WW, Dowling JN, Wagner MM. Generating a reliable reference standard set for syndromic case classification. J Am Med Inform Assoc 2005;12(6):618-629 [FREE Full text] [doi: 10.1197/jamia.M1841] [Medline: 16049227]
45. Cleveland WS, Devlin SJ, Grosse E. Regression by local fitting: methods, properties, and computational algorithms. J Econom 1988 Jan;37(1):87-114. [doi: 10.1016/0304-4076(88)90077-2]
46. R Foundation. The R Project for Statistical Computing. 2018. URL: https://www.r-project.org/ [WebCite Cache ID 74fxFOWWm]
47. Lozano J. Zenodo. lozalojo/mem: Second release of the MEM R library (Version v2.0) URL: https://zenodo.org/record/165983 [accessed 2018-12-15] [WebCite Cache ID 74fxMQ95O]
48. Deiner M, Lietman T, McLeod S, Chodosh J, Porco TJ. Surveillance tools emerging from search engines and social media data for determining eye disease patterns. JAMA Ophthalmol 2016 Sep 1;134(9):1024-1030 [FREE Full text] [doi: 10.1001/jamaophthalmol.2016.2267] [Medline: 27416554]
49. Scarpino SV, Meyers LA, Johansson MA. Design strategies for efficient arbovirus surveillance. Emerg Infect Dis 2017 Dec;23(4):642-644 [FREE Full text] [doi: 10.3201/eid2304.160944] [Medline: 28322711]
50. Fisher BL. National Vaccine Information Center. 2016. CDC admits flu shots fail half the time URL: https://www.nvic.org/nvic-vaccine-news/april-2016/cdc-admits-flu-shots-fail-half-the-time.aspx [accessed 2019-01-31] [WebCite Cache ID 75ppfFL6v]
51. de Jong JC, Claas EC, Osterhaus AD, Webster RG, Lim WL. A pandemic warning? Nature 1997;389(6651):554 [FREE Full text]
52. Atreja A, Gordon SM, Pollock DA, Olmsted RN, Brennan PJ, Healthcare Infection Control Practices Advisory Committee. Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control. Am J Infect Control 2008 Apr;36(3 Suppl):S37-S46. [doi: 10.1016/j.ajic.2008.01.002] [Medline: 18374211]

Abbreviations

CDC: Centers for Disease Control and Prevention
CPOE: computerized physician order entry
ED: emergency department
EMR: electronic medical record
ILI: influenza-like illness
IMU: influenza medication use
MEM: Moving Epidemic Method
NHI: National Health Insurance
OPD: outpatient department
RIDT: rapid influenza diagnostic test
RITP: rapid influenza laboratory tests with positive results
RT-PCR: reverse transcription polymerase chain reaction
TMUHcS: Taipei Medical University Health Care System
TWCDC: Taiwan Centers for Disease Control and Prevention
WHO: World Health Organization

Edited by G Eysenbach; submitted 29.09.18; peer-reviewed by YJ Tseng; comments to author 25.10.18; revised version received 18.12.18; accepted 20.01.19; published 01.02.19.

Please cite as:
Yang CY, Chen RJ, Chou WL, Lee YJ, Lo YS
An Integrated Influenza Surveillance Framework Based on National Influenza-Like Illness Incidence and Multiple Hospital Electronic Medical Records for Early Prediction of Influenza Epidemics: Design and Evaluation
J Med Internet Res 2019;21(2):e12341
URL: http://www.jmir.org/2019/2/e12341/
doi:10.2196/12341
PMID:30707099

Original Paper

Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model

Chi-Shiang Wang[1], MS; Pei-Ju Lin[1], MS; Ching-Lan Cheng[2,3,4], PhD; Shu-Hua Tai[4], MS; Yea-Huei Kao Yang[2,3], BSPharm; Jung-Hsien Chiang[1,5], PhD

[1]Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan

[2]School of Pharmacy, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[3]Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[4]Department of Pharmacy, National Cheng Kung University Hospital, National Cheng Kung University, Tainan, Taiwan

[5]Institute of Medical Informatics, National Cheng Kung University, Tainan, Taiwan

Corresponding Author:
Jung-Hsien Chiang, PhD
Department of Computer Science and Information Engineering
National Cheng Kung University
1 University Road
Tainan, 701
Taiwan
Phone: 886 6 2757575 ext 62534
Email: jchiang@mail.ncku.edu.tw

Abstract

Background: Adverse drug reactions (ADRs) are common and are the underlying cause of over a million serious injuries and deaths each year. The most familiar method to detect ADRs is relying on spontaneous reports. Unfortunately, the low reporting rate of spontaneous reports is a serious limitation of pharmacovigilance.

Objective: The objective of this study was to identify a method to detect potential ADRs of drugs automatically using a deep neural network (DNN).

Methods: We designed a DNN model that utilizes the chemical, biological, and biomedical information of drugs to detect ADRs. This model aimed to fulfill two main purposes: identifying the potential ADRs of drugs and predicting the possible ADRs of a new drug. For improving the detection performance, we distributed representations of the target drugs in a vector space to capture the drug relationships using the word-embedding approach to process substantial biomedical literature. Moreover, we built a mapping function to address new drugs that do not appear in the dataset.

Results: Using the drug information and the ADRs reported up to 2009, we predicted the ADRs of drugs recorded up to 2012. There were 746 drugs and 232 new drugs, which were only recorded in 2012 with 1325 ADRs. The experimental results showed that the overall performance of our model with mean average precision at top-10 achieved is 0.523 and the rea under the receiver operating characteristic curve (AUC) score achieved is 0.844 for ADR prediction on the dataset.

Conclusions: Our model is effective in identifying the potential ADRs of a drug and the possible ADRs of a new drug. Most importantly, it can detect potential ADRs irrespective of whether they have been reported in the past.

(J Med Internet Res 2019;21(2):e11016) doi:10.2196/11016

KEYWORDS

adverse drug reactions; deep neural network; drug representation; machine learning; pharmacovigilance

Introduction

An adverse drug reaction (ADR) [1,2] is a serious problem that refers to side effects that occur despite the administration of a regular dose of a drug. It is estimated that over 2 million serious ADRs occur among hospitalized patients, which causes >100,000 deaths each year [3,4]. Unfortunately, it is difficult to identify or predict potential ADRs owing to insufficient data.

Spontaneous reporting in pre- and postmarket stages are the most familiar methods to identify ADRs early on. Specifically, safety reports from clinical trials are used to list common ADRs

in the premarket stage [5], while data collected and analyzed from various databases and marketing surveys, such as pharmacovigilance and risk management, are used in the postmarket stage. Although most new ADRs are identified from spontaneous reports, >90% go unreported [6,7]; this is recognized to be a major limitation. Pharmaceutical companies are trying to avoid side effects in the development stage of drugs. However, although they can identify and address common side effects, it is generally not feasible to identify or predict rare and serious side effects. To overcome these limitations, several studies have utilized a substantial amount of data and various information sources to predict ADRs using statistical methods [8] and machine learning approaches [9].

The fundamental method for identifying ADRs pertains to identifying the relationship between drugs and their side effects from diverse sources of information [10-12] such as clinical trials, electronic medical records (EMRs), social media, and biomedical literature. For instance, PubMed contains valuable information that could aid ADR detection. Karimi et al [13] reviewed data mining and techniques related to computer science, which have been studied in the area of drug safety to identify reports of ADR from different sources. Tatonetti et al [14] proposed a novel algorithm for building a predictive model that can detect hidden interactions in adverse event reports to infer unreported adverse events. Wang et al [15] developed a model for identifying ADRs using data mining to extract information from millions of EMRs. It used clinical notes with information on specific drugs and known adverse drug events (ADEs) that have been preprocessed using statistical methods to compute the probability that a given drug-disorder pair represents a valid ADE association. This method automatically determines whether a specific adverse event is caused by a specific drug based on the content of PubMed citations [16]. Finkelstein et al [17] developed a tool to automatically detect and summarize information on ADRs from journal papers. It then ranked the ADRs of a drug on a user-friendly interface for physicians.

Several studies have utilized either chemical or molecular pathways of drugs to predict ADRs [18]. Cami et al [19] developed a novel approach to predict ADEs by using information on specific drugs and the adverse event to predict likely unknown ADEs. Lorberbaum et al [20] hypothesized that systems biology and chemical genomics data can improve drug safety surveillance by highlighting drugs with a mechanistic connection to the target phenotype and by filtering those which do not. They presented an algorithm, the modular assembly of drug safety subnetworks, to combine systems pharmacology and pharmacovigilance data. The algorithm markedly improved drug safety monitoring for 4 clinically relevant ADRs. Huang et al [21] proposed a framework for predicting ADR profiles by integrating protein-protein interaction networks with drug structures. Some researchers utilized the chemical, biological, and phenotypic characteristics of drugs to predict the ADRs. Liu et al [22] proposed a machine learning approach for predicting the ADRs by integrating the phenotypic characteristics, which included chemical structure, biological properties, and protein target and pathway information.

However, most of those approaches rely on heavily handcrafted features and treat ADR identification as a classification problem, which does not take the order of the ADRs discovered into consideration. Therefore, the process tends to be more expensive and leads to the loss of significant information on drug-ADR relationships in the model training phase. Furthermore, these approaches are unable to predict the ADR of new drugs, thus rendering the detection of ADR more difficult [19].

To address these limitations, we used a deep neural network (DNN) model for the detection of ADRs of drugs. The model has 2 purposes: the identification of ADRs, which entailed the discovery of potential ADRs of a drug from known ADR records, and the prediction of ADRs, which pertained to predicting the possible ADRs for a new drug. We used the word-embedding approach and mapping function to process new drugs that did not appear in the dataset. Furthermore, we examined the overall performance of the model with various feature combinations and the number of hidden layers in the DNN architecture.

Methods

Data Description

To develop and evaluate a DNN model, we used data from Side Effect Resource (SIDER) [23], a database of drugs with side effects, which contains information on medicines in the market since 2009 and their recorded ADRs [24]. We collected the ADR information from 2009 and 2012 from SIDER to represent the simulated prospective approach. In total, 746 drugs and 1325 side effect terms related to these drugs were recorded in both years. Additional 232 drugs appeared only in the 2012 dataset as new drugs. It is important to monitor the ADRs of drugs throughout their life cycle, from the preclinical research phase to postmarket surveillance. The fundamental properties of drugs rely on preclinical *in vitro* safety profiling that involves the testing of compounds with chemical and biological properties. Therefore, we extracted these properties as a part of the features in the model. We extracted the 17 molecular descriptors of drugs from PubChem [25] (Textbox 1). We utilized the biological features from DrugBank [26] to represent the biomolecular interactions and pathways. These features contain the targets, enzymes, transporters, and carriers of each drug and their actions.

For enriching the scientific evidence and enhancing the detection of ADRs, we collected millions of papers from the Medical Literature Analysis and Retrieval System Online (MEDLINE) [27] to be used as auxiliary data to enrich the information about each drug. We used the name of each drug as the query term and selected all the papers related to the drug, published before 2009, such as case reports, clinical trials, and observational studies. The reason for collecting the papers published before 2009 is that we wanted to simulate the progress of drug surveillance from 2009 to 2012.

Textbox 1. The 17 molecular descriptors for the chemical features identified in this study.

1. Molecular Weight
2. XLogP3
3. Hydrogen Bond Donor Count
4. Hydrogen Bond Acceptor Count
5. Rotatable Bond Count
6. Exact Mass
7. Monoisotopic Mass
8. Topological Polar Surface Area
9. Heavy Atom Count
10. Formal Charge
11. Complexity
12. Isotope Atom Count
13. Defined Atom Stereocenter Count
14. Undefined Atom Stereocenter Count
15. Defined Bond Stereocenter Count
16. Undefined Bond Stereocenter Count
17. Covalently- Bonded Unit Count

Features of Drug Description

We treated ADR identification as an information retrieval problem, such that our model could discover the potential relationships between each drug and the 1325 side effects recorded. We represented the prediction target of 1325 dimensions with a binary profile of elements corresponding to the presence or absence of side effects with 1 or 0, $Y \in {}^{n\times1325}$ with n being the number of drugs. Each drug was associated with 3 types of features: the chemical properties, biological properties, and information from the literature. In addition, the known ADR records of drugs were included. After preprocessing, we filtered out 2 empty properties: the Isotope Atom Count and the Undefined Bond Stereocenter Count. Subsequently, the feature of chemical properties was represented using a 15-dimensional vector, with $X_{\text{Chem}} \in {}^{n\times15}$ for each element. The biological properties, extracted from DrugBank, contained 4 phases of information including the carriers, enzymes (for drug metabolism), protein targets, and transporters (for drug transportation). After preprocessing, we utilized the biological information to represent each drug with the 1048-dimensional vector, which included 788 protein targets, 162 enzymes, 85 transporters, and 13 carriers, with $X_{\text{Bio}} \in {}^{n\times1048}$ for different action types in each element. The known ADR records of a drug played an important role in identifying potential ADRs. Thus, we leveraged this information to predict potential ADRs that appeared in 2012.

The biomedical literature played an important role in this study because it contains a large amount of information related to drugs and ADRs such as clinical notes and case reports. However, one of the issues in extracting the drug information from biomedical literature is the uncertainty regarding which words or documents represent the drug. Therefore, we trained the model to understand the semantic features of drugs from 2.3 million biomedical papers on 764 drugs introduced before 2009 by utilizing one of the most popular embedding methods Word2Vec [28-30] to model it using the skip-gram model. Subsequently, we used the vector of the drug name as the drug vector (drug2vec, D2V), $X_{\text{D2V}} \in {}^{n\times400}$, with n being the number of drugs. We observed that the features represented by D2V were more comprehensive than the intrinsic features in the experiments. However, as we used drug names as the query term, we could not identify papers related to 232 new drugs. To address drugs that were not observed during the embedding training step, we expanded the D2V by introducing a drug description mapping function. Using V_{D2V} to denote the word-embedding space of the drug vector existing in the training and V_{DDV} to denote the new drug description vector the summation of each word vector related to the new drug in the papers, the mapping function $f(v)$: $V_{\text{DDV}} \rightarrow V_{\text{D2V}}$ was developed and parameterized by a W, such that $v' = Wv$ for $v \in V_{\text{DDV}}$ and $v' \in V_{\text{D2V}}$ [31] using the least absolute shrinkage and selection operator regression for training the W [32]. This expansion method enabled the model to process new drugs, making it more flexible.

Figure 1. The architecture of the deep neural network model for predicting and identifying the possible adverse drug reactions (ADRs) of a drug. After predicting, we generated a list of possible ADRs of a drug by ranking the probability of ADRs from the output in the model.

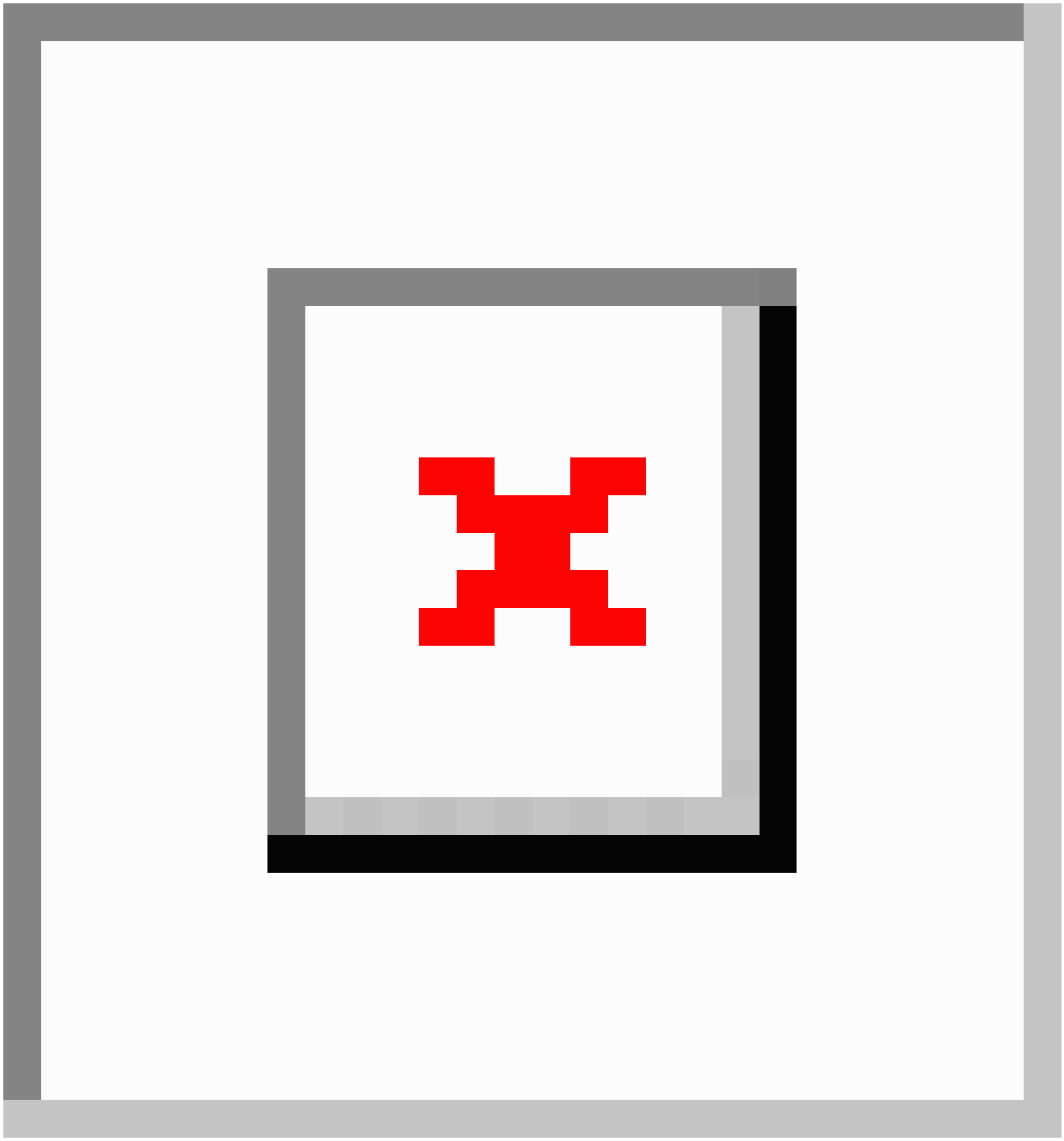
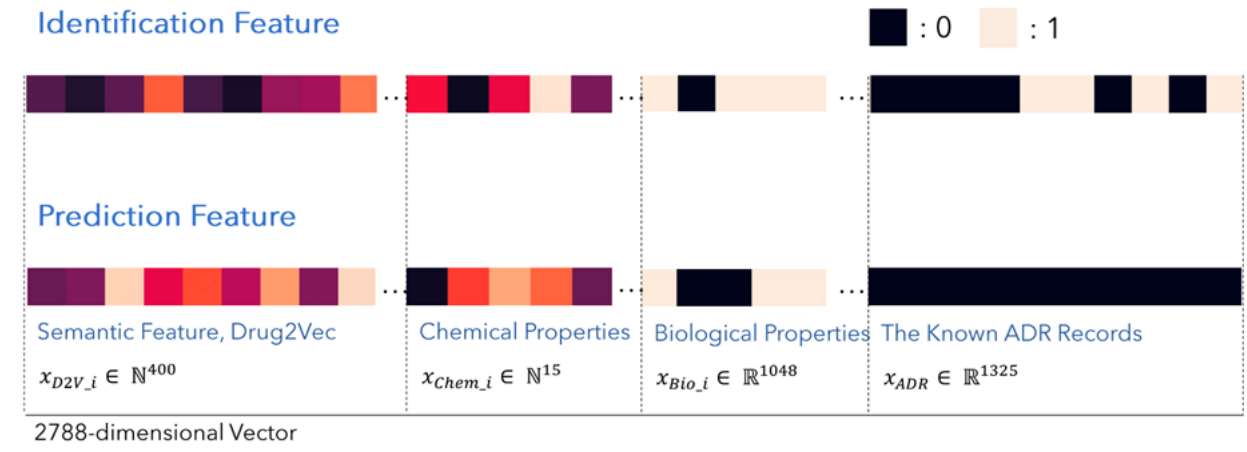


Figure 2. Feature representation of adverse drug reaction (ADR) identification and prediction.



Adverse Drug Reaction Detection Deep Neural Network Model Description

We designed a DNN model that can identify and predict the ADRs of a drug with different requirements. This model (Figure 1) was based on one of the common DNN architectures, multilayer perceptron [33], which has been successfully applied in several prior studies. We added the dropout layer between each dense layer to avoid model overfitting in the training step [34,35]. The process of the nonlinear transformation in each layer ensured that the model could learn more information from the input data. The model aimed at identifying and predicting potential ADRs of a drug. The identification function is fulfilled by seeking the potential ADRs of drugs by using known records of ADRs and various features of drugs. The prediction function pertains to the detection of potential ADRs of new drugs. We designed 2 kinds of feature representations (Figure 2) to distinguish the tasks of identification and prediction. In prediction, we assumed that the ADR record for a new drug was empty. Therefore, the feature of known ADR records with zero indicated that the possible ADRs of new drugs relied only on the semantic feature (drug2vec), chemical properties, and biological properties.

We treated the identification task as an information retrieval problem because drugs may have more than one ADR. Therefore, we designed the last layer with 1325 hidden nodes, which was equal to the number of ADRs in the dataset. Evaluating the probability of ADR y of a given drug x, we defined the $p(y \mid x) = \sigma(Wh + b)$, σ as the sigmoid function, to transfer the hidden vector h to the value between 0 and 1. We learned the $p(y \mid x)$ by minimizing the cross-entropy with D. When θ denoted all parameters of the model, the objective function (D, θ) was formulated as follows:

$$(D, \theta) = -Summation\ (y_i\ log\ [p\ (y|x)] + [1-y_i]\ log\ [1-p\ (y|x]) / N$$

Results

In this study, we present a detailed analysis of the performance of our DNN model. Let Q denote the number of drugs in the dataset. We evaluated the model using the area under the receiver operating characteristic curve, shown in Table 1, and the mean average precision (MAP) *AvePrecision* $(q)/Q$, which has been widely used in multilabel problem and information retrieval evaluation. First, we assessed the abilities of different feature combinations to detect the ADRs of drugs. We examined the performance of our model with reference to the features presented in Figure 3 (image on the left). The drug features included the biological, chemical, and D2V features. The chemical feature was found to perform poorly because of the duplicate and indistinguishable chemical properties extracted.

Moreover, we removed the D2V and kept the other features to train the model. The results showed that the D2V was most informative, possibly because the D2V learned the valuable information from millions of papers. We then focused on method comparison with several common methods. We compared the abilities of 3 machine learning methods, namely, probability matrix factorization (PMF), Linear Support Vector Classifier, and Gaussian Naïve Bayes [36,37], to predict and identify the ADRs of drugs. Figure 3 (image on the right) shows the performance of different models based on 5-fold cross-validation using all biological, chemical, and D2V properties as features, except PMF. The PMF exhibited the worst performance because it considered the relationship between drugs and ADRs only based on latent information. One of the reasons why our model outperforms others is that the features of a drug enrich the information via the nonlinear transformation in deep learning.

Subsequently, we investigated whether our model could process the specific tasks of prediction and identification. The performance on the prediction task (Figure 4, image on the left) exceeded that on the identification task. The identification task is more difficult than prediction because the former entails the detection of potentially rare ADRs. Although the identification function of this model could be improved, its overall performance revealed its capacity to address both tasks simultaneously.

In addition, we plotted the performance of the model with a different number of hidden layers (Figure 4, image on the right). The performance of the model did not improve with an increase in the number of hidden layers. The model with 2 hidden layers was better than the others. Specifically, with the limited data size for 3 hidden layers, the model was unable to learn the good

parameters from the data. Evidently, the number of hidden layers relies on data properties and the amount of data in the DNN.

To evaluate our mapping function, we examined the drug expansion through the transfer of drug description to the D2V. The results, shown in Table 2, indicated that our mapping function could fit the performance of the D2V. The performance of the drug description through the mapping function was slightly better than that of D2V, possibly because some drugs did not exist in the space of D2V. This finding indicates that this model cannot predict the ADRs of such drugs without the mapping function. Accordingly, the mapping function was found to render the model more flexible to address new drugs.

Table 1. The result showing the performance of model evaluated by area under the receiver operating characteristic curve (AUC).

Model	AUC
Probability matrix factorization	0.500
Linear Support Vector Classifier	0.523
Gaussian Naïve Bayes	0.597
Deep neural network adverse drug reaction (DNN ADR) without hidden layer	0.641
DNN ADR with 1 hidden layer	0.823
DNN ADR with 2 hidden layers	*0.844*[a]
DNN ADR with 3 hidden layers	0.814
DNN ADR without Bio features	0.823
DNN ADR without Chem features	0.837
DNN ADR without drug2vec features	0.803
DNN ADR	*0.844*

[a]The italicized values indicate the best results in this comparison.

Figure 3. Left: Effects of different feature combinations to detect the adverse drug reactions (ADRs) of drugs; right: A comparison of our deep neural network (DNN) model with various machine learning approaches. PMF: probability matrix factorization; LinearSVC: Linear Support Vector Classifier; GaussianNB: Gaussian Naïve Bayes.

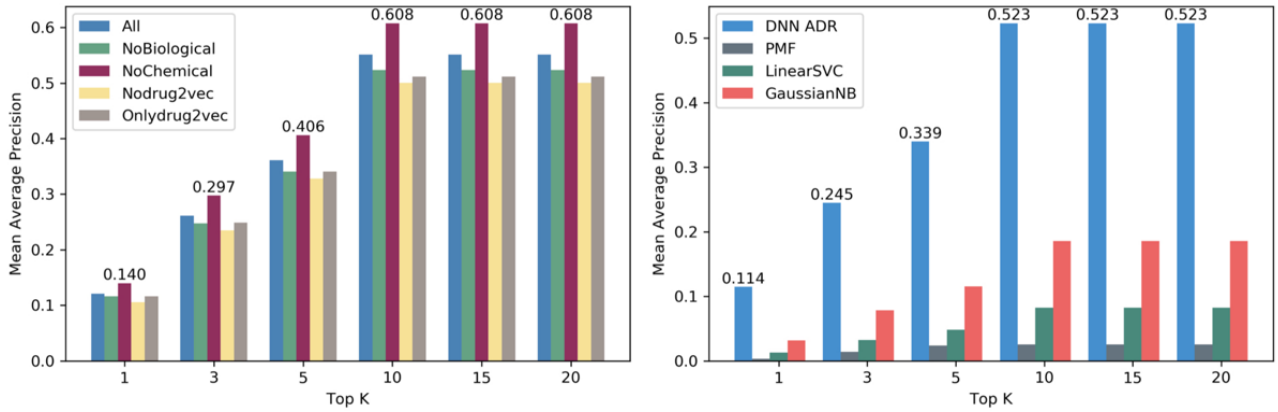


Figure 4. Left: Performance of the deep neural network (DNN) model on the adverse drug reaction (ADR) identification and prediction tasks and the overall performance; right: In this experiment, we showed the performance of the model with several different layers. GaussianNB: Gaussian Naïve Bayes; LinearSVC: Linear Support Vector Classifier.

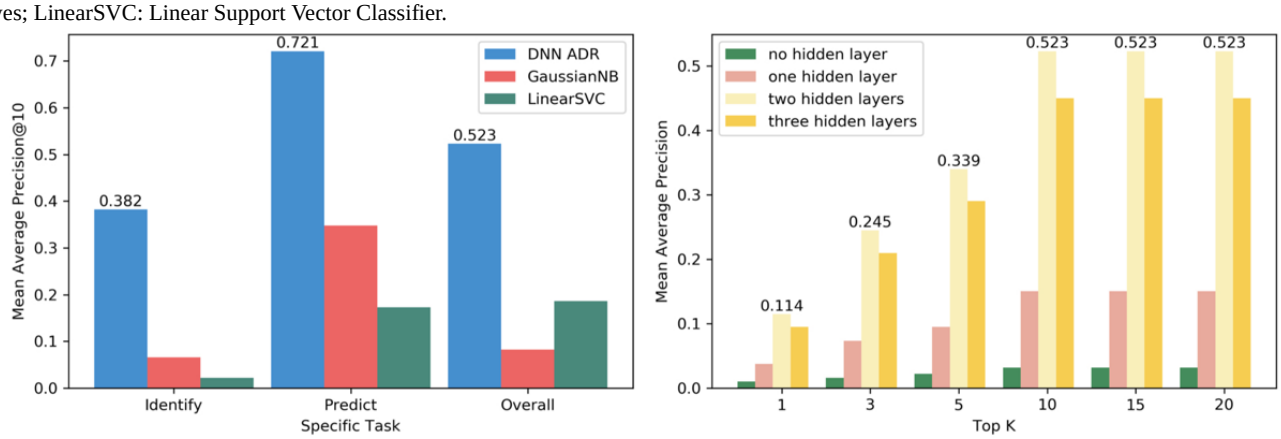


Table 2. The results showing the ability of the mapping function to transfer the drug description to drug2vec with Mean Average Precision at Top N (MAP@N).

MAP@N	1	3	5	10	15	20
Mapping function	*0.068*[a]	*0.179*	*0.272*	*0.462*	*0.462*	*0.462*
drug2vec	0.065	0.174	0.267	0.453	0.453	0.453

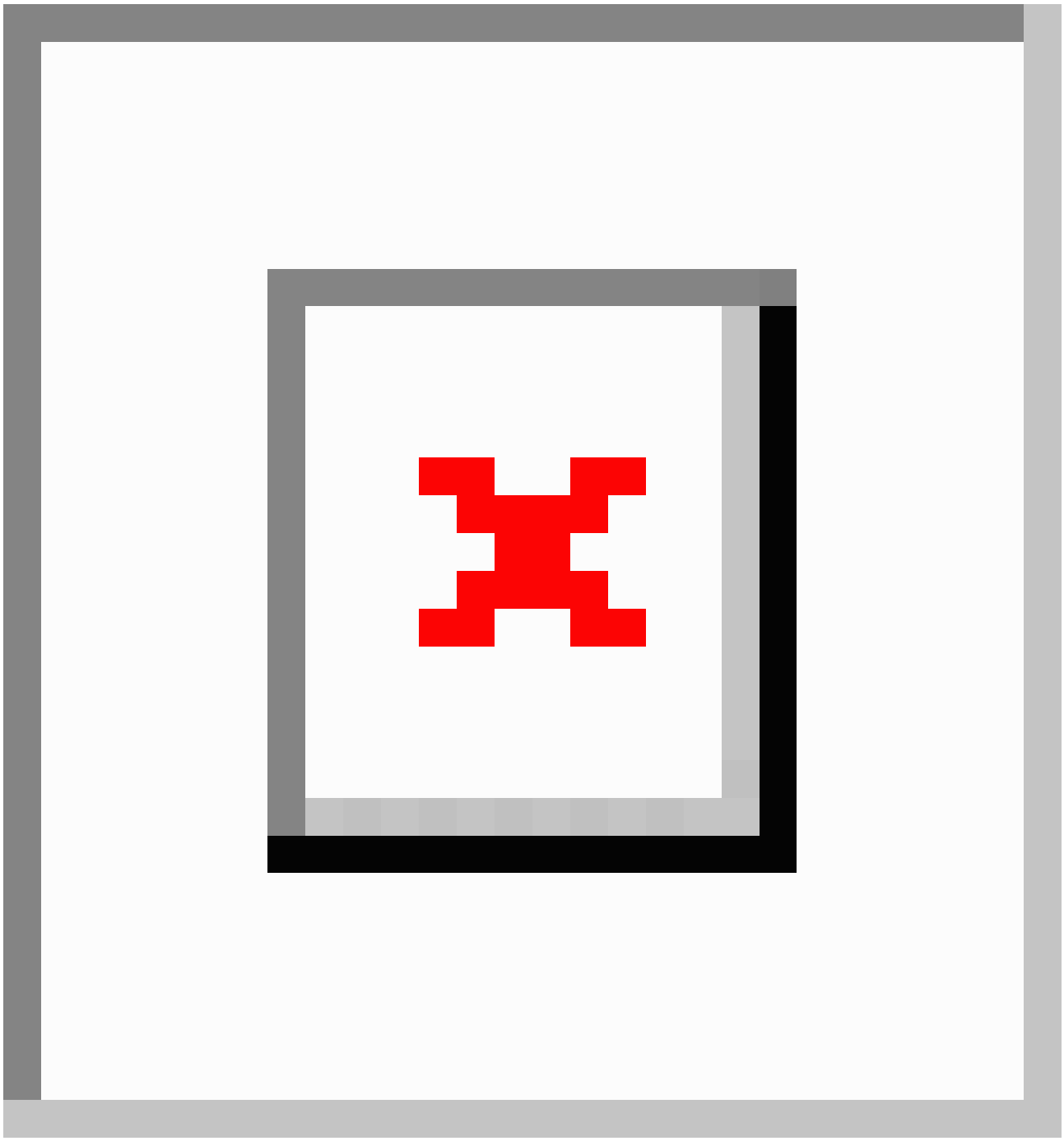
[a]The italicized values indicate the best results in this comparison.

Discussion

Principal Findings

In this study, we aimed to increase the diversity of information on drugs to improve our ability to detect ADRs. Accordingly, we extracted information from the chemical and biological properties of drugs and from the existing biomedical literature. The MEDLINE was selected as the source for biomedical literature to identify important auxiliary data because it contains several types of biomedical papers, such as clinical trials, case reports, and observational studies, related to drugs. However, it was difficult to use keywords to identify specific drugs from millions of papers and words. Therefore, we utilized 2.3 million biomedical papers to identify the semantic features of drug using the skip-gram model in Word2Vec. In particular, for a central word w_t under consideration, the probability $p\ (w_{t\pm i} \mid w_t)$ of predicting the surrounding word $w_{t\pm i}$ depended on the w_t. The subscript t indicates a target word, such as the drug name "Dantrolene," and the i represented the windows size of the target word. This feature helped us extract the latent information of each word, including the words comprising the drug's name. After this word-embedding training, we investigated whether the semantic feature (D2V) could represent the properties of drugs, such as biomedical and chemical properties. We visualized the relationships between the learned drug-embedding vectors based on their similarity (Figure 5). This graph included 746 drugs and presented the relationship of these drugs with those of the latent information learned from the semantic feature. Each node presents a drug and the edge represents the similarity between other drugs. The more similar the drugs in each pair were, the closer they were in the graph. The larger nodes represented the drugs that were more similar to other drugs. We found that the model seems to cluster the drugs used in specific treatments.

Figure 5. Relationship between drugs using the semantic feature (drug2vec) of the deep neural network model. There were 746 nodes in this graph, each representing a drug. The clusters indicated the drugs with a specific treatment. Top: The cluster comprised antidepressants; middle: The cluster contained antibiotics; bottom: The cluster included ophthalmic medications.



For instance, the drugs in the cluster with the blue circle shown in Figure 5 (image on the top) comprised antidepressants such as trazodone, citalopram, clomipramine, and paroxetine. The cluster to the right contained antibiotics such as ampicillin, ceftazidime, cefpodoxime, and cefotaxime shown is Figure 5 (image in the middle). Moreover, the cluster at the bottom included ophthalmic medications such as fluorometholone, levocabastine, brinzolamide, and dipiverfrin (Figure 5, image at the bottom). There were other small clusters with their own specific treatment. Thus, we learned the relationships among drugs and the latent information from the papers included in the embedding function at the text-level. Accordingly, the semantic feature had a great effect on the performance of our model.

Subsequently, we examined the ability of this model to perform its identification and prediction functions with reference to serious ADRs defined by the Micromedex. Using the identification function of the model, we ranked the potential ADRs in a list by the probability of their occurrence (Table 3). One of the reasons why the probability of ADR presented in Table 3 was not highly prominent was that the positive samples of these ADRs were rarely reported in this dataset. However, our model could identify that hydroxychloroquine led to muscle cramps, which is a serious ADR that occurs in severe

neuromuscular disease. In addition, we extracted 5 drugs that were recorded only in 2012 to examine the prediction function of this model.

Table 3. The adverse drug reaction (ADR) prediction and identification results of the model.

Drug	Serious ADR	Rank	Probability
Identification results of drugs with known ADR records			
Dantrolene	Anemia	12	0.012
Dantrolene	Congestive heart failure	15	0.009
Hydroxychloroquine	Muscle Cramp	1	0.997
Hydroxychloroquine	Photophobia	16	0.017
19-nortestosterone	Serum cholesterol raised	4	0.150
Carbachol	Retinal detachment	3	0.690
Prediction results of drugs without ADR records			
Atazanavir	Anemia	17	0.920
Carbinoxamine maleate	Agranulocytosis	14	0.453
Carbinoxamine maleate	Anemia, Hemolytic	16	0.340
Darunavir	Hyperglycemia	20	0.750
Temsirolimus	Infection	20	0.974
Zoladex	Myocardial infarction	7	0.961
Zoladex	Hypersensitivity	12	0.920

Findings revealed that our model has the capacity to predict the serious ADRs of new drugs. For instance, the model predicted that Zoladex could lead to a serious ADR, myocardial infarction, which is one of the commonest causes of death in developing countries.

Limitations

This study has several limitations that need to be addressed in future studies. First, data diversity plays an important role in the model. We only used the data published in SIDER. Our model will be more persuasive and reliable if we can include more data from different datasets. Because the chemical and biological properties of drugs contribute most to their effects on human, the more the databases of drug properties included, the better the performance of our model. On the other hand, if we have access to more open-source data, including clinical trials, spontaneous reporting systems, and EMRs with support from government and pharmaceutical industry, our model will have better prediction. Furthermore, our model focused on the ADR prediction and identification. To identify the probability of occurrence of each ADR, we set 1325 hidden nodes and the total number of ADRs in the dataset in the output layer. In other words, although we had a mapping function to address new drugs, this model could only predict existing ADRs. Therefore, in the future work, we plan to utilize more detailed features such as drug-ADR interaction [13], drug-drug interaction, and ADR-ADR interaction networks for the prediction of ADRs. Furthermore, we also plan to investigate other embedding approaches to represent the ADRs to help predict the relationships between drugs and new ADRs.

Conclusions

We developed a novel ADR detection model based on the biological and chemical properties of drugs and the D2V (the semantic feature). After discussing the drug similarities with domain experts from the National Cheng Kung University Hospital and the Institute of Clinical Pharmacy and Pharmaceutical Sciences, we found out that the D2V can represent a characteristic of the drug. Our model could not only discover the potential ADRs of drugs but also predict the possible ADRs of new drugs. To discover potential ADRs based on the previous records, our model could identify the hidden relationship between ADR-ADR interactions. Furthermore, to predict the possible ADRs of a new drug without any previous ADR records, using the D2V feature, our mapping function exhibited good profiling for transferring the drug description into the D2V. The model exhibited good performance on both tasks and generated the most suitable results. It will help pharmacists and health care providers to understand the potential risk of side effect of drugs and address the issue of underreporting of spontaneous reports. Above all, our model will aid pharmacovigilance by identifying and predicting potential ADRs.

Acknowledgments

The authors would like to express their great appreciation to the Ministry of Science and Technology (MOST) and the Ministry of Health and Welfare (MOHW) for their support (grant numbers: MOST 104-2923-E-006-003-MY3, MOST 107-2634-F-006-006,

and MOHW105-FDA-D-113-000416). The authors would also like to appreciate Yung-Hsin Tseng, a master's student, who was involved in conventional test and case study.

Conflicts of Interest

None declared.

References

1. Coleman JJ, Pontefract SK. Adverse drug reactions. Clin Med (Lond) 2016 Oct;16(5):481-485. [doi: 10.7861/clinmedicine.16-5-481] [Medline: 27697815]
2. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet 2000 Oct 07;356(9237):1255-1259. [doi: 10.1016/S0140-6736(00)02799-9] [Medline: 11072960]
3. Couper M. World Health Organization. 2002. Safety of Medicines: A guide to detecting and reporting adverse drug reactions URL: http://apps.who.int/iris/bitstream/handle/10665/67378/WHO_EDM_QSM_2002.2.pdf;jsessionid=A03B5B56E0F1BCB404D0E3315FCA27F5?sequence=1 [accessed 2018-11-30] [WebCite Cache ID 74JWvMVtQ]
4. Lazarou J, Pomeranz B, Corey P. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998 Apr 15;279(15):1200-1205. [Medline: 9555760]
5. Yao B, Zhu L, Jiang Q, Xia HA. Safety monitoring in clinical trials. Pharmaceutics 2013 Jan 17;5(1):94-106 [FREE Full text] [doi: 10.3390/pharmaceutics5010094] [Medline: 24300399]
6. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions : a systematic review. Drug Saf 2006;29(5):385-396. [doi: 10.2165/00002018-200629050-00003] [Medline: 16689555]
7. Brown J, Kulldorff M, Chan K, Davis R, Graham D, Pettus P, et al. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. Pharmacoepidemiol Drug Saf 2007 Dec;16(12):1275-1284. [doi: 10.1002/pds.1509] [Medline: 17955500]
8. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther 1981 Aug;30(2):239-245. [Medline: 7249508]
9. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 1998 Jun;54(4):315-321. [Medline: 9696956]
10. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
11. Yang M, Kiang M, Shang W. Filtering big data from social media--Building an early warning system for adverse drug reactions. J Biomed Inform 2015 Apr;54:230-240 [FREE Full text] [doi: 10.1016/j.jbi.2015.01.011] [Medline: 25688695]
12. Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. J Am Med Inform Assoc 2014;21(2):308-314 [FREE Full text] [doi: 10.1136/amiajnl-2013-001718] [Medline: 23907285]
13. Karimi S, Wang C, Metke-Jimenez A, Gaire R, Paris C. Text and Data Mining Techniques in Adverse Drug Reaction Detection. ACM Comput Surv 2015 Jul 21;47(4):1-39. [doi: 10.1145/2719920]
14. Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. J Am Med Inform Assoc 2012;19(1):79-85 [FREE Full text] [doi: 10.1136/amiajnl-2011-000214] [Medline: 21676938]
15. Wang G, Jung K, Winnenburg R, Shah NH. A method for systematic discovery of adverse drug events from clinical notes. J Am Med Inform Assoc 2015 Nov;22(6):1196-1204 [FREE Full text] [doi: 10.1093/jamia/ocv102] [Medline: 26232442]
16. Wang W, Haerian K, Salmasian H, Harpaz R, Chase H, Friedman C. A drug-adverse event extraction algorithm to support pharmacovigilance knowledge mining from PubMed citations. AMIA Annu Symp Proc 2011;2011:1464-1470 [FREE Full text] [Medline: 22195210]
17. Finkelstein J, Chen Q, Adams H, Friedman C. Automated Summarization of Publications Associated with Adverse Drug Reactions from PubMed. AMIA Jt Summits Transl Sci Proc 2016;2016:68-77 [FREE Full text] [Medline: 27570654]
18. Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. J Am Med Inform Assoc 2014 Oct;21(e2):e278-e286 [FREE Full text] [doi: 10.1136/amiajnl-2013-002512] [Medline: 24644270]
19. Cami A, Arnold A, Manzi S, Reis B. Predicting adverse drug events using pharmacological network models. Sci Transl Med 2011 Dec 21;3(114):114ra127 [FREE Full text] [doi: 10.1126/scitranslmed.3002774] [Medline: 22190238]
20. Lorberbaum T, Nasir M, Keiser MJ, Vilar S, Hripcsak G, Tatonetti NP. Systems pharmacology augments drug safety surveillance. Clin Pharmacol Ther 2015 Feb;97(2):151-158 [FREE Full text] [doi: 10.1002/cpt.2] [Medline: 25670520]
21. Huang L, Wu X, Chen JY. Predicting adverse drug reaction profiles by integrating protein interaction networks with drug structures. Proteomics 2013 Jan;13(2):313-324. [doi: 10.1002/pmic.201200337] [Medline: 23184540]

22. Liu M, Wu Y, Chen Y, Sun J, Zhao Z, Chen X, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. J Am Med Inform Assoc 2012 Jun;19(e1):e28-e35 [FREE Full text] [doi: 10.1136/amiajnl-2011-000699] [Medline: 22718037]
23. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 2010;6:343 [FREE Full text] [doi: 10.1038/msb.2009.98] [Medline: 20087340]
24. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res 2016 Jan 04;44(D1):D1075-D1079 [FREE Full text] [doi: 10.1093/nar/gkv1075] [Medline: 26481350]
25. Chen B, Wild D, Guha D. PubChem as a source of polypharmacology. J Chem Inf Model 2009 Sep;49(9):2044-2055. [doi: 10.1021/ci9001876] [Medline: 19708682]
26. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018 Jan 04;46(D1):D1074-D1082 [FREE Full text] [doi: 10.1093/nar/gkx1037] [Medline: 29126136]
27. Srinivasan P. Text mining: Generating hypotheses from MEDLINE. J Am Soc Inf Sci 2004 Mar;55(5):396-413. [doi: 10.1002/asi.10389]
28. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781 2013 Jan 16:-. [doi: 10.1109/TNN.2003.820440] [Medline: 18244602]
29. Le QV, Mikolov T. Distributed Representations of Sentences and Documents. 2014 Presented at: ICML'14 Proceedings of the 31st International Conference on International Conference on Machine Learning; June 21-26, 2014; Beijing, China p. 1188-1196.
30. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics 2017 Dec;5:135-146 [FREE Full text] [doi: 10.1162/tacl_a_00051]
31. Sugawara E, Nikaido H. Properties of AdeABC and AdeIJK efflux systems of Acinetobacter baumannii compared with those of the AcrAB-TolC system of Escherichia coli. Antimicrob Agents Chemother 2014 Dec;58(12):7250-7257 [FREE Full text] [doi: 10.1128/AAC.03728-14] [Medline: 25246403]
32. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J R Statist Soc B 2011;73:273-282.
33. Singhal A, Simmons M, Lu Z. Text mining for precision medicine: automating disease-mutation relationship extraction from biomedical literature. J Am Med Inform Assoc 2016 Dec;23(4):766-772 [FREE Full text] [doi: 10.1093/jamia/ocw041] [Medline: 27121612]
34. Hinton G, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors. 2012 Jun 3. URL: https://arxiv.org/pdf/1207.0580.pdf [WebCite Cache ID 75xhdRuMY]
35. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Learning to detect cells using non-overlapping extremal regions. The Journal of Machine Learning Research 2014 Jan;15(1):348-356. [Medline: 23285570]
36. Ruta D. Robust method of sparse feature selection for multi-label classification with Naive Bayes. 2014 Presented at: Federated Conference on Computer Science and Information Systems, FedCSIS; September 2014; Warsaw, Poland p. 380. [doi: 10.15439/2014F502]
37. Pinto D, Costa P, Camacho R, Costa V. Predicting drugs adverse side-effects using a recommender-system. In: Springer. 2015 Presented at: International Conference on Discovery Science; 4-6 October; Banff, Canada p. 201-208.

Abbreviations

ADE: adverse drug events
ADR: adverse drug reaction
AUC: area under the receiver operating characteristic curve
DNN: deep neural network
EMR: electronic medical records
MAP: mean average precision
MEDLINE: Medical Literature Analysis and Retrieval System Online
PMF: probability matrix factorization
SIDER: Side Effect Resource

Edited by G Eysenbach; submitted 10.05.18; peer-reviewed by HT Yang, Y Li; comments to author 02.09.18; revised version received 03.10.18; accepted 04.11.18; published 06.02.19.

Please cite as:
Wang CS, Lin PJ, Cheng CL, Tai SH, Kao Yang YH, Chiang JH
Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model
J Med Internet Res 2019;21(2):e11016
URL: http://www.jmir.org/2019/2/e11016/
doi:10.2196/11016
PMID:30724742

Original Paper

Patient-Level Prediction of Cardio-Cerebrovascular Events in Hypertension Using Nationwide Claims Data

Jaram Park[1*], PhD; Jeong-Whun Kim[2,3*], MD, PhD; Borim Ryu[1], MS; Eunyoung Heo[1], MS; Se Young Jung[1,4], MD, MPH; Sooyoung Yoo[1], PhD

[1]Office of eHealth Research and Business, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[2]Department of Otorhinolaryngology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[3]Department of Otorhinolaryngology, Seoul National University College of Medicine, Seoul, Republic of Korea

[4]Department of Family Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Sooyoung Yoo, PhD
Office of eHealth Research and Business
Seoul National University Bundang Hospital
173beon-gil 82, Gumi-ro, Bundang-gu
Seongnam, 13620
Republic of Korea
Phone: 82 31 787 8980
Fax: 82 31 787 4054
Email: yoosoo0@gmail.com

Abstract

Background: Prevention and management of chronic diseases are the main goals of national health maintenance programs. Previously widely used screening tools, such as Health Risk Appraisal, are restricted in their achievement this goal due to their limitations, such as static characteristics, accessibility, and generalizability. Hypertension is one of the most important chronic diseases requiring management via the nationwide health maintenance program, and health care providers should inform patients about their risks of a complication caused by hypertension.

Objective: Our goal was to develop and compare machine learning models predicting high-risk vascular diseases for hypertensive patients so that they can manage their blood pressure based on their risk level.

Methods: We used a 12-year longitudinal dataset of the nationwide sample cohort, which contains the data of 514,866 patients and allows tracking of patients' medical history across all health care providers in Korea (N=51,920). To ensure the generalizability of our models, we conducted an external validation using another national sample cohort dataset, comprising one million different patients, published by the National Health Insurance Service. From each dataset, we obtained the data of 74,535 and 59,738 patients with essential hypertension and developed machine learning models for predicting cardiovascular and cerebrovascular events. Six machine learning models were developed and compared for evaluating performances based on validation metrics.

Results: Machine learning algorithms enabled us to detect high-risk patients based on their medical history. The long short-term memory-based algorithm outperformed in the within test (F1-score=.772, external test F1-score=.613), and the random forest-based algorithm of risk prediction showed better performance over other machine learning algorithms concerning generalization (within test F1-score=.757, external test F1-score=.705). Concerning the number of features, in the within test, the long short-term memory-based algorithms outperformed regardless of the number of features. However, in the external test, the random forest-based algorithm was the best, irrespective of the number of features it encountered.

Conclusions: We developed and compared machine learning models predicting high-risk vascular diseases in hypertensive patients so that they may manage their blood pressure based on their risk level. By relying on the prediction model, a government can predict high-risk patients at the nationwide level and establish health care policies in advance.

(J Med Internet Res 2019;21(2):e11757) doi:10.2196/11757

KEYWORDS

health risk appraisal; risk; hypertension; chronic disease; clustering and classification; decision support systems

Introduction

Nationwide health maintenance programs are aimed at the prevention of chronic diseases. South Korea has a single-payer national health insurance system in which all health care providers must participate to claim for their medical expenses [1]. In this efficient health care system, the Korean government provides a nationwide health maintenance program to all national health insurance members aged 40 years and above on a biennial basis [2].

Hypertension is one of the most important chronic diseases requiring management via the nationwide health maintenance program because the burden of this condition is enormous. Approximately 10% of the total medical expenditure is associated with hypertension and its attendant complications, resulting in high economic costs [3]. In particular, East Asia and the Pacific region have the highest absolute burden of 439 million hypertensive patients [4]. Moreover, increased blood pressure leads to 9.4 million deaths associated with ischemic heart disease, stroke, and heart failure. However, among the general population of hypertensive patients, just 46.5% were found aware of their existing condition, 36.9% received treatment, and only 13.8% actively controlled their blood pressure [5]. Hypertension and its complications are attributed to modifiable risk factors, such as high salt diets, physical inactivity, and obesity. Therefore, health care providers should inform patients about their risks of complications caused by hypertension, so that they can improve the modifiable risk factors [6,7].

Health risk appraisal (HRA) is one of the most widely used screening tools for increasing both, the awareness and treatment levels of hypertension [8,9]. From 2009, it has been provided for all patients included in the national health maintenance program in South Korea to outline the importance of controlling high blood pressure. However, there are some limitations in using the HRA for predictive purposes. First, the predictability of cardio-cerebrovascular events is not very reliable. Second, it is based on a static statistical model that is not dynamically improvable on a regular basis. Finally, HRA is usable only when a patient is included in the health maintenance program. Due to these limitations, patients with hypertension tend to overlook their risks of developing cardio-cerebrovascular complications, in turn lowering the treatment rate of hypertension. In this study, we aimed to develop and compare machine learning models predicting high-risk vascular diseases for hypertensive patients, so that they can manage their blood pressure based on their risk level.

To develop these models, we used the longitudinal dataset of the nationwide sample cohort, which allowed tracking patients' medical history across all health care providers in Korea (N=51,920) spanning 12 years. Furthermore, to ensure the generalizability of our models [10-12], we conducted an external validation using another national sample cohort dataset, comprised of different patients, published by National Health Insurance Service (NHIS).

Methods

Data Description

Based on the mandatory social insurance system, the Korean National Health Insurance Service (NHIS) has achieved universal coverage of the population since the mid-1970s [1]. The NHIS is a single-insurer system, and the system has paid health care providers based on fee-for-service. Therefore, almost all health care data are centralized in the large-scale database of NHIS [13]. In this study, we used two distinct datasets published independently by NHIS. The first dataset was used to develop and test machine learning models, and the second was used to conduct external validation of developed models.

To develop and test machine learning models, we used a sample cohort of national health check-up programs [2,14]. The NHIS provides a biennial health check-up program to all national health insurance members over 40 years of age free of charge. The dataset contains health records of a total of 514,866 patients, randomly selected from the health insurance members who have been served the health check-up program. For external validation, we used a national sample cohort dataset, which includes the data of one million patients who are randomly selected comprising 2.2% of the total Korean population in 2002. The dataset includes health records of patients from infants to elderly people over 85 years old. Both datasets contain patients' social and economic qualification variables, the status of medical resource utilization, statement, details of treatment, type of disease, and details of prescription, and the status of the clinic [13,14]. The detailed information of the variables in both datasets is described in Multimedia Appendix 1. With the benefits of a 12-year longitudinal dataset reflecting a nationwide sample cohort, we were able to track each patient's medical history from all types of health care providers (N=51,920), including a tertiary hospital (number of beds ≥300), general hospitals (number of beds ≥30), and clinics. To protect patient privacy, the personal information and clinical institution information were deidentified. The statistics of the NHIS datasets used for building models are presented in Table 1.

The study was approved by the Seoul National University Bundang Hospital Institutional Review Board (B-1512/326-102).

Table 1. Statistics of National Health Insurance Service dataset (2002-2013). The precise percentage of the numbers in this table cannot be provided because the official total numbers are unavailable. However, we believe that each dataset contains almost all the medical records of the sampled patients, since South Korea has a mandatory social insurance system that meets the universal coverage of the population and medical institutions.

Description	Health check-up cohort (n)	National sample cohort (n)
Hospitals	51,920	52,483
Patients	514,866	1,113,656
Prescriptions	83,935,395	83,935,395
Visits	96,534,359	119,362,188
Diagnostic codes (full code name)	17,385	19,626
Diagnostic codes (first 3-digits)	2160	2319
Annual patient visits, mean	15.6	8.9
Diagnostic codes/visit, mean	2.4	2.5
Drugs/prescription, mean	4.4	4.4

Study Population Definition

We focused on patients with essential hypertension and developed models to predict cardio-cerebrovascular events. Therefore, we identified patients with confirmed essential hypertension and cardio-cerebrovascular events based on the results of previous research [15,16]. Patients with hypertension were defined as the subjects newly diagnosed with essential hypertension (International Classification of Diseases, Tenth Revision ICD-10: I10, I100, or I109) and newly treated with at least one Anatomical Therapeutic Chemical (ATC) code that is related to hypertension between June 2004 and December 2008 (the ATC codes used in this study are presented in Multimedia Appendix 2). Patients with cardio-cerebrovascular were defined as the subjects newly diagnosed with ischemic heart disease (ICD-10: I20-I25), cerebrovascular diseases (ICD-10: G45, I60-I64, I65-I69), or chronic heart failure (ICD-10: I42, I50), or newly treated with at least one ATC code that is related to cardio-cerebrovascular medication (the ATC codes are presented in Multimedia Appendix 2).

We excluded patients with any previous record of antihypertensive medication, essential hypertension, ischemic heart disease (ICD-10: I20-I25), cerebrovascular diseases (ICD-10: G45, I60-I64, I65-I69), or chronic heart failure (ICD-10: I42, I50) during the washout period between May 2003 and May 2004.

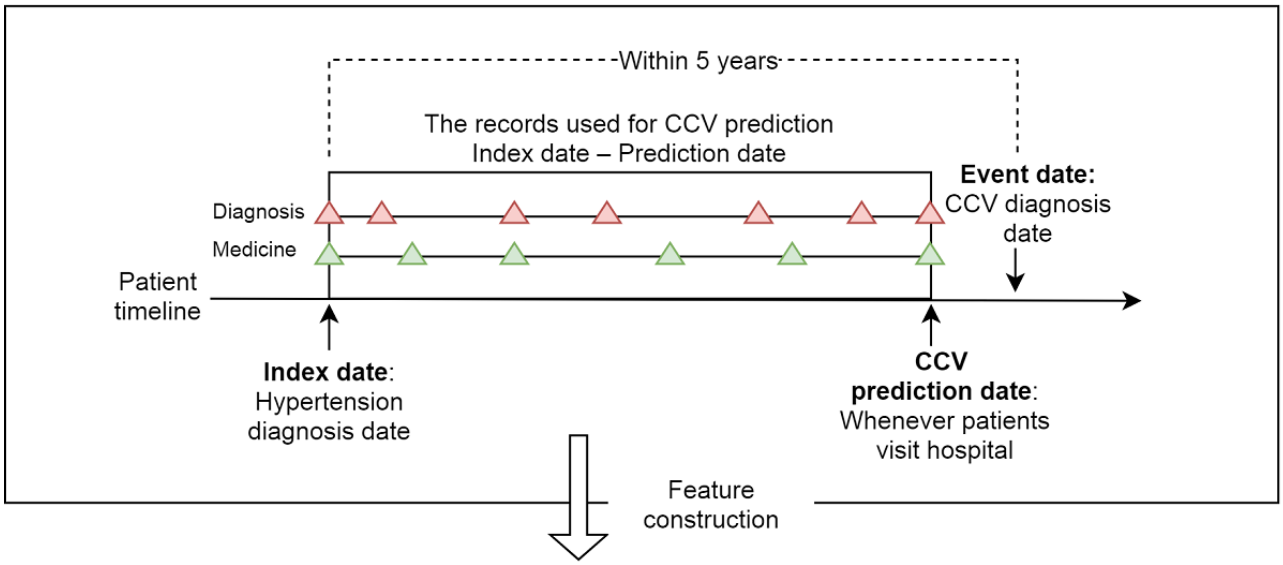
Input Features and Algorithms

The index date of each patient was defined as the time point of the first hypertension diagnosis or the time point of the first medication records related to hypertension. The event date of each patient was defined as the time point of the first cardio-cerebrovascular diagnosis within five years (see Figure 1). The prediction model application scenario was that when a patient visits a hospital, the models predict current or near-term risks which we defined as the duration from current/today to the next hospital visit (one month on average) high-risk vascular disease events in a patient to enable patients to manage blood pressure immediately. We limited the prediction periods to five years from the first hypertension diagnosis. Therefore, the algorithms predict current high-risk vascular disease events with medical records from a maximum of five years ago.

Of all medical records, we selected the following main features. First, the basic information of the patient, including age, gender, and hospital visit type (inpatient, outpatient, emergency) at the index date (n=3). Second, all diagnosis records from the index date to the event date (n=1252). All diagnosis records (ie, ICD-10 codes) were grouped by the first 3 digits, which comprise the main disease category. Each dimension represents the total number of occurrences of a specific code between the index date and the event date (of note, we excluded ICD-10 A, B, C, L, P, V, W, X, Y, and Z disease categories because of their low relevance for cardiovascular and cerebrovascular events). Third, all medication records from the index date to the event date (n=130). The first 3 digits of the ATC codes for medication records were used for feature construction. Regarding the medication records, we used both the total number of occurrences of a medication (n=65) and the number of days of a medication (n=65). As we used features based on the total number of occurrences in the medical records (ie, ICD-10 codes for diagnosis and ATC codes for medication), null values in a feature mean no occurrence of the disease or medicine for each patient. Therefore, we replaced null values with zeros (eg, the third patient in Figure 1).

After processing the features, we finally obtained 1385 of them for each patient. All the aggregated input vectors were linearly normalized to the range [0,1]. To investigate the advantage of time series characteristics in a longitudinal dataset for predicting diseases, we also developed algorithms based on a Recursive Neural Network (RNN) which is able to capture temporal patterns present in temporal sequenced data. For prediction models based on RNN, each time steps t comprised of all medical codes c_1, c_2, …, $c_{|C|}$ was converted to a binary vector $x_t=\{0,1\}^{|C|}$, and recent 50, 100, or 150 time steps of the binary vector (ie, hospital visits) were used for prediction.

Figure 1. Feature matrix construction. CCV: cardio-cerebrovascular.



	Basic Information of patient at the index date			The total number of occurrences of a diganosis			The total number of occurrence of a medication			The number of days of a medication	
Patient ID	Age	Sex	Visit type	I10	M54	...	ATC_C09	ATC_M09	...	ATC_C09_days	ATC_M09_days
10000001	65	F	inpatient	30	40	...	35	62	...	100	150
10000002	75	M	outpatient	10	0	...	9	0	...	60	20
10000003	61	M	emergency	0	0	...	0	0	...	0	0
10000004	85	F	outpatient	50	80	...	40	10	...	180	200

Before attempting machine learning algorithms, a univariate feature selection was performed to remove irrelevant features to the outcome variable. Based on the chi-square statistics scores, 555 significant features (P<.05) out of 1385 features were obtained. To avoid model overfitting, the top 55 and 278 features were selected, which represented 10% and 50% of the number of significant features, respectively. Subsequently, we compared the model performances according to the number of features (ie, 55, 278, and 555, respectively). The detailed information for the top 55 selected features and a list with all the features is in Multimedia Appendix 3.

We developed the models based on 6 common machine learning algorithms and compared the performance of the algorithms for predicting the cardio-cerebrovascular outcome of hypertensive patients: logistic regression (LR), support vector machine (SVM) [17], decision tree (DT) [18], random forest (RF) [19], multilayer perceptron (MLP) [20], and long short-term memory (LSTM) for time series prediction. The hyper-parameters used for training models are in Multimedia Appendix 4.

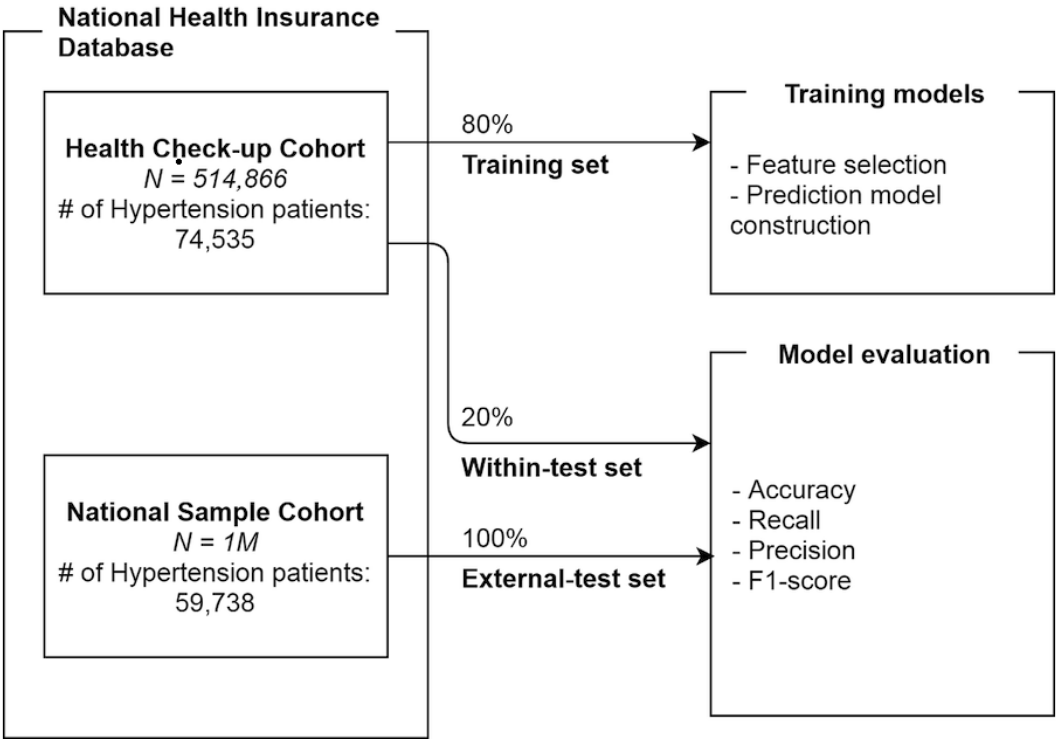
Model Evaluation Strategy

The following 3 experiments were used as outcomes: (1) predicting a cardiovascular event, (2) predicting a cerebrovascular event, and (3) predicting a cardio-cerebrovascular event (ie, vascular disease). Furthermore, we compared the performances of each algorithm using F1-score according to the number of features. We used 4 evaluation metrics commonly used in classification tasks to evaluate the performance of our models as follows:

- Accuracy: the proportion of patients who were predicted as their actual status
- Precision: the proportion of patients that actually had the diseases out of patients that were predicted as having diseases
- Recall: the proportion of patients that were predicted as having diseases out of all patients who actually had the diseases
- F1-score: the harmonic mean of precision and recall

To compare the evaluation metrics of each machine learning algorithm, the study population extracted from the Health check-up cohort was randomly split into 80% training and 20% test sets. Based on the 80% training set, key features were selected and the prediction models were trained (see Figure 2). We then tested our prediction models using the 20 % test sets (termed within test). After that, to confirm the external validity of the prediction models, the entire study population of the National Sample Cohort was used for external evaluation (termed external test). The training set was once more randomly split into 10-fold to conduct stratified 10-fold cross-validation. The 10-evaluation metrics from the folds were averaged to produce a single estimation.

Figure 2. Flowchart of the evaluation strategy.



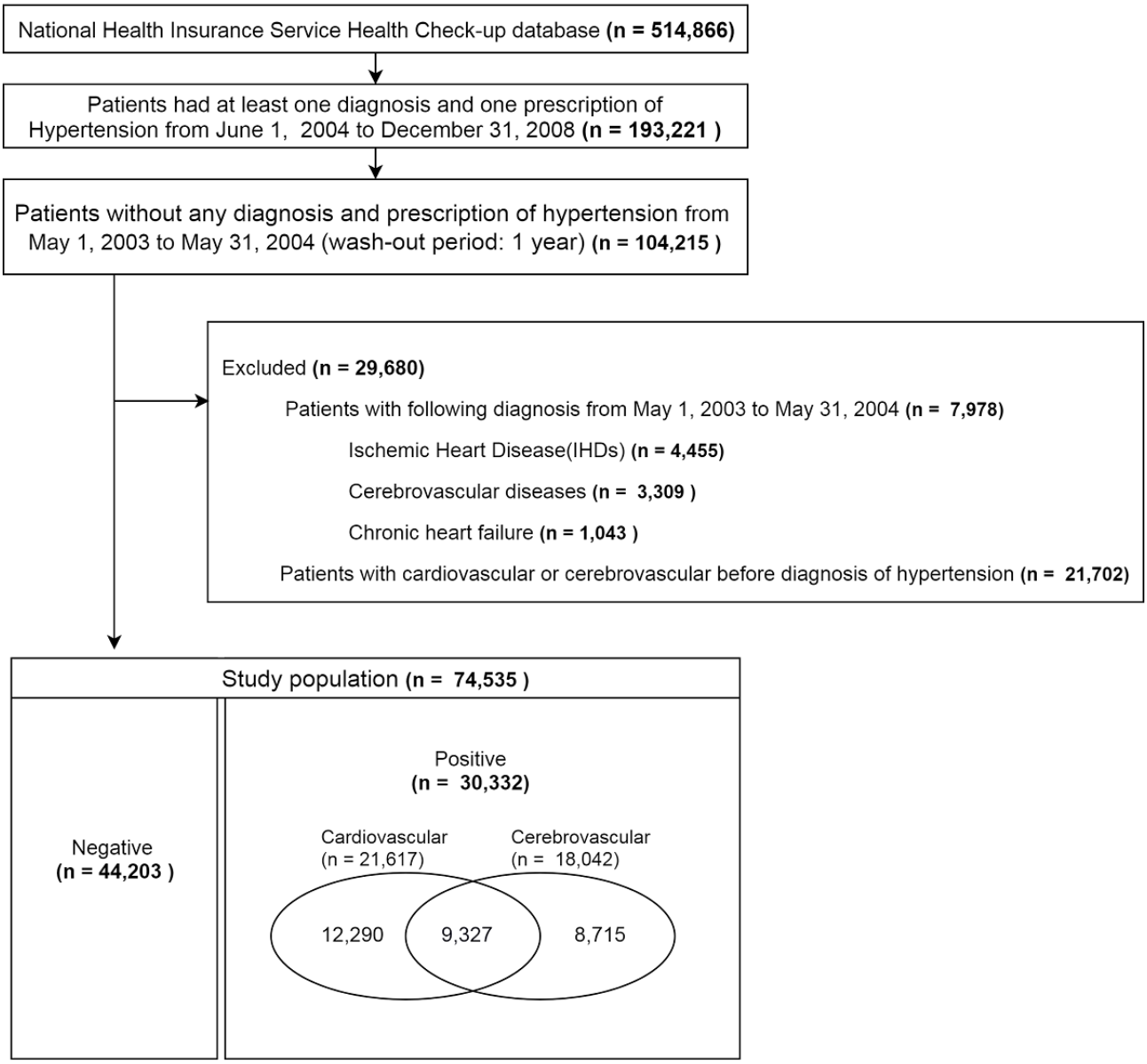
Results

Study Population

On the basis of the subject definition, we obtained the data of 74,535 patients from the first dataset, for whom a total of 136,843,003 medical records over 12 years were retrieved. As shown in Figure 3, of all hypertensive patients 59% (44,203/74,535) were diagnosed with hypertension only. Among the patients with hypertension, 29% (21,617/74,535) and 24% (18,042/74,535) were diagnosed with cardiovascular and cerebrovascular diseases, respectively.

For the external test, we extracted the data of 67,696 patients with hypertension from the second dataset. To match the age distribution, we only considered patients aged over 40 years. Finally, we obtained the data of 59,738 patients with hypertension. Of all hypertensive patients from the external test set, 60% (36,248/59,738) were diagnosed with hypertension only. Among the patients with hypertension, 28% (16,605/59,738) and 23% (13,828/59,738) were diagnosed with cardiovascular and cerebrovascular diseases, respectively. The disease outcomes were similarly distributed between the within test and external test set. Further characteristics of patients with hypertension only and cardiovascular and cerebrovascular events that were used in training models are presented in Table 2. The variables in the table present patient data at the time point of the first hypertension diagnosis and the patients were grouped by outcome events. The most characteristics were similarly distributed among the groups.

Figure 3. Flowchart describing the study population.



Performance of Cardio-Cerebrovascular Prediction Models

Table 3 presents the performance of the algorithms with 555 significant features predicting our target outcomes. The results showed that the LSTM algorithm outperforms the other algorithms, except for the recall score in within test, resulting in an accuracy score of .831. While the MLP, LR, and SVM algorithms outperformed RF algorithm in within test with F1 scores of .771, .762, and .760 respectively, the MLP, LR, and SVM algorithms performance in external test dropped dramatically to .065, .013, and .019 respectively, decreasing the recall score to ≤.034. However, DT, RF, and LSTM algorithms showed relatively small performance degradation in the external test. Therefore, in terms of a generalization of the model, the RF algorithm performed best by achieving an F1-score of .705 in the external test for all disease outcomes. The results confirmed that the RF algorithm is the most robust algorithm against overfitting [19,21]. In addition, the results showed the importance of the external validity of disease prediction models. Even though the performance of a model was reliable in within test, we confirmed that it can be vulnerable to new patients from the external test dataset. Therefore, as Damen et al [11] argued, the future research directions of disease prediction should focus on external validation.

The results of each outcome presented in Table 3 show that predicting a single outcome (ie, cardiovascular or cerebrovascular event) showed poor performance compared to predicting the cardio-cerebrovascular events of 5 algorithms. Specifically, predicting cerebrovascular events showed higher performance degradation than predicting cardiovascular events. This fact implies that it could be a good strategy to develop machine learning models predicting similar disease groups as a single outcome, such as cardio-cerebrovascular events in this study. This method will help balance the labels that are mostly imbalanced disease outcomes, and will finally improve the performance of the models as shown by the results of this study.

Table 2. General characteristics of the study population.

Variable	Hypertension only (n=44,203)	Cardio-cerebrovascular (n=30,332)	Cardiovascular (n=21,617)	Cerebrovascular (n=18,042)
Age (years), mean (SD)	57.1 (9.1)	60.5 (9.6)	60.3 (9.5)	61.8 (9.4)
Gender, n (%)[a]				
Female	19,036 (43.1)	14,253 (47.0)	9911 (45.7)	8933 (49.5)
Male	25,167 (56.9)	16,079 (53.0)	11,706 (54.3)	9109 (50.5)
Body mass index, mean (SD)	24.5 (2.9)	24.5 (3.0)	24.6 (3.0)	24.4 (3.0)
Smoking, n (%)[a]				
None	29,144 (65.9)	20,645 (68.1)	14,583 (67.5)	12,558 (69.6)
Past	4035 (9.1)	2453 (8.1)	1799 (8.3)	1326 (7.3)
Current	8647 (19.6)	5716 (18.8)	4174 (19.3)	3249 (18.0)
Drinking, n (%)[a]				
Nondrinker	23,645 (53.5)	17,825 (58.8)	12,611 (58.3)	10,979 (60.9)
Drinker	19,647 (44.4)	11,869 (39.1)	8568 (39.6)	6652 (36.9)
Systolic blood pressure, mean (SD)	137.5 (18.6)	137.2 (18.6)	137.2 (18.7)	137.5 (18.6)
Diastolic blood pressure, mean (SD)	85.3 (12.0)	84.6 (11.9)	84.6 (12.0)	84.4 (11.8)
Total cholesterol, mean (SD)	202.2 (38.4)	203.8 (39.5)	204.2 (39.7)	203.7 (39.0)
Fasting blood sugar level, mean (SD)	102.3 (35.1)	104.5 (39.4)	105 (40.8)	105.0 (39.6)
Diabetes, n (%)[a]	2616 (5.9)	2250 (7.4)	1863 (8.6)	1543 (8.6)
Hyperlipidemia, n (%)[a]	3784 (8.6)	3026 (10.0)	2495 (11.5)	1982 (11.0)

[a]The percent of this variable may not add up to 100% due to the missing value.

Figure 4 shows the F1-scores of each model in both test sets according to the number of features. In the within test, LSTM algorithm outperformed other algorithms, indicating the importance of considering the order of medical records and time information. In the external test, RF algorithm was the best regardless of the number of features. This result clearly indicates that an increasing number of features leads to the model overfitting. For example, while the machine learning algorithms with 55 features showed relatively consistent performance in both within test and external test to predict cardio-cerebrovascular events, the F1-scores of LR, SVM, and MLP algorithms with 278 and 555 features dropped down dramatically even under the same conditions (ie, number of samples, training dataset) except for the number of features. Specifically, the sensitivity of the models was extremely poor (see Table 4). This fact implies that the trained probability threshold does not work properly on new patients' medical history. Interestingly, RF, DT, and LSTM algorithms showed relatively consistent performance even with changes in the number of features. Therefore, our results indicated that RF is the most robust machine learning algorithm to predict diseases in both external validity and the changes in the number of features.

Table 3. Performance of prediction of each outcome across the models with all significant features (N=555).

Prediction outcome algorithms	Within test				External test			
	Accuracy	Recall	Precision	F1-score	Accuracy	Recall	Precision	F1-score
Cardio-cerebrovascular								
Logistic regression	.797	.807	.721	.762	.609	.007	.869	.013
Support vector machine	.796	.803	.722	.760	.610	.009	*.877* [a]	.019
Decision tree	.780	*.818*	.691	.749	.740	.737	.650	.691
Random forest	.793	.799	.718	.757	*.744*	*.779*	.644	*.705*
Multilayer perceptron	.806	.803	.742	.771	.616	.034	.754	.065
Long short-term memory	*.831*	.772	*.790*	*.772*	.681	.716	.553	.613
Cardiovascular								
Logistic regression	.748	.784	.540	.640	.732	.048	*.807*	.091
Support vector machine	.743	.797	.533	.639	*.735*	.068	.747	.125
Decision tree	.707	*.814*	.492	.613	.673	*.788*	.449	.572
Random forest	.723	.798	.509	.622	.685	.787	.461	*.582*
Multilayer perceptron	*.757*	.782	*.559*	*.652*	.727	.098	.547	.166
Cerebrovascular								
Logistic regression	.741	.757	.471	.581	*.769*	.002	*.821*	.005
Support vector machine	.733	.776	.463	.580	.769	.002	.795	.004
Decision tree	.672	*.828*	.405	.544	.662	.735	.381	.501
Random forest	.698	.812	.427	.560	.674	*.793*	.397	*.529*
Multilayer perceptron	*.749*	.787	*.486*	*.601*	.769	.001	.833	.001

[a]The highest scores are presented in italics.

Figure 4. Model evaluation (F1-score) results based on the number of features across 6 models. (LR: logistic regression, SVM: support vector machine, DT: decision tree, RF: random forest, MLP: multilayer perceptron, LSTM: long short-term memory).

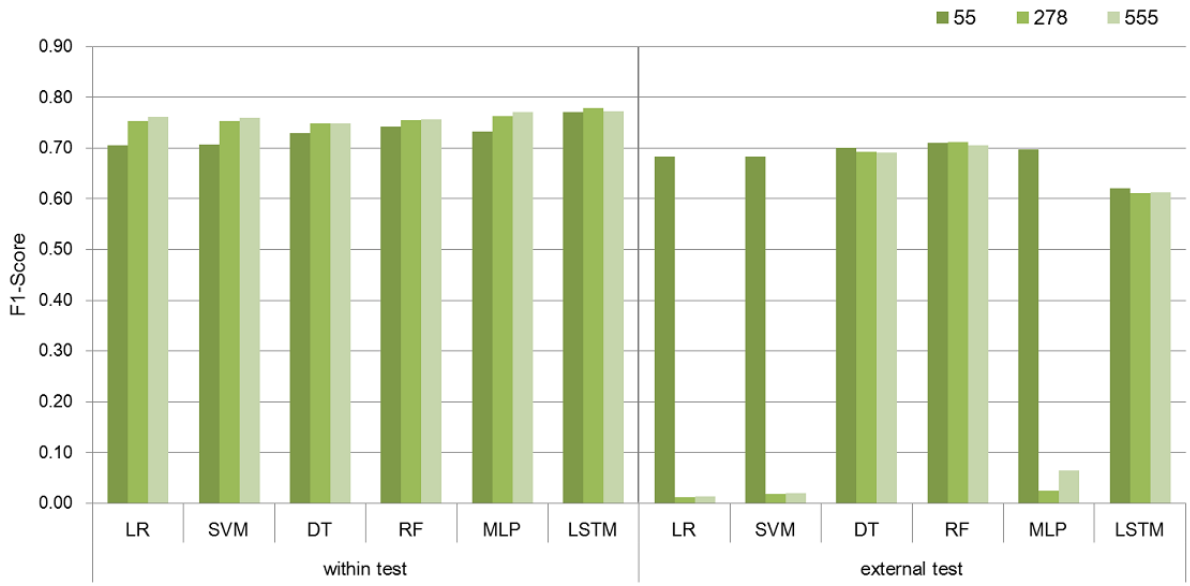


Table 4. Prediction for cardio-cerebrovascular according to the number of features across 6 models.

Number of features per algorithms	Within test				External test			
	Accuracy	Recall	Precision	F1-score	Accuracy	Recall	Precision	F1-score
55 features								
Logistic regression	.742	.762	.658	.706	.707	.801	.595	.683
Support vector machine	.742	.764	.657	.707	.707	.803	.594	.683
Decision tree	.752	*.822* [a]	.656	.730	.712	.858	.593	.701
Random forest	.771	.813	.684	.743	*.726*	.858	*.607*	*.711*
Multilayer perceptron	.766	.778	.691	.732	.722	.815	.610	.698
Long short-term memory	*.837*	.752	*.810*	*.771*	.618	*.874*	.491	.620
278 features								
Logistic regression	.788	.793	.717	.753	.609	.006	*.819*	.011
Support vector machine	.790	.789	.721	.753	.609	.009	.814	.017
Decision tree	.774	*.828*	.684	.749	.735	.758	.637	.692
Random forest	.786	.813	.706	.756	*.746*	*.798*	.643	*.712*
Multilayer perceptron	.802	.785	.742	.763	.609	.012	.690	.024
Long short-term memory	*.834*	.796	*.778*	*.779*	.653	.770	.521	.611
555 features								
Logistic regression	.797	.807	.721	.762	.609	.007	*.869*	.013
Support vector machine	.796	.803	.722	.760	.610	.009	.877	.019
Decision tree	.780	*.818*	.691	.749	.740	.737	.650	.691
Random forest	.793	.799	.718	.757	*.744*	*.779*	.644	*.705*
Multilayer perceptron	.806	.803	.742	.771	.616	.034	.754	.065
Long short-term memory	*.831*	.772	*.790*	*.772*	.681	.716	.553	.613

[a]The highest scores are presented in italics.

Discussion

Management of hypertension should be conducted according to the patient's risk level. In the present study, we developed and compared machine learning models predicting high-risk vascular diseases for hypertensive patients, so that they can manage their blood pressure based on their risk level. The results of this study suggest that machine learning algorithms predict which patients have high risks based on their medical history. To confirm the usefulness of the models that were developed in this study, we conducted the external validation [11,22] using another nationwide claim dataset. The LSTM algorithm outperformed in the within test, and the RF-based algorithm of risk prediction showed better performance over other machine learning algorithms in terms of generalization [21]. The results also confirmed that the models with fewer variables are more generalizable [22].

In recent years, many studies on using machine learning to predict these diseases have been actively conducted with the emergence of large-volume data, such as electronic medical records (EMRs) [23-27]. Previous disease prediction models have used variables from a range of sources, including patient diagnosis and medication, and the models demonstrated better disease predicting performance than more established methods, such as those included in the American Heart Association/American College of Cardiology guidelines [25]. In addition, lab test data and wearable sensor data were also used. Because both these types of data directly reflect the health status of patients, they are very useful for predicting future health complications. However, this fact causes a practical difficulty with using the models, because these data sources involve time-consuming and expensive processes, and not all patients have access to them. Moreover, even with such patient datasets acquired, most developed prediction models are not used in real practice due to the lack of external validity [10-12]. The models were developed by using the datasets of a single institution or of multiple centers and therefore are restricted to predicting the diseases of patients visiting that local site only [22]. We mainly focused on patients' diagnoses and medication records, so that our models could predict the complications of hypertension with easily obtainable features, and conducted external validation to ensure the generalizability of our models. The prediction model in this study is based on medical history after the hypertension was diagnosed, not a snapshot of the patient's health status; therefore, patients may obtain their risk levels more appropriately and in a more cost-effective way.

The univariate feature selection method allowed for the improved prediction of outcomes, without overfitting problem.

Interestingly, among the top 55 (or 10%) of significant features, a total of 80% of features were related to medication. The result of this feature selection suggests that medication information plays a major role in predicting cardio-cerebrovascular events in hypertensive patients. These results may be related to the fact that the initial treatment may vary according to the patient's status and clinical decisions [28,29], and medication switching is more frequent than diagnosis switching. Therefore, medication information contains more various and complex information about hypertensive patients compared to diagnostic information [30]. Furthermore, we found a difference between patient groups according to outcome events. Patients without cardio-cerebrovascular outcome events had more prescriptions and medication days. As this study was focused on patient-level prediction through the development and comparison of machine learning algorithms and not on population-level estimations, we did not investigate these treatment patterns in detail. However, this may be an interesting future target for research. We have, therefore, reported the average and median values of top 55 features for the groups in Multimedia Appendix 3.

Moreover, we manually investigated the selected top 55 feature list and found that a considerable number of features seemed irrelevant. Even though the variables seemed not directly related to cardio-cerebrovascular events, we have not removed the features from the model because (1) the variables were selected based on the 12-year treatment characteristics of half million Korean patients and could be associated with omitted variables and (2) the developed model will be applied to Korean patients with the same medical care behavior characteristics. This result provides future research insight investigating the relationship between the target outcome and those uncovered important diagnoses or medications so far based on population-level research using various analysis methods, such as statistical analysis and network analysis [31,32].

We used nationwide claims data including each patient's treatment and medication history for 12 years to train the machine learning algorithms. There are some limitations in using claims data for predicting diseases. First, there is an issue of the accuracy of disease code due to the purpose of billing [33]. Second, it is difficult to identify the uncovered services and the use of generic medicines that are not prescribed by health care providers. Finally, as we defined clinical events using diagnostic codes, the rates of the event could be underestimated [34].

Despite these limitations, the nationwide claims dataset provides good opportunities to apply the models. First, the dataset contains medical records from almost all medical services (ie, national level), not from a single institution or multiple centers. Furthermore, for scalability of the models, only the diagnosis and medication records of patients were used for training the models. Therefore, the pretrained machine learning algorithm could be used for developing prediction models of extending diseases and improving other prediction models by transferring knowledge, as suggested by Choi et al [35]. Second, the developed model could be applied to manage national-level disease risk. This study is, to the best of our knowledge, the first disease prediction model for patients based on nationwide claims data. Therefore, by relying on the prediction model, a government can predict high-risk patients at a nationwide level and establish health care policies in advance.

There are several future research directions. First, we can refine the models using up-to-date nationwide health insurance data and obtain feedback from physicians about the feasibility and predictability of the models at every single clinic. Second, clinical trials on the effect of controlling hypertension can be conducted by applying dynamic prediction models to patients in every single visit.

Acknowledgments

This work was supported by the Creative Industrial Technology Development Program (10053249, Development of Personalized Health Care System exploiting User Life-Log and Open Government Data for Business Service Model Proof on Whole Life Cycle Care) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea). This study used NHIS-HEALS data (NHIS-2017-2-326).

Authors' Contributions

JP and J-WK equally contributed to the work as first authors. J-WK initiated this study as a principal investigator of the project. JP worked out technical details, processed and analyzed the data. BR, EH, and SY contributed to data preparation and discussion. SY and SYJ equally supervised the entire process as corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of the dataset.

[PDF File (Adobe PDF File), 293KB - jmir_v21i2e11757_app1.pdf]

Multimedia Appendix 2

Anatomical therapeutic chemical codes for subject definition.

[PDF File (Adobe PDF File), 213KB - jmir_v21i2e11757_app2.pdf]

Multimedia Appendix 3

List of features.

[PDF File (Adobe PDF File), 453KB - jmir_v21i2e11757_app3.pdf]

Multimedia Appendix 4

Hyper-parameters used for training models.

[PDF File (Adobe PDF File), 323KB - jmir_v21i2e11757_app4.pdf]

References

1. Kwon S. Thirty years of national health insurance in South Korea: lessons for achieving universal health care coverage. Health Policy Plan 2009 Jan;24(1):63-71. [doi: 10.1093/heapol/czn037] [Medline: 19004861]
2. Seong SC, Kim Y, Park SK, Khang YH, Kim HC, Park JH, et al. Cohort profile: the National Health Insurance Service-National Health Screening Cohort (NHIS-HEALS) in Korea. BMJ Open 2017 Sep 24;7(9):e016640 [FREE Full text] [doi: 10.1136/bmjopen-2017-016640] [Medline: 28947447]
3. Gaziano TA, Bitton A, Anand S, Weinstein MC, International SOH. The global cost of nonoptimal blood pressure. J Hypertens 2009 Jul;27(7):1472-1477. [doi: 10.1097/HJH.0b013e32832a9ba3] [Medline: 19474763]
4. Pereira M, Lunet N, Azevedo A, Barros H. Differences in prevalence, awareness, treatment and control of hypertension between developing and developed countries. J Hypertens 2009 May;27(5):963-975. [Medline: 19402221]
5. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, et al. Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries. Circulation 2016 Aug 09;134(6):441-450. [doi: 10.1161/CIRCULATIONAHA.115.018912] [Medline: 27502908]
6. Williamson JD, Supiano MA, Applegate WB, Berlowitz DR, Campbell RC, Chertow GM, SPRINT Research Group. Intensive vs Standard Blood Pressure Control and Cardiovascular Disease Outcomes in Adults Aged ≥75 Years: A Randomized Clinical Trial. JAMA 2016 Jun 28;315(24):2673-2682 [FREE Full text] [doi: 10.1001/jama.2016.7050] [Medline: 27195814]
7. Whitworth JA, World Health Organization, International Society of Hypertension Writing Group. 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. J Hypertens 2003 Nov;21(11):1983-1992. [doi: 10.1097/01.hjh.0000084751.37215.d2] [Medline: 14597836]
8. Yen L, McDonald T, Hirschland D, Edington DW. Association between wellness score from a health risk appraisal and prospective medical claims costs. J Occup Environ Med 2003 Oct;45(10):1049-1057. [doi: 10.1097/01.jom.0000088875.85321.b9] [Medline: 14534445]
9. Lim SS, Carnahan E, Nelson EC, Gillespie CW, Mokdad AH, Murray CJL, et al. Validation of a new predictive risk model: measuring the impact of the major modifiable risks of death for patients and populations. Popul Health Metr 2015;13:27 [FREE Full text] [doi: 10.1186/s12963-015-0059-8] [Medline: 26435702]
10. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol 2014;14:40 [FREE Full text] [doi: 10.1186/1471-2288-14-40] [Medline: 24645774]
11. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016 May 16;353:i2416 [FREE Full text] [Medline: 27184143]
12. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208. [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
13. Lee J, Lee JS, Park S, Shin SA, Kim K. Cohort Profile: The National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. Int J Epidemiol 2017 Dec 01;46(2):e15. [doi: 10.1093/ije/dyv319] [Medline: 26822938]
14. National Health Insurance Sharing Service. Details of DB and cost: Sample cohort DB. Seoul, South Korea: National Health Insurance Sharing Service URL: https://nhiss.nhis.or.kr/bd/ab/bdaba022eng.do [WebCite Cache ID 71DHtmfsp]
15. Shin S, Song H, Oh S, Choi KE, Kim H, Jang S. Effect of antihypertensive medication adherence on hospitalization for cardiovascular disease and mortality in hypertensive patients. Hypertens Res 2013 Nov;36(11):1000-1005. [doi: 10.1038/hr.2013.85] [Medline: 23966057]
16. Kim S, Shin DW, Yun JM, Hwang Y, Park SK, Ko Y, et al. Medication Adherence and the Risk of Cardiovascular Mortality and Hospitalization Among Patients With Newly Prescribed Antihypertensive Medications. Hypertension 2016 Mar;67(3):506-512. [doi: 10.1161/HYPERTENSIONAHA.115.06731] [Medline: 26865198]

17. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems and their applications 1998;13(4):18-28. [doi: 10.1109/5254.708428]
18. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Trees. FL: Chapman and Hall/CRC; 1984.
19. Breiman L. Random forests. Machine learning 2001;45(1):5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
20. Rumelhart D, Hinton G, Williams R. Learning internal representations by error propagation. In: Parallel distributed processing: explorations in the microstructure of cognition, vol. 1. Cambridge, MA, USA: MIT Press; 1986:318-362.
21. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? The Journal of Machine Learning Research 2014;15(1):3133-3181 [FREE Full text]
22. Callahan A, Shah N. Machine Learning in Healthcare. In: Key Advances in Clinical Informatics. Amsterdam: Elsevier; 2018:279-291.
23. Ballinger B, Hsieh J, Singh A, Sohoni N, Wang J, Tison G, et al. DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction. 2018 Presented at: AAAI International Conference on Artificial Intelligence; 2018; New Orleans, LA, USA.
24. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc 2017 Dec 01;24(2):361-370 [FREE Full text] [doi: 10.1093/jamia/ocw112] [Medline: 27521897]
25. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017;12(4):e0174944 [FREE Full text] [doi: 10.1371/journal.pone.0174944] [Medline: 28376093]
26. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. J Med Internet Res 2018 Jan 30;20(1):e22 [FREE Full text] [doi: 10.2196/jmir.9268] [Medline: 29382633]
27. Sakr S, Elshawi R, Ahmed A, Qureshi WT, Brawner C, Keteyian S, et al. Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project. PLoS One 2018;13(4):e0195344 [FREE Full text] [doi: 10.1371/journal.pone.0195344] [Medline: 29668729]
28. Baggarly SA, Kemp RJ, Wang X, Magoun AD. Factors associated with medication adherence and persistence of treatment for hypertension in a Medicaid population. Res Social Adm Pharm 2014;10(6):e99-112. [doi: 10.1016/j.sapharm.2014.02.002] [Medline: 24731547]
29. Lauffenburger JC, Landon JE, Fischer MA. Effect of Combination Therapy on Adherence Among US Patients Initiating Therapy for Hypertension: a Cohort Study. J Gen Intern Med 2017 Jun;32(6):619-625 [FREE Full text] [doi: 10.1007/s11606-016-3972-z] [Medline: 28050754]
30. Risson V, Saini D, Bonzani I, Huisman A, Olson M. Patterns of Treatment Switching in Multiple Sclerosis Therapies in US Patients Active on Social Media: Application of Social Media Content Analysis to Health Outcomes Research. J Med Internet Res 2016;18(3):e62 [FREE Full text] [doi: 10.2196/jmir.5409] [Medline: 26987964]
31. Goh K, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. In: Proc Natl Acad Sci U S A. 2007 May 22 Presented at: The human disease network. Proceedings of the National Academy of Sciences. ;104(21); 2007; Washington, DC p. 8685-8690 URL: http://www.pnas.org/cgi/pmidlookup?view=long&pmid=17502601 [doi: 10.1073/pnas.0701361104]
32. Jeong E, Ko K, Oh S, Han HW. Network-based analysis of diagnosis progression patterns using claims data. Sci Rep 2017 Nov 14;7(1):15561 [FREE Full text] [doi: 10.1038/s41598-017-15647-4] [Medline: 29138438]
33. Kimm H, Yun JE, Lee S, Jang Y, Jee SH. Validity of the diagnosis of acute myocardial infarction in korean national medical health insurance claims data: the korean heart study (1). Korean Circ J 2012 Jan;42(1):10-15 [FREE Full text] [doi: 10.4070/kcj.2012.42.1.10] [Medline: 22363378]
34. Psaty BM, Delaney JA, Arnold AM, Curtis LH, Fitzpatrick AL, Heckbert SR, et al. Study of Cardiovascular Health Outcomes in the Era of Claims Data: The Cardiovascular Health Study. Circulation 2016 Jan 12;133(2):156-164 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.018610] [Medline: 26538580]
35. Choi E, Bahadori M, Schuetz A, Stewart W, Sun J. Doctor AI: Predicting clinical events via recurrent neural networks. Proceedings of Machine Learning for Healthcare Conference. 2016 Presented at: Machine Learning for Healthcare Conference; 2016.08.19-20; LA, California.

Abbreviations

ATC: Anatomical Therapeutic Chemical
DT: decision tree
EMR: electronic medical records
HRA: health risk appraisal
ICD-10: International Classification of Diseases, Tenth Revision
LR: logistic regression
LSTM: long short-term memory
MP: multilayer perceptron

NHIS: National Health Insurance Service
RNN: Recursive Neural Network
RF: random forest
SVM: support vector machine

Edited by G Eysenbach; submitted 02.08.18; peer-reviewed by SY Shin, Y Park; comments to author 09.10.18; revised version received 11.11.18; accepted 22.11.18; published 15.02.19.

Please cite as:
Park J, Kim JW, Ryu B, Heo E, Jung SY, Yoo S
Patient-Level Prediction of Cardio-Cerebrovascular Events in Hypertension Using Nationwide Claims Data
J Med Internet Res 2019;21(2):e11757
URL: http://www.jmir.org/2019/2/e11757/
doi:10.2196/11757
PMID:30767907

Original Paper

Novel Approach to Inpatient Fall Risk Prediction and Its Cross-Site Validation Using Time-Variant Data

Insook Cho[1], PhD; Eun-Hee Boo[2], PhD; Eunja Chung[3], MS; David W Bates[4], PhD; Patricia Dykes[4], PhD

[1]Inha University, Incheon, Republic of Korea

[2]National Health Insurance Service Ilsan Hospital, Gyeonggi-do, Republic of Korea

[3]Bundang Seoul National University Hospital, Gyeonggi-do, Republic of Korea

[4]The Center for Patient Safety Research and Practice, Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, United States

Corresponding Author:
Insook Cho, PhD
Inha University
Nursing Department (5E 301)
Inharo 100, Nam-gu
Incheon, 22212
Republic of Korea
Phone: 82 328608201
Fax: 82 328745880
Email: insook.cho@inha.ac.kr

Abstract

Background: Electronic medical records (EMRs) contain a considerable amount of information about patients. The rapid adoption of EMRs and the integration of nursing data into clinical repositories have made large quantities of clinical data available for both clinical practice and research.

Objective: In this study, we aimed to investigate whether readily available longitudinal EMR data including nursing records could be utilized to compute the risk of inpatient falls and to assess their accuracy compared with existing fall risk assessment tools.

Methods: We used 2 study cohorts from 2 tertiary hospitals, located near Seoul, South Korea, with different EMR systems. The modeling cohort included 14,307 admissions (122,179 hospital days), and the validation cohort comprised 21,172 admissions (175,592 hospital days) from each of 6 nursing units. A probabilistic Bayesian network model was used, and patient data were divided into windows with a length of 24 hours. In addition, data on existing fall risk assessment tools, nursing processes, Korean Patient Classification System groups, and medications and administration data were used as model parameters. Model evaluation metrics were averaged using 10-fold cross-validation.

Results: The initial model showed an error rate of 11.7% and a spherical payoff of 0.91 with a c-statistic of 0.96, which represent far superior performance compared with that for the existing fall risk assessment tool (c-statistic=0.69). The cross-site validation revealed an error rate of 4.87% and a spherical payoff of 0.96 with a c-statistic of 0.99 compared with a c-statistic of 0.65 for the existing fall risk assessment tool. The calibration curves for the model displayed more reliable results than those for the fall risk assessment tools alone. In addition, nursing intervention data showed potential contributions to reducing the variance in the fall rate as did the risk factors of individual patients.

Conclusions: A risk prediction model that considers longitudinal EMR data including nursing interventions can improve the ability to identify individual patients likely to fall.

(J Med Internet Res 2019;21(2):e11505) doi:10.2196/11505

KEYWORDS

across sites validation; electronic medical records; inpatient falls; nursing dataset; predictive model

Introduction

A considerable body of literature exists on fall prevention and reduction, yet despite many attempts by hospitals to reduce fall rates, significant and sustained reductions have proved elusive [1]. Risk assessment tools have been developed for several decades, and many risk factor identification studies have been published. St. Thomas' Risk Assessment Tool in Falling Elderly Inpatients (STRATIFY) and the Hendrich II are 2 examples consisting of 5-7 subscales [2-4]. However, most predate the broad use of electronic medical records (EMRs), and the tools were largely developed using limited data collected by researchers. The Cochrane reviews of Cameron et al [5] and Gillespie et al [6] imply that there is a significant lack of evidence on the efficacy of tools used to assess the risk of falling.

EMRs contain a considerable amount of information about patient histories and patient information conveyed both for discrete events and in narratives such as nursing notes. The increasing adoption of EMRs makes such clinical documentation a potentially rich and underutilized source of information for supporting nursing decisions [7]. Two types of data in the EMRs, in particular, present an opportunity for automated risk prediction: (1) structured longitudinal data and (2) semistructured or narrative data of nursing statements conveyed in clinical notes. Nursing assessment data are often recorded in a structured form or using a predefined template. Nursing notes contain rich nursing-process information about identified nursing problems, provided interventions, and patients' response. Several studies have investigated inpatient prediction models using EMR data. One study [8] used physician orders, nursing assessments and care plans, progress notes, and the intensity of nursing care needs to predict inpatient falls. Another study [9] conducted in 13 nursing homes used a minimum dataset(MDS) and structured data from EMRs, such as medications and nursing problems, at 1 week after admission and 1 week after a room change to predict resident falls. In addition, Tescher et al [10] and Giles et al [11] used EMR data to identify risk factors for the development of pressure ulcers and inpatient falls, respectively. Nevertheless, these studies are limited by their use of summary metrics rather than time-varying variables and did not consider the nursing interventions provided in an attempt to prevent falls.

The rapid adoption of EMRs and the integration of nursing data into clinical repositories have made large quantities of clinical data available for both clinical practice and research [7]. The aims of this study were to incorporate longitudinal EMR nursing-process data as a novel feature in calculating the risk for falls and to validate the findings at an external site. Intended nursing activities contribute to decreasing the risk and, thus, controlling for these will facilitate the ability to predict fall risk at a specific time-point. In addition, external validation is important for generalizability and discrimination when a model is applied at other sites or when using other EMR systems [12]. This research team noted several points that Goldstein et al [12] addressed in their systematic review of EMR-based prediction models: (1) it is easier to predict the short-term risk of events, as the data are observed more frequently; (2) patient populations included in EMRs may be more reflective of the real world than the data collected for research purposes; and (3) prediction models based on EMR data can often be implemented more easily than traditional algorithms that need to be translated before being applied in a clinical setting.

This study investigated the following research questions: (1) How can longitudinal data from nursing records be incorporated into fall risk modeling, which predicts daily risk at the patient-level? (2) How can electronic EMR data be incorporated into a fall risk modeling paradigm, focusing on 2 types of data elements of the EMR (structured data and semistructured data)? and (3) Does the fall risk model developed at a particular site or using a particular EMR system environment work at another site with a different EMR system and a different fall risk assessment tool?

This research team cast the problem of risk modeling as a probabilistic Bayesian network, which has several advantages for capturing and reasoning with uncertainty [13]. These methods are capable of producing 2 valuable outcomes as follows: (1) an interpretable set of concept variables associated with the risk of falling at the population level and (2) an actionable model to estimate the risk of falling for individual patients.

Methods

Study Sites and Cohorts

The 2 study cohorts were derived from the clinical data repositories of 2 institutions. One tertiary hospital was the "development site," while the other tertiary hospital was the "validation site"; both are located near Seoul, South Korea. Both hospitals have approximately 1000 beds and have used EMR systems for >10 years. The development site had 24,000 coded nursing statements mapped to the International Classification for Nursing Practice (ICNP) terminology. These statements are used for documenting nursing notes with free-text entries. The validation site has coded nursing statements represented by 3N (North American Nursing Diagnosis Association, Nursing Intervention Classification, and Nursing Outcome Classification). The 2 study sites have different EMR systems with 2 different terminology standards and 2 different fall risk assessment tools.

The development cohort consisted of hospitalized inpatients who were admitted to 6 nursing units with high fall rates from September 1, 2014, to August 31, 2015. Patients were mainly registered in cardiovascular, hematology-oncology, and neurology medical departments. Inclusion criteria included adults aged ≥18 years and admitted for at least 24 hours. Exclusion criteria included admission to a psychiatric, obstetric, emergency, or pediatric medical department. Patients who died or had received resuscitation treatment were excluded. We identified 14,307 admissions (122,179 hospital days) that conformed with the inclusion criteria. We identified 220 events by analyzing the hospital's event-reporting system, and an additional 18 cases were found through chart reviews conducted after prefiltering the free-text entries.

The validation cohort included 21,172 (172,592 hospital days) admissions from 6 medical-surgical nursing units. The units were selected on the basis of consistent nurse staffing and a case-mix with high fall rates in the hospital. The eligibility criteria applied to the development cohort were also applied to the validation cohort. As the fall rate on nursing units was estimated to be lower in the validation site, we extended data collection to a 2-year period from June 1, 2014, to May 31, 2016. A total of 292 falls were identified after analyzing the reporting system and chart reviews. We adopted the NDNQI operational definition of falls and level of injury [14].

Each cohort was divided randomly into model training and testing sets. For both training and testing, the patient stays were divided into windows with a length of 24 hours because nurses' fall risk assessments can be conducted on a daily basis. For example, a patient hospitalized for 4 days can have a maximum of 4 fall risk assessments performed and documented in the EMR. A sliding-window approach was used to generate multiple windows covering patients' data during their hospital stay by shifting the window to consecutive fall events. For fallers, only data that applied to within 24 hours before a fall were considered; data obtained prior to this were eliminated because it remains unclear whether they should receive a positive or negative label. For nonfallers, all of their data were included and labeled as negative. Samples were split into the training and testing sets while including samples from a given patient only in one of these sets. This approach was used to mirror the end-use situation more closely, where the system is evaluated on patients who are different from those on whom the model was trained. The imbalance between positive and negative labels was removed by oversampling the positives based on the ratio of positive-to-negative examples. According to a study [15] on machine learning using imbalanced data, the oversampling method is better than intelligent sampling techniques such as SMOTE (synthetic minority oversampling technique) and borderline SMOTE. For assuring the model validation, we applied the 10-fold validation method, which reuses the training dataset, randomly generating 90 (training) to 10 (testing) splits 10 times.

This retrospective study was reviewed and approved by the institutional review boards at the 2 hospitals, and the need for patients' informed consent was waived because the study involved the collection of deidentified data.

Identifying Model Concepts and Mapping Into Local Data Elements

Variables were selected on the basis of a literature review focusing on clinical guidelines published within the past 5 years (2012-2017). We adopted the following 8 fall prevention guidelines recommended by the Joint Commission [1], including the guideline of the Korean Hospital Nurses Association [16]: Agency for Healthcare Research and Quality [17]; ECRI Institute [18]; Institute for Clinical Systems Improvement [19]; Institute for Healthcare Improvement [20]; Joint Commission Center for Transforming Healthcare [21]; Veterans Affairs [22]; and Veterans Affairs National Center for Patient Safety [23].

Table 1 lists the concepts identified according to category and care components. We used the concepts to build a predictive Bayesian network structure for falls. We standardized the concepts by mapping to standard nursing terminologies in the current releases of the Logical Observation Identifiers Names and Codes and the ICNP. Then, the standard concepts were semantically mapped to local data elements of each EMR environment. To find available coded and structured or semistructured data, we explored the 2 hospitals' EMR systems. We found that only nursing data available from nursing records and nursing notes met these criteria. The mapping process was conducted by a project team consisting of 6 experts in the following relevant domains: nursing informatics, terminology, quality management, and patient safety. We have previously described the mapping process [24,25].

Figure 1 shows the 4 steps used in this study to develop and validate the fall risk prediction model: (1) review of guidelines and literature; (2) represent the concepts in a standardized terminology; (3) train and evaluate the model; and (4) cross-validate the model.

Table 1. Concepts derived from the literature review and local data elements mapped to concept variables in the prediction model.

Category and care component	Model concept	EMR[a] data element in development site	EMR data element in validation site
Patients' characteristics			
Demographics	Age	Age	Age
Diagnosis or procedure	Primary and secondary dx[b], surgical operation	Medical dx. (ICD[c] code), dates of surgical operation	Medical dx. (ICD code), dates of surgical operation
Administrative	Discharge unit, medical department, hospital days	Discharge unit, medical department, length of stay	Discharge unit, medical department, length of stay
Contributing factors: patient			
Physiological or disease-related factors	Visual and hearing impairment, elimination impairment, gait, mobility impairment, use of walking aids or devices, presence of dizziness, general weakness, orthostatic hypertension, and pain	Nursing assessment and dx.; physiologic evaluation and problem (eg, impaired mobility, incontinence, etc), KPCS[d]	Nursing assessment and dx.; physiologic evaluation and problem (eg, impaired mobility, incontinence, etc), KPCS
Cognitive factors	Dementia, delirium, disorientation, level of consciousness, fear, irritability, noncompliance	Nursing assessment or dx.; cognitive function (eg, acute confusion, disorientation, noncompliance, etc)	Nursing assessment or dx.; cognitive function (eg, acute confusion, disorientation, noncompliance, etc)
Behavioral factors	Fall history, sleep impairment	Presence of past falls, nursing dx. related to sleep	Presence of past falls, nursing dx. related to sleep
Therapeutics	Medications, adverse reaction to medications, catheter (IV[e]-line, tube, Foley), use of restraints	Medication list by class (sedatives, antidepressant, antiemetics, antipsychotics, antianxiety drugs, diuretics, antiepileptics, antihypertensives, analgesics, antiarrhythmics and NSAIDs[f]), Physician order of fluid injection, tube, Foley and restraints.	Medication list by class (sedatives, antidepressant, antiemetics, antipsychotics, antianxiety drugs, diuretics, antiepileptics, antihypertensives, analgesics, antiarrhythmics and NSAIDs), Physician order of fluid injection, tube, Foley and restraints.
Mitigating factors			
Universal fall precautions	Fall precautions on admission, regular rounds	Nursing interventions; safety education on admission, rounds per 2 hours	Nursing interventions; safety education on admission, rounds per 2 hours
Education and communication	Patient and caregiver education, presence of bedsitter, use of visual indicators, communicating fall risk status to care team	Nursing interventions; fall prevention education, presence of bedsitter, use of visual indicators, and activities communicating fall risk status to care team	Nursing interventions; fall prevention education, presence of bedsitter, use of visual indicators, and activities communicating fall risk status to care team
Observation and surveillance	Fall risk assessment tool	Hendrich II score and subscores [2]	STRATIFY[g] score and subscores [1],
Risk-target intervention	Cognitive and mental function	Nursing interventions: repeatedly provision of orientation, hourly rounding, assigning room close to nursing station, keep caregivers or family members on bed-side, etc.	Nursing interventions: repeatedly provision of orientation, hourly rounding, assigning room close to nursing station, keep caregivers or family members on bed-side, etc.
	Toileting problem	Nursing interventions: provision toilet scheduling, assist toileting, provision comodo or bed-pan, etc.	Nursing interventions: provision toilet scheduling, assist toileting, provision comodo or bed-pan, etc.
	Impaired mobility	Nursing interventions: provision of mobility devices, walking aids, and assistance, etc.	Nursing interventions: provision of mobility devices, walking aids, and assistance, etc.
	Medication review	Nursing interventions: rearranging medication time, provision side-effect precaution, etc.	Nursing interventions: rearranging medication time, provision side-effect precaution, etc.
	Sleep disturbance	Nursing interventions: attention to night movement and noise, inducing sleep pattern changes, etc.	Nursing interventions: attention to night movement and noise, inducing sleep pattern changes, etc.
Environmental intervention	Keeping paths clear, inspect furniture, equipment, lighting, floor, room arrangement	Nursing interventions; environmental targeted	Nursing interventions; environmental targeted

[a]EMR: electronic medical record.

[b]dx: diagnoses.

[c]ICD: International Classification of Diseases.

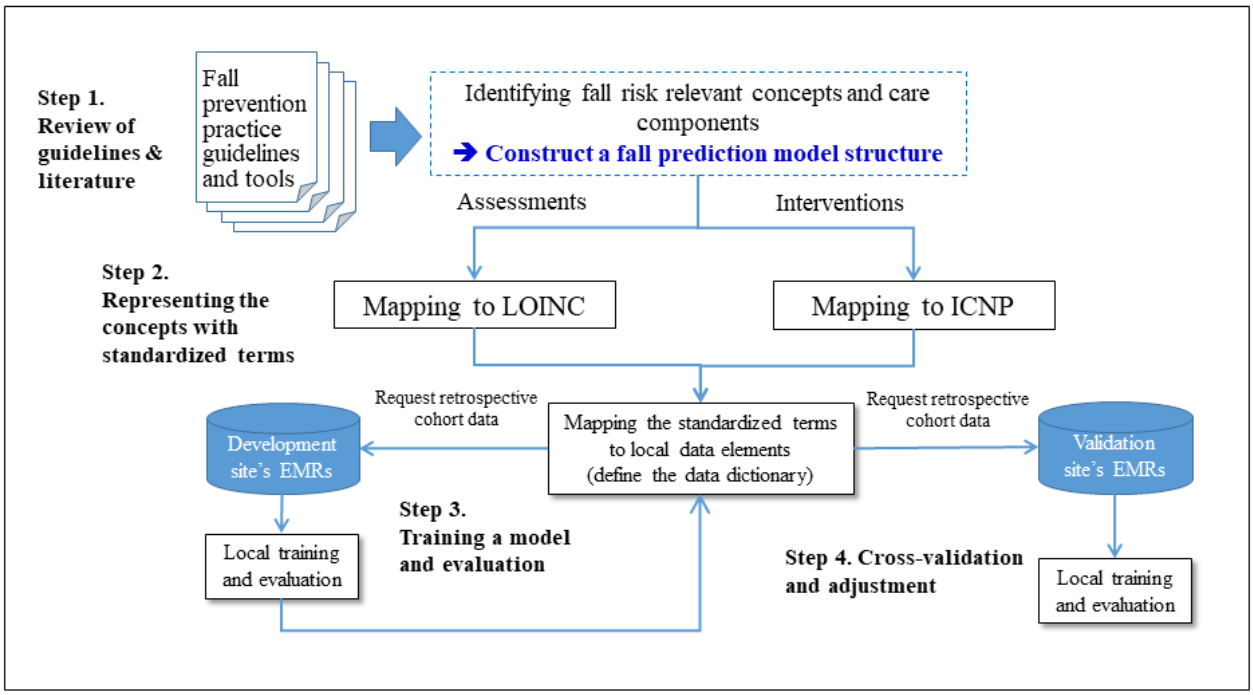
[d]KPCS: Korean Patient Classification System.

[e]IV: intravenous.

[f]NSAIDs: nonsteroidal anti-inflammatory agents.

[g]STRATIFY: St. Thomas' Risk Assessment Tool in Falling Elderly Inpatients.

Figure 1. The 4 steps of building a predictive Bayesian network model. LONC: Logical Observation Identifiers Names and Codes; ICNP: International Classification for Nursing Practice; EMR: electronic medical record.



Modeling Strategy

Our research team used the following principles to enable the prediction model translation into practice: (1) based on the existing nursing knowledge or clinical guidelines; (2) interpretable to users; and (3) parameterized to be adjusted and refined based on the target population's characteristics changing over time and sites. At the development site, we first constructed a concept model and, then, mapped the concept variables to local data elements, which followed by training with local cohort data. The same concept model was then applied to the validation site, and the model parameters were trained and tested by the local cohort.

The Bayesian network model was specified as follows. A Bayesian network or probability network B=(Pr, G) is a model of a multivariate probability distribution over a set of selected concept variables and consists of a graphical structure G and an associated distribution Pr [26]. The graphical structure takes the form of a directed acyclic graph G=(V(G), A(G)) with nodes $V(G)=\{v_1, v_2,\ldots, v_n\}$ and arcs $A(G) \subseteq V(G)\times V(G)$, where G represents a random variable that takes one of a finite set of values. The arcs in the graph present the probabilistic influences between the variables.

To build the Bayesian network model structure, we identified relationships between the concepts derived from the 8 fall prevention guidelines. The relationships, expressed with arcs in the network graph, were determined based on physiological, chronological, and logical processes. For example, the items of visual impairment, frequent toileting, transfer, and mobility from the STRATIFY 5 subscales closely relate to the data from nursing assessments. Furthermore, the Hendrich II 7 subscales have close relationships with medications, gender, medical diagnosis, as well as nursing assessments. These relationships were expressed in the network structure. The local conditional

probability distributions $Pr(V_i \mid \pi(V_i))$ (we call it parameters) for each variable V_i were obtained from each local (training) dataset. For the identified networks, the conditional probability distributions were computed on the basis of the weighted averages of probability estimates from the local dataset and a prior Dirichlet distribution, that is, multinomial distributions whose parameters can be interpreted as counts on the dataset:

where  is the probability distribution estimated from a given dataset D, Θ is the Dirichlet prior over the possible values of V_i, n is the size of the dataset D, and n_0 is the number of past cases on which the contribution of Θ is based.

Model Evaluation and Cross-Site Validation

The model prediction performance was assessed using sensitivity, specificity, receiver operating characteristics (ROC) curves, 10-fold cross-validation, and performance indices such as the spherical payoff [27]. In addition, the model was compared with the performance of 2 fall risk assessment tools (Hendrich II and STRATIFY) using calibration curves and ROC curves. A calibration curve does not quantitatively measure the reliability of probability predictions, but instead gives a graphical representation to capture the intuitive meaning of the calibration of a given system [28].

We performed a sensitivity analysis to establish the quality and clinical utility of the fully specified Bayesian network. We observed the output of the network to detect possible inaccuracies in the underlying probability distribution. We determined the degree to which variations in the posterior probability distributions were explained by other variables. The model sensitivity was calculated as the variance reduction with continuous variables and the entropy reduction with ordinal-scale or categorical variables. We used Netica modeling software (version 3.2, Norsys Software Corporation, Vancouver, Canada) to complete the analysis.

Statistical Analysis

Descriptive statistics on population profiles are presented as mean and SD or frequency and percentage values. Each cohort was compared using chi-square test or *t* test to quantify differences in the population characteristics. Statistical analyses were performed using R software (version 3.3, R Foundation for Statistical Computing, Vienna, Austria).

Results

Cohort Description

The 2 cohort populations had some differences in their characteristics (Table 2). The development-site patients were distributed almost equally across the age groups, but they had a longer length of stay than those in the validation site. The majority of the development-site patients (10309/14307, 72.05%) had a neoplasm or circulatory disease, and most of them also had secondary diagnoses. The validation-site patients were older and had more admissions for respiratory and gastrointestinal diseases and surgical procedures. However, no significant difference was noted in the frequency of falling; the total falls per 1000 hospital days were 1.95 and 1.69 at the development and validation sites, respectively (χ^2_1 =2.6; *P*=.11); the corresponding rates for injurious falls per 1000 hospital days were 0.44 and 0.40, respectively (χ^2_1=0.3; *P*=.58). As the rates of injurious falls were calculated only on the basis of data from the event-reporting system, they could have been underestimated due to missing reports [25]. Among the injurious falls at the development and validation sites, 91% (49/54) and 75% (52/69) were minor, respectively (χ^2_1=4.9; *P*=.03). No major injuries were reported.

Table 2. Characteristics of the two cohorts.

Characteristic	Development site (n=14,307)	Validation site (n=21,172)	χ^2 or *t* (*df*)	*P* value
Females, n (%)	6157 (43.03)	11,199 (52.90)	332.20[a]	<.001
Age in years, n (%)			629.0 (*4*)[b]	<.001
<50	3165 (22.12)	5593 (26.42)	N/A[c]	N/A
50-60	3251 (22.72)	3844 (18.16)	N/A	N/A
60-70	3356 (23.46)	3517 (16.61)	N/A	N/A
70-80	3281 (22.93)	5039 (23.80)	N/A	N/A
>80	1254 (8.76)	3179 (15.02)	N/A	N/A
Length of stay in days, mean (SD)	8.54 (11.52)	8.15 (11.28)	3.14[a]	.002
Medical diagnosis, n (%)			11,701.0 (*7*)[b]	<.001
Neoplasm	4639 (32.4)	4869 (23.00)	N/A	N/A
Benign	385 (2.7)	1066 (5.03)	N/A	N/A
Circulatory disorder	5670 (39.6)	769 (3.63)	N/A	N/A
Respiratory and gastrointestinal disorders	655 (4.6)	5630 (26.60)	N/A	N/A
Surgical procedure	517 (3.6)	2163 (10.22)	N/A	N/A
Neurological disorder	998 (7.0)	263 (1.24)	N/A	N/A
Infectious disorder	115 (0.8)	813 (3.84)	N/A	N/A
Other	1328 (9.3)	5599 (26.45)	N/A	N/A
Presence of secondary diagnosis, n (%)	14,242 (99.6)	13,421 (63.40)	6497.45[a]	<.001
Korean Patient Classification System[d] , n (%)			52.8 (*4*)[b]	<.001
Group 1	227 (1.59)	377 (1.78)	N/A	N/A
Group 2	8197 (57.29)	11,349 (53.60)	N/A	N/A
Group 3	3898 (27.25)	5630 (26.59)	N/A	N/A
Group 4	1627 (11.37)	1332 (6.29)	N/A	N/A
Groups 5 and 6	262 (1.83)	0 (0)	N/A	N/A
Number of medications daily, mean (SD)	2.5 (6.8)	18.6 (9.9)	−1835.04[a]	<.001
Total number of medications, mean (SD)	24.4 (75.7)	172.3 (317.7)	−63.07[a]	<.001
Fall events			4.7 (*1*)[b]	.09
One	231 (1.61)	284 (1.34)	N/A	N/A
Multiple	7 (0.05)	8 (0.04)	N/A	N/A

[a]χ^2.

[b]*t* (*df*).

[c]N/A: not applicable.

[d]Group 1 has the lowest nursing needs, while group 6 has the highest nursing needs.

Prediction Modeling at the Development Site

The fall prediction model identified at the development site consisted of 56 nodes and 82 links. The error rate of the prediction model was 11.7%, and the spherical payoff was 0.91. The calibration curves showing the relationship between observed and predicted outcome event rates divided into deciles revealed that the prediction reliability differed between the prediction model and the Hendrich II tool (Figure 2). The prediction model was imprecise at the 2 extreme probability ranges, with high probabilities underestimated and low probabilities overestimated; the Hendrich II tool (for a high-risk score of ≥5) showed a similar pattern.

Figure 3 (left side) shows the ROC curves created to determine the ability of the model to discriminate between at-risk and no-risk patients. The area under the ROC curve was 0.96 for the prediction, demonstrating almost perfect discrimination, while it was only 0.69 for the Hendrich II tool.

In the model development site, the sensitivity test showed that Hendrich II data reduced the variance the most (Figure 4, dark blue bars), followed in order by nursing assessments and diagnoses, nursing interventions, Korean Patient Classification System (KPCS) [29], and medications. The demographics and administrative data made virtually zero contributions. However, for the validation site, medication and KPCS contributed better to the variance reduction than nursing-process data.

Cross-Site Validation

The validation model consisted of 48 nodes and 80 links. The error rate was 4.87%. The logarithmic loss and spherical payoff were 0.13 and 0.96, respectively. These scores indicate the classification abilities of the model [30]. The logarithmic loss is a cross-entropy estimate that measures the additional penalty for using an approximation instead of the true model. Closer to 0 indicates a lower penalty [31]. The spherical payoff indexes performance of classification models, with 1 representing best classifier performance. The calibration curves in Figure 5 show that the lowest projected risk decile accounted for only 3.16% (448/14176) of the observed falls. The proportion of observed falls increased steadily with the projected risk, to reach 84.8% (480/566) in the highest-risk decile, while the curve for the STRATIFY tool did not exhibit a consistent increase. The prediction model showed a good calibration curve with better precision at extreme probability ranges; the STRATIFY has blunt calibration with a cutoff score of 2.

The area under the ROC curve was 0.99 and slightly higher than that for the development site model, which implies that the model performance was >30% higher than that of the STRATIFY tool (Figure 3, right graph). The results of the sensitivity analysis (Figure 4, light gray bars) showed that the medication and KPCS data had a greater influence on the occurrence of falls at the validation site.

Figure 2. Calibration curves for the prediction and Hendrich II models at the development site. The data are mean and 95% CIs.

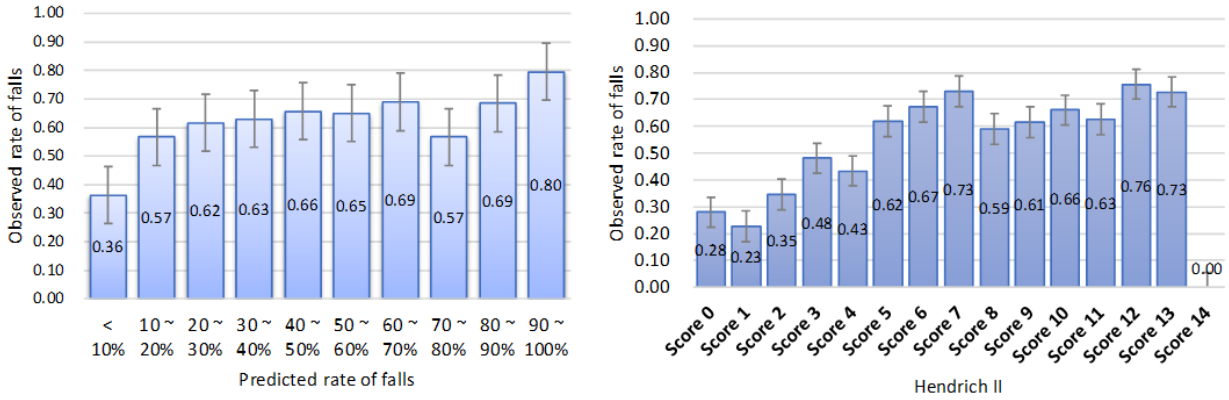


Figure 3. The receiver operating characteristics curves showing the discrimination ability in the fall prediction. AUC: area under the curve. STRATIFY: St. Thomas' Risk Assessment Tool in Falling Elderly Inpatients.

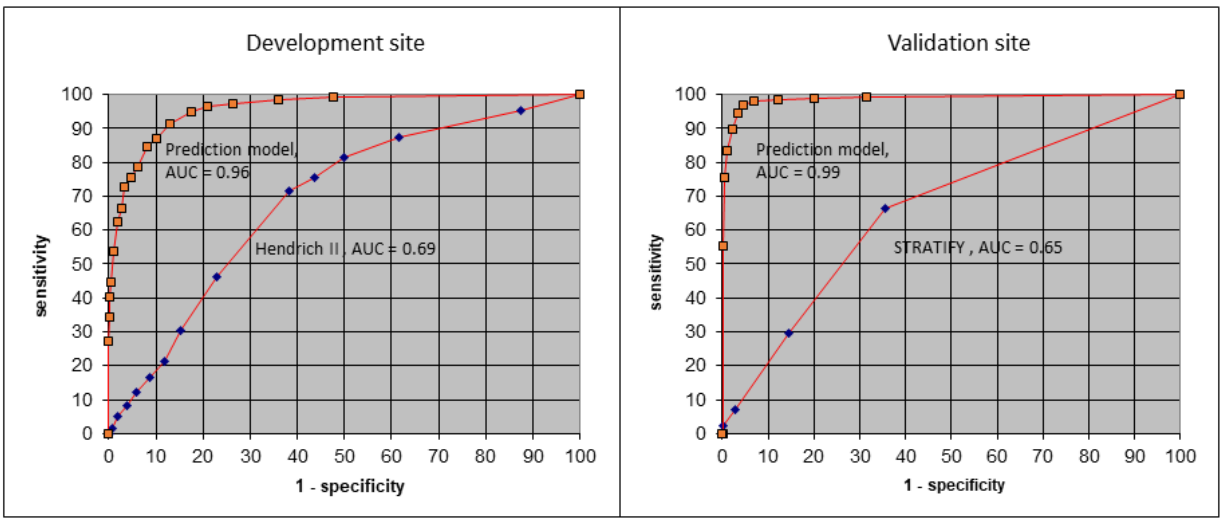


Figure 4. Results of the sensitivity analysis for subgroup summations of the prediction models. Dark-gray and light-gray bars correspond to the development and validation sites, respectively.

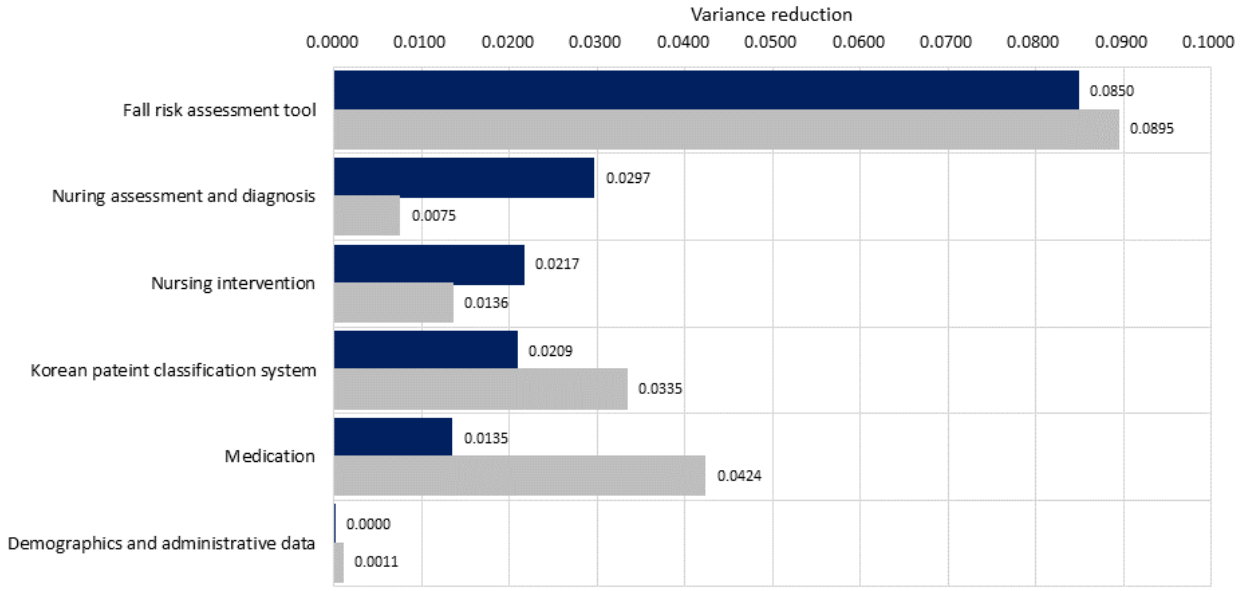
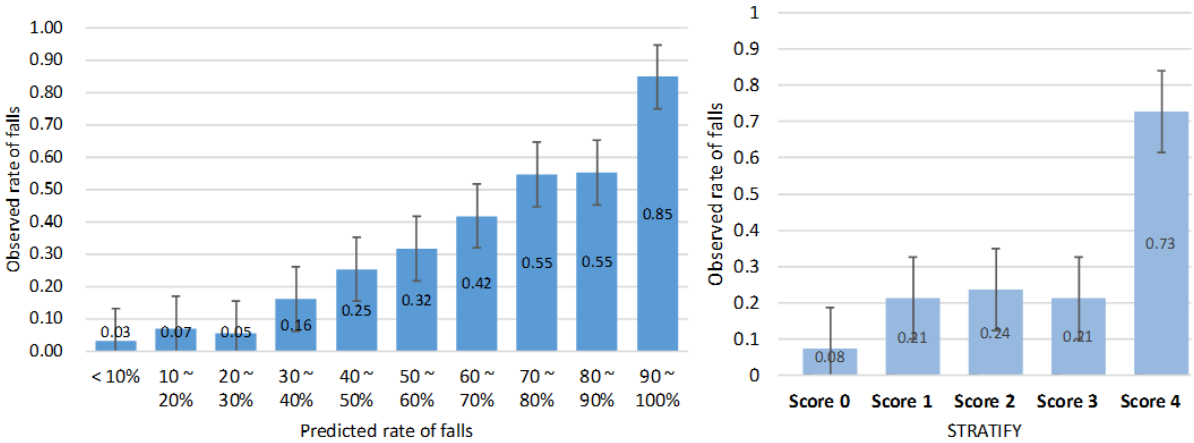


Figure 5. Calibration curves for the prediction model and St. Thomas' Risk Assessment Tool in Falling Elderly Inpatients (STRATIFY) tool at the validation site. The data are mean and 95% CIs.



Discussion

Principal Findings

We found that longitudinal EMR data could be incorporated successfully into a prediction model, which performed better at discriminating at-risk and no-risk patients than did the existing fall risk assessment tools alone. The EMR data included in the model were medication, patient classification (KPCS), the fall risk assessment tool, and the nursing-process (assessment, diagnoses, and intervention), demographics, and administrative data. The model exhibited acceptable performance at the 2 sites with different EMR systems, patient populations, fall risk assessment tools, and nursing terminology standards. In particular, semistructured EMR data (mostly nursing-process data) were semantically incorporated into a prediction model. These results imply that evidence-based prediction models that incorporated all relevant and time-variant data elements from an EMR system can be used as a more reliable guide for the fall risk assessment tools alone.

The 2 sites involved in this study have different patient profiles in terms of age, primary diagnosis, and medication distributions. However, the rates of falling and injurious falls at the 2 sites were similar. This finding is consistent with a study from the National Institutes of Health in 2013 [32] that involved 1263 hospitals across the United States. The authors [32] found no trend in the rates of falling or injurious falls according to the hospital size and staffing level. Except for unit type, the differences in fall rates within each organization characteristic ranged from 0.17 to 0.33 falls per 1000 hospital days.

We used all the data available in the EMR systems that are known to be relevant to inpatient falls based on clinical guidelines. One of the challenges in EMR-based studies is the presence of missing data [12,33]. For example, at the validation site we observed about 64.65% (124254/192183) missing data for the risk assessment tool score and 8.87% (17051/192183) for the KPCS score and subscores. The missing risk assessment tool score data were attributed to the hospital's local policy that specified reassessment period at 2 or 3 times a week for at-risk

patients and once a week for no-risk patients, as well as changes in the status. These local policies varied by each hospital's structural factors such as staffing level and patient-nurse ratio. In practice, a previous risk score is assumed to be valid until the day before the next reassessment. However, implicit assumptions have limitations from a data integrity perspective. The Bayesian inference is greatly advantageous for handling missing data and can produce accurate predictions even when complete data are not available [13]. The expectation-maximization algorithm that we used in the learning network performed automatic inference based on *a-priori* probabilities [13].

A key challenge when building predictive models from EMR data is handling nursing interventions. These interventions are confounders in that they can reduce the likelihood of a fall and, thereby, make it difficult to distinguish between patients who are at risk for falls based on their fall risk assessment score and those who are at risk, but their fall risk is mitigated by preventive interventions. Paxton et al [34] highlighted that not taking this masking into account may lead to models that are useless in practice. We, therefore, adopted a prognostic Bayesian network and noticed that the occurrence of falling for a specific patient is generally influenced by the sequence of preventive actions performed by nurses, which, in turn, may depend on the information that is available about the patient before any interventions aimed at preventing falls are implemented. Often, falls are also influenced by the underlying condition of a patient. We, therefore, formally defined a prediction as a probability distribution, *Pr* (falls | E, T), where E are the available patient data and T denotes nursing interventions provided by nurses.

The model developed in this study could be used to evaluate the performance and uncertainty of the Bayesian network. The c-statistic values of 0.96 and 0.99 found in this study were much higher than those found in studies of prediction models for mortality and clinical outcomes based on the EMR data (c-statistic=0.84 and 0.83, respectively) [12]. Our c-statistic values were assured through the testing set and 10-fold cross-validation, which supports the reliability of the performance of the models. In addition, the present c-statistic values were much higher than that in the study of Yokota and Ohe [8] (c-statistic=0.72), which developed a model for predicting the risk of falling based on EMR data. Yokota and Ohe's study [8] included physician-order items such as treatment directions, laboratory test and imaging findings, therapies, medications, and nursing assessment and plans. However, only items related to the intensity of nursing care needs with age and sex remained in their final regression model.

Another comparable study is that of Marier et al, who investigated fall prediction using the MDS and EMR data of 13 nursing home residents [9]. They compared 4 regression models and found that the rate of observed falls increased from 28.6% to 32.3% among residents in the highest-risk decile when EMR data were added to an MDS-only model. However, the report of that study did not include any model performance metrics such as c-statistic values.

The approach adopted in this study has several advantages over previously proposed methods for estimating the risk of falling. The first advantage relates to external validation, which is uncommon given that almost all studies have validated performance within the same EMR environment [12]. We conducted an external validation of the developed model at a second site with a different EMR system, patient population, fall risk assessment tool, and nursing terminology standard. For fall risk assessment, a substantial number of tools are readily available and widely used in hospitals. These tools assess many of the same areas of risk [35]. These findings suggest that our model is highly portable and comprehensive.

Second, this study incorporated >50 concepts mapped to 70 time-varying data elements, which represents a relatively large number of variable sets. We found only a small number of studies that used longitudinal EMR data, and they did not fully utilize the depth of information on patients available in the nursing records to identify predictor variables [8,10,11]. Instead, those studies used summary metrics or opted for smaller predefined lists. Considering the advantage that the size of EMR data is not limited to the number of patients or the number of potential predictor variables, integrating repeated observations over time is a key strength of this study's use of EMR data.

A third advantage of our approach relates to the incorporation of nursing-process data, including the fall prevention interventions provided to patients. It is difficult to find an EMR-based study that has integrated the nursing activities of assessments, diagnoses, and interventions—this was possible in this study because the 2 EMR systems included complete electronic nursing notes consisting of coded and standardized statements using locally developed data dictionaries [36]. In addition, we identified how the nursing activities captured by the EMR system affect the reduction in the variance of fall events. This finding showed that to accurately predict falls, the nursing data in EMR systems are as important as the individual risk factors of patients, implying that using readily available data for risk prediction may simplify computation. Early identification and more precise prediction of at-risk patients has the potential to improve outcomes by facilitating the timely initiation of appropriate and targeted attention, interventions, and monitoring.

Finally, using our model, we calculated for each patient, the daily estimate of their risk of falling. As the estimated probability ranges from 0% to 100%, users could set a cutoff of risk depending on an appropriate level of sensitivity and specificity.

The next steps involve implementing this approach more broadly and performing a prospective evaluation of the net benefits obtained by providing fall prevention nursing decision support in practice, as well as validating the model at other sites. For example, interventions tailored to patients' individual fall risk factors could be recommended in real time to them. We plan to incorporate a tailored intervention guide according to the individual risk factors of at-risk patients. This will be a great opportunity to explore how the algorithms impact the clinical decision making of nurses.

Conclusion

We found that a risk prediction model that utilizes longitudinal EMR data on nursing assessments, diagnoses, and interventions can improve the ability to identify individual patients who are at a high risk of falling. The prediction model has demonstrated portability and reliability and can, therefore, be applied across hospitals with different EMR environments. Current EMR systems—even suboptimal ones—can be leveraged for the secondary use of clinical data to prevent patients from falling.

Acknowledgments

We would like to thank Keoung-Hee Choi, Jungmee Han, and Won-Hee Park of medical information department of the 2 hospitals for helping us retrieve the raw data efficiently. In addition, we appreciate the clinical staff of Nursing Departments and graduate students involved in the concept mapping process of local data elements with standard vocabularies. This study was supported by a grant from the Korea Healthcare Technology R&D Project, Ministry for Health and Welfare, Republic of Korea (No. HI15C1089 and HI17C0809).

Conflicts of Interest

None declared.

Authors' Contributions

IC conceived and designed the study, supervised and contributed to the data analysis, interpreted the results, and drafted and revised the paper. EHB and EJC contributed to the study design, data acquisition, results interpretation, and paper revision. DB and PD substantially contributed to data interpretation and made critical revisions regarding the intellectual content.

References

1. The Joint Commission. 2015 Sep 28. Sentinel Event Alert 55: Preventing falls and fall-related injuries in health care facilities URL: https://www.jointcommission.org/sea_issue_55/ [accessed 2018-07-06] [WebCite Cache ID 70hNWlJGJ]
2. Oliver D, Britton M, Seed P, Martin F, Hopper A. Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: case-control and cohort studies. BMJ 1997 Oct 25;315(7115):1049-1053 [FREE Full text] [doi: 10.1136/bmj.315.7115.1049]
3. Hendrich AL, Bender PS, Nyhuis A. Validation of the Hendrich II Fall Risk Model: a large concurrent case/control study of hospitalized patients. Appl Nurs Res 2003 Feb;16(1):9-21. [doi: 10.1053/apnr.2003.YAPNR2] [Medline: 12624858]
4. Morse J, Morse R, Tylko S. Development of a Scale to Identify the Fall-Prone Patient. Can J Aging 2010 Nov 29;8(04):366-377 [FREE Full text] [doi: 10.1017/S0714980800008576]
5. Cameron ID, Gillespie LD, Robertson MC, Murray GR, Hill KD, Cumming RG, et al. Interventions for preventing falls in older people in care facilities and hospitals. Cochrane Database Syst Rev 2012 Dec 12;12:CD005465. [doi: 10.1002/14651858.CD005465.pub3] [Medline: 23235623]
6. Gillespie LD, Robertson MC, Gillespie WJ, Sherrington C, Gates S, Clemson LM, et al. Interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2012 Sep 12(9):CD007146. [doi: 10.1002/14651858.CD007146.pub3] [Medline: 22972103]
7. Westra BL, Latimer GE, Matney SA, Park JI, Sensmeier J, Simpson RL, et al. A national action plan for sharable and comparable nursing data to support practice and translational research for transforming health care. J Am Med Inform Assoc 2015 May;22(3):600-607. [doi: 10.1093/jamia/ocu011] [Medline: 25670754]
8. Yokota S, Ohe K. Construction and evaluation of FiND, a fall risk prediction model of inpatients from nursing data. Jpn J Nurs Sci 2016 Apr;13(2):247-255. [doi: 10.1111/jjns.12103] [Medline: 27040735]
9. Marier A, Olsho LEW, Rhodes W, Spector WD. Improving prediction of fall risk among nursing home residents using electronic medical records. J Am Med Inform Assoc 2016 Mar;23(2):276-282. [doi: 10.1093/jamia/ocv061] [Medline: 26104743]
10. Tescher AN, Branda ME, Byrne TJO, Naessens JM. All at-risk patients are not created equal: analysis of Braden pressure ulcer risk scores to identify specific risks. J Wound Ostomy Continence Nurs 2012;39(3):282-291. [doi: 10.1097/WON.0b013e3182435715] [Medline: 22552104]
11. Giles LC, Whitehead CH, Jeffers L, McErlean B, Thompson D, Crotty M. Falls in hospitalized patients: can nursing information systems data predict falls? Comput Inform Nurs 2006;24(3):167-172. [Medline: 16707948]
12. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017 Dec;24(1):198-208 [FREE Full text] [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
13. Lucas PJF, van der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. Artif Intell Med 2004 Mar;30(3):201-214. [doi: 10.1016/j.artmed.2003.11.001] [Medline: 15081072]
14. NDNQI Staff. National Database of Nursing Quality Indicators. 2010 May. Guidelines for Data Collection on the American Nurses Association's National Quality Forum Endorsed Measures: Nursing Care Hours per Patient Day Skill Mix Falls

Falls with Injury URL: http://www.k-hen.com/Portals/16/Topics/Falls/ANAsNQFspecs.pdf [accessed 2016-10-25] [WebCite Cache ID 70hROcY18]
15. Van Hulse J, Khoshgoftaar T, Napolitano A. Experimental perspectives on learning from imbalanced data. 2007 Presented at: Proceedings of the 24th international conference on Machine learning; June 20-24, 2007; Corvalis, Oregon, USA p. 935-942. [doi: 10.1145/1273496.1273614]
16. Korean Hospital Nurses Association. 2008. Fall prevention nursing practice guide and summary URL: http://khna.or.kr/bbs/linkfile/resource/khna_Fcare.pdf [accessed 2018-01-15] [WebCite Cache ID 70hRn7JeG]
17. Ganz D, Huang C, Saliba D, Shier V. Agency for Healthcare Research and Quality. Preventing Falls in Hospitals: A Toolkit for Improving Quality of Care URL: http://www.ahrq.gov/professionals/systems/hospital/fallpxtoolkit/index.html [WebCite Cache ID 70hRqm9lg]
18. ECRI Institute. Healthcare Risk Control: Actionable tools, evidence-based guidance, and practical strategies to improve patient safety and quality URL: https://www.ecri.org/components/HRC/Pages/default.aspx [accessed 2018-07-06] [WebCite Cache ID 70hRw3uHr]
19. Degelau JB, Bungum L, Flavin P, Harper C, Leys K, Lundquist L, et al. Institute for Clinical Systems Improvement. 2012. Prevention of Falls (Acute Care) URL: http://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_patient_safetyreliability_guidelines/falls/ [WebCite Cache ID 70hS1ai8O]
20. Boushon B, Nielsen G, Quigley P, Rutherford P, Taylor J, Shannon D, et al. Institute for Healthcare Improvement. 2012. How-to Guide: Reducing Patient Injuries from Falls URL: https://www.ihconline.org/media/cms/HowtoGuideReducingPatientInjuriesfromFalls_copy.pdf [accessed 2018-07-06] [WebCite Cache ID 70hSPoR0q]
21. The Joint Commission Center for Transforming Healthcare. Targeted Solutions Tool for Preventing Falls URL: http://www.centerfortransforminghealthcare.org/tst_pfi.aspx [WebCite Cache ID 70hSaMlyj]
22. U.S. Department of Veterans Affairs. VA National Center for Patient Safety URL: https://www.patientsafety.va.gov/professionals/onthejob/falls.asp [accessed 2018-07-06] [WebCite Cache ID 70hSfli5I]
23. Neily J, Quigley P, Essen K. U.S. Department of Veterans Affairs. 2015. Implementation Guide for Fall Injury Reduction: VA National Center for Patient Safety Reducing Preventable Falls and Fall-Related Injuries URL: https://www.patientsafety.va.gov/docs/fallstoolkit14/falls_implementation_%20guide%20_02_2015.pdf [accessed 2018-07-06] [WebCite Cache ID 70hQvmMRR]
24. Jung M, Cho I, Chae J, Kim J, Kim Y. Comparing Inpatient Falls Guidelines to Develop an ICNP. Stud Health Technol Inform 2017;245:1365. [Medline: 29295444]
25. Cho I, Boo E, Lee S, Dykes P. Automatic population of eMeasurements from EHR systems for inpatient falls. J Am Med Inform Assoc 2018 Jun 01;25(6):730-738 [FREE Full text] [doi: 10.1093/jamia/ocy018] [Medline: 29659868]
26. Pearl J. Probabalistic reasoning in intelligent systems. San Mateo, CA: Morgan Kaufmann Publishers; 1997.
27. Marcot B. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecological Modelling 2012 Apr;230:50-62 [FREE Full text] [doi: 10.1016/j.ecolmodel.2012.01.013]
28. Diamond G. What price perfection? Calibration and discrimination of clinical prediction models. Journal of Clinical Epidemiology 1992 Jan;45(1):85-89 [FREE Full text] [doi: 10.1016/0895-4356(92)90192-P]
29. Song KJ, Kim EH, Yoo CS, Park HO, Park KO. Development of KPCS (Korean Patient Classification System for Nurses) based on nursing needs. J Korean Clin Nurs Res 2009;15(1):5-17.
30. Morgan MG, Henrion M. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. New York, NY: Cambridge University Press; 1992.
31. Cowell R, Dawid A, Spiegelhalter D. Sequential model criticism in probabilistic expert systems. IEEE Trans Pattern Anal Machine Intell 1993;15(3):209-219. [doi: 10.1109/34.204903]
32. Bouldin ELD, Andresen EM, Dunton NE, Simon M, Waters TM, Liu M, et al. Falls among adult patients hospitalized in the United States: prevalence and trends. J Patient Saf 2013 Mar;9(1):13-17 [FREE Full text] [doi: 10.1097/PTS.0b013e3182699b64] [Medline: 23143749]
33. Kharrazi H, Wang C, Scharfstein D. Prospective EHR-based clinical trials: the challenge of missing data. J Gen Intern Med 2014 Jul;29(7):976-978 [FREE Full text] [doi: 10.1007/s11606-014-2883-0] [Medline: 24839057]
34. Paxton C, Niculescu-Mizil A, Saria S. Developing predictive models using electronic medical records: challenges and pitfalls. AMIA Annu Symp Proc 2013;2013:1109-1115 [FREE Full text] [Medline: 24551396]
35. Perell K, Nelson A, Goldman R, Luther S, Prieto-Lewis N, Rubenstein L. Fall Risk Assessment Measures: An Analytic Review. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 2001 Dec 01;56(12):M761-M766 [FREE Full text] [doi: 10.1093/gerona/56.12.M761]
36. Cho I, Park H. Evaluation of the expressiveness of an ICNP-based nursing data dictionary in a computerized nursing record system. J Am Med Inform Assoc 2006;13(4):456-464 [FREE Full text] [doi: 10.1197/jamia.M1982] [Medline: 16622170]

Abbreviations

EMR: electronic medical record
ICNP: International Classification for Nursing Practice

KPCS: Korean Patient Classification System
MDS: minimum dataset
ROC: receiver operating characteristics
SMOTE: synthetic minority oversampling technique
STRATIFY: St. Thomas' Risk Assessment Tool in Falling Elderly Inpatients

Edited by G Eysenbach; submitted 06.07.18; peer-reviewed by M Immonen, M Stein, L Montesinos; comments to author 06.09.18; revised version received 06.10.18; accepted 26.10.18; published 19.02.19.

Please cite as:
Cho I, Boo EH, Chung E, Bates DW, Dykes P
Novel Approach to Inpatient Fall Risk Prediction and Its Cross-Site Validation Using Time-Variant Data
J Med Internet Res 2019;21(2):e11505
URL: https://www.jmir.org/2019/2/e11505/
doi:10.2196/11505
PMID:30777849

Original Paper

Evaluating the Quality of Health Information in a Changing Digital Ecosystem

Alla Keselman[1], PhD; Catherine Arnott Smith[2], PhD; Anita C Murcko[3], MD, FACP; David R Kaufman[3], PhD, FACMI

[1]Division of Specialized Information Services, National Library of Medicine, National Institutes of Health, Bethesda, MD, United States

[2]University of Wisconsin-Madison, Madison, WI, United States

[3]Arizona State University, Scottsdale, AZ, United States

Corresponding Author:
Alla Keselman, PhD
Division of Specialized Information Services
National Library of Medicine
National Institutes of Health
6707 Democracy Boulevard
Bethesda, MD, MSC 5467
United States
Phone: 1 301 827 5671
Email: allagkeselman@gmail.com

Abstract

Background: Critical evaluation of online health information has always been central to consumer health informatics. However, with the emergence of new Web media platforms and the ubiquity of social media, the issue has taken on a new dimension and urgency. At the same time, many established existing information quality evaluation guidelines address information characteristics other than the content (eg, authority and currency), target information creators rather than users as their main audience, or do not address information presented via novel Web technologies.

Objective: The aim of this formative study was to (1) develop a methodological approach for analyzing health-related Web pages and (2) apply it to a set of relevant Web pages.

Methods: This qualitative study analyzed 25 type 2 diabetes pages, which were derived from the results of a Google search with the keywords “diabetes,” “reversal,” and “natural.” The coding scheme, developed via a combination of theory- and data-driven approaches, includes 5 categories from existing guidelines (resource type, information authority, validity of background information sources, objectivity, and currency) and 7 novel categories (treatment or reversal method, promises and certainty, criticisms of establishment, emotional appeal, vocabulary, rhetoric and presentation, and use of science in argumentation). The coding involves both categorical judgment and in-depth narrative characterization. On establishing satisfactory level of agreement on the narrative coding, the team coded the complete dataset of 25 pages.

Results: The results set included “traditional” static pages, videos, and digitized versions of printed newspapers or magazine articles. Treatments proposed by the pages included a mixture of conventional evidence-based treatments (eg, healthy balanced diet exercise) and unconventional treatments (eg, dietary supplements, optimizing gut flora). Most pages either promised or strongly implied high likelihood of complete recovery. Pages varied greatly with respect to the authors’ stated background and credentials as well as the information sources they referenced or mentioned. The majority included criticisms of the traditional health care establishment. Many sold commercial products ranging from dietary supplements to books. The pages frequently used colloquial language. A significant number included emotional personal anecdotes, made positive mentions of the word *cure*, and included references to nature as a positive healing force. Most pages presented some biological explanations of their proposed treatments. Some of the explanations involved the level of complexity well beyond the level of an educated layperson.

Conclusions: Both traditional and data-driven categories of codes used in this work yielded insights about the resources and highlighted challenges faced by their users. This exploratory study underscores the challenges of consumer health information seeking and the importance of developing support tools that would help users seek, evaluate, and analyze information in the changing digital ecosystem.

(J Med Internet Res 2019;21(2):e11129) doi:10.2196/11129

KEYWORDS

eHealth; eHealth literacy; type 2 diabetes mellitus; consumer health information; health literacy; information evaluation; information quality; information literacy

Introduction

Background

Evaluating the quality of online consumer health information has been a central issue in consumer health informatics for many years. However, with the emergence of new Web media platforms and the ubiquity of social media, the need for critical evaluation has taken on a new dimension and urgency. Individuals living with life-threatening and chronic diseases search the internet for treatment alternatives. Many such searches lead to sites containing nonevidence-based advice with targeted marketing and clickbait headlines.

Type 2 diabetes is a chronic disease that, according to the American Diabetes Association (ADA), affects 25 million Americans. Another 7 million may be undiagnosed, whereas approximately 80 million have impaired glucose tolerance, also known as *prediabetes*. These individuals have access to a plethora of online resources of highly variable quality. The objectives of this qualitative formative study are to develop a methodological approach for analyzing health-related Web pages likely to be viewed by individuals with type 2 diabetes and apply it to a set of relevant pages [1].

Quality of Online Health Information

For health information seekers, the World Wide Web can be a source for both valuable information and misinformation. In the early 1990s, Gordon Guyatt was credited with coining the term *evidence-based medicine* or EBM, reflecting the hierarchy of scientific evidence employed in the development of clinical advice [2]. In 1997, MEDLINE indexed the first study evaluating consumer health Web pages [3]. Published in British Medical Journal, the study reviewed 41 pages with advice on managing children's fever at home, concluding with an alarm that "only a few web sites provided complete and accurate information for this common and widely discussed condition."

Today, the problem persists, and the rapidly growing popularity of social media is making the problem of identifying quality information more pressing. For example, although Facebook has recently made an effort to reduce unsolicited commercial content that appears as news headlines, these are not typically vetted, checked for accuracy, or monitored in any way [4,5]. Many sites prey on vulnerable populations who may be receptive to promises of a quick and easy cure or an alternative to medical establishment recommendations.

Health-related misinformation may be considered on a continuum that ranges from deliberately deceitful with the intent to promote specious products to sites that may more benignly endorse a product or claims that lack scientific credibility. The epidemic of fake or controversial health news presents formidable challenges for consumers and health educators. It also provides interesting research opportunities for the consumer health informatics community.

Type 2 Diabetes Reversal as a Theme of Consumer-Targeting Websites

Type 2 diabetes is a serious disease that, over time, can damage small blood vessels and nerves, causing serious problems in the eyes, heart, brain, kidneys, skin, and feet, ranking in the top ten "killer" diseases in the United States [1]. This complex chronic disease starts with silent metabolic changes that precede symptoms and frank hyperglycemia (elevated blood glucose) by 7 to 10 years. With its ubiquitous and growing prevalence, type 2 diabetes is the focus of many consumer-targeting health information websites. A common theme is *reversal* of type 2 diabetes.

From the perspective of evidence-based medical authority (eg, as exemplified by the clinical practice recommendations of the ADA [6]), type 2 diabetes can be prevented, postponed, and placed into remission by lifestyle measures (diet, exercise, stress reduction) and therapy. However, except in rare and extreme circumstances, it cannot be cured or reversed. Medical literature references to diabetes *reversal* are sparse and limited to those on very low carbohydrate diets, postbariatric surgery, or in experimental animal models. Therefore, frank *reversal* of the complex metabolic derangements of type 2 diabetes is uncommon and quite difficult to accomplish. The lay use of the term *diabetes reversal* observed in the websites of interest, therefore, does not signal true medical reversals; instead, these are descriptions of diseases in remission characterized by reduction or discontinuation of medication. However, the ambiguity can lead readers to assume that reversal and cure are synonymous.

Barriers to Effectively Negotiating the World of Digital eHealth

The danger of inaccurate health information is heightened by the public's potential vulnerability to it. Factors that increase an individual's vulnerability vary from a desperate desire for cure to dissatisfaction with traditional health care and limited health literacy. Health literacy can be defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" [7]. The magnitude of the problem, affecting more than 90 million Americans, and its impact on health care and public health have been well established. Health literacy-related knowledge and skills are particularly deficient among vulnerable populations such as the elderly [8], disadvantaged youth [9], or people with lower levels of education. When individuals with lower health literacy conduct Web searches, they rely on criteria that do not correspond to the commonly cited quality guidelines when evaluating online health information [10]. For example, Mackert et al [11] showed that individuals low in health literacy used "position in search results, quality of pictures, celebrity endorsement, and website authorship as criteria to evaluate online health information."

Newer concepts of health literacy frame it in terms of complex multidimensional models with broad applications in medicine and public health. For example, Osborne et al [12] used a validity-driven approach to develop a 9-factor Health Literacy Questionnaire (HLQ) model of health literacy. Examples of the model's scales include abilities to establish relationships with health care providers, take responsibility for one's own health, and obtain social support. Sørensen et al [13] developed an integrative conceptual model of health literacy drawing on a systematic literature review of 17 definitions and 12 conceptual models. This Health Literacy Survey (HLS) model integrates medical and public health as well as individual and population views of health literacy [14]. The new models underscore the complex array of individual and sociocultural factors underlying health information functioning.

Whereas health literacy enablers and barriers operate across a range of situations, electronic health (eHealth) literacy has emerged as both a distinct construct and an area of research into competencies needed for successful functioning in the world of digital health information [15]. Norman and Skinner introduced an influential eHealth literacy model, which comprised the following 6 types of literacy:

1. *Computer literacy*: the skills to use computers productively.
2. *Information literacy*: the skills to articulate information needs; to locate, evaluate, and use information; and to apply information to create and communicate knowledge.
3. *Media literacy*: the ability to select, interpret, evaluate, contextualize, and create meaning from resources presented in a variety of visual or audio forms [16].
4. *Conventional literacy and numeracy*: reading comprehension and quantitative skills for interpreting information artifacts such as graphs, scales, and forms.
5. *Scientific literacy*: familiarity with basic biological concepts and the scientific method as well as the ability to understand, evaluate, and interpret health research findings using appropriate scientific reasoning.
6. *Health Literacy*: the acquisition, evaluation, and appropriate application of relevant health information as described previously.

Chan and Kaufman extended and applied this model to analyze how individuals use Web resources to answer questions across different health topics [17-19]. They noted that the most frequently encountered barriers related to information literacy. These resulted in failures to identify relevant links and cues on websites, locate relevant information, and evaluate the trustworthiness and credibility of health information [9,20,21]. On the other hand, the skills associated with media literacy, especially in the context of new media or social media are not well understood [22-24]. Similarly, health information studies typically do not address scientific literacy.

Scientific literacy, as it pertains to matters of health, has been an area of inquiry in educational research for some time [25]. The concept extends beyond knowledge of specific scientific concepts, also involving knowledge *about* science and attitudes *toward* science. Knowledge about science is essential for guiding individuals in recognizing questions that are suitable for scientific investigation (eg, what treatment works best for a given condition) as well as important features of investigations (eg, randomization, control, sample size). Understanding what science is about, referred to as *the nature of science*, and positive attitude toward science also lead laypeople to differentiate between theory and evidence, favor systematic evidence as a source of knowledge, understand possible causes of scientific controversy, and appreciate the importance of logical consistency in explanations [26]. Although scientific literacy competencies used in this broad definition are relevant to health information seeking and evaluation, the concept has not been given much attention in consumer health research.

While health literacy is often discussed in research as a characteristic of individuals, HLQ and HLS perspectives suggest that it can also be conceived as characterizing the relationship between an individual and a set of resources. For example, different ways of presenting numerical data or scientific explanations may impact a reader's ability to evaluate information, intentionally or unintentionally, for better or worse.

Information Quality Evaluation Guidelines

Concern over health information quality online has been present from the dawn of the World Wide Web era. The Health on the Net Foundation's code of conduct, launched in 1996, offered a set of best practice guidelines for website maintainers to follow [27]. To evaluate content itself, the DISCERN instrument was developed in 1996 and 1997 as a joint collaboration between the National Health Service and the British Library. DISCERN was a product of stakeholders chosen from across health care: generalist and specialist physicians, but also librarians and health communications specialists, self-help patient group representatives, medical publishers and journalists, and health services researchers [28].

DISCERN's creators designed it to support websites' evaluation by health information providers, serve as a checklist for content creators and a training tool for health care professionals, and most important, as a decision support for consumers who want to know more about a treatment they are using [29,30]. Although DISCERN was originally designed to target paper-based patient-facing leaflets, it can be used to evaluate any text-based information pertaining to treatment. This freely available instrument measures 16 items pertaining to markers of information quality (eg, reliability, relevance, balance, description of a treatment's risks and benefits). Today this Web 1.0 tool remains in use, with results reported in over 150 published studies suggesting that DISCERN rankings are similar regardless of whether they are given by patients or consumers, or health care professionals.

Study Objectives

This study aims to characterize health information sources in our ever-expanding digital ecosystem, including nonevidence-based pages presenting information about type 2 diabetes reversal. Specific objectives involve (1) reviewing top results pages in response to query about natural reversal of type 2 diabetes, (2) developing a methodological approach for capturing their essential content and informational characteristics, and (3) evaluating the utility of the approach with the above set of pages.

Methods

Page Selection

We started with reviewing a collection of non-EBM type 2 diabetes health advice Web pages that we have accumulated over the years of exploratory interest in the topic, noting frequent references to *reversal* and *natural* remedies. We then performed a search on Google using Firefox 61.0.2 browser on the first author's machine using the keywords *diabetes*, *reversal*, and *natural* and collected the first 3 pages of results (31 pages). After reviewing these links, we excluded those that did not pertain to type 2 diabetes, focused on animals, or required creating a password-protected log-in. This resulted in 24 pages. The 25th page, hyperlinked from one of the search results, came from our initial review cluster.

Coding Scheme Development and the Final Scheme

The coding scheme was developed using a combination of theory-driven and data-driven approaches [31]. First, the authors reviewed existing core information evaluation principles and criteria underlying various information science instruments, including the DISCERN instrument described above. The authors also attempted to code the pages using DISCERN, but found that it could not be justly applied across diverse information formats that included newspaper articles and patient testimonials. Afterwards, the authors conducted several rounds of review of 3 pages from the study set, noting and discussing perceived relevant characteristics. This resulted in the final coding scheme that included both pre-established categories (ie, represented in existing guidelines) and novel categories, as described in Table 1.

Our aim was to conduct detailed descriptive analysis that frames the user's information-seeking experience. Although most questions are phrased to require binary (yes or no) responses, coding also involved writing short narrative responses (eg, "although citations are not provided, studies are described with partial information that would enables studies to be found eventually"). Three team members coded the data. To establish qualitative intercoder agreement [32,33], 13 pages were reviewed by 2 coders (in different permutations), which was followed by iterative in-depth team review and discussion of the coding. While the narrative data were not amenable to inferential statistical analysis, the team found the degree of narrative agreement [33] satisfactory; disagreements were resolved via discussion and the narratives were merged. The remaining pages were each reviewed by 1 coder (AK). With the exception of the *information authority* and *objectivity* coding that involved reviewing the parent site's "About" page, all the coding was done based on the information within the page only. The pages were coded during July to November 2017, as they appeared at the time.

Table 1. Final coding scheme.

Source and category	Subcategories
Existing guidelines	
Resource type	• Digitized content (simultaneously published in traditional mass media) • Static Web pages • Web 2.0 content (Wikipedia, blogs, support groups, online communities, listservs, social networking sites, RSS feeds, and YouTube videos)
Information authority	• Is it clear who is responsible for the contents of the page? • What are the author's academic or professional credentials? • Is there a way of verifying the legitimacy of the organization, group, company, or individual authoring the content? • Is there any indication of the author's qualifications for writing on a particular topic? • Is there a sponsoring or hosting organization that is separate from the author?
Validity of background information sources	• Is there content that needs to be cited, but is not? • Are the sources for factual information clearly listed or cited so they can be verified in another source? (subsumes: are authors of testimonials and stated support verifiable?) • Is the information from sources known to be reliable? • Do citations or references actually support the information presented on the page? • Are there endorsements by celebrity nonexperts? • Who is referred to as "Dr" or "physician"? • Is there mention of "secret recipe" ("virtually unknown method") known only to the page's owners or promoters? • Is there a disclaimer on the page (what does it state)?
Objectivity	• Does the content appear to contain any evidence of bias? 1. Is the page selling a product? 2. Does the page encourage a certain action? 3. Does the page-supporting organization engage in lobbying or advocacy or encourage lobbying or advocacy? • Is there a link to a page describing the goals or purpose of the sponsoring organization or company? • If there is any advertising on the page, is it clearly differentiated from the informational content?
Currency	• Are there dates on the page to indicate when the page was written, when the page was first placed on the Web, or when the page was last revised?
Emotional appeal	• Does the page contain emotional testimonies or personal anecdotes? • Does the page contain disturbing photos or images of health care professionals and procedures?
Data-driven	
Treatment or reversal method	• What is the proposed diabetes treatment or reversal method?
Promises and certainty	• Does the page make a claim of having a solution (approach or product) producing results that are: 1. Quick 2. Painless or noninvasive or implemented via a simple procedure or with simple ingredients 3. Relatively inexpensive • Is there a promise of complete recovery for a condition that is known to be chronic or incurable?
Criticisms of establishment	• Is there implication or statement of conspiracy or purposeful misleading on the part of: 1. Pharmaceutical companies? 2. Doctors or conventional health care providers? 3. Government agencies? • Are there suggestions of media bias in covering relevant health issues? • Are there implications or statements that the reader's or viewer's doctor is incompetent? • Are there criticisms of biomedical research supporting the establishments' guidelines? (eg, methodology and research focus because of funding).
Vocabulary	• Does the page refer to *cure* or other words that are unlikely to be used in evidence-based medical literature (eg, *proven*)?

Source and category	Subcategories
Rhetoric and presentation	• Are there cliff-hangers in the content? For example, "In the next few minutes, I'm going to share with you the little-known natural remedy that will help you leave pills and needles behind forever..." • Is there an appeal to buy something right away? • Is the language very colloquial? • Is there a long speech that culminates in a request for money?
Use of science in argumentation	• Are biological mechanisms of diseases and treatments presented? • Are there claims that the coder perceives as: Exaggerated? False? Unverifiable? Mentioning controversial or not quite scientific concepts? Otherwise problematic? • Is there a contrast between claims about the complexity or uncertainty of the condition or treatment and the simplicity and certainty of the proposed solution? (statements that are too good to be true?)

Results

Types of Web-Based Information Resources

Of the 25 pages, 13 were "traditional" static pages, 8 were Web 2.0 sources or static pages with video components. Four pages were digitized versions of printed newspaper or magazine articles. Three were patient testimonials.

Diabetes Treatment or Reversal Methods

The number of remedies proposed by each page ranged from 1 to 7, with the mean of 2.24 (Table 2). The most commonly mentioned remedy involved taking dietary supplements. Recommended supplements included a range of herbs, vitamins, and minerals, with several mentions of cinnamon, turmeric, and chromium picolinate. These recommendations conflict with the ADA statement that "research has not been able to prove that dietary or herbal supplements (including omega-3 supplements, cinnamon, and other herbs) help to manage diabetes" [34]. The second most common recommendation was adhering to some general nutritional guidelines or healthy eating. General nutritional guidelines typically mentioned avoiding refined sugars and grains and eating more fiber and healthy fats. One page stated that a vegan diet was essential, describing an uncited study linking diabetes risk with consuming animal products.

Several pages promoted "superfoods." Unlike general healthy eating guidelines, *superfoods* recommendations focused on specific "healing properties" of a particular food (eg, grapefruit). A special nutritional protocol advertised by one of the pages promised "to kill the microbes and parasites (eg, pancreatic flukes) identified by the consultation." The undescribed protocol, apparently available from clinics promoted by the page, had to be followed by *electromedicine* said to "use gentle electrical waves to do the things necessary to rebuild the immune system."

Promises and Certainty

The specific promises made by the pages varied greatly. Fourteen of 25 pages either promised complete recovery or strongly implied that it was highly possible. In doing so, they often referred to *reversing* diabetes (eg, "type 2 diabetes is almost always reversible and this is almost ridiculously easy to prove"). Some expressed very high level of certainty: "If you follow our recommendations to the letter we guarantee that you will eventually be able to throw your medication away and never need it again!" Only 1 article, authored by a registered dietitian, discussed what it meant by *diabetes reversal*, explaining how *remission* is a more accurate term than *cure*.

Pages often promoted their approaches to "reversing" diabetes as quick, easy, and low cost. For example, 11 pages claimed to have a solution guaranteed to work within a specified period, from 11 days to 3 months. While some articles described difficult, extremely low-calorie regimens, 10 touted the ease of reversing diabetes (eg, "ridiculously simple"). Finally, 9 stressed that the treatments they proposed were inexpensive ("so inexpensive it might as well be free").

Table 2. Numbers of pages (N=25) proposing specific remedies.

Remedy	Pages mentioning the remedy, n (%)
Supplements	14 (56)
General nutrition guidelines or healthy eating	13 (52)
Exercise	8 (32)
Special nutritional protocol: specific *superfoods*	7 (28)
Caloric reduction or intermittent fasting	5 (20)
Stress reduction	3 (12)
Improved sleep	2 (8)
Electromedicine	1 (4)
Optimizing gut flora	1 (4)
Weight loss	1 (4)

Information Authority

Key information authority characteristics are summarized in Table 3. While 7 pages were authored or verified by physicians, other authors self-reported a range of qualifications for addressing the topic. Two authors were naturopaths. Several pages were written by journalists or patients giving testimonials. Of the remaining authors, 1 was a registered dietitian, 1 a self-described "clinical nutritionist" (certification not stated), and 1 an "ex-pharmaceutical chemist." In addition, 4 authors were identified as health coaches or health experts without listed medical credentials. For example, 1, a former professional athlete, self-described as "one of the most trusted health and fitness experts." Another self-identified as a health coach and a popular health and lifestyle reporter. However, another was described as "a catalyst voice" for alternative treatments and a founder of an independent health research foundation. Finally, 2 of the stated authors emphasized their lay relationship to the content, referring to themselves as average folk or concerned parents.

Validity of Background Information Sources

The 25 sources varied in how they cited and validated the information they presented. As standards of providing citations differ greatly across various information types, it is not surprising that 20 pages contained uncited mentions of studies and data that could not be easily found based on those mentions. For example, phrases such as "studies show" were made without references or hyperlinked pointers to the studies. A representative example is the unreferenced statement that "a number of clinical studies have been carried out in recent years that show potential links between herbal therapies and improved blood glucose control." Sources also made unreferenced statements that are not currently endorsed by leading relevant authority such as the ADA. For example, 1 page stated that "the solution to curing type 2 diabetes lies with killing the microbes and parasites inside the organs," and that diabetes can be caused by "hepatitis c virus." Another claimed that "cinnamon can curb the current epidemic of type 2 diabetes," without references to any existing studies.

Sixteen out of 25 pages explained some of their information sources, either in the form of citations and references or by providing enough descriptive details so that the sources could be located with relative ease. Of these, 12 included sources that were deemed to be authoritative or reliable (eg, ADA, National Institute of Diabetes and Digestive and Kidney Diseases, and publications in journals listed in MEDLINE). When authoritative or reliable sources were included, they were used to support the specific statements to which they were linked in all but 1 case. Three of the pages not explaining their background sources claimed to have some sort of secret or "virtually unknown" recipe for treating diabetes.

Objectivity

All pages were in the dot-com domain. Nineteen included a link to a page describing the site's or the sponsoring organization's goal. These goals varied in specificity, but typically had to do with information provision. Only 1 page called for legislative advocacy and encouraged readers to take actions such as petitioning the Department of Veteran's Affairs to "employ licensed naturopathic physicians."

Ten sites sold products ranging from alternative treatments and supplements to books and films. One had a paid access section. Some others did not sell products directly but contained links to the author's books for sale or fee-for-service practice. Fourteen pages contained advertisements, which, in 4 cases, were not clearly differentiated from the page's content.

Currency

In 21 out of 25 cases, the page included a date indicating when the page was written, the information was first placed on it, or the page was copyrighted.

Criticisms of the Establishment

Of the 25 sources, 14 made critical remarks about the pharmaceutical and health care establishment. Of these, 13 suggested malevolent intent or conspiracy on the part of various establishment agents (Table 4). Pharmaceutical companies received the greatest amount of criticism. For example, 1 page stated that "the pharmaceutical industry is a gigantic machine which has to sustain itself" and asked, "why would these companies be at all interested in truly reversing diabetes? How would that benefit them financially?" Another wrote, "Most big pharma companies don't know squat about how to reverse your diabetes."

Table 3. The pages' information authority characteristics (N=25).

Information authority characteristic	Pages, n (%)
Content has identifiable author(s)	22 (88)
Existence and legitimacy (accuracy of self-identification) of the author verified[a]	17 (68)
Sponsoring or hosting organization separate from the author[b]	11 (44)
Content authored or verified by someone described as a credentialed physician[c]	7 (28)

[a]As evidenced by a detailed on-site biography and/or external Web presence (eg, profiles in LinkedIn and online directories and business listings).

[b]For example, a newspaper or magazine, an association, and a public television channel.

[c]Stated MD (Medical Doctor) or DO (Doctor of Osteopathy) degree.

Table 4. Alleged malevolent intent or conspiracy agents (N=25).

Malevolent intent or conspiracy agents	Pages, n (%)
Pharmaceutical companies	7 (28)
Doctors	5 (20)
Biomedical research	3 (12)
Media	1 (4)
Other or unspecified	6 (24)

Two pages, 1 with a disclaimer, made explicit claims that diabetes medications recommended by health professionals are extremely dangerous and should be avoided (eg, "Avoid these 3 doctor recommended treatments: oral medication, insulin therapy, [and] other injectables"). One page suggested that the ADA dietary guidelines "seem to serve the medical practitioners more than the patients." Another criticism warned readers that they had been "lied to for years." In additions to claims of malevolence, 5 pages suggested medical doctors' incompetence in diabetes management because of their lack of familiarity with nutrition and herbal medicine (eg, "doctors get little if any formal nutrition training in medical school").

Emotional Appeal, Vocabulary, Rhetoric, and Presentation

Fourteen pages included language that was judged by the coders as *very colloquial*, appearing to aim for establishing commonalities and rapport with the reader or viewer. For example, 1 page stated, "First I want you to know that here at [Company Name] we're really a lot like you. Average people who just so happened to be committed to helping people." Several pages included emotional personal anecdotes such as stories of family members suffering from diabetes complications. As *natural* was part of the query used to identify pages for the study, it is not surprising that positive mentions of *natural remedies* and *natural treatment* were common. Fourteen pages included these terms; 10 made positive references to *cure*. Pages referred to nature and "Mother Nature" as a wise positive "healing" force, regaling the reader with "natural remedies," "natural healing," "natural processes," and "natural cures." Three pages (videos) included the infomercial approach of a long speech culminating in a direct request for an immediate purchase.

Use of Science in Argument—Explanatory Mechanisms Behind Treatments

Of the 25 pages, 22 presented some biological explanations of their proposed treatment mechanism(s), with most pages including more than one. Often, pages combined mechanisms and methods widely accepted in standard care (eg, described by the ADA) with more uncertain and controversial ones. Depth of explanations ranged widely, from simple statements that a specific method (eg, a supplement) "improves sugar metabolism" to detailed explanations of intracellular molecular mechanisms, in our view, well beyond the level of comprehension of an educated layperson (eg, "Acetic acid protects the liver by increasing tolerance of lipogenesis and fatty acid synthesis responsible for improving cholesterol levels").

The following explanatory mechanisms were particularly prominent:

- "*Unclogging" liver and pancreas for normal insulin production*: A number of pages recommending low-calorie diets or intermittent fasting explained that this method "unclogs" fat from liver and pancreas, thus restoring them to normal functioning essential for insulin production and glycemic control.
- *Reducing blood glucose and improving glucose metabolism (without mentioning insulin)*: Many pages explained their treatment methods (eg, specific foods and supplements) by stating that these methods "reduce[d] blood glucose," "improve[d] glucose tolerance factor," or "help[ed] metabolize glucose," without mentioning insulin. Three pages that promoted exercise mentioned that it builds muscle that burns more glucose. In addition, a number of pages included a biological explanation of how foods high in sugar or refined carbohydrates created spikes in blood glucose (sugar) levels and needed to be avoided. These pages also often explained that foods high in fiber were beneficial because they slowed down glucose absorption.
- *Improving insulin secretion and insulin sensitivity*: A number of pages stated that foods or supplements implicated in their treatment methods influenced production or secretion of insulin, improving insulin resistance or sensitivity, and, in some cases, "mimicked insulin."
- *Reducing inflammation*: Several pages related diabetes to inflammation, stating that food or supplements described by them fought it or increased "good bacteria" in the intestinal lining.
- *Strengthening cells and organs*: Several pages related diabetes to weakened immune system, "weakened organs" (in particular, liver and pancreas), and "weakened cells." They proposed that their methods "strengthened" cells and organs. Mechanisms ranged from supplements that "help strengthen the cellular signal" to "electrical waves" that "kill parasites and microbes that weaken organs."

Discussion

Web pages about natural treatment of diabetes analyzed in this study proposed a number of reversal methods and differed greatly in terms of their alignment with accepted standard of care recommendations, promises, levels of certainty, authors' background, transparency of sources, rhetoric, style, and attitude toward pharmaceutical and medical establishment. This exploratory study only looked at a small sample of pages pertaining to a single topic collected from 1 search on 1 machine. However, it underscores the challenges of consumer

health information seeking and the importance of developing support tools that would help users seek, evaluate, and analyze information in the changing digital ecosystem. Future work within this research program will focus on extending the approach to a number of domains, developing more robust evaluation criteria, and exploring computational approaches to pages' analysis.

Is the Consumer Health Web Universe a Dangerous Place?

Consequences of following recommendations promoted by the sources analyzed in this study are likely to vary. The core lifestyle modifications recommended by many pages, namely, exercise, weight loss, stress reduction, and a healthy balanced diet are evidence-based components of conventional type 2 diabetes regimes. Very low–calorie diets and intermittent fasting may be beneficial, but because of the known risks (including death), they require medical supervision, especially for those with diabetes requiring medication. Supplements (including cinnamon), superfoods, and optimization of gut flora, promoted by many pages, have a low health risk but little-to-no proven benefit and can be financially draining. However, as illustrated in Table 2, most pages repackage the core lifestyle recommendations and add-on *essential* product purchases, muddying the water. The stakes for harm are even higher when pages promising to reverse diabetes undermine the use of medications, including recommending unsupervised medication discontinuation, and promote an antiscience attitude.

Fit With Existing and Added Health Information Evaluation Criteria

One of the objectives of this work was to develop a methodological approach for analyzing the digital health information sources in the era of online videos and social media. The study suggests that both traditional and data-driven categories of codes (see Table 1) yielded insights about the resources and highlighted challenges faced by their users.

Existing Evaluation Criteria Categories

Information Authority

Assessing authority of information authors and sponsors was straightforward, except for the case of determining the ownership of a YouTube channel. In the majority of cases, pages analyzed in this study had clear authorship indicators, with the authors having sufficient Web presence to lend credibility to their stated identity and credentials. While a sizable minority of pages (7 of 25) was authored or verified by credentialed physicians, the majority were created by noncredentialed individuals.

Validity of Background Information Sources

Pages in this sample frequently described or mentioned scientific studies without providing references that would allow their unambiguous identification. They also typically did not reference their biological explanations or statements about treatments that were not aligned with ADA guidelines, developed upon an extensive review of the scientific evidence from peer-reviewed sources [35]. However, the same pages typically included citations of some external sources with background information about type 2 diabetes. These were often high-quality authoritative sources.

While assessing validity of cited information sources was straightforward, the expression of this criterion varied for different publication formats because of their differing conventions. For example, bibliography style references are difficult to present in videos and uncommon in newspaper articles where a detailed description of a background study is a more likely quality indicator. Still, the criterion of validity remains highly relevant and the lack of credible citations or pointers, or a mismatch between citations and their purported claims, raises concerns about information quality.

Objectivity

Across the range of information source types, this criterion was unambiguous. Page's or sponsor's goals, typically stated on the site's About page, as well as the information about sales of relevant products and services, provided information helpful for judging objectivity. Selling services and products such as supplements by default indicated their endorsement.

Information Currency

Currency turned out to be a challenging criterion because of the range of events that could be time-stamped on the pages. Although most of the pages had a time stamp, these were more likely to be the dates of the page's copyright than of information authorship.

Novel Evaluation Criteria Categories

Treatment or Reversal Method and Use of Science in Argumentation

These categories are discussed together because scientific argumentation usually explained the treatment methods. For this sample, the *use of science in argumentation* code proved valuable for elucidating the challenges facing health information seekers. The pages typically provided some biological information, claiming effects of substances or procedures on insulin production, glucose metabolism, and cells/organs/microorganisms. Often, these explanations blended widely accepted biological mechanisms with controversial ones. While many pages limited their biology to simple causal statements such as "cinnamon improves insulin sensitivity," others employed complexity well beyond the level of comprehension of an educated layperson. Many of the non-ADA-aligned treatment methods and biological mechanisms mentioned on the pages had corresponding coverage in peer-reviewed science literature, albeit scientific literature described them as more controversial, less certain, and limited to a narrow range of application (eg, demonstrated effects limited to animal studies or lacking adequate controls). Applying the *use of science in argumentation* code underscores the formidable challenge of supporting lay assessment of plausibility of online health information. It also suggests that the science literacy component of eHealth literacy deservers greater attention.

Promises

A promise of recovery from a chronic disease, with a high level of certainty and implied treatment simplicity, was a useful indicator of concern about a page's quality. Promises were

especially troubling when presented as guaranteed within a specific time frame.

Criticisms of Establishment

As a criterion, this one is easy to apply, revealing, and provides disconcerting information. The importance of the code is illustrated by the disturbing number of pages making critical comments in our sample (14 of 25). Criticisms of health care–related establishment are disconcerting because they attempt to discredit primary sources of evidence-based care, positioning nonevidence-based methods as primary, rather than complementary. In making critical claims, pages often had to tread the line between denouncing some official sources and yet conveying respect for science and evidence and support from some studies. The public's response to the representation of science, doctors, and pharmaceutical organizations in the digital ecosystem merits further examination.

Emotional Appeal, Vocabulary, Rhetoric, and Presentation

General level of colloquialism and informality and the use of personal or emotional anecdotes seemed to be most related to document type, with more informal language used in videos and newspaper articles. Specific high certainty words such as *cure* and *guaranteed*, on the other hand, are potential quality signifiers that merit further research.

Scientific Literacy as a Dimension of eHealth Literacy Revisited

As mentioned earlier, this study underscores the importance of scientific literacy as a component of eHealth literacy. The role of science knowledge in daily life has long been debated in the fields of public health and science education. This study illustrates that scientific literacy, although important, should not be equated with content knowledge. It is often unrealistic to expect laypeople to have biomedical knowledge necessary to analyze the argument behind controversial treatment methods. Although some claims may be refuted by high school biology (eg, diabetes is a disease of "weak organs"), many remain difficult to evaluate even after a thorough analysis of a PubMed search. In such cases, the relevant aspects of science literacy are not specific content knowledge, but understanding the nature of science and scientific evidence, uncertainty, and the process of biomedical discovery. Such knowledge is likely to trigger skepticism about overgeneralizations, oversimplifications, and exaggerations inherent in many consumer-targeting pages that promise quick and easy fixes for complex health problems. Complexity of science literacy also underscores the importance of promoting traditional information evaluation criteria such as source authority and objectivity.

Implications for Research and Practice: Ways to Support Researchers and Consumers

The methodological approach, described in this study, is a coding scheme designed expressly for research purposes, rather than as a tool for evaluating Web pages. Much more work, aimed at expanding, fine-tuning, simplifying, and validating the criteria, is needed before this approach can produce a numerical score that could be used to assess a Web page. Such a tool could be very valuable to both researchers and every day information seekers. It would be particularly beneficial if developed and validated for a wide range of online information sources, including blogs, message boards, videos, and other social media platforms.

In addition to providing an evaluation guide, medical and informatics organizations may investigate developing resources that address common controversial claims. The establishment often ignores nonevidence-based treatment recommendations, despite their visibility in the public domain. For example, the consumer portion of the ADA site dedicates very little space to a discussion about cinnamon. A thorough respectful explanation of why the use of cinnamon should be treated with caution may be a more effective way to help consumers. New models of health literacy, such as HLQ and HLS [12-14], that consider complex sociocultural determinants of information behaviors (eg, root causes of different attitudes toward medical establishment) could surface explanations as to why certain groups find specific messages to be compelling. It could also inform the presentation and content of explanations that counter potentially deleterious messages. Computation-based informatics tools may also play a role in helping users evaluate Web pages. This study suggests that certain terms and phrases, particularly those indicating high confidence and rejection of traditional medicine, may be alert markers. Research and development into automated language-based categorization may be useful in flagging suspect pages. Finally, this study suggests the importance of science education for the development of science literacy and the potential synergy between classroom science and health informatics.

Acknowledgments

The work was in part supported by the intramural research program of the National Library of Medicine.

Authors' Contributions

AK, CAS, and DRK participated in conceptualizing the study, developing the coding scheme, coding the data, and writing the manuscript. ACM participated in conceptualizing the study, reviewing the coding scheme, and writing the manuscript.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. 2017. National Diabetes Statistics Report, 2017: Estimates of Diabetes and its Burden in the United States URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2018-12-04] [WebCite Cache ID 6zU9QIxRP]
2. Guyatt GH. Evidence-based medicine. ACP J Club 1991;114:A16. [doi: 10.7326/ACPJC-1991-114-2-A16]
3. Impicciatore P, Pandolfini C, Casella N, Bonati M. Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home. Br Med J 1997;314(7098):1875-1879 [FREE Full text] [Medline: 9224132]
4. Lazer D, Baum M, Grinberg N, Friedland L, Joseph K, Hobbs W, et al. Harvard Kennedy School, Shorenstein Center on Media, Politics and Public Policy. 2017. Combating fake news: An agenda for research and action URL: https://shorensteincenter.org/wp-content/uploads/2017/05/Combating-Fake-News-Agenda-for-Research-1.pdf [accessed 2018-12-04] [WebCite Cache ID 74Puz2SZr]
5. Schwitzer G. Pollution of health news. Br Med J 2017 Dec 15;356:j1262. [doi: 10.1136/bmj.j1262] [Medline: 28298320]
6. American Diabetes Association. Clinical Practice Recommendations URL: http://www.diabetes.org/research-and-practice/we-support-your-doctor/clinical-practice-recommendations.html [accessed 2018-09-17] [WebCite Cache ID 72VNNw4FW]
7. Institute of Medicine (US) Committee on Health Literacy. In: Kindig DA, Panzer AM, Nielsen-Bohlman L, editors. Health literacy: a prescription to end confusion. Washington, DC: National Academies Press; 2004.
8. Sharit J, Hernández MA, Czaja SJ, Pirolli P. Investigating the roles of knowledge and cognitive abilities in older adult information seeking on the web. ACM Trans Comput Hum Interact 2008;15(1):3 [FREE Full text] [doi: 10.1145/1352782.1352785] [Medline: 20011130]
9. Subramaniam M, St Jean B, Taylor NG, Kodama C, Follman R, Casciotti D. Bit by bit: using design-based research to improve the health literacy of adolescents. JMIR Res Protoc 2015;4(2):e62 [FREE Full text] [doi: 10.2196/resprot.4058] [Medline: 26025101]
10. 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;99(6):1017-1025. [doi: 10.1016/j.pec.2016.01.007] [Medline: 26817407]
11. Mackert M, Kahlor L, Tyler D, Gustafson J. Designing e-health interventions for low-health-literate culturally diverse parents: addressing the obesity epidemic. Telemed J E Health 2009;15(7):672-677 [FREE Full text] [doi: 10.1089/tmj.2009.0012] [Medline: 19694596]
12. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
13. Sørensen K, van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, HLS-EU Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
14. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European Health Literacy Survey (HLS-EU). Eur J Public Health 2015;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
15. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
16. Thoman E. Media literacy education can address the problem of media violence. In: Dudley W, editor. Media violence: Opposing viewpoints. San Diego: Greenhaven Press; 1999:131-136.
17. Chan CV, Matthews LA, Kaufman DR. A taxonomy characterizing complexity of consumer eHealth literacy. AMIA Annu Symp Proc 2009 Nov 14;2009:86-90 [FREE Full text] [Medline: 20351828]
18. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
19. Chan C, Mirkovic J, Furniss S, Kaufman DR. eHealth literacy demands and cognitive processes underlying barriers in consumer health information seeking. KM&EL 2015;7(4):550-575.
20. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
21. Manafò E, Wong S. Assessing the eHealth literacy skills of older adults: a preliminary study. J Consum Health Internet 2012;16(4):369-381. [doi: 10.1080/15398285.2012.701163]
22. Sørensen K. The need for “Health Twitteracy” in a postfactual world. Health Lit Res Pract 2017;1(2):e86-e89. [doi: 10.3928/24748307-20170502-01]
23. Hayat TZ, Brainin E, Neter E. With some help from my network: supplementing eHealth literacy with social ties. J Med Internet Res 2017;19(3):e98 [FREE Full text] [doi: 10.2196/jmir.6472] [Medline: 28360024]
24. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]

25. Roberts DA. Scientific literacy, science literacy, and science education. In: Lederman NG, Abell SK, editors. Handbook of Research on Science Education, Volume II. New York: Routledge; 2014:545-559.
26. Bybee R, McCrae B, Laurie R. PISA 2006: an assessment of scientific literacy. J Res Sci Teach 2009;46(8):865-883. [doi: 10.1002/tea.20333]
27. Boyer C, Baujard V, Nater T, Scherrer J, Appel R. QOI6/416: the Health On the Net Code of Conduct for medical and health-related web sites: three years on. J Med Internet Res 1999 Sep 19;1(Suppl 1):e99. [doi: 10.2196/jmir.1.suppl1.e99]
28. Discern. The DISCERN Project 1996-7 URL: http://www.discern.org.uk/original_discern_project.php [accessed 2018-05-17] [WebCite Cache ID 6zUNKTprC]
29. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999;53(2):105-111 [FREE Full text] [Medline: 10396471]
30. Dueñas-Garcia OF, Kandadai P, Flynn MK, Patterson D, Saini J, O'Dell K. Patient-focused websites related to stress urinary incontinence and pelvic organ prolapse: a DISCERN quality analysis. Int Urogynecol J 2015;26(6):875-880. [doi: 10.1007/s00192-014-2615-3] [Medline: 25578872]
31. Whitney W, Dutcher GA, Keselman A. Evaluation of health information outreach: theory, practice, and future direction. J Med Libr Assoc 2013;101(2):138-146 [FREE Full text] [doi: 10.3163/1536-5050.101.2.009] [Medline: 23646029]
32. Smith CA, Hundal S, Keselman A. Knowledge gaps among public librarians seeking vaccination information: a qualitative study. J Consum Health Internet 2014;18(1):44-46. [doi: 10.1080/15398285.2014.869168] [Medline: 26550002]
33. Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. Int J Qual Meth 2002;1(2):13-22. [doi: 10.1177/160940690200100202]
34. American Diabetes Association. Herbs, Supplements and Alternative Medicines URL: http://www.diabetes.org/living-with-diabetes/treatment-and-care/medication/other-treatments/herbs-supplements-and-alternative-medicines/ [accessed 2018-05-17] [WebCite Cache ID 6zUPuDyXu]
35. American Diabetes Association. Introduction: Standards of Medical Care in Diabetes—2018. Diabetes Care 2018;41(Supplement 1):S1-S2 [FREE Full text] [doi: 10.2337/dc18-Sint01]

Abbreviations

ADA: American Diabetes Association
EBM: evidence-based medicine
eHealth: electronic health
HLQ: Health Literacy Questionnaire
HLS: Health Literacy Survey

Edited by M Focsa; submitted 24.05.18; peer-reviewed by L Kayser, P Ghezzi; comments to author 20.07.18; revised version received 20.09.18; accepted 25.10.18; published 08.02.19.

Please cite as:
Keselman A, Arnott Smith C, Murcko AC, Kaufman DR
Evaluating the Quality of Health Information in a Changing Digital Ecosystem
J Med Internet Res 2019;21(2):e11129
URL: https://www.jmir.org/2019/2/e11129/
doi:10.2196/11129
PMID:30735144

Original Paper

Toward an Objective Assessment of Implementation Processes for Innovations in Health Care: Psychometric Evaluation of the Normalization Measure Development (NoMAD) Questionnaire Among Mental Health Care Professionals

Christiaan Vis[1,2], MA; Jeroen Ruwaard[3,4], PhD; Tracy Finch[5], PhD; Tim Rapley[6], PhD; Derek de Beurs[7], PhD; Henk van Stel[8], PhD; Britt van Lettow[9], PhD; Mayke Mol[3,4], MSc; Annet Kleiboer[1,2], PhD; Heleen Riper[1,2,4], PhD; Jan Smit[3,4], PhD

[1]Department of Clinical, Neuro-, & Developmental Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Mental Health, Amsterdam Public Health Research Institute, Amsterdam, Netherlands

[3]Amsterdam UMC, Vrije Universiteit, Psychiatry, Amsterdam Public Health Research Institute, Amsterdam, Netherlands

[4]Research and Innovation, GGZ inGeest Specialized Mental Healthcare, Amsterdam, Netherlands

[5]Department of Nursing, Midwifery & Health, Northumbria University, Northumbria, United Kingdom

[6]Department of Social Work, Education & Community Wellbeing, Northumbria University, Northumbria, United Kingdom

[7]Mental Health, Netherlands Institute For Health Services Research (NIVEL), Utrecht, Netherlands

[8]Julius Center Research Program Methodology, Department of Public Health, Healthcare Innovation & Evaluation and Medical Humanities, University Medical Center Utrecht, Utrecht, Netherlands

[9]Nictiz, The Hague, Netherlands

Corresponding Author:
Christiaan Vis, MA
Department of Clinical, Neuro-, & Developmental Psychology
Faculty of Behavioural and Movement Sciences
Vrije Universiteit Amsterdam
Van der Boechorststraat 7
Amsterdam, 1081 BT
Netherlands
Phone: 31 646098094
Email: p.d.c.vis@vu.nl

Abstract

Background: Successfully implementing eMental health (eMH) interventions in routine mental health care constitutes a major challenge. Reliable instruments to assess implementation progress are essential. The Normalization MeAsure Development (NoMAD) study developed a brief self-report questionnaire that could be helpful in measuring implementation progress. Based on the Normalization Process Theory, this instrument focuses on 4 generative mechanisms involved in implementation processes: coherence, cognitive participation, collective action, and reflexive monitoring.

Objective: The aim of this study was to translate the NoMAD questionnaire to Dutch and to confirm the factor structure in Dutch mental health care settings.

Methods: Dutch mental health care professionals involved in eMH implementation were invited to complete the translated NoMAD questionnaire. Confirmatory factor analysis (CFA) was conducted to verify interpretability of scale scores for 3 models: (1) the theoretical 4-factor structure, (2) a unidimensional model, and (3) a hierarchical model. Potential improvements were explored, and correlated scale scores with 3 control questions were used to assess convergent validity.

Results: A total of 262 professionals from mental health care settings in the Netherlands completed the questionnaire (female: 81.7%; mean age: 45 [SD=11]). The internal consistency of the 20-item questionnaire was acceptable (.62≤alpha≤.85). The theorized 4-factor model fitted the data slightly better in the CFA than the hierarchical model (Comparative Fit Index=0.90, Tucker Lewis Index=0.88, Root Mean Square Error of Approximation=0.10, Standardized Root Mean Square Residual=0.12,

χ^2_2=22.5, P≤.05). However, the difference is small and possibly not outweighing the practical relevance of a total score and subscale scores combined in one hierarchical model. One item was identified as weak ($\lambda_{CA.2}$=0.10). A moderate-to-strong convergent validity with 3 control questions was found for the Collective Participation scale (.47≤r≤.54, P≤.05).

Conclusions: NoMAD's theoretical factor structure was confirmed in Dutch mental health settings to acceptable standards but with room for improvement. The hierarchical model might prove useful in increasing the practical utility of the NoMAD questionnaire by combining a total score with information on the 4 generative mechanisms. Future research should assess the predictive value and responsiveness over time and elucidate the conceptual interpretability of NoMAD in eMH implementation practices.

(J Med Internet Res 2019;21(2):e12376) doi:10.2196/12376

KEYWORDS

implementation science; eHealth; psychometrics; eMental health; normalization process theory; implementation assessment

Introduction

Background

More than 2 decades of research has shown that psychotherapy delivered through the internet, also referred to as eMental Health (eMH) interventions, can be an effective way to treat patients with common mental disorders such as depression and anxiety disorder [1]. Several examples of clinics routinely offering innovative and new eMH services exist, such as the Australian MindSpot clinic [2], GGZ InGeest Mindway [3] and Interapy in the Netherlands [4], Internetpsykiatr in Sweden [5], and Internetpsykatrien in Denmark [6,7]. Despite these examples, and although the technical infrastructure seems to be in place, large-scale use of eMH interventions in routine care is still lower than expected [8]. Knowledge on factors hindering or facilitating implementation is maturing [9,10]. However, measuring implementation outcomes reliably remains a challenge [11,12]. We conducted a psychometric validation study of a recently developed theory-informed implementation measurement instrument: the Normalization MeAsure Development (NoMAD) questionnaire.

Theoretical Underpinning

Various frameworks and theories for understanding implementation processes and evaluating outcomes exist [13,14]. For example, models such as the Knowledge-to-Action model [15] have been specifically designed to describe and guide implementation processes. Determinant frameworks such as the Consolidated Framework for Implementation Research (CFIR) [16] provide taxonomies of barriers and hindering factors to aid the evaluation of implementation outcomes. Similarly, the Reach Effectiveness-Adoption Implementation Maintenance framework [17] summarizes key indicators for implementation success to inform policy and decision making. Classic psychological behavior change theories such as the Theory of Planned Behavior [18] have been used to study the role of attitudes and intentions in the behavior of individuals involved in and affected by implementation processes. Although such theories can be useful in describing behavior change mechanisms and explaining how change in individuals involved in implementation processes occurs, they do not necessarily consider what people actually do when implementing innovations in health care practice but rather focus on beliefs and attitudes. The Normalization Process Theory (NPT) [19,20] aims to fill this void by specifically looking at the process of implementation.

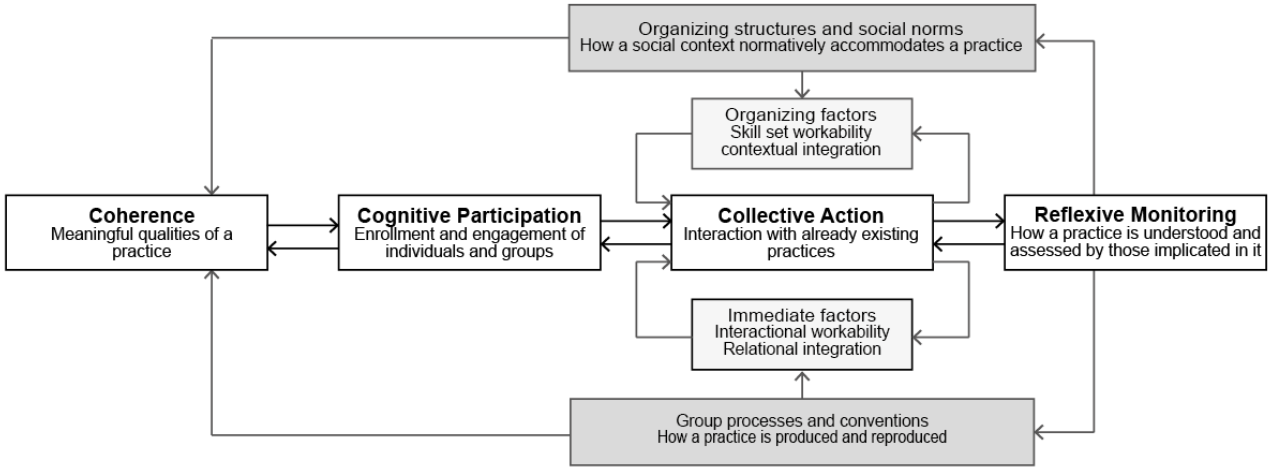
NPT (Figure 1 [19]) states that a normalization process is a process of embedding and integrating health care innovations in routine care as a product of action of individuals and groups. It focuses on the things that people individually and collectively do to normalize an innovation, that is, for it to become part of routine health care practice. NPT is a heuristic tool to understand the work of implementation, embedding, and integration of new practice and the contribution and roles of individuals and groups to this work. According to the theory, 4 mechanistic constructs play a central role in generating the work of implementation:

- Coherence (CO) of the innovation with the goals of daily routine. Individuals and groups go through a process of sense-making to establish the meaningfulness of the innovation for normal service delivery goals and practices.
- Cognitive participation (CP) as a process of enrollment and engagement of individual participants and groups involved in the implementation processes, through which they become committed to the normalization of the innovation.
- Collective action (CA) by individuals and groups to apply the innovation in daily routine. Here, applying an innovation has certain implications as to what and how normalization should be achieved, which requires investments of effort by the participants.
- Reflexive monitoring (RM) through which participants in the implementation process evaluate and appraise the use of the innovation in practice.

These 4 constructs are influenced by group processes and social conventions as well as the organizational factors and social structures people operate in. In turn, this social and organizational context defines factors that promote or inhibit the work of individuals and collectives in implementing innovations in daily routines.

Earlier work showed that NPT has good face validity in designing and evaluating implementation processes of innovations [21]. A recent literature review of 108 studies indicated that NPT successfully aids in the conceptual understanding of implementation processes and outcomes across a wide variety of health care settings [22].

Figure 1. Conceptual model of Normalization Process Theory (NPT): 4 constructs situated in a social and organizational context.



In alignment with the general approach of NPT, the NoMAD study developed a brief self-report questionnaire for the purpose of determining factors likely to affect normalization processes [23-25]. Ultimately, the questionnaire aims to enable (1) assessment of progress toward normalization over time in an implementation project and (2) comparison of normalization (progress or outcomes) between sites in multicenter studies. The NoMAD is intended to be used by people involved in the implementation of innovations in a health care setting and aims to be neutral to the implementation object. The target populations of the instrument are the deliverers and facilitators of the innovation being implemented, such as medical specialists, general practitioners, therapists, nurses, administrators, and managers.

A pool of 46 construct items was generated, appraised, and validated in 5 UK and 1 Australian samples of health care staff (n_{pooled}=413) involved in 6 different implementation projects [23-25]. A psychometric evaluation of the initial item pool resulted in a 20-item questionnaire of which the theoretical model approximated the data acceptably and appeared to have good internal consistency (total Normalization Process Scale (NPS): alpha=.89, CO: alpha=.71, CA: alpha=.78, CP: alpha=.81, RM: alpha=.65) [25].

Objective

We translated the questionnaire into Dutch and aimed to confirm the theoretical factor structure in mental health professionals working to implement eMH in Dutch mental health care settings. We tested 3 factor structures: (1) A 4-factor model to confirm the theorized model, summarizing item scores per construct; (2) A unidimensional model to test whether the items in the questionnaire can be summarized by 1 single factor score; and (3) A hierarchical model to test whether the 4 first-order factors can be expressed in 1 second-order factor. Where the first model aims to capture a more detailed view on implementation processes, the second model might lend support for practical comparison of those processes. The third model might provide a more detailed understanding of normalization processes on the construct level combined with the practical value of the overall total normalization score in 1 measurement model. Conforming to the English validation study [23-25], we explored potential improvements and the questionnaire's convergent validity with 3 control questions.

Methods

Sample and Recruitment

Using a cross-sectional design, mental health professionals with various occupational backgrounds involved in implementing eMH in Dutch routine mental health care practices were invited to complete the NoMAD questionnaire. We defined involvement in implementation as the situation in which respondents were in the early stages of using eMH in their occupational tasks. By this, novelty to the respondent in applying such interventions in routine care was assumed. Following the English NoMAD study, an open sampling strategy was applied to obtain a sample of 300 respondents. Considering the commonly applied rule of thumb of 7 to 10 complete cases per item with a minimum of 100 complete cases, we expected this target sample size to provide satisfactory statistical power and precision for estimating the model's parameters [26,27]. Recruitment targeted mental health professionals involved in using novel eMH interventions in (1) primary care for patients with mild symptomatology (general practitioners or general practice–based mental health nurse specialists), (2) basic care for patients with moderate symptomatology, and (3) specialized care provided by specialists to patients with severe mental health complaints. A total of 3 groups of Dutch mental health professionals were identified as suitable for recruitment:

- Group 1: mental health care professionals in 4 large regional mental health organizations for common mental disorders and post-traumatic stress disorders.
- Group 2: general practice–based mental health nurse specialists, in the context of the national electronic health (eHealth) Monitor survey conducted in 2016 for which panels and profession associations were sampled [28].
- Group 3: attendees at the annual cognitive behavioral therapy (CBT) congress held in the Netherlands in 2016, which attracted a nationwide audience of mental health professionals.

A total of 3 different recruitment strategies were applied. Sample 1 was obtained through convenience sampling by which

participants were recruited through key contact persons in various mental health organizations. Sample 2 was obtained through existing respondent panels and professional associations in the context of the national eHealth survey. Participants for samples 1 and 2 were invited by email providing general information about the study, a link to more in-depth information, and an anonymous link to the Web-based survey. Sample 3 was recruited through an information kiosk and leaflets at the annual CBT conference.

Translation

The classical Brislin approach to questionnaire translation [29] was used to translate the English NoMAD questionnaire into Dutch. A small (N=3) sample of experts in implementing and using eMH interventions were asked to verbalize their thoughts while interpreting the translated items in a cognitive group interview [30]. The interview focused on the interpretation of the questions, the response scales, and the identification of terms that needed to be adjusted and/or rephrased. Problematic items were rephrased to form the final version of the Dutch NoMAD instrument. Back translation by a blinded professional translator confirmed equivalence of semantic meaning of the corresponding individual items by the principal investigator (TF) of the English NoMAD. The final Dutch translation of the questionnaire is included in Multimedia Appendix 1.

Data Collection

The questionnaire was administered via a commercial Web-based survey system (NETQ Internet Surveys 6.5 [31]). The research team tested the survey for sequencing of the items, technical reliability, and data export procedures. Participants were asked to provide consent for using their (anonymized) data in this study. They provided this digitally through the survey platform before they were allowed access to the survey.

Normalization MeAsure Development Questionnaire

The NoMAD questionnaire in this study consisted of 3 parts: Part A tapping basic demographic information, Part B collecting general normalization ratings about the current use and likelihood of using the intervention in the future, and Part C comprising 20 items measuring the 4 NPT constructs. Users of the questionnaire are required to tailor the implementation object (ie, intervention) to the context of its application. In this study, the terms (the intervention) were replaced with “eMental health.”

Part A: Demographic variables. In line with the English NoMAD, basic demographic variables were included in the first part of the questionnaire, including gender, age, years of working experience, professional job category, and relevant care sector.

Part B: General normalization items. Part B contained 3 questions addressing perceptions of respondents regarding past, current, and future normality of the intervention. The 3 questions were scored on a 1 to 10 Visual Analogue Scale [32]. To increase comparability to the UK study, these 3 items were added to the questionnaire as control questions to assess its convergent validity, that is, the 3 questions are not to be regarded as an integral part of the core of the NoMAD questionnaire [23-25].

Part C: NPT constructs. Part C consisted of the 20 items representing the NPT constructs in 4 subscales with the following allocation: CO: 4 items; CP: 4 items; CA: 7 items; and RM: 5 items. The 20 original items are listed in Textbox 1.

The items were rated on a 5-point Likert scale (1=completely agree to 5=completely disagree), with an additional response option to indicate if a statement was applicable (0=not applicable). Item 2 (CA.2) in the CA scale is negatively formulated; all other items were formulated in a positive sense. Respondents were required to rate all statements.

The Dutch translations are in Multimedia Appendix 1.

Scoring

Scale scores were calculated by taking the mean of answered items of a scale. A minimum of 2 items within a scale had to be rated to calculate a scale score. Items rated as “not applicable” were excluded from the calculation. The total NPS score was calculated by taking the mean of all answered items for which complete cases were considered to have less than 15% missing data.

Data Analyses

Descriptive statistics were calculated to summarize the item and scale scores. Internal consistency of the total score and the 4 theoretical constructs were analyzed by calculating the Cronbach alpha [27] for the pooled dataset. The quality of the construct structure was further assessed by applying a confirmatory factor analysis (CFA) using Structural Equation Modelling (SEM). A total of 3 models were evaluated: (1) the theorized 4-factor model, (2) a unidimensional model, and (3) a hierarchical model. All 3 models included the 20 items from Part C of the questionnaire. The items were scored on a 5-point Likert scale resulting in an ordinal ordering of the data. The sum scale score of the items approximates a continuous scale by which we expected the latent constructs to be normally distributed. The CFA was run with the robust Weighted Least Square Means and Variances (WLSMV) estimator using polychoric correlation matrices [26]. Model fit was assessed by estimating the misfit between the observed and implied covariance matrices using the chi-squared test ($\chi^2 \leq 3df$). This was supplemented with 4 other fit estimators to strengthen the basis for our conclusions: the Standardized Root Mean Square Residual (SRMR≤0.08) as an absolute index of the average discrepancy between the correlations in the implied model and the observed data; the Root Mean Square Error of Approximation (RMSEA≥0.95) providing a population-based goodness-of-fit indication corrected for model complexity; the Comparative Fit Index (CFI≥0.95) providing an index of goodness-of-fit relative to a null model (ie, no covariances between items); and the Tucker Lewis Index (TLI≥0.95) as an index of goodness-of-fit relative to a null model corrected for model complexity [26,33,34]. The 3 models under evaluation are expected to be nested. We applied the scaled chi-square difference test (χ^2_{diff} test, analysis of variance) to compare the fit of the 3 models [26].

Textbox 1. Normalization MeAsure Development (NoMAD) questionnaire part C items.

Coherence (CO):

1. CO.1. I can distinguish [the intervention] from usual ways of working.
2. CO.2. Staff in this organization have a shared understanding of the purpose of [the intervention].
3. CO.3. I understand how [the intervention] affects the nature of my own work.
4. CO.4. I can see the potential value of [the intervention] for my work.

Cognitive participation (CP):

1. CP.1. There are key people who drive [the intervention] forward and get others involved.
2. CP.2. I believe that participating in [the intervention] is a legitimate part of my role.
3. CP.3. I'm open to working with colleagues in new ways to use [the intervention].
4. CP.4. I will continue to support [the intervention].

Collective action (CA):

1. CA.1. I can easily integrate [the intervention] into my existing work.
2. CA.2. [the intervention] disrupts working relationships.
3. CA.3. I have confidence in other people's ability to use [the intervention].
4. CA.4. Work is assigned to those with skills appropriate to [the intervention].
5. CA.5. Sufficient training is provided to enable staff to implement [the intervention].
6. CA.6. Sufficient resources are available to support [the intervention].
7. CA.7. Management adequately support [the intervention].

Reflexive monitoring (RM):

1. RM.1. I am aware of reports about the effects of [the intervention].
2. RM.2. The staff agree that [the intervention] is worthwhile.
3. RM.3. I value the effects [the intervention] has had on my work.
4. RM.4. Feedback about [the intervention] can be used to improve it in the future.
5. RM.5. I can modify how I work with [the intervention].

Potential improvements to the factor structure were explored by identifying low item-factor loadings (λ<0.3) to ensure that items are meaningfully related to the respective factors [26]. Modification indices (modification index [MI], χ^2_{diff}≥3.84) were assessed to identify item-item error covariances that might improve the model fit.

In the absence of a gold standard for the assessment of normalization, we exploratively used the 3 general normalization items (part B) to assess the convergent validity of the theorized model. We assessed the Pearson correlation coefficients for all 4 constructs and general normalization items and applied the following strength indicators for the correlations: 0≤*r*<.3 is weak, .3≤*r*<.5 is moderate, and *r*≥.5 is strong [35]. These quality indicators were applied in all correlation assessments.

Data cleaning and analyses were performed in RStudio [36,35] using the following packages: psych [37], ggplot2 [38], sjPlot [39], lavaan [40], semPlot [41], and semTools [42].

Ethical Approval and Consent to Participate

Ethical and scientific approval was granted by the Scientific and Ethical Review Board of the Faculty of Behavioural and Movement Sciences at the VU Amsterdam (file number: VCWE-2016-006).

Results

Sample

Over a period of 10 months (May 2016 to February 2017), 262 respondents completed the questionnaire. Table 1 provides an overview of the samples and participant characteristics. On a pooled level, participants were middle-aged (*M*= 45, SD=11), female (81.7%), and had over 11 years working experience in their respective fields (52.9%). The response rate for group 2 was 22.8% (125 out of 547) [28]. For sample groups 1 and 3, response rates are not available because of the convenience and open sampling approach. The time required to complete the questionnaire was 7.56 min on average (SD=6.48, n=134, based on questionnaire log files).

Scale scores

Figure 2 shows the distributional characteristics of the scale scores for the combined samples. The 4 subconstructs (CO, CP, CA, and RM) and the NPS follow similar response patterns.

Considering the length of the boxplot for the scales, respondents vary less in responses to items for the CO construct and more for CP and CA. The distributions of 3 subscales appear to have a slight tendency toward agreement with item statements where CA received mostly neutral responses. Most outliers are in the disagreement end of the scales.

Table 2 shows the mean scale scores, indicating that respondents on average agreed with the item statements. Respondents disagreed considerably with item CA.2, indicating that they did not find the intervention disruptive to working relations (Figure 3).

Table 1. Sample composition and demographics of respondents of the Dutch Normalization MeAsure Development questionnaire.

Variable	Pooled	Group 1[a]	Group 2[b]	Group 3[c]
Cases, n (%)	262 (100.0)	115 (43.9)	125 (47.7)	22 (8.4)
Age (years), mean (SD)	46.4 (11)	41.5 (10.7)	48.6 (10.1)	43.1 (11)
Gender, n (%)				
Female	214 (81.7)	91 (79.1)	108 (86.4)	15 (68.2)
Work experience (years), n (%)				
<1	4 (1.7)	3 (1.2)	1 (0.4)	0 (0)
1-2	16 (6.6)	3 (1.2)	13 (5.4)	0 (0)
3-5	46 (19.0)	19 (7.9)	27 (11.2)	0 (0)
6-10	48 (19.8)	21 (8.7)	19 (7.9)	8 (3.3)
11-15	32 (13.2)	17 (7.0)	13 (5.4)	2 (0.8)
>15	96 (39.7)	36 (14.9)	52 (21.5)	8 (3.3)
Sector[d], n (%)				
PC-MH[e]	135 (51.5)	12 (4.6)	122 (46.6)	2 (0.4)
BC-MH[f]	35 (13.4)	20 (7.6)	8 (3.1)	7 (2.7)
SC-MH[g]	114 (43.5)	97 (37.0)	0 (0.0)	17 (6.5)

[a]Group 1: mental health care professionals in large regional mental health organizations.

[b]Group 2: general practice–based mental health nurse specialists.

[c]Group 3: mental health professionals attending the annual national cognitive behavioral therapy (CBT) congress.

[d]Sector: respondents could choose multiple answers: primary care-mental health services, basic care-mental health, and specialist care-mental health.

[e]PC-MH: primary care-mental health services.

[f]BC-MH: basic care-mental health.

[g]SC-MH: specialist care-mental health.

Figure 2. Boxplot of the scale scores for the combined mental health samples. CA: collective action; CO: coherence; CP: cognitive participation; NPS: normalization process scale; RM: reflexive monitoring.

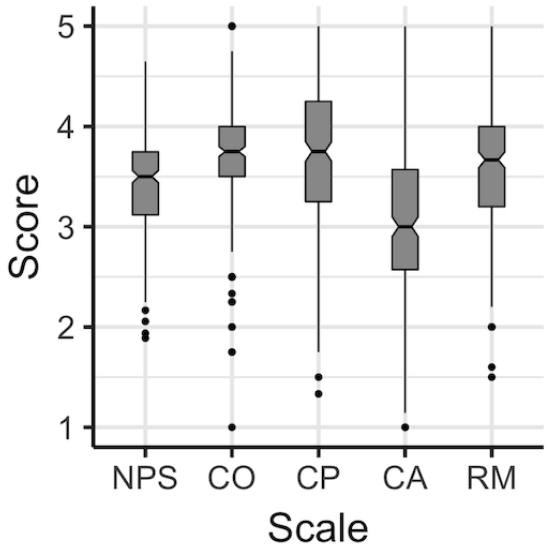


Table 2. Mean scale scores.

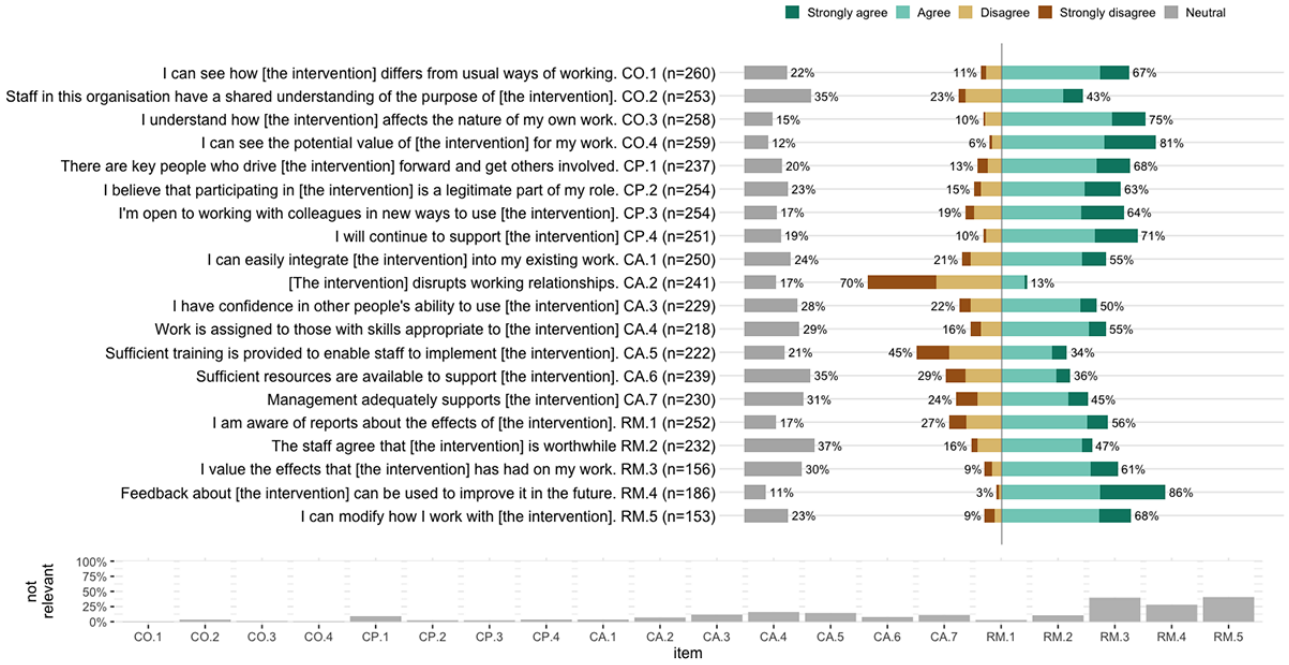
Scale[a]	n[b]	Mean (SD)	Low[c]	High[c]
Normalization process scale (NPS)[c]	221	3.54 (0.51)	2.11	4.85
Coherence (CO)	259	3.70 (0.67)	1.00	5.00
Cognitive participation (CP)	256	3.69 (0.73)	1.33	5.00
Collective action (CA)	227	3.30 (0.69)	1.29	5.00
Reflexive monitoring (RM)	181	3.55 (0.62)	1.50	5.00

[a]For the total NPS scale, a maximum of 15% missingness was allowed. For the sub-scales, a minimum of 2 rated items were needed to calculate a mean.

[b]n varies because of item nonresponse.

[c]Low and High represent the lowest (1) and highest (5) score, respectively, rated by the respondents.

Figure 3. Frequency distribution of item responses. The upper part of the figure shows the percentage of respondents reporting strongly disagree, disagree, agree, or strongly agree. The gray bar coupled to the y-axis indicates the percentage of participants rating an item as "neutral." The lower part of the figure shows the percentage of respondents who chose to not to rate a specific item (ie, not relevant). CO: coherence, CP: cognitive participation. CA: collective action. RM: reflexive monitoring.



Internal Consistency

Considering the number of items, the internal consistency of the translated NoMAD questionnaire is good for the total score ($alpha_{NPS}$=.85) and ranges from questionable to acceptable for the subscales (.62≤alpha≤.75; Table 3). Internal consistency improved to good when items were dropped.

Factor Structure

Table 4 summarizes the CFA results and the fit indices for the 3 models: (1) the first order 4-factor model in which normalization is defined by 4 correlated constructs, (2) the first order unidimensional model, and (3) the hierarchical model in which a second-level factor accounts for the correlations among the 4 first-order factors. Considering the number of items, all 3 models fitted the data reasonably well. Both the 4-factor model and the hierarchical model represented the observed data significantly better than the unidimensional model (respectively: χ^2_6=220.7, $P \leq .05$, and χ^2_4=198.1, $P \leq .05$). The 4-factor model performed better than the hierarchical model (χ^2_2=22.5, $P \leq .05$) with less discrepancy between the obtained and implied data (χ^2_{164}=559.7, SRMR=0.12), better fit per variable (RMSEA=0.10), and better fit relative to a baseline model (CFI=0.90, TLI=0.88). Notwithstanding the significance, the difference for the chi-square test statistic and the fit indices is small and potentially not outweighing the practical relevance of a total summary score and subscale scores combined in one measurement model. Therefore, the factor structure of the hierarchical model is displayed in Figure 4.

Table 3. Internal consistency calculated by using Cronbach alpha.

Scale	Cronbach alpha UK[a]	Cronbach alpha NL[b] (95% CI)	Cronbach alpha, if item dropped	Item-rest correlation
Normalization process scale	.89	.85 (0.82-0.89)	.86 (CA.2[c])	.03 (CA.2)
Coherence	.71	.71 (0.61-0.81)	.80 (CO.2[d])	.25 (CO.2)
Cognitive participation	.81	.62 (0.51-0.73)	.75 (CP.1[e])	.10 (CP.1)
Collective action	.78	.75 (0.69-0.82)	.81 (CA.2)	.00 (CA.2)
Reflexive monitoring	.65	.64 (0.54-0.74)	—[f]	.36 (RM.1[g])

[a]UK: English validation study results [25].

[b]NL: current Dutch study sample.

[c]CA.2: collective action item 2.

[d]CO.2: coherence item 2.

[e]CP.1: cognitive participation item 1.

[f]No improvement of alpha found.

[g]RM.1: reflexive monitoring item 1.

Table 4. Results of the confirmatory factor analysis (CFA). A fourth model is included in the CFA to explore potential improvements only.

Model	n_{par}[a]	χ^{2}[b]	df	CFI[c]	TLI[d]	RMSEA[e]	SRMR[f]
Four-factor	106	559.7	164	0.90	0.88	0.10	0.12
Unidimensional	100	837.3	170	0.82	0.80	0.12	0.15
Hierarchical	104	580.9	166	0.89	0.87	0.10	0.12
Hierarchical modified	101	426.1	146	0.93	0.91	0.09	0.11

[a]n_{par}: number of parameters estimated in the CFA.

[b]χ^{2}: scaled chi-squared test.

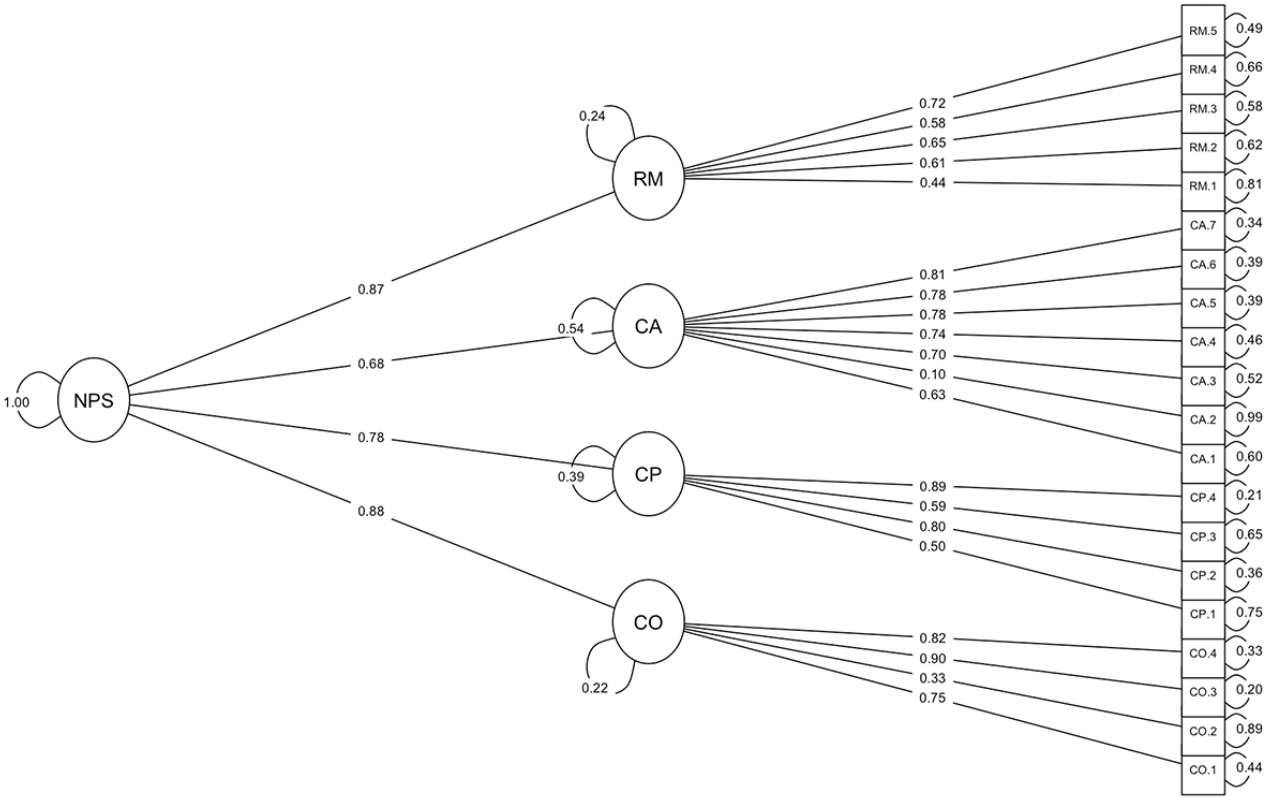
[c]CFI: Comparative Fit Index.

[d]TLI: Tucker Lewis Index.

[e]RMSEA: Root Mean Square Error of Approximation.

[f]SRMR: Standardized Root Mean Square Residual.

Figure 4. Factor structure of the hierarchical model including item factor loadings and residuals. CA: collective action; CO: coherence; CP: cognitive participation; NPS: normalization process scale; RM: reflexive monitoring.



Potential Model Improvements

We explored possibilities to improve the measurement accuracy and reliability of the proposed hierarchical model. From evaluating the item-factor loadings, it can be concluded that item CA.2 has a weak relationship with CA (λ=0.10), indicating that less than 1% of the variance in this item is explained by this factor. This is confirmed by the "Cronbach alpha-if-item-dropped" statistic indicating an improvement in the measured internal consistency of the theorized model when this item is removed (Table 3). MIs were consulted for the 4-factor model and 2 error term covariances—CA.3 to CA.4 (MI=76.9, δ=0.56) and CP.3 to CP.4 (MI=51.1, δ=0.59)—were identified as potential improvements to the model. For indicative purposes, the CFA was performed for an adapted hierarchical model in which the weakest item (CA.2) was removed and the 2 error terms were added. The modified model performed slightly better than the unmodified models (Table 4).

Convergent Validity

Following the UK study, we also explored the convergent validity of the original 20-item 4-factor model by correlating the observed mean factor scores with the mean scores for the 3 general normalization questions. Table 5 summarizes the findings. Weak correlations with the general normalization items were found for CO, CA, and RM ($-.02 \leq r \leq .27$, $0.09 \leq P \leq .81$). The CP subscale had a moderate-to-strong correlation with the normalization items ($.42 \leq r \leq .59$, $P \leq .05$).

Table 5. Convergent validity: correlations of the total score and 4 factors with the general normalization items (Part B of the questionnaire).

General item	NPS[a] (95% CI)	CO[b] (95% CI)	CP[c] (95% CI)	CA[d] (95% CI)	RM[e] (95% CI)
No. 1 Feels familiar	.26 (0.14 to 0.38)	.04[f] (−0.09 to 0.17)	.50 (0.40 to 0.59)	.14 (0.01 to 0.26)	−.02[g] (0.15 to 0.11)
No. 2 Is normal	.35 (0.23 to 0.46)	.13 (0.01 to −0.26)	.42 (0.31 to 0.52)	.27 (0.15 to 0.39)	.18 (0.05 to 0.30)
No. 3 Becomes normal	.32 (0.21 to 0.42)	.10[h] (−0.03 to 0.22)	.59 (0.51 to 0.66)	.10[i] (−0.02 to 0.22)	.03[j] (−0.09 to 0.15)

[a]NPS: normalization process scale.

[b]CO: coherence.

[c]CP: cognitive participation.

[d]CA: collective action.

[e]RM: reflexive monitoring.

[f]*P*=.52.

[g]*P*=.81.

[h]*P*=.12.

[i]*P*=.09.

[j]*P*=.63; all other correlations are significant.

Discussion

Principal Findings

Accurate and reliable instruments for measuring implementation factors and progress are currently few but required to improve the uptake of eMH interventions in routine care [11,43,44]. For this study, we translated NoMAD from English to Dutch and sought to confirm its theorized 4-factor structure in mental health care settings.

Our suggests that the NoMAD can be used reliably in assessing normalization processes in Dutch mental care settings. Using structural equation modelling, the CFA showed that the 4-factor model fitted the observed data best. This finding points in the same direction as the English psychometric study (CFI=0.95, TLI=0.93, RMSEA=0.08, SRMR=0.03, estimator: maximum likelihood) [25]. The hierarchical model might prove useful in increasing the practical utility of the NoMAD questionnaire. It offers implementation practitioners and researchers with an overall normalization score enabling comparisons across implementation projects. In addition, the subscales scores can provide a more fine-grained understanding of normalization processes and aid in identifying specific areas for improvement.

Considering the factor loadings of both the 4-factor and hierarchical models and the measured internal consistency, improvement of the theorized model seems desirable. Most notably, the explained variance in item CA.2: [the intervention] disrupts working relationships, was below validation standards (see Figure 3). Even though the extent to which people are using eMH interventions in practice might influence their perceived disruptive nature, a possible explanation might be found in the limited variance in ratings for this item as 70% of the respondents strongly disagreed with this item. This might stem from differences in linguistic interpretation by the respondents. For example, it could be that the translation of the term disrupt in CA.2 has a more negative connotation among the Dutch respondents than it has among English native speakers, leading to a tendency toward negative responses in the Dutch sample. However, this is speculative, and we feel it is too early to discard the item. We suggest further deliberation on the theorized meaning of the latent and observed variables to determine the influence of sample characteristics, implementation objects, and linguistic differences in the item formulation before conclusive decisions about possible improvements to the theorized model can be made [26,33]. In addition, we observed an increase of responses rating items in the RM scale as not applicable. Given the novelty of eMH to the care setting, it might be that the respondents have had limited exposure to the intervention to reflect on its implementation. This corresponds with a sequential interpretation of the NPT constructs but is not the only approach to the ordering of NPT mechanisms [22], and we did not measure the stage or type of implementation trajectory that respondents were currently engaged with, making it too early to draw any conclusions on the item response patterns at this stage.

Limitations

In view of the heterogeneity in implementation objects and health care settings included in current and previous NoMAD validation studies, the relevance of items should be taken in to account when administering the questionnaire to specific groups of respondents [24,25,45]. Although an open recruitment strategy was used for this study, it may be that the respondents had a natural inclination to partake in research or had a pioneering standpoint toward implementing eMH. This could have led to certain trends in the data that are not necessarily representative of the wider mental health care community involved in implementing eMH interventions. In this respect, it must be noted that sample groups 1 and 3 (mental health specialists and attendees of a national annual CBT congress, respectively) were convenience sampled. Sample 2 (48% of the pooled sample used in the CFA) was obtained through surveying members of a national professional association of general practice–based mental health nurse specialists [28]. We aimed to reduce selection bias by including these 3 sampling sources but results need to be interpreted with care because of the open design.

For the questionnaire, a 5-point Likert scale has been used for scoring the items. It goes beyond the scope of this study to

discuss the consequences of this choice in depth, but it is important to note that we approached the individual responses as ordinal data because the numbers in Likert scales represent verbal statements and not numeric entities. The mean is commonly applied to summarize data allowing for calculating SDs and CIs. However, these indicators can be biased by outliers in non-normal distributions, possibly resulting in a distorted indication of the centrality of the data [46]. In addition, the approach to item nonresponse (or missingness) should be considered. Item nonresponse means that even though the respondent has participated in the study, data for certain items are unavailable for analysis [47]. In this study, respondents needed to either rate their agreement with the NoMAD items or indicate the item as not applicable.

For calculating a scale score, 2-rated items per scale were required. This approach might be limited in informing normalization processes as 50% (more for scales of more than 4 items) of the items in the same scale could be rated as not applicable. One possibility to reduce this nonignorable form of nonresponse is to apply a forced-choice approach by removing the option for respondents to rate an item as “not applicable” from the questionnaire. However, there is a risk that forcing a rating might lead to an artificial response when a respondent feels they do not have a choice. Another possibility is to apply theoretically defined and empirically confirmed cutoff for allowable missingness in calculating the scale scores.

A further methodological limitation of this study relates to the fit indices used to evaluate the models in the CFA. As the fit indices we used were developed for maximum likelihood–based parameter estimators, they should be interpreted with caution for ordinal data using robust Weighted Least Square estimators such as the WLSMV that we applied. It is argued that the distribution of the data and sample size have a consistent influence that might lead to overestimation of fit indices with ordinal data [48].

Future Research

With this study, we have successfully contributed to the ambition of NoMAD in delivering a generic implementation measurement instrument for measuring normalization processes across different health care settings, including mental health [11,12,23-25,49,50]. Future research should assess relative predictive value and add to the practical interpretability and utility of the questionnaire. The hierarchical model provides the added value of a singular score for situations that require comparative evaluations of different implementation processes, while retaining the possibility to assess context-specific implementation processes at the construct level for understanding where implementation challenges exist in the development of effective and efficient implementation activities. However, and although interpretability of the sub-scale scores and the total NPS score does make sense from a mathematical perspective, the meaning and normativity of the scores in practice need to be established before these scores can serve implementation research and practice meaningfully. Future research should establish normative data and assess the implied factor structure of the hierarchical model in different datasets.

To increase comparability with the UK psychometric study, the 3 general normalization items were added to the questionnaire solely for assessing convergent validity [23-25]. Although this gives some indications of correlation of the NoMAD items with similar scales, the status of these 3 items is unclear. However, they do not constitute to the core of the questionnaire, and users are advised to disregard them. Instead, different measures of comparable constructs should be examined to establish a stronger assessment of convergent validity. Preferably, a multitrait-multimethod matrix should be used to strengthen conclusions about construct validity by using different methods such as organizational data on normalization success [34,51,52].

Test-retest reliability should be assessed to examine responsiveness of the questionnaire over time, to establish the ability of the questionnaire to measure changes when they occur. Responsiveness can be considered a measure of longitudinal validity and can be assessed by testing the predefined hypothesis about expected differences in changes between known samples at different time points [27]. As the duration to achieve implementation success can vary across context implementation object and implementation activities, careful consideration is needed regarding an appropriate time frame for repeat testing to assess responsiveness of the NoMAD questionnaire [53]. Applying a large-scale, stepped-wedge randomized controlled trial, NoMAD is used to measure change in normalization processes over time in the ImpleMentAll project (study protocol forthcoming) to test the effectiveness of tailored implementation compared with usual implementation activities for eMH interventions.

Conclusions

Accurate and reliable assessment of implementation processes are needed to advance the implementation of eMH interventions in routine care. The translated NoMAD questionnaire proves to be a promising instrument in measuring implementation processes of innovative interventions in Dutch mental health care settings. The theorized 4-factor model approached the observed data acceptably, but there is room for improvement. The hierarchical model might prove useful in increasing the practical utility of the NoMAD questionnaire. Future research should add to the practical utility of the questionnaire by establishing normative data and assess the relative predictive value and responsiveness of the questionnaire over time.

Acknowledgments

The authors would like to thank Dr Els Dozeman and Dr Anneke van Schaik who collaborated in the translation process and part of the data collection. Ms Victoria Thijssen provided support in setting up and executing the practicalities of the study for which the authors are also grateful.

The authors would like to express their gratitude to the MasterMind and the ImpleMentAll projects for supporting this work. The MasterMind project was partially funded under the ICT Policy Support Programme (ICT PSP) as part of the Competitiveness and Innovation Framework Programme (CIP) by the European Community (Grant Agreement number: 621000). The ImpleMentAll project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 733025. These funding bodies had no influence on the design, execution, analysis, or interpretation of the results of this study.

Authors' Contributions

JS, HR, JR, TF, TR, and CV originated the idea for conducting this study. CV and JR designed the study protocol and CV, JR, DdB, HvS, BvL, and MM collected the data. The analysis was performed by JR and CV, and CV authored the study. All authors provided feedback and suggestions for this manuscript and read and approved the final manuscript. Coauthor HvS passed away at the time the manuscript was written.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Translated Normalization MeAsure Development-NL Questionnaire.

[DOCX File, 24KB - jmir_v21i2e12376_app1.docx]

References

1. Andersson G, Carlbring P, Lindefors N. History and Current Status of ICBT. In: Lindefors N, Andersson G, editors. Guided Internet-Based Treatments in Psychiatry. Switzerland: Springer International Publishing; 2016:1-16.
2. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]
3. Mol M, Dozeman E, Provoost S, van Schaik A, Riper H, Smit JH. Behind the scenes of online therapeutic feedback in blended therapy for depression: mixed-methods observational study. J Med Internet Res 2018 May 03;20(5):e174 [FREE Full text] [doi: 10.2196/jmir.9890] [Medline: 29724708]
4. Ruwaard JJ, Lange A, Schrieken B, Emmelkamp PM. Efficacy and effectiveness of online cognitive behavioral treatment: a decade of interapy research. Stud Health Technol Inform 2011;167:9-14. [Medline: 21685634]
5. Hedman E, Ljótsson B, Rück C, Bergström J, Andersson G, Kaldo V, et al. Effectiveness of internet-based cognitive behaviour therapy for panic disorder in routine psychiatric care. Acta Psychiatr Scand 2013 Dec;128(6):457-467. [doi: 10.1111/acps.12079] [Medline: 23406572]
6. Folker A, Mathiasen K, Lauridsen SM, Stenderup E, Dozeman E, Folker MP. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: A comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018 Mar;11:60-70 [FREE Full text] [doi: 10.1016/j.invent.2018.02.001] [Medline: 30135761]
7. Mathiasen K, Riper H, Andersen TE, Roessler KK. Guided internet-based cognitive behavioral therapy for adult depression and anxiety in routine secondary care: observational study. J Med Internet Res 2018 Nov 28;20(11):e10927 [FREE Full text] [doi: 10.2196/10927] [Medline: 30487118]
8. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED consortium. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
9. Titzler I, Saruhanjan K, Berking M, Riper H, Ebert DD. Barriers and facilitators for the implementation of blended psychotherapy for depression: A qualitative pilot study of therapists' perspective. Internet Interv 2018 Jun;12:150-164 [FREE Full text] [doi: 10.1016/j.invent.2018.01.002] [Medline: 30135779]
10. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of eMental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
11. Proctor EK, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
12. Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. Implement Sci 2015 Nov 04;10:155 [FREE Full text] [doi: 10.1186/s13012-015-0342-x] [Medline: 26537706]

13. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. Am J Prev Med 2012 Sep;43(3):337-350 [FREE Full text] [doi: 10.1016/j.amepre.2012.05.024] [Medline: 22898128]
14. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
15. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
16. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
17. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [Medline: 10474547]
18. 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]
19. May C, Finch TL. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
20. May CR, Mair F, Finch T, MacFarlane A, Dowrick CF, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
21. Murray CE. Diffusion of innovation theory: a bridge for the research–practice gap in counseling. J Couns Dev 2009;87:108-116 [FREE Full text] [doi: 10.1002/j.1556-6678.2009.tb00556.x]
22. May CR, Cummings A, Girling M, Bracher M, Mair FS, May C, et al. Using normalization process theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018 Dec 07;13(1):80 [FREE Full text] [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]
23. Finch TL, Rapley T, Girling M, Mair FS, Murray E, Treweek S, et al. Improving the normalization of complex interventions: measure development based on normalization process theory (NoMAD): study protocol. Implement Sci 2013 Apr 11;8:43 [FREE Full text] [doi: 10.1186/1748-5908-8-43] [Medline: 23578304]
24. Rapley T, Girling M, Mair FS, Murray E, Treweek S, McColl E, et al. Improving the normalization of complex interventions: part 1-development of the NoMAD instrument for assessing implementation work based on normalization process theory (NPT). BMC Med Res Methodol 2018 Nov 15;18(1):133 [FREE Full text] [doi: 10.1186/s12874-018-0590-y] [Medline: 30442093]
25. Finch T, Girling M, May CR, Mair FS, Murray E, Treweek S, et al. Improving the normalization of complex interventions: part 2 - validation of the NoMAD instrument for assessing implementation work based on normalization process theory (NPT). BMC Med Res Methodol 2018 Nov 15;18(1):135 [FREE Full text] [doi: 10.1186/s12874-018-0591-x] [Medline: 30442094]
26. Brown TA. Confirmatory factor analysis for applied research. In: Kenny DA, editor. Methodology in the Social Sciences (Series). New York: Guilford Publications, Inc; 2006.
27. Terwee CB, Bot SD, de Boer M, 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]
28. Krijgsman J, Swinkels I, van Lettow B, de Jong J, Out K, Friele R, et al. Meer dan Techniek: eHealth monitor 2016. In: eHealth monitor 2016. Den Haag: Nictiz en het NIVEL; Oct 2016:1-176.
29. Brislin RW. Back-translation for cross-cultural research. J Cross Cult Psychol 1970;1(3):185-216. [doi: 10.1177/135910457000100301]
30. Willis G. Cognitive Interviewing - A Tool for Improving Questionnaire Design. California: Sage Publications; 2005:1-335.
31. NetQ. NetQL. 2009. NETQ Internet Surveys 6 URL: http://www.netq.nl [accessed 2016-02-25] [WebCite Cache ID 75rJ35nES]
32. Couper M, Tourangeau R, Conrad FG, Singer E. Evaluating the effectiveness of visual analog scales. Social Science Computer Review 2016 Aug 19;24(2):227-245. [doi: 10.1177/0894439305281503]
33. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
34. Hinkin TR. A brief tutorial on the development of measures for use in survey questionnaires. Organ Res Methods 2016 Jun 29;1(1):104-121. [doi: 10.1177/109442819800100106]
35. Field AZ, Miles J, Field Z. Discovering Statistics Using R. London: SAGE Publications; 2012:1-993.
36. RStudio Team. RStudio. Boston, MA: RStudio, Inc; 2016. URL: http://www.rstudio.com/ [accessed 2019-02-01] [WebCite Cache ID 75rJkjybf]
37. Revelle W. The Comprehensive R Archive Network.: R-project.org; 2018. psych: Procedures for Psychological, Psychometric, and Personality Research [Software] URL: https://cran.r-project.org/web/packages/psych [accessed 2019-02-01] [WebCite Cache ID 75rKG82FO]

38. Wickham H. The Comprehensive R Archive Network.: R-project.org; 2009. ggplot2: Elegant Graphics for Data Analysis [software] URL: https://cloud.r-project.org/web/packages/ggplot2 [accessed 2019-02-01] [WebCite Cache ID 75rKTKSpO]
39. Decke D. The Comprehensive R Archive Network.: R-project.org; 2018. sjPlot: Data Visualization for Statistics in Social Science URL: https://cloud.r-project.org/web/packages/sjPlot [accessed 2019-02-01] [WebCite Cache ID 75rKmlnaY]
40. Rosseel Y. The Comprehensive R Archive Network.: R-project.org; 2012. lavaan: Latent Variable Analysis URL: https://cran.r-project.org/web/packages/lavaan/ [accessed 2019-02-01] [WebCite Cache ID 75rKvKil3]
41. Epskamp S. The Comprehensive R Archive Network. 2017. emPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output URL: https://cran.r-project.org/web/packages/semPlot [accessed 2019-02-01] [WebCite Cache ID 75rL1dkq1]
42. Contributors. The Comprehensive R Archive Network. 2018. semTools: Useful Tools for Structural Equation Modeling URL: https://cran.r-project.org/web/packages/semTools/ [accessed 2019-02-01] [WebCite Cache ID 75rLCDexp]
43. Finch TL, Mair FS, O'Donnell C, Murray E, May CR. From theory to 'measurement' in complex interventions: methodological lessons from the development of an e-health normalisation instrument. BMC Med Res Methodol 2012 May 17;12:69 [FREE Full text] [doi: 10.1186/1471-2288-12-69] [Medline: 22594537]
44. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving Implementation of eMental Health for Mood Disorders in Routine Practice: Systematic Review of Barriers and Facilitating Factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
45. Jacobs SR, Weiner BJ, Bunger AC. Context matters: measuring implementation climate among individuals and groups. Implement Sci 2014 Apr 17;9:46 [FREE Full text] [doi: 10.1186/1748-5908-9-46] [Medline: 24742308]
46. Jamieson S. Likert scales: how to (ab)use them. Med Educ 2004 Dec;38(12):1217-1218. [doi: 10.1111/j.1365-2929.2004.02012.x] [Medline: 15566531]
47. Bernaards C, Sijtsma K. Influence of Imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. Multivariate Behav Res 2000 Jul 01;35(3):321-364. [doi: 10.1207/S15327906MBR3503_03] [Medline: 26745335]
48. Nye C, Drasgow F. Assessing goodness of fit: simple rules of thumb simply do not work. Organ Res Methods 2010 Jun 07;14(3):548-570. [doi: 10.1177/1094428110368562]
49. Clinton-McHarg T, Yoong SL, Tzelepis F, Regan T, Fielding A, Skelton E, et al. Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the Consolidated Framework for Implementation Research: a systematic review. Implement Sci 2016 Dec 08;11(1):148 [FREE Full text] [doi: 10.1186/s13012-016-0512-5] [Medline: 27821146]
50. McEvoy R, Ballini L, Maltoni S, O'Donnell CA, Mair F, Macfarlane A. A qualitative systematic review of studies using the normalization process theory to research implementation processes. Implement Sci 2014 Jan 02;9(1):2 [FREE Full text] [doi: 10.1186/1748-5908-9-2] [Medline: 24383661]
51. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959;56(2):81-105. [doi: 10.1037/h0046016]
52. Schmitt N, Stults DM. Methodology review: analysis of multitrait-multimethod matrices. Appl Psychol Meas 2016 Jul 26;10(1):1-22. [doi: 10.1177/014662168601000101]
53. Rogers E. Diffusion of Innovations: Fifth Edition. New York: Free Press: a division of Simon & Schuster, Inc; 2003:1-551.

Abbreviations

BC-MH: basic care-mental health
CA: collective action
CBT: cognitive behavioral therapy
CFA: confirmatory factor analysis
CFI: Comparative Fit Index
CO: coherence
CP: cognitive participation
eHealth: electronic health
eMH: eMental health
MI: modification index
NoMAD: Normalization MeAsure Development
NPS: normalization process scale
NPT: Normalization Process Theory
PC-MH: primary care-mental health
RE-AIM: Reach Effectiveness-Adoption Implementation Maintenance
RM: reflexive monitoring
RMSEA: Root Mean Square Error of Approximation
SC-MH: specialist care-mental health

SRMR: Standardized Root Mean Square Residual
TLI: Tucker Lewis Index
WLSMV: Weighted Least Square Means and Variances

Edited by G Eysenbach; submitted 01.10.18; peer-reviewed by E Wouters, M Feijt; comments to author 01.11.18; revised version received 21.12.18; accepted 20.01.19; published 20.02.19.

Please cite as:
Vis C, Ruwaard J, Finch T, Rapley T, de Beurs D, van Stel H, van Lettow B, Mol M, Kleiboer A, Riper H, Smit J
Toward an Objective Assessment of Implementation Processes for Innovations in Health Care: Psychometric Evaluation of the Normalization Measure Development (NoMAD) Questionnaire Among Mental Health Care Professionals
J Med Internet Res 2019;21(2):e12376
URL: http://www.jmir.org/2019/2/e12376/
doi:10.2196/12376
PMID:30785402

Original Paper

QuikLitE, a Framework for Quick Literacy Evaluation in Medicine: Development and Validation

Jiaping Zheng[1], MS; Hong Yu[2,3], PhD

[1]College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, United States

[2]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States

[3]Center for Healthcare Organization and Implementation Research, Bedford Veterans Affairs Medical Center, Bedford, MA, United States

Corresponding Author:
Hong Yu, PhD
Department of Computer Science
University of Massachusetts Lowell
One University Avenue
Lowell, MA, 01854
United States
Phone: 1 978 934 3620
Email: yu_hong@uml.edu

Abstract

Background: A plethora of health literacy instruments was developed over the decades. They usually start with experts curating passages of text or word lists, followed by psychometric validation and revision based on test results obtained from a sample population. This process is costly and it is difficult to customize for new usage scenarios.

Objective: This study aimed to develop and evaluate a framework for dynamically creating test instruments that can provide a focused assessment of patients' health literacy.

Methods: A health literacy framework and scoring method were extended from the vocabulary knowledge test to accommodate a wide range of item difficulties and various degrees of uncertainty in the participant's answer. Web-based tests from Amazon Mechanical Turk users were used to assess reliability and validity.

Results: Parallel forms of our tests showed high reliability (correlation=.78; 95% CI 0.69-0.85). Validity measured as correlation with an electronic health record comprehension instrument was higher (.47-.61 among 3 groups) than 2 existing tools (Short Assessment of Health Literacy-English, .38-.43; Short Test of Functional Health Literacy in Adults, .34-.46). Our framework is able to distinguish higher literacy levels that are often not measured by other instruments. It is also flexible, allowing customizations to the test the designer's focus on a particular interest in a subject matter or domain. The framework is among the fastest health literacy instrument to administer.

Conclusions: We proposed a valid and highly reliable framework to dynamically create health literacy instruments, alleviating the need to repeat a time-consuming process when a new use scenario arises. This framework can be customized to a specific need on demand and can measure skills beyond the basic level.

(J Med Internet Res 2019;21(2):e12525) doi:10.2196/12525

KEYWORDS

health literacy; psychometrics; crowdsourcing

Introduction

Background

The past few decades have seen a proliferation of health literacy instruments. Recent reviews have identified dozens of tools [1-5], ranging from general measurements to disease-, content-, or population-specific ones. These instruments aim to measure a variety of skills necessary to function in the health care system. For example, 1 study [5] categorized 51 instruments based on 11 dimensions, including the ability to perform basic reading tasks, to communicate on health matters, and to derive meaning from sources of information. The ability to understand information is 1 of the 4 skills of health literacy identified in a systematic review [6]. It is also one of the most measured skills in the instruments. Those that measure this skill are widely used in research.

Designing an instrument measuring reading ability, or print literacy, is a time- and effort-intensive process. It usually starts with experts curating passages of text or word lists, followed by psychometric validation and revision based on test results obtained from a sample population. Once validated, the instruments stay static.

There are a few potential drawbacks of reusing instruments designed long in the past. First, language use patterns evolve over time. Health literacy, reading ability in particular, needs to adapt to these changes. Instruments that were designed from early text sources may be out of date when employed decades later. Although we are not aware of reports of this nature in the health literacy literature, researchers working on general vocabulary estimation tools have seen the need to update old tests [7].

Moreover, the public's reading abilities may also change because of increased exposure to print material. Statistics of educational attainment show that the population is receiving more education. Degrees conferred at various postsecondary levels all rose more than 30% over the decade between 2004-05 and 2014-15 according to a recent US national report [8]. More exposure to advanced text material at or above college level may improve one's reading ability. Older instruments that tend to use low-grade-level text may struggle to distinguish readers proficient above the very basic level that is required to function in the health care system. This ceiling effect, many test takers obtaining perfect scores [9], can be more pronounced when such tests are administered to groups in the general population, reflecting that many were developed with convenience samples of patients in a health care setting. Therefore, they function well as screening tools to detect low health literacy but may fail to properly separate advanced readers.

In this work, we aimed to develop a test framework that can be customized to a specific need on demand and can measure skills beyond the basic level.

Prior Work

We highlight a few instruments in this section that measure the individual skills and abilities of understanding written text. For a more complete review of instruments that measure both reading and other skills, we refer the reader to a recent review [5].

Numerous instruments have been developed to test health literacy since the 1990s. There are 2 such frequently used instruments: the Rapid Estimate of Adult Literacy in Medicine (REALM) [10] and the Test of Functional Health Literacy in Adults (TOFHLA) [11], with its shortened form Short Test of Functional Health Literacy in Adults (S-TOFHLA) [12].

REALM is a tool based on word pronunciation. A list of 66 common medical terms is organized into 3 columns according to the number of syllables and pronunciation difficulty. The administrator records the number of terms correctly pronounced by the test taker, and the raw count can be converted to 1 of the 4 grade levels: 0 to 3, 4 to 6, 7 to 8, and 9 and above. Criterion validity of REALM is established with Wide Range Achievement Test-Revised (WRAT-R) and other tests in the general domain. Estimate of administration time is under 3 min, making it easy to fit in a busy clinical workflow.

TOFHLA is designed to measure patients' ability to read and understand what they commonly encounter in the health care setting. It consists of 17 numeracy items and 3 prose passages. The passages are drawn from actual materials a patient may need to read, including instructions for preparation for an upper gastrointestinal series, the patient *Rights and Responsibilities* section of a Medicaid application, and a standard informed consent form. They are converted to a Cloze test with 50 items. Total scores are divided into 3 levels: inadequate, marginal, and adequate. TOFHLA's correlations with WRAT-R, REALM were tested to establish validity. TOFHLA takes up to 22 min to administer.

Aiming to reduce the administration time, TOFHLA was abridged to an abbreviated version, S-TOFHLA, which takes a maximum of 12 min [12]. A total of 2 passages with 36 items were selected from the full version. S-TOFHLA's validity is compared with the long version of the TOFHLA and the REALM.

Since the publication of REALM and TOFHLA, many new instruments were derived from them, for different use cases. They were often used as the reference to test for criterion validity. The development process remains largely the same, requiring expert curation and time-consuming validation. For instance, Literacy Assessment for Diabetes [13], Rapid Estimate of Adult Literacy in Vascular Surgery [14], and Arthritis-Adapted REALM [15] were examples in the REALM family. Oral Health Literacy Instrument [16] and Nutritional Literacy Scale [17] followed the design of TOFHLA.

New instruments are constantly developed for particular use scenarios. Examples of specific disease or condition included tests on asthma [18], hypertension [19], diabetes [20], colon cancer [21], and heart failure [22]. Tools for a specific population such as adolescents [23,24] were also developed. In different health domains, Rapid Estimate of Adult Literacy in Dentistry (REALD)-30 [25], REALD-99 [26], Test of Functional Health Literacy in Dentistry [27], Health Literacy in Dentistry (HeLD) [28], and short - form HeLD-14 [29] targeted dentistry, and Rapid Estimate of Adult Literacy in Genetics [30] measured literacy in genetics.

Another line of research used self-reported comprehension assistance seeking–behavior, as opposed to testing an underlying reading ability, to identify patients with inadequate health literacy. One such study presented 3 questions that can each screen for low literacy [31]. An instrument with a single item was evaluated in a primary care setting to rule out patients with limited health literacy [32].

Among the menagerie of instruments, Medical Term Recognition Test (METER) [33] bears the most similarity to our framework. It included 40 actual medical words and 40 nonwords and required the participant to mark the actual words. This format is generally known as a Yes-No test in the language testing research community. It was proposed in the 1980s as a simple alternative to the traditional multiple-choice method of testing vocabulary knowledge [34]. Scoring of the METER test

suffers from a problem that is common to this type of tests: ambiguity in unmarked items. It is not clear whether the participant was uncertain about the item or genuinely did not know it. Our work addressed this problem by explicitly giving various degrees of familiarity with an item as answer options. A second drawback of this tool is that it reused many of the REALM words, rendering the test somewhat redundant.

Methods

Study Approval

This study was approved by the Institutional Review Board at the University of Massachusetts Medical School.

Instrument Framework

We modeled our test framework after the Yes-No vocabulary test. Vocabulary is critical to text comprehension [35]. A meta-analysis showed that vocabulary knowledge most likely played a causal role in comprehension [36]. Another work showed that self-reported comprehension scores improved after lay definitions were provided for medical jargon [37].

In psycholinguistic research, the Yes-No test for vocabulary knowledge usually comprises words at different frequency levels and pseudowords to calibrate for random guessing. Pseudowords are strings of letters that follow the phonotactic and morphological rules of a language but are generally not actual words. The participants are asked to indicate whether they know each of the items.

Although this test format seems simple, creating them is not. Our framework generalized this format by relaxing the need to curate a new set of word and pseudoword items each time a new test is required. Moreover, it can account for uncertainty in the participant's familiarity with a word. Our framework can also be customized to a particular domain of interest such as dentistry or hypertension.

There are 2 parts to generating a test set under our framework. We start from a vocabulary with their associated occurrence frequencies in a large corpus. The vocabulary is first divided into 10 equally sized tiers based on their frequency. A total of 5 words are then randomly selected from each tier. Next, 2 pseudowords are generated from 2 random words in each tier. The 50 words and 20 pseudowords constitute a complete instantiation of the framework. The options a test taker has for each item are a 4-level Likert scale:

1. I have never seen this word and do not know its meaning.
2. I have seen this word but do not know its meaning.
3. I think I know the word's meaning, but I am not sure.
4. I am sure I know the word's meaning.

Scoring Method

To calculate a score, we measure the agreement between a user and a master. A master perfectly answers all the true words with the most confident value and all pseudowords with the lowest value on the Likert scale. We generalized Cohen kappa (κ) as a measure of agreement, which calculates the observed and chance disagreement:

$$\kappa = 1 - q_o/q_e \quad (1)$$

where q_o is the observed disagreement proportion and q_e is the expected disagreement by chance. In an ordinal scale like ours, the proportion can be weighted to account for varying degrees of disagreement [38].

When all the items are considered equal, as in weighted κ, the ratings from the 2 raters can be summarized in a $K \times K$ contingency table, where K is the number of categories into which a test item can be assigned. The disagreement proportions can be found from this table by multiplying the different degrees of disagreement v_{ij}, where v_{ij} is the weight indicating the disagreement when 1 rater assigned i whereas the other assigned j to an item.

We generalized this agreement by allowing the test items to carry different weights, thus accounting for their prevalence in a corpus and a person's likelihood of knowing them. We calculate the observed disagreement proportion by summing the individual item's disagreement, weighted by an item weight. Let $u=[u_1, u_2,..., u_N]$ denote the item weights for N test items. Note that the weights are normalized such that $0 \leq u_i \leq 1$ and $\sum_{i=1}^{N} u_i = 1$. Let $k=[k_1, k_2,..., k_N]$ and $l=[l_1, l_2,..., l_N]$ denote the category assignments given to the test items by the 2 raters, respectively. Finally, let $v(i, j)$ denote a function that returns the disagreement weight between categories i and j. The observed disagreement can be found in equation 2 (Figure 1).

Figure 1. Observed and chance disagreement calculation.

$$q_o = \mathbf{u}^{\mathsf{T}} v(\mathbf{k}, \mathbf{l}) \quad (2)$$

$$q_e = \sum_{i=1}^{K} \sum_{j \neq i} v(i,j) P_k(i) P_l(j) \quad (3)$$

$$P_k(x) = \mathbf{u}^{\mathsf{T}} [\mathbf{k} = x] \quad (4)$$

$$P_l(x) = \mathbf{u}^{\mathsf{T}} [\mathbf{l} = x] \quad (5)$$

The chance disagreement follows from weighted κ, with the distribution of category assignments for each rater weighted by u as shown in equations 3-5 in Figure 1.

Our generalized κ can be found by substituting the 2 disagreement proportions in equation 1 with equations 2 and 3. The score still has a value range between 0 and 1. An example of the detailed calculations is provided in Multimedia Appendix 1.

Assessment of Reliability

Test Format

In total, 2 parallel instantiations of our framework were created using the same corpus, and scores were calculated using the same disagreement weight and scheme. The 2 tests were shown back to back to participants without demarcations. The test takers were not informed that they were taking 2 equivalent tests.

Test Administration

We administered the 2 parallel instantiations of our framework to 100 Amazon Mechanical Turk (AMT) users. They were screened to be from the United States and had an approval rate of at least 90%. We then eliminated answers from users that were not native speakers of English.

Several quality control items were randomly embedded in the test. They were simple and unambiguous questions with only 1 clear and correct answer. They served to identify users that attempted to game our test.

Assessment of Validity

Test Format

We designed a 4-part online questionnaire to validate our health literacy framework. The test consisted of S-TOFHLA, Short Assessment of Health Literacy-English (SAHL-E) [39], our QuikLitE framework, the short form ComprehENotes test [40], and 5 self-reported document difficulty questions.

To generate a set of test items from our framework, we used the Google Books Ngram Corpus [41] as our starting vocabulary. This corpus is a large multilingual collection of digitized books, which were automatically annotated with syntactic information. The English corpus contains approximately 4.5 million volumes and close to half a trillion words. As the earliest volumes date from the 1800s, we selected a subset of books from 2000 and onward to ensure the vocabulary frequencies reflect current language usage patterns. Due to digitization errors, there were non-English words and nonletter symbols in the resulting vocabulary list. We filtered this list to only keep those that appear in WordNet [42]. The required pseudowords were generated by Wuggy [43]. Wuggy's algorithm operates by building a chain of subsyllabic elements from a large lexicon and then iterates through this chain to search for possible pseudowords. Given a template word, Wuggy can generate pseudowords that match the template's subsyllabic structure and transition frequency between them.

The ComprehENotes test is an instrument to assess electronic health record (EHR) notes comprehension. It includes 55 snippets of EHR notes from 6 common diseases or conditions and questions generated using the Sentence Verification Technique. In our online setup, we employed the 14-item short-form test.

Texts in the document difficulty questions were randomly selected from Wikipedia articles in the Medicine category. As the writing quality and style vary among Wikipedia articles, we limited our article selection to those that were marked as feature articles. These featured articles, according to Wikipedia editors, are "professional, outstanding, and thorough" and are "a definitive source for encyclopedic information." Furthermore, only articles designated with top or high importance were considered to eliminate obscure topics. These designations signify "extremely important" or "clearly notable" articles, and there are "strong interests from nonprofessionals around the world" or "many average readers." Finally, to control for document length, the first few paragraphs of the selected articles were used, and all documents were approximately 300 words long. For each document, the users were asked to rate its difficulty from 1 (easiest to understand) to 10 (most difficult to understand).

Similar to the parallel form reliability test, quality control items that were designed to resemble real test questions were also randomly inserted to filter out cheating test takers.

Test Administration

We recruited AMT users to take 3 versions of our online test. The tests differed in the instantiation of our framework and the document difficulty self-assessment. We generated 2 sets of word items from our framework. A total of 2 sets of Wikipedia article excerpts were selected for the document difficulty questions. The 3 versions of the test included different combinations of the vocabulary test and document difficulty test.

A power analysis projected a sample size of 158 to achieve a power of 0.8 with a medium effect size. Published instruments such as S-TOFHLA and SAHL-E, with which we compared in this study, used data from approximately 200 users for validation. We, therefore, recruited 200 users for each of our test version. They were screened in the same fashion as in the reliability assessment.

When scoring our literacy test, we adopted a linear disagreement weight, that is, $v(i, j)=|i-j|$. Item weights for true words were based on their transformed frequency in the Google Books Ngram Corpus. Specifically, the word frequencies were converted to a logarithmic scale and standardized. These transformed frequencies were then passed through a logistic function to obtain the item weights. This item weight scheme emphasizes words with high frequencies and applies minimum weight on the rare words. We expect high-frequency words to be known by most native speakers, and unfamiliarity indicates lower language ability and literacy. At the other end of the frequency spectrum, rare words may pose a challenge for most people, holding little power to distinguish the test takers' vocabulary knowledge. Pseudowords were each assigned a weight equal to the average weight of the true words.

Results

Score Distribution

We first present a distribution of health literacy scores as assessed by our framework in Figure 2. Mean scores among users in the 3 groups were 0.514 (SD 0.114), 0.498 (SD 0.154), and 0.528 (SD 0.101).

Reliability

Of the 100 users that participated in the parallel form test, 90 responses were legitimate. Demographic information of the users is shown in Table 1. The correlation between scores of the 2 equivalent forms was .78 (95% CI 0.69-0.85; P<.001), suggesting a high level of reliability.

Validity

Demographic information of the AMT users is shown in Table 2.

Correlation measured between user score and ComprehenENotes on the 3 groups of users were moderate to decent, shown in Table 3. The correlation coefficients were .61 (95% CI 0.51-0.69), .49 (95% CI 0.38-0.59), and .47 (95% CI 0.35-0.57).

We also measured polyserial correlation between our score and the self-reported document difficulty. The document difficulty scores were reverse coded in the analysis and treated as an ordinal variable. The correlations of the 3 groups were .30 (95% CI 0.17-0.43), .21 (95% CI 0.07-0.34), and .29 (95% CI 0.15-0.41). The weak correlations may be partially explained by the fact that despite given a range of 1 to 10, the AMT users on average rated the document difficulty at 3.8, with an SD of 2.0. As the document excerpts were taken from well-written articles for a wide readership, and over 70% of the users had at least an associate degree, the actual document difficulty ratings concentrated in a narrow range.

Our framework achieved higher correlation with both ComprehENotes and self-reported document difficulty than the 2 existing instruments.

Subpopulation Differences

We compared the score differences between the subpopulations in our validation data. We divided the data based on gender, race, and age to test differences in the subpopulations. Analysis of variance (ANOVA) showed that there was no significant difference between males and females ($F_{1,79}$=2.895; P=.09). Older users (>45 years) tended to score higher ($F_{1,579}$=21.182; P<.001). White users achieved better scores than non-white users ($F_{1,579}$=15.462; P<.001).

Ceiling Effect

Existing health literacy instruments may exhibit a ceiling effect, as shown in our data. A total of 52.8% of the users received the full score in SAHL-E and 55.1% in S-TOFHLA, whereas 32.4% scored perfectly in both tests. Furthermore, an overwhelming majority (94.3%) of the users made at most 1 error in either 1 of the tests. This phenomenon was also reported in other studies [44,45]. In contrast, our framework can accommodate a large variation of user health literacy levels. Among different educational attainment levels (high school or less, college, graduate), ANOVA analysis showed that scores under our framework were significantly different ($F_{2,578}$=5.605; P<.01).

Administration Time

The median time the AMT users finished our test is reported in Table 4. The majority (90.36%) of users completed the test in less than 5 min. On average, they finished the test 1.5 min faster than S-TOFHLA. Compared with SAHL-E, users took an additional 1.5 min. Among the 30 health literacy instruments with a reported administration time from a catalog [5], our test time is smaller than or equal to 23 measures.

Figure 2. Boxplot of Amazon Mechanical Turk users' health literacy score according to our framework.

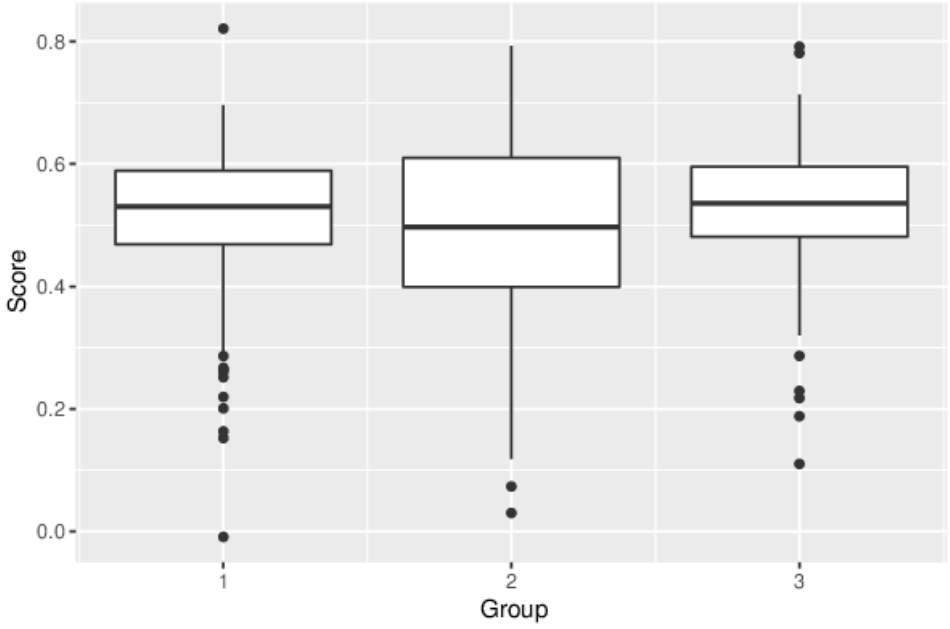


Table 1. Demographic information of Amazon Mechanical Turk users in reliability assessment (N=91).

Characteristic	Users, n (%)
Sex	
Female	50 (55)
Male	41 (45)
Race	
White	74 (81)
Black	8 (9)
Hispanic	5 (5)
Asian	3 (3)
American Indian/Alaska Native/Native Hawaiian/Other Pacific Islander	1 (1)
Age (years)	
18-24	6 (7)
25-34	44 (48)
35-44	24 (26)
45-54	9 (10)
55-64	8 (9)
Education	
High school diploma	27 (30)
Associate	27 (30)
Bachelor	30 (33)
Master or higher	7 (8)

Table 2. Demographic information of Amazon Mechanical Turk users in validity assessment.

Characteristic	Group 1 (N=192), n (%)	Group 2 (N=196), n (%)	Group 3 (N=193), n (%)
Sex			
Female	89 (46.4)	104 (53.1)	109 (56.5)
Male	103 (53.7)	92 (46.9)	84 (43.5)
Race			
White	136 (70.8)	154 (78.6)	141 (73.1)
Black	21 (10.9)	19 (9.7)	20 (10.4)
Hispanic	12 (6.3)	7 (3.6)	16 (8.3)
Asian	15 (7.8)	13 (6.6)	14 (7.3)
American Indian/Alaska Native/Native Hawaiian/Other Pacific Islander	4 (2.1)	1 (0.5)	0 (0)
Other	4 (2.1)	2 (1)	2 (1)
Age (years)			
18-24	27 (14.1)	22 (11.2)	24 (12.4)
25-34	83 (43.2)	72 (36.7)	75 (38.9)
35-44	45 (23.4)	58 (29.6)	58 (30.1)
45-54	16 (8.3)	26 (13.3)	22 (11.4)
55-64	9 (4.7)	12 (6.1)	9 (4.7)
>65	12 (6.3)	6 (3.1)	5 (2.6)
Education			
Less than high school	0 (0)	4 (2)	0 (0)
High school diploma	53 (27.6)	62 (31.6)	50 (25.9)
Associate	48 (25)	30 (15.3)	36 (18.7)
Bachelor	75 (39.1)	74 (37.8)	80 (41.5)
Master or higher	16 (8.3)	26 (13.3)	27 (14)

Table 3. Validity measured by correlation with ComprehENotes and self-reported document difficulty.

Test instrument	ComprehENotes			Document difficulty		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
QuikLitE	.61	.49	.47	.30	.21	.29
SAHL-E[a]	.42	.38	.43	.23	.10	.11
S-TOFHLA[b]	.34	.46	.40	.23	.14	.11

[a]SAHL-E: Short Assessment of Health Literacy-English

[b]S-TOFHLA: Short Test of Functional Health Literacy in Adults.

Table 4. Median administration time in seconds.

Test instrument	Group 1	Group 2	Group 3
QuikLitE	173.5	180.5	189
SAHL-E[a]	64	63	64
S-TOFHLA[b]	194.5	199.5	192
ComprehENotes	376	432.5	376

[a]SAHL-E: Short Assessment of Health Literacy-English.

[b]S-TOFHLA: Short Test of Functional Health Literacy in Adults.

Discussion

Principal Findings

The responses from AMT users showed high reliability using a parallel form test of our framework. Validity as assessed by correlation with ComprehENotes, an EHR comprehension instrument, was higher than 2 existing health literacy instruments. Our framework also displayed a higher correlation with AMT users' self-reported document difficulty than S-TOFHLA and SAHL-E. Furthermore, QuikLitE is among the easiest to administer and does not exhibit a ceiling effect.

Administration

Unlike REALM and its derivatives that rely on word pronunciation checks, our framework can be used in a waiting room without the presence of an administrator or even at home, where the test taker may experience less anxiety. In a clinic, a test can be administered by a nurse with minimal interference to the clinical workflow as it takes less than 5 min. For patients uncomfortable with an electronic device, a paper format can be used, either in a clinic or at home.

Our test can be useful for patients who have seen the material in other instruments. For patients with more exposure to written material, our test can still measure their literacy level. Moreover, if an instrument does not exist for a particular domain of interest, a test can be prepared using our framework.

Flexibilities

Many aspects of our framework can be easily adjusted to a test designer's focus. This has several advantages over existing instruments that are static. First, our framework allows for easy instantiation to suit the test designer's emphasis on a particular subject matter or health care domain. The test may be customized to a particular health care domain or personalized for a specific patient's need. For example, the education material given to a diabetes patient is different from that given to a cancer patient. Separate vocabularies can be compiled from source texts of various subjects, and subject-specific tests can be created to target patients' particular needs.

Furthermore, administering the same test over time to monitor a patient's health literacy can be problematic because repeated testing may result in memorization of the test items, making the measurement unreliable. Creating a unique test on demand with our framework can reduce item repetition, while maintaining similar coverage of word knowledge in vocabulary.

In addition, there is no inherent limit to the number of items that can be included in a test under our framework. The only limit is a test taker and the administrator's available time. Therefore, to get a broader coverage test of health literacy, a health practitioner can use more words and pseudowords. The same scoring method can be applied without modification.

Finally, our framework can be adapted to other languages. This is especially helpful in languages that REALM style pronunciation tests are difficult to develop because of the phonemic orthographic rules such as in Spanish.

Limitations

As large corpora are readily available, it is straightforward to create a test set with our framework electronically. However, scoring our test is challenging manually. This may limit its utility when a test is administered in a paper format and a score is needed immediately.

As a test can be generated dynamically, there may be discrepancies with each administration if a new set is created, making comparison difficult. Nevertheless, in our reliability assessment, the median score difference between the 2 equivalent forms is only 0.06. This difference may have little impact on the overall health literacy assessment of a test taker.

In our dataset, the samples were biased toward educated white users. More tests may be needed to assess reliability and validity on underrepresented population in future studies.

Finally, our framework focuses on print literacy. Numeracy and other skills are also recognized as important for managing one's health. Reading and understanding health-related text is, however, still a critical component to successful engagement with the health care system.

Conclusions

Over the decades, a plethora of health literacy instruments were published. Designing such instruments are often time-consuming. When a new need arises, such as a new health context or a specific disease or condition, the laborious development process has to be repeated. We, therefore, proposed QuikLitE, a novel framework that can dynamically generate and score a word recognition–based health literacy instrument. Test results with online AMT users showed high parallel form reliability, and it correlated well with ComprehENotes, an EHR comprehension assessment instrument.

Acknowledgments

Research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under Award Number R01LM012817. 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.

Multimedia Appendix 1

An illustrative example of scoring a user's answer.

[PDF File (Adobe PDF File), 112KB - jmir_v21i2e12525_app1.pdf]

References

1. Duell P, Wright D, Renzaho AM, Bhattacharya D. Optimal health literacy measurement for the clinical setting: a systematic review. Patient Educ Couns 2015 Nov;98(11):1295-1307. [doi: 10.1016/j.pec.2015.04.003] [Medline: 26162954]
2. O'Neill B, Gonçalves D, Ricci-Cabello I, Ziebland S, Valderas J. An overview of self-administered health literacy instruments. PLoS One 2014 Dec 5;9(12):e109110 [FREE Full text] [doi: 10.1371/journal.pone.0109110] [Medline: 25478813]
3. Altin SV, Finke I, Kautz-Freimuth S, Stock S. The evolution of health literacy assessment tools: a systematic review. BMC Public Health 2014 Dec 24;14:1207 [FREE Full text] [doi: 10.1186/1471-2458-14-1207] [Medline: 25418011]
4. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol 2011 Apr;64(4):366-379. [doi: 10.1016/j.jclinepi.2010.04.005] [Medline: 20638235]
5. Haun JN, Valerio MA, McCormack LA, Sørensen K, Paasche-Orlow MK. Health literacy measurement: an inventory and descriptive summary of 51 instruments. J Health Commun 2014;19(Suppl 2):302-333. [doi: 10.1080/10810730.2014.936571] [Medline: 25315600]
6. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
7. McLean S, Kramer B. The creation of a new vocabulary levels test. Shiken 2015;19(2):1-11 [FREE Full text]
8. Snyder TD, de Brey C, Dillow SA. National Center for Education Statistics. Washington, DC: Institute of Education Sciences, US Department of Education; 2016. Digest of Education Statistics 2015 URL: https://nces.ed.gov/pubs2016/2016014.pdf [accessed 2019-02-13] [WebCite Cache ID 76AIngKUk]
9. 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]
10. Davis TC, Crouch MA, Long SW, Jackson RH, Bates P, George RB, et al. Rapid assessment of literacy levels of adult primary care patients. Fam Med 1991 Aug;23(6):433-435. [Medline: 1936717]
11. 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]
12. 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 [FREE Full text] [doi: 10.1016/S0738-3991(98)00116-5] [Medline: 14528569]
13. Nath CR, Sylvester ST, Yasek V, Gunel E. Development and validation of a literacy assessment tool for persons with diabetes. Diabetes Educ 2001;27(6):857-864. [doi: 10.1177/014572170102700611] [Medline: 12211925]
14. Wallace LS, Ergen WF, Cassada DC, Freeman MB, Grandas OH, Stevens SL, et al. Development and validation of the Rapid Estimate of Adult Literacy in Vascular Surgery (REAL_VS). Ann Vasc Surg 2009;23(4):446-452. [doi: 10.1016/j.avsg.2008.10.005] [Medline: 19059757]
15. Swearingen CJ, McCollum L, Daltroy LH, Pincus T, Dewalt DA, Davis TC. Screening for low literacy in a rheumatology setting: more than 10% of patients cannot read. J Clin Rheumatol 2010 Dec;16(8):359-364. [doi: 10.1097/RHU.0b013e3181fe8ab1] [Medline: 21085021]
16. Sabbahi DA, Lawrence HP, Limeback H, Rootman I. Development and evaluation of an oral health literacy instrument for adults. Community Dent Oral Epidemiol 2009 Oct;37(5):451-462. [doi: 10.1111/j.1600-0528.2009.00490.x] [Medline: 19740249]
17. Diamond JJ. Development of a reliable and construct valid measure of nutritional literacy in adults. Nutr J 2007 Feb 14;6:5 [FREE Full text] [doi: 10.1186/1475-2891-6-5] [Medline: 17300716]
18. Williams MV, Baker DW, Honig EG, Lee TM, Nowlan A. Inadequate literacy is a barrier to asthma knowledge and self-care. Chest 1998 Oct;114(4):1008-1015. [doi: 10.1378/chest.114.4.1008] [Medline: 9792569]
19. Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. Patient Educ Couns 2003 Nov;51(3):267-275. [doi: 10.1016/S0738-3991(02)00239-2] [Medline: 14630383]
20. Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. Arch Intern Med 1998 Jan 26;158(2):166-172. [doi: 10.1001/archinte.158.2.166] [Medline: 9448555]
21. Pendlimari R, Holubar SD, Hassinger JP, Cima RR. Assessment of colon cancer literacy in screening colonoscopy patients: a validation study. J Surg Res 2012 Jun 15;175(2):221-226. [doi: 10.1016/j.jss.2011.04.036] [Medline: 21737097]

22. Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. Am J Public Health 2002 Aug;92(8):1278-1283 [FREE Full text] [doi: 10.2105/AJPH.92.8.1278] [Medline: 12144984]
23. Davis TC, Wolf MS, Arnold CL, Byrd RS, Long SW, Springer T, et al. Development and validation of the Rapid Estimate of Adolescent Literacy in Medicine (REALM-Teen): a tool to screen adolescents for below-grade reading in health care settings. Pediatrics 2006 Dec;118(6):e1707-e1714. [doi: 10.1542/peds.2006-1139] [Medline: 17142495]
24. Wu AD, Begoray DL, Macdonald M, Wharf HJ, Frankish J, Kwan B, et al. Developing and evaluating a relevant and feasible instrument for measuring health literacy of Canadian high school students. Health Promot Int 2010 Dec;25(4):444-452 [FREE Full text] [doi: 10.1093/heapro/daq032] [Medline: 20466776]
25. Lee JY, Rozier RG, Lee SD, Bender D, Ruiz RE. Development of a word recognition instrument to test health literacy in dentistry: the REALD-30--a brief communication. J Public Health Dent 2007;67(2):94-98. [doi: 10.1111/j.1752-7325.2007.00021.x] [Medline: 17557680]
26. Richman JA, Lee JY, Rozier RG, Gong DA, Pahel BT, Vann WF. Evaluation of a word recognition instrument to test health literacy in dentistry: the REALD-99. J Public Health Dent 2007;67(2):99-104 [FREE Full text] [doi: 10.1111/j.1752-7325.2007.00022.x] [Medline: 17557681]
27. Gong DA, Lee JY, Rozier RG, Pahel BT, Richman JA, Vann WF. Development and testing of the Test of Functional Health Literacy in Dentistry (TOFHLiD). J Public Health Dent 2007;67(2):105-112. [Medline: 17557682]
28. Jones K, Parker E, Mills H, Brennan D, Jamieson LM. Development and psychometric validation of a Health Literacy in Dentistry scale (HeLD). Community Dent Health 2014 Mar;31(1):37-43. [doi: 10.1922/CDH_3269Jones07] [Medline: 24741892]
29. Jones K, Brennan D, Parker E, Jamieson L. Development of a short-form Health Literacy Dental scale (HeLD-14). Community Dent Oral Epidemiol 2015 Apr;43(2):143-151. [doi: 10.1111/cdoe.12133] [Medline: 25388410]
30. Erby LH, Roter D, Larson S, Cho J. The rapid estimate of adult literacy in genetics (REAL-G): a means to assess literacy deficits in the context of genetics. Am J Med Genet A 2008 Jan 15;146A(2):174-181. [doi: 10.1002/ajmg.a.32068] [Medline: 18076116]
31. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
32. Morris NS, MacLean CD, Chew LD, Littenberg B. The Single Item Literacy Screener: evaluation of a brief instrument to identify limited reading ability. BMC Fam Pract 2006 Mar 24;7:21 [FREE Full text] [doi: 10.1186/1471-2296-7-21] [Medline: 16563164]
33. Rawson KA, Gunstad J, Hughes J, Spitznagel MB, Potter V, Waechter D, et al. The METER: a brief, self-administered measure of health literacy. J Gen Intern Med 2010 Jan;25(1):67-71 [FREE Full text] [doi: 10.1007/s11606-009-1158-7] [Medline: 19885705]
34. Meara P, Buxton BF. An alternative to multiple choice vocabulary tests. Language Testing 2016 Aug 17;4(2):142-154. [doi: 10.1177/026553228700400202]
35. Joshi RM. Vocabulary: a critical component of comprehension. Read Writ Q 2005;21(3):209-219. [doi: 10.1080/10573560590949278]
36. Stahl SA, Fairbanks MM. The effects of vocabulary instruction: a model-based meta-analysis. Rev Educ Res 1986;56(1):72-110. [doi: 10.3102/00346543056001072]
37. Polepalli RB, Houston T, Brandt C, Fang H, Yu H. Improving patients' electronic health record comprehension with NoteAid. Stud Health Technol Inform 2013;192:714-718 [FREE Full text] [doi: 10.3233/978-1-61499-289-9-714] [Medline: 23920650]
38. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968 Oct;70(4):213-220. [Medline: 19673146]
39. Lee SD, Stucky BD, Lee JY, Rozier RG, Bender DE. Short assessment of Health Literacy-Spanish and English: a comparable test of health literacy for Spanish and English speakers. Health Serv Res 2010 Aug;45(4):1105-1120 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01119.x] [Medline: 20500222]
40. 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 Apr 25;20(4):e139 [FREE Full text] [doi: 10.2196/jmir.9380] [Medline: 29695372]
41. Lin Y, Michel J, Aiden EL, Orwant J, Brockman W, Petrov S. Syntactic annotations for the google books ngram corpus. In: Proceedings of the ACL 2012 System Demonstrations. 2012 Presented at: ACL'12; July 10, 2012; Jeju Island, Korea p. 169-174.
42. Miller GA. WordNet: a lexical database for English. Commun ACM 1995 Nov;38(11):39-41 [FREE Full text] [doi: 10.1145/219717.219748]
43. Keuleers E, Brysbaert M. Wuggy: a multilingual pseudoword generator. Behav Res Methods 2010 Aug;42(3):627-633. [doi: 10.3758/BRM.42.3.627] [Medline: 20805584]

44. Lindau ST, Tomori C, Lyons T, Langseth L, Bennett CL, Garcia P. The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. Am J Obstet Gynecol 2002 May;186(5):938-943. [doi: 10.1067/mob.2002.122091] [Medline: 12015518]
45. Chesser AK, Keene Woods N, Wipperman J, Wilson R, Dong F. Health literacy assessment of the STOFHLA: paper versus electronic administration continuation study. Health Educ Behav 2014 Feb;41(1):19-24. [doi: 10.1177/1090198113477422] [Medline: 23444322]

Abbreviations

AMT: Amazon Mechanical Turk
ANOVA: analysis of variance
EHR: electronic health record
HeLD: Health Literacy in Dentistry
METER: Medical Term Recognition Test
REALD: Rapid Estimate of Adult Literacy in Dentistry
REALM: Rapid Estimate of Adult Literacy in Medicine
SAHL-E: Short Assessment of Health Literacy-English.
S-TOFHLA: Short Test of Functional Health Literacy in Adults
TOFHLA: Test of Functional Health Literacy in Adults
WRAT-R: Wide Range Achievement Test-Revised

Edited by G Eysenbach; submitted 23.10.18; peer-reviewed by A Chesser, Y Wang; comments to author 28.11.18; revised version received 04.12.18; accepted 14.12.18; published 22.02.19.

Please cite as:
Zheng J, Yu H
QuikLitE, a Framework for Quick Literacy Evaluation in Medicine: Development and Validation
J Med Internet Res 2019;21(2):e12525
URL: http://www.jmir.org/2019/2/e12525/
doi:10.2196/12525
PMID:30794206

Original Paper

Cost-Effectiveness of Telemedicine in Remote Orthopedic Consultations: Randomized Controlled Trial

Astrid Buvik[1], MD; Trine S Bergmo[2], PhD; Einar Bugge[3], MD, PhD; Arvid Smaabrekke[1], MD; Tom Wilsgaard[3,4], PhD; Jan Abel Olsen[4,5], PhD

[1]Department of Orthopaedic Surgery, University Hospital of North Norway, Tromsø, Norway

[2]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[3]Centre for Quality Improvement and Development, University Hospital of North Norway, Tromsø, Norway

[4]Department of Community Medicine, The University of Tromsø - The Arctic University of Norway, Tromsø, Norway

[5]Centre for Health Economics, Monash University, Melbourne, Victoria, Australia

Corresponding Author:
Astrid Buvik, MD
Department of Orthopaedic Surgery
University Hospital of North Norway
PO Box 4,
Tromsø, 9038
Norway
Phone: 47 95255148
Email: Astrid.buvik@unn.no

Abstract

Background: Telemedicine consultations using real-time videoconferencing has the potential to improve access and quality of care, avoid patient travels, and reduce health care costs.

Objective: The aim of this study was to examine the cost-effectiveness of an orthopedic videoconferencing service between the University Hospital of North Norway and a regional medical center in a remote community located 148 km away.

Methods: An economic evaluation based on a randomized controlled trial of 389 patients (559 consultations) referred to the hospital for an orthopedic outpatient consultation was conducted. The intervention group (199 patients) was randomized to receive video-assisted remote orthopedic consultations (302 consultations), while the control group (190 patients) received standard care in outpatient consultation at the hospital (257 consultations). A societal perspective was adopted for calculating costs. Health outcomes were measured as quality-adjusted life years (QALYs) gained. Resource use and health outcomes were collected alongside the trial at baseline and at 12 months follow-up using questionnaires, patient charts, and consultation records. These were valued using externally collected data on unit costs and QALY weights. An extended sensitivity analysis was conducted to address the robustness of the results.

Results: This study showed that using videoconferencing for orthopedic consultations in the remote clinic costs less than standard outpatient consultations at the specialist hospital, as long as the total number of patient consultations exceeds 151 per year. For a total workload of 300 consultations per year, the annual cost savings amounted to €18,616. If costs were calculated from a health sector perspective, rather than a societal perspective, the number of consultations needed to break even was 183.

Conclusions: This study showed that providing video-assisted orthopedic consultations to a remote clinic in Northern Norway, rather than having patients travel to the specialist hospital for consultations, is cost-effective from both a societal and health sector perspective. This conclusion holds as long as the activity exceeds 151 and 183 patient consultations per year, respectively.

Trial Registration: ClinicalTrials.gov NCT00616837; https://clinicaltrials.gov/ct2/show/NCT00616837 (Archived by WebCite at http://www.webcitation.org/762dZPoKX)

(J Med Internet Res 2019;21(2):e11330) doi:10.2196/11330

KEYWORDS

telemedicine; orthopedics; videoconferencing; remote consultation; outpatients; randomized controlled trial; economic evaluation; cost-effectiveness analysis; QALY

Introduction

Similar to many other countries' publicly funded national health services, a key principle in Norway is that people should have equal access for equal need irrespective of their income or region of residence [1,2]. Thus, patients' travel expenditures on public transportation are reimbursed, except a small user fee. In 2015, total reimbursement of patients' travel expenditures accounted for 2.4% of the total budget for the specialist health services [3]. In particular, patients in the northern and western part of Norway have to travel long and often burdensome journeys to receive specialist care.

Musculoskeletal injuries are the most common causes of disability and chronic pain. Surgery for orthopedic conditions is witnessing some of the greatest growth rates in developed nations across the world [4]. Decentralizing orthopedic outpatient consultations is of special interest when a large number of patients live in remote areas, many of whom are not able to use public transport, or they need assistance by accompanying persons.

Decentralized services using outreach clinics or modern information and communication technologies have the potential to improve access, avoid patient travels, and reduce health sector costs. One such technology is telemedicine consultations using real-time videoconferencing. Today, the use of telemedicine to facilitate treatment and care over a distance has been investigated in almost all clinical specialties [5-7]. Several studies have demonstrated the feasibility of using telemedicine to provide orthopedic consultations to patients living in remote areas [8-11]. Teleorthopedics involve the delivery of specialist services across a distance, usually between an orthopedic surgeon and a patient [12]. It has been reported that teleorthopedics in an outpatient setting is safe and without serious adverse events [13], and that it has increased patient satisfaction [14,15], reduced travels and saved time for the patients [16,17], and reduced costs [18]. Teleorthopedics can also improve the effectiveness of rehabilitation after orthopedic surgery [4]. A study of pediatric orthopedic patients found that even greater benefit can be obtained from telemedicine consultation for patients with a disability where the cost and inconvenience of patient transport are considerably increased [19]. Videoconferencing has also successfully been used for distance training and educational purposes in the field of orthopedics [20,21]. Despite positive reporting of telemedicine studies, the uptake in clinical practice remains low [5,12,22,23].

There exist few randomized controlled trials (RCTs) evaluating telemedicine used in orthopedic outpatient clinics [24], and even fewer that have analyzed if teleorthopedic services are cost-effective compared with traditional outpatient consultations. The main arguments for introducing telemedicine services have been to reduce costs, improve efficiency, and increase quality of and access to health care services [5,25]. Hence, there is a need to determine the extent to which teleorthopedics proves to be cost-effective. Economic evaluation provides information about the costs and benefits of the alternatives under consideration [26]. Health care costs represent the value of resources used, such as staff, equipment, and consumables. Resources outside the health system can also be included such as the patients' travel time and costs. Benefits refer to the value of changes in health outcomes. These changes can be negative and worsen health or positive and improve health [25].

In this study, we report the results from an economic evaluation. The primary objective of this study was to examine the cost-effectiveness of the telemedicine service, compared with standard in-person consultations at the hospital from a societal perspective. The secondary objective was to assess the robustness of the results by conducting sensitivity analysis. The costs included were health care costs, patient costs, and time costs measured as production loss. Health outcomes were measured as quality-adjusted life years (QALYs) gained.

Methods

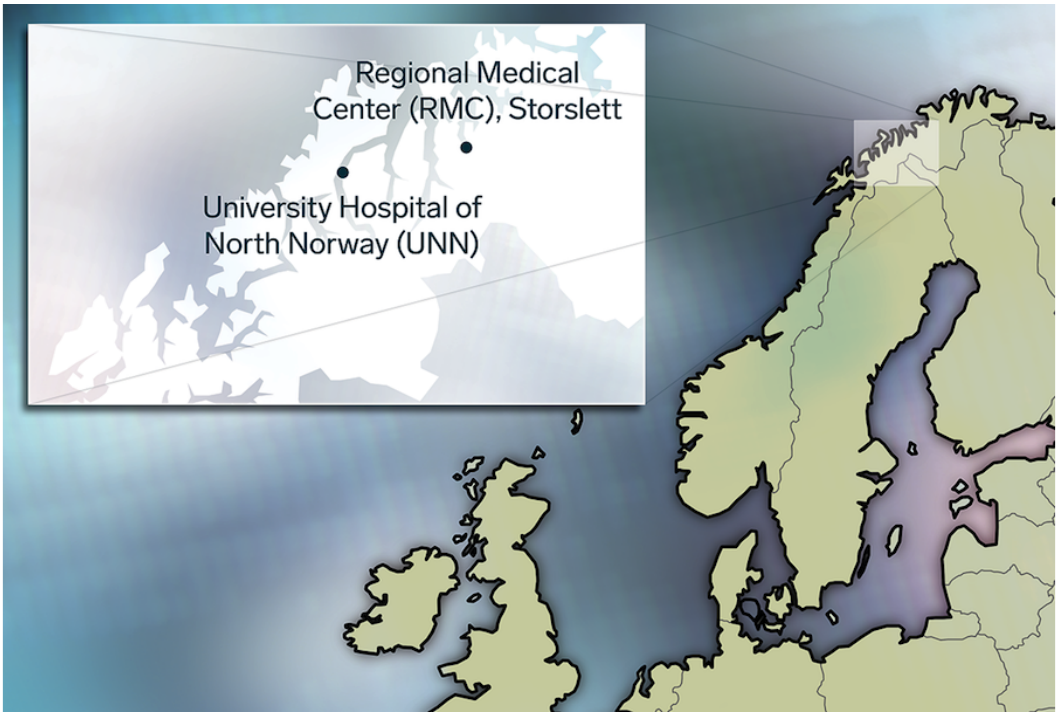
Study Design and Recruitment

The economic study was based on a prospective RCT. This trial was conducted to assess if remote consultations using videoconferencing in orthopedic consultation was as safe and effective as in usual in-person care. Patients were recruited from the 4 northernmost municipalities in Troms County (Figure 1). All participants had been referred to or had scheduled a visit at the orthopedic outpatient clinic at the University Hospital of North Norway (UNN) and found to be suitable for a video-consultation. Patients were excluded if they were in need of any of the following: an advanced physical examination, a computed tomography scan, an ultrasound, an interpreter, seeing a specific surgeon, or if unable to give informed consent. Of the 402 patients who were randomized into 2 groups, 13 withdrew from the study or did not meet for the consultations. This left 389 patients in the study. Patients randomized to the intervention group received remote orthopedic video-consultations at the Regional Medical Center (RMC) (n=199). Patients randomized to the control group received standard outpatient consultations at the hospital (n=190). Informed consent was obtained from all the participants. The study was approved by the Regional Committee for Medical and Health Research Ethics.

Equipment and Training

The remote consultations were performed through real-time videoconferencing. Both the remote center and the hospital used the Cisco TelePresence System and the Norwegian Health Network for secure data transmission (Screen: ViewSonic, Modl nr VS10946-Ie; at the remote center: Tandberg 990MXP; at the hospital: Tandberg 1500MXP). The orthopedic surgeons had some initial training and technical assistance in the beginning of the trial. Registered nurses were operating the service at the remote center. Before the trial, 2 nurses from the remote clinic received intensive training both to operate the technical equipment and to assist in treating the patients locally. They attended casting courses and were trained in clinical examination techniques.

Figure 1. Norway and the area where the study patients were recruited and location of the University Hospital of North Norway and Regional Medical Center (inserted).



The Remote Consultation

The patients were scheduled for an appointment at the local center by the surgeon at the hospital. The orthopedic surgeons (3 consultants and 2 experienced registrars) were randomly selected to conduct the video-consultations if they were available at the specific time. The surgeon made the videoconference call to the remote center.

The patients showed up and were welcomed by one of the trained nurses who set up the videoconferencing at the remote site. The nurses assisted during the consultation and performed physical tasks, for example, changed a cast or removed stitches. No physician was present during the video consultations at the remote site. An existing digital X-ray lab served by a local radiographer was available at the remote clinic. Digital X-rays were, if needed, available and shown to the patients at the time of the consultation.

Usual Care

In the control group, patients received standard consultations at the hospital outpatient clinic. In 32% of the standard consultations, the orthopedic surgeons needed assistance from a nurse [13].

In both the standard and the video consultation alternatives, the usual mandatory registration and documentation in patients' medical records were carried out by the orthopedic surgeon. This involves the conclusion of the consultation, agreement between surgeon and patient regarding any follow-up appointments, prescriptions, referrals for operation, further investigations, physiotherapy, and an application for orthopedic aid if needed. The average number of consultations per patient was 1.5 (range 1-6). For more details of the trial method, see Buvik et al [13].

Economic Evaluation

The economic evaluation consisted of trial-based analyses following the guidelines for health economic evaluation [27,28]. The cost and effectiveness data used in the economic study were based on actual investments, personnel costs, patient travels, and health outcomes collected during the trial described above. A societal perspective was adopted for calculating costs including health care costs, private costs, and production loss. Effectiveness was measured in terms of QALYs gained.

Data on costs and QALYs gained were collected alongside the trial at baseline and at 12 months follow-up using questionnaires, patient charts, and consultation records. These were valued using externally collected data on unit costs and utilities. To increase generalizability and make the cost-effectiveness result useful for decision making, the resources used in the trial were valued using equipment prices, unit costs, travel fares, and salaries from 2017/2018. An extended sensitivity analysis was conducted to address the robustness of the results.

Costs

Three types of costs were included: (1) costs associated with implementing and running the telemedicine service in clinical practice, (2) travel costs, and (3) production losses.

Costs Associated With Implementing and Running the Service

The implementation costs included the costs related to the investment in videoconferencing equipment (codec, screen, and camera) at the remote center and the hospital, and an extra computer and printer at the remote center. The remote center already had a broadband connection for other purposes. The computer at the remote center provided the nurse access to the patients' hospital records, and the printer was used to give patients a paper copy of the records on request. In addition,

costs related to running the service at the remote center were estimated. This included a registered nurse in a 20% position. Other costs associated with setting up the service included initial training sessions and travel costs related to these activities, line rent, and rent associated with the extra space needed at the local center. No extra technical support was needed as they used existing resources at the hospital. The costs estimated were only those that differed between telemedicine and hospital consultations, that is, the incremental costs. The time costs for the orthopedic surgeons were the same for both consultation forms [13]. The consumables, X-ray, and administrative costs were assumed to be similar for both groups. The cost of the nursing assistance during the standard consultations was also included. Equipment prices and line rent for both the remote clinic and the hospital were collected from the purchasing department at the hospital. A one-time equipment cost can be spread over the expected lifetime of the equipment by annuitizing the cost using a discount factor. The investment costs including equipment, installation, and training were annuitized into an equivalent annual cost assuming a 3% discount rate and a 5-year lifespan for the equipment. The costs of the extra space and other facilities at the local center were collected from financial and administrative records at the hospital. Official salary for nurses was used to estimate the costs of the extra nurse position. The costs are presented as total annual costs and costs per patient consultation (unit costs).

Travel Costs

Travel costs were collected directly from the patients during the trial. Data on traveling time, distance, and mode of transport to the consultation were collected using a questionnaire that was handed to the patient directly after each consultation. Main occupation, if they were on sick leave, and the need of overnight stay were also included. Additionally, Google map was used to estimate the travel distance from the patients' home to the consultation site either at the remote center or at the hospital (shortest and fastest). The orthopedic surgeon decided if the patient needed a companion or extra transportation on health-related grounds, reported the patients' main occupation, and if they needed sick leave. The travel costs were calculated using regulations and official travel fare rates by the Norwegian Patient Travel Agency in 2018 [29].

Production Losses

Production losses were estimated for patients in full- or part-time employment who had to take time off from work to attend the orthopedic consultations. Part-time employment was set to 50%. Time costs for the patients who were unemployed or on sick leave benefits were not included. If the information about working status was missing from the self-reported questionnaires, the orthopedic surgeons' registration forms were used. Official Norwegian average wages were used to value absence from work to estimate the production losses.

Only 3 of the 199 patients who were offered a video consultation at the remote center had a new consultation at the hospital, because of their need for a face-to-face consultation to carry out examination that is not possible over the video link. The cost of these second consultations was also included in the analysis. Based on the resources available at the hospital and the experiences from the trial, we assumed that for 300 patient consultations annually, 5 would need a second consultation.

Quality-Adjusted Life Years Gained

Health outcomes were measured in QALYs gained, a composite measure incorporating both quantity and quality-of-life impacts of treatment [26]. As a patient-reported outcome measure, we applied the EQ-5D which is the most widely used generic preference-based instrument for valuing QALYs [30]. EQ-5D questionnaires were collected at baseline and at 12 months follow-up. The questionnaires were handed to the patients immediately after the first consultation and sent by mail 12 months after the last consultation during the trial. The scoring algorithm estimated for a sample of the general population in United Kingdom was used to calculate utility values from the utility scored in the EQ-5D instrument (the EuroQol health states) [31]. Utility values were calculated only if all 5 of the EQ-5D dimensions were answered. Finally, QALYs were then calculated by multiplying the change in utility value with the duration of the health state (1 year) [26].

Statistical Analysis

Results are presented as means (SDs) or numbers (percentages). Differences between the groups were analyzed using 2 sample *t* tests, chi-square tests, or generalized estimating equations (GEEs). GEEs were used with an exchangeable covariance structure to control for dependence between 2 or more repeated consultations for some participants. All statistical analyses were performed using STATA version 14.0 (StataCorp LP Texas, USA).

Sensitivity Analysis

One-way sensitivity analyses were conducted to assess the robustness of the results. Parameters have been varied one at a time to assess the effect on the cost-effectiveness and to determine breakeven values. We recalculated the cost analysis in 3 separate scenarios: one included a less costly Skype for Business solution, the second assumed a shorter distance to the main hospital, and a third scenario includes the expenditure to the hospital excluding patients' own travel expenditures and production losses. The number of consultations needed to break even was calculated for all scenarios.

Results

Table 1 provides characteristics of the participants at baseline.

Table 1. Descriptive baseline characteristics from first consultation according to location.

Baseline characteristics of the participants (n)	UNN[a] standard consultation (n=190)	RMC[b] telemedicine consultation (n=199)
Males, n (%)	75 (39.5)	82 (41.2)
Age (years), mean (SD)	46.7 (24.9)	48.8 (24.0)
Age group (years), n (%)		
1-15	42 (22.1)	29 (14.6)
16-66	94 (49.5)	117 (58.8)
67-90	54 (28.4)	53 (26.6)
Patient residential municipality, n (%)		
Kvænangen	25 (13.2)	26 (13.0)
Nordreisa	82 (43.2)	90 (45.2)
Skjervøy	47 (24.7)	45 (22.6)
Kåfjord	36 (18.9)	38 (19.1)
Cause of consultation, n (%)		
New referral	69 (36.3)	81 (40.7)
Control after elective surgery	25 (13.2)	22 (11.1)
Control after trauma, surgery	33 (17.4)	35 (17.6)
Control after trauma, no surgery	55 (28.9)	50 (25.1)
Chronic disease	8 (4.2)	11 (5.5)
Employment status (n=177+190)[c], n (%)		
Full-time worker	45 (25.4)	56 (29.5)
Part-time worker	23 (13.0)	20 (10.5)
Homemaker	12 (6.8)	19 (10.0)
Unemployed	2 (1.1)	2 (1.1)
Retired or disability benefit	55 (31.1)	61 (32.1)
Student or pupil	40 (22.6)	32 (16.8)
EQ-5D-3L index (n=165+178)[c], mean (SD)	0.70 (0.25)	0.68 (0.26)
EQ VAS 1-100 (n=140+150)[c], mean (SD)	75 (18)	73 (19)

[a]UNN: University Hospital of North Norway

[b]RMC: Regional Medical Center.

[c]Number of item responses in UNN and RMC, respectively.

Costs

Costs Associated With Implementing and Running the Service

The costs of setting up the teleorthopedic service are presented in Table 2. Total costs of investing in standard videoconferencing units at both sites were €16,511 (1 Euro=9.60 Norwegian krone, April 10, 2018). The total annual costs including annuitized investment costs (equipment and initial training), line rent, extra personnel costs, and rent for extra office space at the regional center were €20,684. The largest cost component is the extra nursing costs at the local health center. Nearly, two-thirds of the total annual cost of the teleorthopedic service are extra personnel costs. If a less costly Skype for Business alternative had been used, the annual costs would have been reduced to €17,535 (see Table 2 for more details).

Table 2. The costs of setting up a video-assisted outpatient clinic (in Euro).

Cost elements	UNN[a] standard consultation	RMC[b] telemedicine consultation	Total (Euro)[c]	Annual cost
Alternative A[d]				
Videoconferencing equipment	5104	6250	11,354	—
Personal computer (PC)	—	463	463	—
Screen	—	156	156	—
Printer	—	114	114	—
Initial training of nurse and physician	—	—	4424	—
Total investment A			16,511	3605[e]
Additional costs alternative A				
Line rental, Norwegian Health Net	—	104[f]	—	1250
Sum alternative A				4855
Alternative B[g]				
Camera	96	937	1033	—
Screen	—	833×2	1666	—
Microphone	111	—	111	—
PC	—	463	463	—
Printer	—	114	114	—
Initial training of nurse and physician	—	—	4424	—
Total investment B			7811	1706[e]
Additional costs alternative A and B				
Technical support[h]	—	—	—	—
Rent for local RMC	—	—	—	3542
Nurse at RMC[i]	—	—	—	12,083
In need of a second consultation at the hospital[j]				204
Total additional costs				15,829
Total annual cost alternative A				20,684
Total annual cost alternative B				17,535

[a]RMC: Regional Medical Center, remote location.

[b]UNN: University Hospital of North Norway, standard consultation.

[c]1 Euro=9.60 Norwegian krone, exchange rate from the Norwegian Bank on April 10, 2018.

[d]Alternative A: Videoconferencing units: UNN—Cisco TelePresence System EX90; RMC—Cisco TelePresence MX200 G2 (prices obtained from the purchasing department at the hospital).

[e]Annual cost has been calculated using a 3% discount factor and a 5-year lifetime of the equipment.

[f]Per month.

[g]Alternative B: Skype for Business: UNN—Camera Logitech: Webcam C930e—net camera; Tablemicrofon: Jabra SPEAK 510+MS (already installed 1 PC and 2 screens for standard consultations); RMC—2 screens Philips Signage Solutions Q-Line BDL5535QL+camera/microphone Logitech GROUP+PC (prices obtained from the purchasing department at the hospital).

[h]Technical support—no extra costs included as this support has been covered by existing support at the hospital.

[i]20% part time, including social costs.

[j]Three patients needed a second consultation at UNN because of an unsatisfactory consultation at the RMC during the trial (out of 199 patients) [13]. Since we have assumed 300 patients a year in the teleconsultation alternative, costs of a second consultation have been included for 5 patients per year.

Travel Costs

Table 3 shows details on patients' modes of transport. Most patients in the remote group traveled by private car. The chosen mode of transportation reflects the lack of available public transportation in the area. In the group of patients traveling to the hospital, 26% needed extra transportation facilities because of their health condition. This number was 30% for the patients in the telemedicine group (*P*=.31). In addition, the need for travel companions was the same in both groups (30% in the hospital consultation group and 27% in the telemedicine group; *P*=.45). The time spent on traveling was 6 times higher for patients traveling to the hospital. Patients in the telemedicine group saved an average 7 hours and 40 minutes on traveling (see Table 4 for more details). The average travel cost per patient is €148.65 for the standard consultations at the hospital, as compared with €40.73 for the video consultations, including user fees for the patients (€31.04; see Multimedia Appendix 1).

Table 3. Patient transport mode to each consultation per allocation. Of the 389 patients participating in this study, some attended more than 1 consultation; consequently, the total number of consultations in this study was 559 (257 at UNN and 302 at RMC).

Transport mode	UNN[a], n (%)	RMC[b], n (%)	*P* value[c]	*P* value, GEE[d]
Taxi[e]	55 (21.4)	60 (19.9)	.57	.77
Taxi, as main transport	47 (18.3)	60 (19.9)	.71	.56
Airplane	3 (1.1)	0	N/A[f]	N/A
Bus[g]	72 (28.0)	6 (2.0)	<.001	<.001
Bus, as main transport	66 (25.7)	6 (2.0)	<.001	<.001
Private car[h]	106 (41.2)	211 (69.9)	<.001	<.001
Private car, as main transport	98 (38.1)	209 (69.2)	<.001	<.001
Express boat	28 (10.9)	0	<.001	N/A
Ferry[i]	19 (7.4)	7 (2.3)	.004	.01
Other[j]	0	15 (5.0)	<.001	N/A
Not reported or Missing	16 (6.2)	13 (4.3)	.31	N/A

[a]UNN: University Hospital, standard consultation.

[b]RMC: Regional Medical Center, remote location.

[c]Test for equality between UNN and RMC using chi-square test.

[d]Test for equality between UNN and RMC using generalized estimating equations (GEEs) with a logit link function and a binary response, transport (yes or no).

[e]Including taxi as shuttle to other transport (bus, express boat, or airplane).

[f]N/A: not applicable, few or no observations.

[g]Including bus as shuttle to other transport (airplane or express boat).

[h]Including private car as shuttle to other transport (bus, express boat, airplane, or taxi).

[i]Always in combination with other transport (bus, private car, express boat, or taxi).

[j]Walking, bicycle, or working car.

Table 4. Patients' travel details and working status.

Patients' travel and working status	UNN[a] standard consultation	RMC[b] telemedicine consultation	*P* value[c]
Travel distance in kilometers, shortest distance[d] (n=257+302)[e], mean (SD)	148 (31)	46 (17)	<.001
Travel distance one way in kilometers[f] (n=224+284)[e], mean (SD)	248 (59)	47 (28)	<.001
Travel time one way in minutes[f] (n=243+293)[e], mean (SD)	277 (94)	47 (43)	<.001
Need of companion, (n=245+294)[e], n (%)	73 (29.8)	79 (26.9)	.45
Need of extra transport[g], (n=249+297)[e], n (%)	64 (25.7)	88 (29.7)	.31
Working full time[h], (n=136+179)[i], n (%)	59 (43.4)	75 (41.9)	.68
Working part time[h], (n=136+179)[i], n (%)	29 (21.3)	28 (15.6)	.68
Sick leave—all[j], (n=138+180)[i], n (%)	60 (43.5)	71 (39.4)	.47
Actual working[k]—full time, (n=76+109)[i], n (%)	20 (26.3)	36 (33.0)	.09
Actual working[k]—part time, (n=76+109)[i], n (%)	15 (19.7)	13 (11.9)	.09

[a]UNN: University Hospital North Norway, standard consultation.

[b]RMC: Regional Medical Center, remote location.

[c]Test for equality between UNN and RMC using *t* test or chi-square test as appropriate.

[d]Calculated road between allocation and municipality center using Google Map, one way. The travel distance for the patient in the municipality, where the RMC is located, is replaced with the mean value of the municipalities' patients reported travel distance.

[e]Number of item responses in UNN and RMC, respectively.

[f]Patients reported distance or time used to travel to the consultation.

[g]Need extra transport, as patient was not able to use public transport.

[h]Patient reported (age between 15 and 67 years), missing value adjusted by doctors reported value.

[i]Number of item responses in UNN and RMC, respectively, age between 15 and 67 years.

[j]Including unemployed and homemakers.

[k]Working—patient not with sick leave.

Production Losses

Production losses for patients who had to be away from work to attend the consultations, the total average costs of the patient transfer amounted to €182.50 per patient for the standard consultations and €51.77 for the teleconsultations. The calculation of travel and time costs for the patients is presented in detail in Multimedia Appendix 1.

Quality-Adjusted Life Years Gained

The average QALYs gained per patient in the telemedicine group was .09 which was not significantly different to the .05 gain in the standard consultation group, *P*=.29.

Cost and Effectiveness

Table 5 presents the costs and effects in each of the 2 alternatives. Among patients in the intervention group, 3 needed a second face-to-face consultation that was not possible to carry out over the video link. The cost of these second consultations was also included in the analysis (Table 5). In total, the telemedicine service costs €65 less per patient than standard consultations at the hospital. Thus, the remote teleorthopedic service is less costly and produced no difference in health outcome, that is, the teleorthopedic service as described in this study is cost-effective. The number of patient consultations needed for telemedicine and standard consultation to be equally costly (breakeven) is 151 patients per year (see Figure 2).

Table 5. Costs and effectiveness for standard and remote consultations (1 Euro=9.60 Norwegian krone, exchange rate from the Norwegian Bank on April 10, 2018).

Costs and effectiveness	UNN[a] standard consultation	RMC[b] telemedicine consultation	Difference
Consultation costs[c] (Euro)			
Investment cost videoconferencing[d]	0	3605	3605
Line rent	0	1250	1250
Room rent	0	3542	3542
Personnel costs (nurse)[e]	906	12,083	11,177
In need of a second consultation at the hospital[f]	0	204	204
Total annual costs	906	20,684	19,778
Cost per consultation[g]	3	69	66
Time and travel costs (Euro)			
Travel costs	149	41	108
Time costs[h]	34	11	27
Total time and travel costs per consultation	183	52	131
Total costs per patient consultation	186	121	65
Effectiveness (QALYs[i] gained)	.05	.09	.04[j]

[a]UNN: University Hospital of North Norway.

[b]RMC: Regional Medical Center.

[c]Consultations cost which are different between the 2 groups.

[d]Total investment costs have been annuitized using 3% discount factor and a 5-year lifetime.

[e]The extra personnel costs at the remote location included a nurse in 20% position. At the standard consultation, a nurse was present in 32% of the consultations, corresponding to 25 hours by 300 consultations a year.

[f]Three patients needed a second consultation at UNN because of an unsatisfactory consultation at the RMC during the trial (out of 199 patients) [13]. Since we have assumed 300 patients a year in the teleconsultation alternative, costs of a 2nd consultation have been included for 5 patients per year.

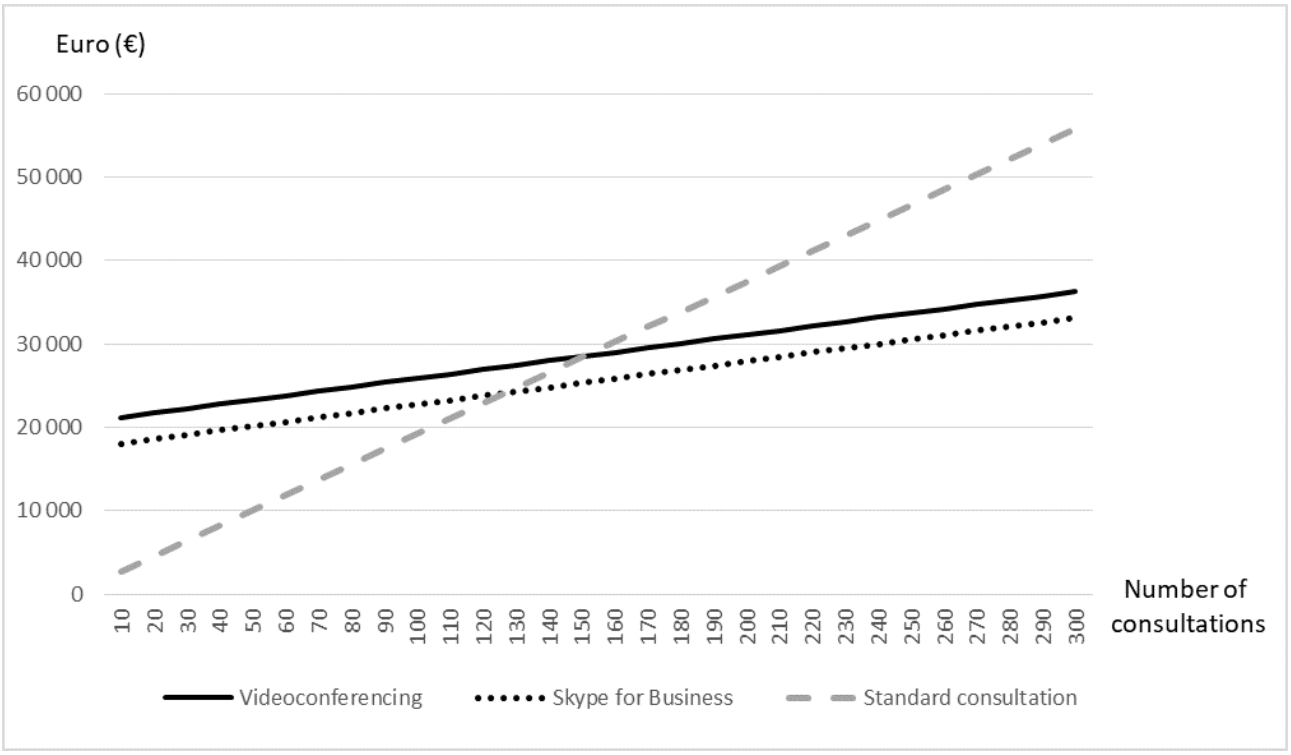
[g]The annual load for this service is estimated to be 300 telemedicine consultations per year.

[h]Production loss because of absence from work to receive orthopedic consultation.

[i]QALYs: quality-adjusted life years.

[j]The difference in QALYs gained was not significant (P=.29) t test.

Figure 2. Total annual costs of the teleorthopedic service including the Skype for Business alternative.



Sensitivity Analysis

The different scenarios in the sensitivity analyses are illustrated in Figures 2 and 3. The main case alternative described above is represented by the solid black curve (videoconferencing) and the gray dotted line (standard consultation at the hospital). The first scenario included a less costly Skype for Business alternative. The total annual cost of this alternative was €17,535, and the number of patient consultations needed to break even was 127 per year (dotted black line). The second scenario, assuming a shorter distance (90 km) between the remote clinic and the hospital, needed 314 patient consultations to breakeven using videoconferencing units (not shown). The third scenario included only expenditures to the hospital excluding production losses and patients' own travel expenditures. This made telemedicine cost-effective for an activity of at least 183 patient consultations a year (not shown).

Figure 3. Cost per patient for base case and the Skype for Business alternative.

Discussion

Principal Findings

The results of this study showed that using videoconferencing to offer orthopedic consultations to patients at the remote clinic costs less than standard outpatient consultations at the specialist hospital, as long as the activity exceeds a minimum of 151 patient consultations per year. For a total workload of 300 patients per year, the annual cost savings amounted to €19,500. With a health care sector cost perspective, the number of patient consultations needed to break even was 183, and the total annual savings amounted to €12,600. Thus, teleorthopedics is cost-effective from both a societal and health care provider perspective. A shorter distance to the hospital was the only scenario that altered the conclusion. Reducing the travel distance by 50 km made the standard consultation more cost effective for up to 314 patients per year.

Assuming a less expensive Skype for Business alternative reduced the cost of telemedicine with €3149 annually, this alternative became cost-effective when including more than 127 patient consultations per year. The reduction in costs by investing in a cheaper videoconferencing solution was relatively modest. One of the reasons is that the equipment cost was less compared with the other cost components such as the extra personnel needed at the remote site. The quality of the videoconferencing might also be reduced using Skype for Business, and more patients would need a second consultation at the hospital making the cost advantage even less. We have not evaluated if a Skype for Business alternative will reduce the quality in picture and/or sound transmission. Other options to reduce the equipment costs are to increase utilization and share the videoconferencing units with other specialties/other use (eg, teaching or meetings) [32]. These possibilities should be considered before setting up video consultations in a specific field.

In most of the published literature, a physician (eg, a general practitioner [GP], a general surgeon, or a resident) has been present at the remote site together with the patient [8,10,11,33,34]. Some studies reported that a nurse at the remote site and a specialist at the hospital could provide satisfactory remote consultations in emergency medicine [35-37]. Wallace et al recommend to include a nurse to host the teleconsultations in place of the GP to reduce the cost of telemedicine [38]. In this study, a nurse was hosting the teleconsultations at the remote site. To our knowledge, no other studies have reported a similar setting when studying the use of teleorthopedics for newly referred patients or follow-up consultations. However, the extra personnel costs still consist of nearly two-thirds of the total annual cost of the teleorthopedic service even if a nurse is hosting the teleconsultations.

Whether to include production losses measured as time off work is controversial [39]. The patients may already be off work, because they are retired or because of their health condition. Health visits of a shorter duration might not represent production losses at all. Some types of work can be postponed until the person is back or one's colleagues can take over. The time costs are important in telemedicine and eHealth, and one should find a way to include these costs [25]. In this study, production loss has only been included for those who reported that they took time off work to attend the consultation. From a societal perspective, these costs are relevant, but not from a health care provider perspective. Excluding the production loss does not change the cost-effectiveness as shown in the sensitivity analysis. Another way is to report the time (hours or days) lost or gained separately without putting a value on it [26]. Our results show that patients receiving standard consultations spent almost 8 hours more per consultation traveling than the patients in the telemedicine group.

Other studies have reported a reduction in the number of referrals to the specialist because of a learning effect and included this as part of the cost savings [32,40]. In this study, the nurse at the remote center reported an increase in the number of patients treated locally mostly because of their newly acquired casting skills. This was seen for patients with a stable fracture (not displaced). If this effect had been included in our analysis, the service would have reduced the need for specialist referrals and incurred additional cost savings.

A third option to prevent the patient from traveling long distances is to have the specialist travel from the hospital to the remote location. However, because of a lack of orthopedic specialists at the central hospital in this region, the opportunity costs of their travel time would be too high. However, in other institutional contexts, this might be another alternative to consider.

In this study, we demonstrated that significantly less public transport such as a bus or boat was used in the remote consultation group (Table 3). This can be explained by the lack of available public transport in the rural area, something that explains the frequent use of taxi and private cars. Expensive and long travels imply that fewer patients are needed to make remote consultations cost-effective.

Strengths and Limitations

The main strength of this study is that the costs and effects have been collected alongside an RCT. The estimated travel costs included in the analysis are based on the actual journeys undertaken by the participants in the trial. The time and travel cost calculation was based on actual travel distances, modes of transportation, how many in need of a companion, time spent on traveling, and the working status of each patient. Official travel fares reimbursed from the Norwegian Patient Travel Agency were used as unit costs. Some of the patients' journeys were organized by the Norwegian Patient Travel Agency (most of the taxi trips and flights). The patients had to apply for reimbursement for additional expenses.

Official travel fares reimbursed by the Norwegian Patient Travel Agency were used to calculate the travel costs. If the patients chose to travel by a more expensive alternative than the travel agency's reimbursement (eg, by plane), the patients had to pay for the difference themselves. This makes the true travel costs for the patients potentially higher than estimated in this study. It is also possible that the actual travel costs for the health care sector are lower than calculated. Some patients did not apply for travel refund, either because they forgot or simply because

they found it too troublesome to fill out the forms. One study from Norway demonstrated that 26% of the patients and 70% of the accompaniers did not apply for a travel refund [41].

Another limitation of this study is that production loss for the persons accompanying the patients on travels was not included. The main reason for this was the lack of information about their working status, and it was considered important to avoid overestimating the benefits of the service. About 28% of the patients needed a travel companion. If these time costs were included, it would have made the teleorthopedic service even more cost-effective.

Training costs have been included as a one-time cost at the startup of the teleorthopedic service. However, training should be included as an ongoing part of the service to promote and sustain use. Training is necessary because of staff turnover, particularly at the hospital. Casting courses and training in clinical examination techniques must also be arranged if there is a change in the nursing staff locally. Telemedicine services are often sensitive to changes in key personnel [42]. Successfully sustaining telemedicine services is about integration and effective change management [43].

Generalizability

One challenge for economic evaluations of telemedicine services is generalizability. High diversity in terms of objectives, technology, application, and context might limit the generalizability of specific evaluations to other settings [44]. The local context will decide the cost parameters such as the need for investment in technology and infrastructure, prices, the costs of extra personnel, and travel and time costs. The results of an economic evaluation are of most value for decision makers in the local area where the evaluation was conducted. It is important to assess if the assumptions, cost parameters, and context can be compared between settings.

To make this study relevant outside of the current institutional setting, we have emphasized a transparent calculation of all cost and effectiveness items, based on 2 costing perspectives. The same methodology can then be generalized, however, based on local prices.

Conclusions

This study showed that providing video-assisted orthopedic consultations at a remote clinic, rather than having patients travel to a centrally located hospital for consultations is cost saving. This conclusion holds from both a societal and a health sector perspective and as long as the activity exceeds 151 and 183 patient consultations per year, respectively.

Acknowledgments

The authors would like to thank Solveig Nørgaard and Anne-Sofie Høgbakk for assisting at the remote consultation; Ann Kristin Hansen and Vegard Fossum for data collection; Kerstin Aune for assisting with the patient questionnaires; physician and staff at the orthopedic outpatient clinic and the department at UNN for recruiting patients; GP in the region for recruiting patients; the Clinical Research Centre at the UNN, Inger Sperstad, Samline Grimsgaard, Inger Sandstad; Norwegian Centre for Integrated Care and Telemedicine, UNN, Lars Rye and Jan Hugo Olsen, and all the patients who participated in this study. The study was funded by a research grant from the North Norway Regional Health Authority (HST 950-10, ID 3543).

Authors' Contributions

AB contributed to the study design, data collection, data analysis, interpretation, and writing of the manuscript. TB contributed to the study design, interpretation, data analysis, and writing of the manuscript. EB contributed to the study design, interpretation, and writing of the manuscript. AS contributed to the study design, data collection, and writing of the manuscript. TW contributed to the study design, data analysis, interpretation, and writing of the manuscript. JAO contributed to the study design, interpretation, and writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Calculation of patients cost for videoconferencing and standard consultations (in Norwegian kroner and Euro).

[PDF File (Adobe PDF File), 228KB - jmir_v21i2e11330_app1.pdf]

References

1. Ministry of Health and Care Services. 2011 Apr 8. Meld. St. 16 (2010–2011). Report to the Storting (white paper). Summary — National Health and Care Services Plan URL: https://www.regjeringen.no/en/dokumenter/meld.-st.-16-2010-2011/id639794/) [accessed 2018-01-03] [WebCite Cache ID 75V2S2LXZ]
2. Patient travels. 2018. Compensation of patient travels URL: https://helsenorge.no/rettigheter/pasientreiser/reimbursement-of-patient-travel [accessed 2018-03-20] [WebCite Cache ID 75VyvDtdT]

3. The Norwegian Directorate of Health. 2015. Data from the specialist health services (Samdata 2014) URL: https://helsedirektoratet.no/publikasjoner/samdata-spesialisthelsetjenesten [accessed 2017-01-17] [WebCite Cache ID 75VyIedh4]
4. Pastora-Bernal JM, Martín-Valero R, Barón-López FJ, Estebanez-Pérez MJ. Evidence of benefit of telerehabitation after orthopedic surgery: a systematic review. J Med Internet Res 2017 Dec 28;19(4):e142 [FREE Full text] [doi: 10.2196/jmir.6836] [Medline: 28455277]
5. Bashshur RL, Shannon G, Krupinski EA, Grigsby J. Sustaining and realizing the promise of telemedicine. Telemed J E Health 2013 May;19(5):339-345 [FREE Full text] [doi: 10.1089/tmj.2012.0282] [Medline: 23289907]
6. Jarvis-Selinger S, Chan E, Payne R, Plohman K, Ho K. Clinical telehealth across the disciplines: lessons learned. Telemed J E Health 2008 Sep;14(7):720-725. [doi: 10.1089/tmj.2007.0108] [Medline: 18817503]
7. Steel K, Cox D, Garry H. Therapeutic videoconferencing interventions for the treatment of long-term conditions. J Telemed Telecare 2011;17(3):109-117. [doi: 10.1258/jtt.2010.100318] [Medline: 21339304]
8. Haukipuro K, Ohinmaa A, Winblad I, Linden T, Vuolio S. The feasibility of telemedicine for orthopaedic outpatient clinics--a randomized controlled trial. J Telemed Telecare 2000;6(4):193-198. [doi: 10.1258/1357633001935347] [Medline: 11027118]
9. Couturier P, Tyrrell J, Tonetti J, Rhul C, Woodward C, Franco A. Feasibility of orthopaedic teleconsulting in a geriatric rehabilitation service. J Telemed Telecare 1998;4(Suppl 1):85-87. [doi: 10.1258/1357633981931597] [Medline: 9640749]
10. Aarnio P, Lamminen H, Lepistö J, Alho A. A prospective study of teleconferencing for orthopaedic consultations. J Telemed Telecare 1999;5(1):62-66. [doi: 10.1258/1357633991932405] [Medline: 10505371]
11. Vuolio S, Winblad I, Ohinmaa A, Haukipuro K. Videoconferencing for orthopaedic outpatients: one-year follow-up. J Telemed Telecare 2003;9(1):8-11. [doi: 10.1258/135763303321159620] [Medline: 12641886]
12. Caffery LJ, Taylor M, North JB, Smith AC. Tele-orthopaedics: a snapshot of services in Australia. J Telemed Telecare 2017 Dec;23(10):835-841. [doi: 10.1177/1357633X17732800] [Medline: 28950754]
13. 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 Dec 8;16:483 [FREE Full text] [doi: 10.1186/s12913-016-1717-7] [Medline: 27608768]
14. Sharareh B, Schwarzkopf R. Effectiveness of telemedical applications in postoperative follow-up after total joint arthroplasty. J Arthroplasty 2014 May;29(5):918-922. [doi: 10.1016/j.arth.2013.09.019] [Medline: 24342278]
15. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Patient satisfaction with remote orthopaedic consultation by using telemedicine: a randomised controlled trial. J Telemed Telecare 2018 Jan 1:1357633X18783921. [doi: 10.1177/1357633X18783921] [Medline: 29973130]
16. Sathiyakumar V, Apfeld JC, Obremskey WT, Thakore RV, Sethi MK. Prospective randomized controlled trial using telemedicine for follow-ups in an orthopedic trauma population: a pilot study. J Orthop Trauma 2015 Mar;29(3):e139-e145. [doi: 10.1097/BOT.0000000000000189] [Medline: 24983434]
17. Jacklin PB, Roberts JA, Wallace P, Haines A, Harrison R, Barber JA, Virtual Outreach Project Group. Virtual outreach: economic evaluation of joint teleconsultations for patients referred by their general practitioner for a specialist opinion. Br Med J 2003 Jul 12;327(7406):84 [FREE Full text] [doi: 10.1136/bmj.327.7406.84] [Medline: 12855528]
18. McGill A, North J. An analysis of an ongoing trial of rural videoconference fracture clinics. J Telemed Telecare 2012 Dec;18(8):470-472. [doi: 10.1258/jtt.2012.gth110] [Medline: 23209275]
19. Rowell PD, Pincus P, White M, Smith AC. Telehealth in paediatric orthopaedic surgery in Queensland: a 10-year review. ANZ J Surg 2014 Dec;84(12):955-959. [doi: 10.1111/ans.12753] [Medline: 25040240]
20. Baruffaldi F, Giangiacomo L, Paltrinieri A, Toni A. Videoconferencing for distance training in orthopaedics. J Telemed Telecare 2003;9(4):241-242. [doi: 10.1258/135763303322225607] [Medline: 12952698]
21. Engle X, Aird J, Tho L, Bintcliffe F, Monsell F, Gollogly J, et al. Combining continuing education with expert consultation via telemedicine in Cambodia. Trop Doct 2014 Apr;44(2):62-68. [doi: 10.1177/0049475513515654] [Medline: 24322763]
22. Zanaboni P, Wootton R. Adoption of telemedicine: from pilot stage to routine delivery. BMC Med Inform Decis Mak 2012;12:1 [FREE Full text] [doi: 10.1186/1472-6947-12-1] [Medline: 22217121]
23. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
24. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015;9:CD002098. [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
25. Bergmo TS. How to measure costs and benefits of eHealth interventions: an overview of methods and frameworks. J Med Internet Res 2015;17(11):e254 [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]
26. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods For The Economic Evaluation Of Health Care Programmes. Oxford: Oxford University Press; Nov 24, 2015:464.
27. The Norwegian Directorate of Health. 2012 Nov 2. Economic evaluation of health services - A guide URL: https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/624/Okonomisk-evaluering-av-helsetiltak-IS-1985.pdf [accessed 2018-01-17] [WebCite Cache ID 75V1jGkcx]

28. Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. Int J Technol Assess Health Care 2005;21(2):165-171. [Medline: 15921055]
29. Patient Travels. 2018. User fees and rates URL: https://helsenorge.no/rettigheter/pasientreiser/rettigheter/egenandel-og-satser [accessed 2018-03-20] [WebCite Cache ID 75V0ynUwS]
30. Wisløff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics 2014 Apr;32(4):367-375 [FREE Full text] [doi: 10.1007/s40273-014-0136-z] [Medline: 24477679]
31. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997 Nov;35(11):1095-1108. [Medline: 9366889]
32. Ohinmaa A, Vuolio S, Haukipuro K, Winblad I. A cost-minimization analysis of orthopaedic consultations using videoconferencing in comparison with conventional consulting. J Telemed Telecare 2002;8(5):283-289. [doi: 10.1177/1357633X0200800507] [Medline: 12396857]
33. Wan AC, Gul Y, Darzi A. Realtime remote consultation in the outpatient clinic--experience at a teaching hospital. J Telemed Telecare 1999;5(Suppl 1):S70-S71. [Medline: 10534849]
34. Wallace P, Haines A, Harrison R, Barber J, Thompson S, Jacklin P, Virtual Outreach Project Group. Joint teleconsultations (virtual outreach) versus standard outpatient appointments for patients referred by their general practitioner for a specialist opinion: a randomised trial. Lancet 2002 Jun 8;359(9322):1961-1968. [Medline: 12076550]
35. Benger JR, Noble SM, Coast J, Kendall JM. The safety and effectiveness of minor injuries telemedicine. Emerg Med J 2004 Jul;21(4):438-445 [FREE Full text] [Medline: 15208226]
36. Brennan JA, Kealy JA, Gerardi LH, Shih R, Allegra J, Sannipoli L, et al. Telemedicine in the emergency department: a randomized controlled trial. J Telemed Telecare 1999;5(1):18-22. [doi: 10.1258/1357633991932342] [Medline: 10505365]
37. Tachakra S, Loane M, Uche CU. A follow-up study of remote trauma teleconsultations. J Telemed Telecare 2000;6(6):330-334. [doi: 10.1258/1357633001936003] [Medline: 11265101]
38. Wallace P, Barber J, Clayton W, Currell R, Fleming K, Garner P, et al. Virtual outreach: a randomised controlled trial and economic evaluation of joint teleconferenced medical consultations. Health Technol Assess 2004 Dec;8(50):1-106 [FREE Full text] [Medline: 15546515]
39. Olsen JA, Richardson J. Production gains from health care: what should be included in cost-effectiveness analyses? Soc Sci Med 1999 Jul;49(1):17-26 [FREE Full text] [Medline: 10414837]
40. Wootton R, Bloomer SE, Corbett R, Eedy DJ, Hicks N, Lotery HE, et al. Multicentre randomised control trial comparing real time teledermatology with conventional outpatient dermatological care: societal cost-benefit analysis. Br Med J 2000 May 6;320(7244):1252-1256 [FREE Full text] [Medline: 10797038]
41. Abelsen B, Gaski M, Pedersen EH, Skipperud M. [Is decentralisation of specialised healthcare beneficial in social and financial terms?]. Tidsskr Nor Laegeforen 2004 May 6;124(9):1256-1259 [FREE Full text] [Medline: 15131710]
42. Whitten P, Adams I. Success and failure: a case study of two rural telemedicine projects. J Telemed Telecare 2003;9(3):125-129. [doi: 10.1258/135763303767149906] [Medline: 12877772]
43. Yellowlees PM. Successfully developing a telemedicine system. J Telemed Telecare 2005 Oct;11(7):331-335. [doi: 10.1258/135763305774472024] [Medline: 16238833]
44. Reardon T. Research findings and strategies for assessing telemedicine costs. Telemed J E Health 2005 Jun;11(3):348-369. [doi: 10.1089/tmj.2005.11.348] [Medline: 16035931]

Abbreviations

GEE: generalized estimating equation
GP: general practitioner
QALY: quality-adjusted life year
RCT: randomized controlled trial
RMC: Regional Medical Center
UNN: University Hospital of North Norway

Edited by G Eysenbach; submitted 19.06.18; peer-reviewed by Y Peng, P Haubruck, J Walsh, T Jiang; comments to author 06.10.18; revised version received 22.11.18; accepted 09.12.18; published 19.02.19.

Please cite as:
Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA
Cost-Effectiveness of Telemedicine in Remote Orthopedic Consultations: Randomized Controlled Trial
J Med Internet Res 2019;21(2):e11330
URL: http://www.jmir.org/2019/2/e11330/
doi:10.2196/11330
PMID:30777845

Original Paper

A Web-Based Intervention (MotivATE) to Increase Attendance at an Eating Disorder Service Assessment Appointment: Zelen Randomized Controlled Trial

James Denison-Day[1*], BSc (Hons), MSc; Sarah Muir[1*], BA (Hons), PhD; Ciarán Newell[2*], BA (Hons), PhD; Katherine M Appleton[1*], BSc, PhD

[1]Research Centre for Behaviour Change, Psychology Department, Bournemouth University, Poole, United Kingdom

[2]Dorset Healthcare University, NHS Foundation Trust, Poole, United Kingdom

[*]all authors contributed equally

Corresponding Author:
James Denison-Day, BSc (Hons), MSc
Research Centre for Behaviour Change
Psychology Department
Bournemouth University
P104, Poole House
Talbot Campus
Poole, BH12 5BB
United Kingdom
Phone: 44 7590989774
Email: jday2@bournemouth.ac.uk

Abstract

Background: Early assessment and treatment of eating disorder patients is important for patient outcomes. However, up to a third of people referred for treatment do not access services and 16.4% do not attend their first scheduled assessment appointment. MotivATE is a fully automated, novel, Web-based program intended to increase motivation to change eating disorder behaviors, designed for delivery at the point of invitation to an eating disorder service, with the aim of increasing service attendance.

Objective: This paper assesses the impact of MotivATE on attendance at assessment when compared with treatment-as-usual.

Methods: A Zelen randomized controlled design was used. All individuals referred to a specialist eating disorder service, Kimmeridge Court in Dorset, UK, over the course of a year (October 24, 2016-October 23, 2017) were randomized to treatment-as-usual only or treatment-as-usual plus an additional letter offering access to MotivATE. Attendance at the initial scheduled assessment appointment was documented. Logistic regression analysis assessed the impact of MotivATE on attendance at assessment. Additional analyses based on levels of engagement with MotivATE were also undertaken.

Results: A total of 313 participants took part: 156 (49.8%) were randomized to treatment-as-usual and 157 (50.2%) were randomized to receive the additional offer to access MotivATE. Intention-to-treat analysis between conditions showed no impact of MotivATE on attendance at assessment (odds ratio [OR] 1.35, 95% CI 0.69-2.66, *P*=.38). Examination of the usage data indicated that only 53 of 157 participants (33.8%) in the MotivATE condition registered with the Web-based intervention. An analysis comparing those that registered with the intervention with those that did not found greater attendance at assessment in those that had registered (OR 9.46, 95% CI 1.22-73.38, *P*=.03).

Conclusions: Our primary analyses suggest no impact of MotivATE on attendance at the first scheduled assessment appointment, but secondary analyses revealed limited engagement with the program and improved attendance in those who did engage. It is unclear, however, if engagement with the program increased motivation and, in turn, attendance or if more motivated individuals were more likely to access the intervention. Further research is required to facilitate engagement with Web-based interventions and to understand the full value of MotivATE for users.

Trial Registration: ClinicalTrials.gov NCT02777944; https://clinicaltrials.gov/ct2/show/NCT02777944 (Archived by WebCite at http://www.webcitation.org/75VDEFZZ4)

(J Med Internet Res 2019;21(2):e11874) doi:10.2196/11874

KEYWORDS

feeding disorders; eating disorders; anorexia nervosa; bulimia nervosa; binge-eating disorder; motivation; early medical intervention; Internet

Introduction

Eating disorders, including anorexia nervosa, bulimia nervosa, binge eating disorder, and other nonspecified feeding or eating disorders, are estimated to affect 725,000 people in the United Kingdom [1]. Eating disorders have the highest mortality rate of all mental health conditions [2] and can have devastating consequences for the affected individual and their family members [3], as well as for the wider community, with estimated costs to the National Health Service (NHS) of between £3.9 billion and £4.6 billion a year [1].

When treating eating disorders, early intervention can be vital, as this reduces the risks of chronicity, which in turn lowers the risks to the individual, the burden of care for families, and the costs to the health care system [4,5]. However, research highlights that up to a third of people referred for specialist psychological treatment do not access services [6].

Nonattendance is frequently due to a complex collection of factors, with individuals with eating disorders often being highly ambivalent about change [7-9]. The role of an individual's eating disorder can be highly valued and perceived as functional, for example, by providing a sense of emotional or social avoidance; in the early stages of an eating disorder, a person may be in denial about the problematic aspects of their behavior [10]. People can also become stuck in their behaviors, leading to low confidence and fears about change, as well as a perception of reduced control over their life and choices, resulting in a passive approach to treatment and recovery [11]. It has also been noted that people with eating disorders can hold negative preconceived expectations of what treatment will entail [11]. These internal factors are also compounded by external and practical issues, such as negative experiences with health care services and professionals, as well as the impact of social stigma. Qualitative research conducted with individuals recovering from eating disorders suggests that these factors among people with eating disorders are potentially modifiable barriers to engagement with services [9].

A service provision survey found that nonattendance at an eating disorders service, as a result of not opting into the service or not turning up to appointments, ranged from 10%-32% of referrals suitable for assessment [11]. As such, a novel Web-based intervention, MotivATE, was developed, which aimed to address some of these barriers and promote treatment engagement [11]. The intervention is intended for use between referral to an eating disorder service and assessment appointment; the intervention focuses on managing expectations of assessment, addressing ambivalence, and increasing users' motivation and confidence to attend their initial appointment. This is achieved through the use of information, motivational tools, interactive activities, and stories from other individuals with eating disorders spread across four brief modules. MotivATE was developed via an iterative process of user evaluation through involvement of people that had experience with the assessment process. This was done within the framework of the intervention mapping process outlined by Bartholomew, Parcel, and Kok [12], which recognizes three phases of intervention development: needs assessment, program development, and evaluation. Full details of the intervention and development process are given elsewhere [11].

Following development, this single-center pilot study was conducted to establish an initial evidence base for the value of MotivATE in a naturalistic clinical setting. The aim of this research was therefore to test the impact of MotivATE on attendance at the first scheduled assessment appointment. The *principal research question* is as follows: Does adding MotivATE to treatment-as-usual impact on attendance at adult eating disorders outpatient services?

Methods

Design

The research was conducted using a two-arm, single-consent Zelen randomized controlled trial. A Zelen randomized consent design [13], which involves randomizing participants prior to consent and then only collecting consent from those in the active condition, was proposed as the most ethical and appropriate approach for answering the research question for several reasons. Firstly, people with eating disorders are often highly ambivalent about recovery [9]. As such, adding trial consent, particularly with the knowledge that they may not receive the active condition, at a time that might be challenging for this group could be deemed unethical. The addition of full consent prior to randomization ran the risk of resentful demoralization within the control group, potentially increasing the rate of nonattendance at assessment appointments. Hawthorne effects, which result in participants changing their behavior as a result of knowing they are being observed, may also occur [14]. Resentful demoralization and Hawthorne effects would not only have implications for the validity of the study, but may also pose significant risks to the health and well-being of the patient and their loved ones. As such, the use of a Zelen design reduces biases and potential negative outcomes. It also allows for a trial that more closely replicates anticipated procedures in usual clinical practice.

Recruitment

No active recruitment took place for the study. All adult referrals to the Kimmeridge Court Eating Disorders service in Dorset, UK, over the period of one year (October 24, 2016-October 23, 2017) were identified for potential inclusion in the study. This time period was chosen to ensure that results were not affected by seasonal variations. No power calculations were done due to both the limited previous literature directly related to this research and the large degree of variation identified in figures for previous attendance rates across eating disorder services [11]. However, a post hoc power analysis was conducted and is presented in the discussion section. This analysis is intended

to inform future similar studies. Participants were referred to the eating disorder service by a health professional as part of the usual referral process. Upon referral to the service, patient information was checked by the eating disorders service staff against the inclusion and exclusion criteria listed in Textbox 1.

All eligible referrals were then randomized into the study using the randomization procedure outlined below.

Control Condition: Treatment-as-Usual

Participants randomized to the control condition received usual care. This consisted of a phone call from service center staff, with details of participants' assessment appointments; a letter; and a compliments, comments, concerns, and complaints leaflet sent to their home address.

Intervention Condition: MotivATE

Participants randomized to the intervention condition received usual care, plus the opportunity to access MotivATE [11]. MotivATE is a fully automated Web-based intervention delivered via four 15-20-minute Web-based modules designed to be used prior to an assessment appointment. Participants were able to access MotivATE from their home computer or via tablet or mobile phone. The content and aims of these modules are briefly outlined in Table 1. Screenshots of different sections of MotivATE can be found in Multimedia Appendix 1.

Access to MotivATE was offered in addition to treatment-as-usual via an invitation letter to access MotivATE. The MotivATE invitation letter included a brief outline of MotivATE, the participant's ID number to be used at registration, and the URL to access the intervention online.

Should a participant not access any modules within seven days of registering, they would receive an automated reminder email, with a second reminder being sent 14 days after registration. Similarly, after completing a module, should a participant not move on to another module, they would receive email reminders after seven and 14 days. Upon completion of all modules, participants received a final email congratulating them on completing MotivATE. The intervention was not altered in any way throughout the course of the trial.

Outcomes

Our primary outcome was attendance at the initial assessment appointment. This was assessed using NHS audit data from the eating disorders service, which provided the number of individuals who did not attend from the MotivATE group versus the control group.

Secondary outcomes were as follows:

1. Engagement with the intervention. Engagement was examined using data on the number of sessions completed by each participant and time spent accessing them; this was generated by the intervention.
2. Participants' perceptions of MotivATE and the perceived impact of MotivATE on their motivation to attend assessment. This was assessed using qualitative data collected from participants from the MotivATE group in semistructured interviews. Participants in this group were given the opportunity to opt in to take part in a semistructured qualitative interview upon registering with the intervention. This triggered an automated email following the participants' assessment appointment that outlined the details of the interview and invited participants to take part. At the midpoint of the study, a second follow-up email, sent out two weeks after the initial invitation, was also added in an attempt to improve uptake.

Due to the nature of the study design, no baseline measures or demographic data were collected.

Initial Assessment Appointment

The initial assessment for both conditions was the same, constituting usual care, with attendance at this appointment being assessed using routine audit data; data regarding assessment attendance was added to the secure participant database by service staff. All treatment following the initial assessment was usual care and beyond the scope of this study.

Randomization, Allocation Concealment, and Blinding

Study-relevant information (ie, control or intervention) was placed in opaque envelopes labeled with participant ID numbers. Participants were prerandomized using block randomization into the intervention or control arms by a member of the research team not directly involved in conducting the study (KMA). This was achieved by generating a random number string from Random.org [15], which was then broken down into consecutive blocks of eight digits and manually balanced to ensure even allocation in each block.

As participants were referred to the service, they were assigned a participant ID incrementally by the service center staff. The opaque envelopes labeled with participant numbers were then included with the invitation-to-assessment letter by service staff. All service center staff and researchers were thus blind to each referral's group allocation, while participants were blind to other possible conditions.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria:

1. Referrals to the eating disorders service during the study period.

Exclusion criteria:

1. Inpatients or cases that are deemed to be an emergency (ie, must be seen within one day) or urgent (ie, must be seen within seven days).
2. Patients who have already been randomized into the study.
3. Non-English speakers.

Table 1. Content of the MotivATE Web-based intervention.

Module	Aim	Description
1: What happens at the first appointment?	Address expectations about the assessment appointment	Provides an interactive quiz to explore common misconceptions about assessment, information about the assessment appointment, and stories and videos about others' experiences.
2: How motivated are you?	Introduce the idea of change	Introduces people to the stages-of-change model with stories of others' experiences. User can choose their stage of change.
3: Arming yourself with information	Help with recognizing problematic behaviors (precontemplation) Address ambivalence	Information about eating disorders that relate to the pros and cons of eating disorders. Those who have selected a contemplation or preparation stage of change can complete their own pros and cons table and complete exercises designed to address ambivalence. Again, stories of others' experiences of an eating disorder are included.
4: Preparing for your assessment	Improve confidence to attend	Includes a video of a clinician welcoming them to the assessment and allows users to make plans to attend their appointment.

Service staff maintained a single, secure record linking the participant's name to their unique participant ID. Once all participants had passed through the study, outcome data was added to the secure record by service staff and any personal identifiers removed prior to this document being delivered to the research team.

Analysis

As outcome data were collected for all participants, due to the nature of the study design, a complete analysis was conducted on an intention-to-treat basis with all data categorized according to the original allocation, using a two-sided 5% significance level. To examine the research question, a logistic regression was conducted. The independent variable for this analysis was the allocated condition (*MotivATE* or *treatment-as-usual*); the dependent variable was attendance at assessment (*attended* or *did not attend*).

Usage data from the MotivATE condition was analyzed using descriptive statistics in order to explore the number of sessions registered by MotivATE, which includes any visit to the intervention, participants who registered with the intervention, as well as how many of the intervention modules were completed by users. Additionally, a follow-up of our primary analysis was conducted to determine the attendance of participants within the MotivATE condition who registered, or did not register, with MotivATE.

Qualitative data were originally to be analyzed using thematic analysis [16]. However, as a result of poor recruitment to this aspect of the study, it was not possible to complete a full analysis of the data. Data were analyzed as case studies instead.

Ethical Approval

The research gained Health Research Authority approval (Reference: 16/SC/0431) from the Hampshire A Ethics Committee and was registered at ClinicalTrials.gov (NCT02777944) prior to commencement. In order to ensure participant safety, all researchers completed Good Clinical Practice (GCP) training and adhered to standard NHS practice guidelines.

Results

Primary Outcome: Attendance at Initial Assessment Appointment

Figure 1 shows the flowchart of the trial. In total, 315 participants were identified for recruitment; however, as a result of a replication in the randomization process, 2 participants were not randomized. This resulted in 313 participants randomized into the trial, of which 157 (50.2%) were randomized to the MotivATE condition and 156 (49.8%) were assigned to the control condition. In total, 274 out of 313 (87.5%) participants attended their assessment appointment, as shown in Table 2.

Binary logistic regression indicated no differences between groups in attendance at the assessment appointment (χ^2_1=0.8, *P*=.38). The odds ratio (OR) for the effect of being offered access to MotivATE on attendance was 1.35 (95% CI 0.69-2.66).

Figure 1. CONSORT flow diagram. GP: general practitioner.

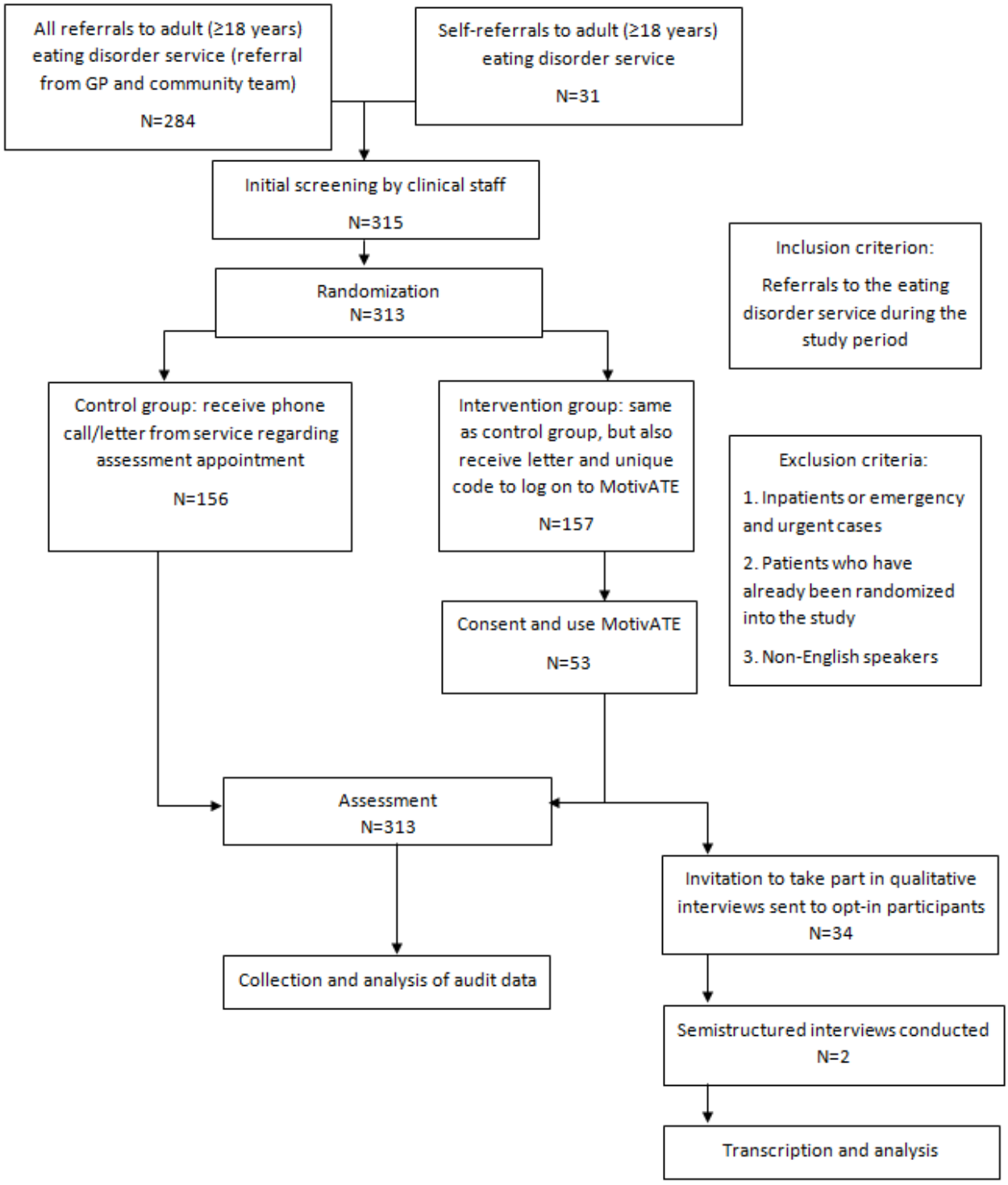


Table 2. Distribution of attendance at assessment appointment across study groups.

Condition	Attended assessment, n (%)	
	Yes	No
MotivATE (N=157)	140 (89.2)	17 (10.8)
Control (N=156)	134 (85.9)	22 (14.1)

Table 3. Number of participants that completed each module.

Module	Number of users that completed the module (N=53), n (%)
1	45 (85)
2	36 (68)
3	27 (51)
4	24 (45)

Table 4. Distribution of attendance at assessment for those that did and did not access MotivATE.

Condition	Attended assessment, n (%)	
	Yes	No
Accessed MotivATE (N=53)	52 (98)	1 (2)
Did not access MotivATE (N=104)	88 (84.6)	16 (15.4)

Secondary Outcome: Engagement With the Intervention

Usage analysis shows that, of the 157 participants assigned to the MotivATE condition, only 53 (33.8%) registered with the intervention. During the trial period, MotivATE registered 1280 separate sessions; however, of these, only 87 (6.80%) were sessions generated by the 53 registered participants. As it was not possible to collect data for nonregistered users, the exact nature of the additional 1193 sessions is difficult to ascertain. However, while some of these are likely to have been participants that visited the intervention but did not register, the high number in relation to recruitment would suggest that this also includes visits by web crawlers, programs, or automated scripts that browse the Internet in a methodical, automated manner. Indeed, 1177 of these sessions consisted of visits to the homepage of one second or less. Of the 53 participants that registered with MotivATE, 8 (15%) completed registration but did not engage with any of the content. The remaining 45 participants (85%) went on to complete, on average, at least half of the available content, with the average participant completing two and a half of the MotivATE modules; 24 of the 53 registered participants (45%) completed all of the four modules. Table 3 shows a breakdown of the number of participants that completed each module. Of the 53 participants that accessed MotivATE, 17 (32%) used the intervention more than once, with the average across all participants being 1.64 visits per person.

Table 4 gives the distribution of attendance at assessment for those that did and did not access MotivATE.

Binary logistic regression indicated that registration with MotivATE did act as a significant predictor of attendance at assessment, where individuals who registered with MotivATE were 9.5 times more likely to attend than those who did not (OR 9.46, 95% CI 1.22-73.38, P=.03). In this instance, the model was found to significantly predict attendance at assessment (χ^2_1=8.5, P=.004).

Secondary Outcome: Perceptions of the Intervention

Only 2 individuals of the 53 (4%) users who registered with MotivATE volunteered for and undertook an interview regarding their perceptions of MotivATE. These interviews have been written up as case studies and are provided in Multimedia Appendix 2.

In brief, the two case studies presented conflicting views of the usefulness of MotivATE to the participants. In one case, MotivATE did appear to have fulfilled its intended purpose; it helped to alleviate anxieties, addressed concerns, and provided practical help to assist in attending the assessment appointment. However, the other case related a much more negative experience, suggesting that the intervention felt patronizing in its approach and reinforced negative preconceptions. Two aspects, however, were common across both participants. First, both interviewees raised concerns that MotivATE represented a “tick-box” exercise rather than a genuine and helpful tool; this was a concern that was alleviated for one participant upon using MotivATE, but was exacerbated for the other. Second was an opinion that, even if MotivATE was not specifically beneficial, the underlying idea was a positive one.

Discussion

This initial research indicates that, at present, MotivATE does not increase attendance at an initial assessment appointment in people with eating disorders. Secondary analyses, however, do give some insight into the potential impact of the intervention. Attendance was 9.5 times more likely for those who registered with MotivATE than those who did not, but registration with the intervention was low (53/157, 33.8%). This may suggest that the lack of overall impact in the study was not a result of the lack of impact of the intervention itself, but rather due to issues of uptake.

Problems with engagement are often encountered when rolling out online interventions [17] and it has been noted that participants of Internet interventions can exhibit lower levels of engagement than program developers originally envision [18]. However, Web-based interventions have been demonstrated to successfully encourage behavior change for a variety of health behaviors and among a variety of populations [19-21]. Lack of engagement, though, may be a particular issue for people with eating disorders. Several recent studies of online motivational interventions for eating disorders similarly found

issues with engagement and high dropout [22-25]. This issue has been attributed to the poor ability of online interventions to attract and retain visitors, relative to other modes of contact [26].

Digital interventions have been identified as a potential approach to improving motivation to change among people with eating disorders [27]; our findings do not rule out this approach to addressing issues regarding low levels of treatment attendance. Indeed, both participants in the case studies suggested that, while MotivATE may not in and of itself be the solution, at least in its current form, the underlying concept remained a positive one. Indeed, once registered, the majority of participants completed more than half of the MotivATE content and a third visited more than once. The key issue faced by MotivATE, which was designed to promote treatment uptake, was itself a lack of uptake. This suggests that approaches addressing low engagement may in fact need to be considered even earlier in the treatment pathway and that simply offering new interventions to address this problem may not be enough. Rather, the way in which these interventions, and indeed treatment more broadly, are presented may need to be more deeply considered.

Furthermore, the observed rate of attendance at initial assessment appointments across this study was unusually high, with nonattendance levels of only 14% recorded for the control group. While within the previously noted range of 10%-32% [11], this figure represented the lower end of this range. Indeed, examination of previous nonattendance rates for the service in the two years prior to the study indicated nonattendance levels of 21% (2014-2015) and 20% (2015-2016). This low level of nonattendance may have impacted the outcome of the primary research question, as this reduced the scope for potential improvements as a result of introducing MotivATE. While it is possible that this change in attendance in comparison to previous years is a result of natural fluctuations, this change may also suggest that other factors present during the study period may have impacted treatment attendance. For example, it is possible that knowledge among service staff that a new intervention was being offered to some patients may have positively influenced the overall service provided, such as the nature of interactions during the phone consultations. It is also possible that this may have been influenced by factors outside of the service, such as changes in local practices relating to the referrals of individuals with eating disorders or initiatives by local charities. However, without further research it is difficult to pinpoint an exact cause.

This study had a number of strengths that support the findings outlined above. The high level of involvement of patients and the public in the development of MotivATE in the early stages ensured that the intervention was designed using a particularly person-based approach. The naturalistic nature of the design means that the research reflects the conditions found in practice. Similarly, conducting the study over one year eliminated the impact of seasonal fluctuations. A further strength of this research was the inclusion of qualitative interviews, which allowed for a greater insight into participants' perspectives of the intervention and its potential impact. The low recruitment to the interviews also provides useful feedback that reliance on email contact to recruit individuals from this population to an interview study appears to be insufficient and more salient approaches should be favored in future research.

Despite these strengths, this study did suffer from a number of limitations. Firstly, the use of only one site limits the applicability of the findings to general practice. This is due to different sites having differing treatment approaches and procedures, such as opt-in programs or longer waiting times, which may impact the effectiveness of the intervention. A second limitation was the lack of a pre-post measure of motivation to change. This was not included due to the naturalistic design of the study; however, without this it is not possible to ascertain if the intervention did improve motivation to change as intended. Motivation to change may have improved, but not sufficiently so to translate into behavior. Alternatively, improved attendance among those who registered for MotivATE may demonstrate increased motivation to change as a result of using MotivATE; it may also suggest that those with a higher motivation to change were more likely to attend anyway and, in turn, more likely to engage with the intervention [28]. Additionally, this study also included individuals who self-referred and these people are likely to have been more motivated to attend than those referred by general practitioners and other health care professionals; however, these individuals constituted less than 10% of the sample and were randomized across both groups; therefore, they are unlikely to have impacted on differences between the conditions. A final limitation is a potential lack of power in the analysis of the primary research question. A post hoc power analysis conducted in G*Power (Heinrich-Heine-Universität Düsseldorf) indicated an achieved power of only .37 in our study. In order to successfully detect an effect at the level observed, at a power of .8 and an alpha of .05, a further power analysis suggests a required sample size of 2559 (OR 1.35, Pr[Y=1 | X=1] H0=0.86). The clinical impact of an effect of this size, however, would also need to be considered.

This work highlights a number of directions for future research. Specifically relating to MotivATE, it is clear that more research is needed to understand the impact that the intervention has on its users, something that should be pursued from both quantitative and qualitative perspectives. This would allow for the investigation of not just whether MotivATE improves outcome measures such as motivation to change, but also some of the personal perspectives and issues surrounding engagement. A better understanding of how to best engage people with eating disorders with interventions, both on- and offline, is needed as early as possible in the treatment pathway. At present, this is not adequately understood and, as such, future research to explore potential ways in which engagement and uptake might be improved would be beneficial. Further research into uptake and engagement with interventions in a general sense is also required. New interventions are developed on a regular basis, but changes to behavior do not always follow.

In conclusion, in its current state, MotivATE cannot be recommended as an intervention to address lack of attendance at eating disorder services. However, with further research and development, this does not rule out the use of digital interventions as a potential approach to addressing this issue.

Acknowledgments

This work was undertaken as part of a PhD being completed by JDD, under the supervision of SM, KMA, and CN. This study was funded by Bournemouth University, United Kingdom.

Conflicts of Interest

All authors were involved in the development of the intervention, MotivATE.

Multimedia Appendix 1

MotivATE screenshots.

[PDF File (Adobe PDF File), 1MB - jmir_v21i2e11874_app1.pdf]

Multimedia Appendix 2

MotivATE case studies.

[PDF File (Adobe PDF File), 37KB - jmir_v21i2e11874_app2.pdf]

Multimedia Appendix 3

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 444KB - jmir_v21i2e11874_app3.pdf]

References

1. BEAT (Beating Eating Disorders). The Costs of Eating Disorders: Social, Health and Economic Impacts.: BEAT; 2015 Feb. URL: https://www.beateatingdisorders.org.uk/uploads/documents/2017/10/the-costs-of-eating-disorders-final-original.pdf [WebCite Cache ID 75bONPguj]
2. 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]
3. Nielsen S, Bará-Carril N. Family, burden of care, and social consequences. In: Treasure J, Schmidt U, van Furth E, editors. Handbook of Eating Disorders. 2nd edition. Chichester, UK: John Wiley & Sons; 2003:191-206.
4. Le Grange D, Loeb KL. Early identification and treatment of eating disorders: Prodrome to syndrome. Early Interv Psychiatry 2007 Feb;1(1):27-39. [doi: 10.1111/j.1751-7893.2007.00007.x] [Medline: 21352106]
5. Vaz AR, Conceição E, Machado PP. Early response as a predictor of success in guided self-help treatment for bulimic disorders. Eur Eat Disord Rev 2014 Jan;22(1):59-65. [doi: 10.1002/erv.2262] [Medline: 24123526]
6. Waller G, Schmidt U, Treasure J, Murray K, Aleyna J, Emanuelli F, et al. Problems across care pathways in specialist adult eating disorder services. Psychiatr Bull 2018 Jan 2;33(1):26-29. [doi: 10.1192/pb.bp.107.018325]
7. Williams S, Reid M. Understanding the experience of ambivalence in anorexia nervosa: The maintainer's perspective. Psychol Health 2010 Jun;25(5):551-567. [doi: 10.1080/08870440802617629] [Medline: 20204933]
8. Colton A, Pistrang N. Adolescents' experiences of inpatient treatment for anorexia nervosa. Eur Eat Disord Rev 2004 Aug 27;12(5):307-316. [doi: 10.1002/erv.587]
9. Leavey G, Vallianatou C, Johnson-Sabine E, Rae S, Gunputh V. Psychosocial barriers to engagement with an eating disorder service: A qualitative analysis of failure to attend. Eat Disord 2011;19(5):425-440. [doi: 10.1080/10640266.2011.609096] [Medline: 21932972]
10. Schoen E, Lee S, Skow C, Greenberg ST, Bell AS, Wiese JE, et al. A retrospective look at the internal help-seeking process in young women with eating disorders. Eat Disord 2012;20(1):14-30. [doi: 10.1080/10640266.2012.635560] [Medline: 22188057]
11. Muir S, Newell C, Griffiths J, Walker K, Hooper H, Thomas S, et al. MotivATE: A pretreatment Web-based program to improve attendance at UK outpatient services among adults with eating disorders. JMIR Res Protoc 2017 Jul 26;6(7):e146 [FREE Full text] [doi: 10.2196/resprot.7440] [Medline: 28747295]
12. Bartholomew LK, Parcel G, Kok G. Intervention mapping: A process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
13. Zelen M. A new design for randomized clinical trials. N Engl J Med 1979 May 31;300(22):1242-1245. [doi: 10.1056/NEJM197905313002203] [Medline: 431682]
14. Payne G, Payne J. Key Concepts in Social Research. London, UK: SAGE Publications; 2004.
15. Random.org. URL: https://www.random.org/ [accessed 2019-01-17] [WebCite Cache ID 75UzT5DgH]

16. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
17. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: An exploratory clinical study of a treatment for depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Auston, TX p. 1421-1430. [doi: 10.1145/2207676.2208602]
18. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
19. 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 Dec 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
20. Watson S, Woodside JV, Ware LJ, Hunter SJ, McGrath A, Cardwell CR, et al. Effect of a Web-based behavior change program on weight loss and cardiovascular risk factors in overweight and obese adults at high risk of developing cardiovascular disease: Randomized controlled trial. J Med Internet Res 2015 Jul 16;17(7):e177 [FREE Full text] [doi: 10.2196/jmir.3828] [Medline: 26183659]
21. 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]
22. Hötzel K, von Brachel R, Schmidt U, Rieger E, Kosfelder J, Hechler T, et al. An Internet-based program to enhance motivation to change in females with symptoms of an eating disorder: A randomized controlled trial. Psychol Med 2014 Jul;44(9):1947-1963. [doi: 10.1017/S0033291713002481] [Medline: 24128818]
23. Leung SF, Ma LC, Russell J. An open trial of self-help behaviours of clients with eating disorders in an online programme. J Adv Nurs 2013 Jan;69(1):66-76. [doi: 10.1111/j.1365-2648.2012.05988.x] [Medline: 22448710]
24. Leung SF, Ma J, Russell J. Enhancing motivation to change in eating disorders with an online self-help program. Int J Ment Health Nurs 2013 Aug;22(4):329-339. [doi: 10.1111/j.1447-0349.2012.00870.x] [Medline: 22882349]
25. Leung SF, Ma JL, Russell J. Enhancing quality of life in people with disordered eating using an online self-help programme. J Eat Disord 2013;1:9 [FREE Full text] [doi: 10.1186/2050-2974-1-9] [Medline: 24999391]
26. Lin JCC. Online stickiness: Its antecedents and effect on purchasing intention. Behav Inf Technol 2007 Nov;26(6):507-516 [FREE Full text] [doi: 10.1080/01449290600740843]
27. Denison-Day J, Appleton KM, Newell C, Muir S. Improving motivation to change amongst individuals with eating disorders: A systematic review. Int J Eat Disord 2018 Sep;51(9):1033-1050. [doi: 10.1002/eat.22945] [Medline: 30189116]
28. Prochaska JO. Transtheoretical model of behavior change. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer Science+Business Media; 2013.

Abbreviations

GCP: Good Clinical Practice
GP: general practitioner
NHS: National Health Service
OR: odds ratio

Edited by G Eysenbach; submitted 13.08.18; peer-reviewed by S Bauer, I Beintner; comments to author 14.10.18; revised version received 12.11.18; accepted 09.12.18; published 27.02.19.

Please cite as:
Denison-Day J, Muir S, Newell C, Appleton KM
A Web-Based Intervention (MotivATE) to Increase Attendance at an Eating Disorder Service Assessment Appointment: Zelen Randomized Controlled Trial
J Med Internet Res 2019;21(2):e11874
URL: http://www.jmir.org/2019/2/e11874/
doi:10.2196/11874
PMID:30810533

Original Paper

Is Blockchain Technology Suitable for Managing Personal Health Records? Mixed-Methods Study to Test Feasibility

Yu Rang Park[1*], PhD; Eunsol Lee[1*], MD; Wonjun Na[2], BS; Sungjun Park[1], BS; Yura Lee[3], MD, PhD; Jae-Ho Lee[3,4], MD, PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Clinical Research Center, Asan Institute of Life Sciences, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[3]Department of Biomedical Informatics, Asan Medical Center, Seoul, Republic of Korea

[4]Department of Emergency Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jae-Ho Lee, MD, PhD
Department of Biomedical Informatics
Asan Medical Center
88 Olympic-ro 43-gil, Songpa-gu
Seoul,
Republic of Korea
Phone: 82 02 3010 3350
Fax: 82 02 3010 8126
Email: jaeholee@amc.seoul.kr

Abstract

Background: There are many perspectives on the advantages of introducing blockchain in the medical field, but there are no published feasibility studies regarding the storage, propagation, and management of personal health records (PHRs) using blockchain technology.

Objective: The purpose of this study was to investigate the usefulness of blockchains in the medical field in relation to transactions with and propagation of PHRs in a private blockchain.

Methods: We constructed a private blockchain network using Ethereum version 1.8.4 and conducted verification using the de-identified PHRs of 300 patients. The private blockchain network consisted of one hospital node and 300 patient nodes. In order to verify the effectiveness of blockchain-based PHR management, PHRs at a time were loaded in a transaction between the hospital and patient nodes and propagated to the whole network. We obtained and analyzed the time and gas required for data transaction and propagation on the blockchain network. For reproducibility, these processes were repeated 100 times.

Results: Of 300 patient records, 74 (24.7%) were not loaded in the private blockchain due to the data block size of the transaction block. The remaining 226 individual health records were classified into groups A (80 patients with outpatient visit data less than 1 year old), B (84 patients with outpatient data from between 1 and 3 years before data collection), and C (62 patients with outpatient data 3 to 5 years old). With respect to mean transaction time in the blockchain, C (128.7 seconds) had the shortest time, followed by A (132.2 seconds) and then B (159.0 seconds). The mean propagation times for groups A, B, and C were 1494.2 seconds, 2138.9 seconds, and 4111.4 seconds, respectively; mean file sizes were 5.6 KB, 18.6 KB, and 45.38 KB, respectively. The mean gas consumption values were 1,900,767; 4,224,341; and 4,112,784 for groups A, B, and C, respectively.

Conclusions: This study confirms that it is possible to exchange PHR data in a private blockchain network. However, to develop a blockchain-based PHR platform that can be used in practice, many improvements are required, including reductions in data size, improved personal information protection, and reduced operating costs.

(J Med Internet Res 2019;21(2):e12533) doi:10.2196/12533

KEYWORDS

personal health record; mobile health; blockchain; Ethereum

Introduction

The ubiquity of mobile phones and rapid spread of wearable devices have greatly increased the amount and accuracy of data directly generated by patients outside of medical facilities [1-3]. These types of data, which are increasingly contributing to the use of personal health records (PHRs), are starting to attract as much attention as the patient data generated by medical institutions. Such data can be linked with the data collected at medical institutions, but it is expected that PHR management will become more personalized. This hope is one of the biggest drivers of change in the medical care paradigm, which is shifting its focus from general, guideline-based treatment to personalized treatment and disease prevention [3-8]. Although there are some advantages to this shift, there are economic, technological, regulatory, and sentimental barriers to the widespread adoption of PHR [9-11].

Recently, several studies have proposed the use of blockchain technology as a potential way to improve current PHR systems, which restrict access to and recording and sharing of data [12-16]. Blockchain databases and platforms are decentralized and irreversible; their advantages include reliability, transparency, and security [17,18]. Blockchain-based platforms are gradually expanding into a range of fields such as administration, insurance, and copyright act [19,20]. Distributed ledger technology (DLT) is the foundation of blockchain. DLT offers a consensus validation mechanism through a network of computers that facilitates peer-to-peer transactions without the need for an intermediary or centralized authority to update and maintain the information generated by the transactions. Each transaction is validated; a group of validated transaction is added and connected as a new “block” to an already existing chain of transactions, giving rise to the term “blockchain” [21]. In the medical field, the use of blockchain in electronic health records (EHRs), clinical trials, and drug tracking has been being proposed [22-24]. In clinical trials and research, there are several ways in which blockchain technology can improve the quality and processing of data [24]. For example, the smart contract function of the blockchain could be used to obtain consent from participants in a trial or track specific clinical trial events to improve the quality of a study [25]. Blockchain-based PHRs have been shown to solve technical and economic problems [12-14]. For example, instead of relying on a trusted third party, individuals may be able to manage their own data and can be assured of trust in the data through the blockchain. For example, after patients receive their medical records from medical institutions, these could be sent to other medical institutions, insurance companies, and research institutes, with the information being verified by the blockchain instead of a trusted third party. The medical information stored in patients’ mobile phones can be used without the help of a medical institution or a company. In addition, individuals can hand over their own data to data utilization parties, receive compensation for the data, and record the transaction details in the blockchain. To our knowledge, no published studies have examined the feasibility, effectiveness, performance, or costs of blockchain-based PHRs. The main questions that this study aimed to address were as follows: (1) Is the blockchain network suitable for PHR management? (2) How long does it take to share and distribute clinical data on a blockchain network? (3) How much does it cost to transmit clinical data in a blockchain network?

Methods

Study Design

To evaluate the usefulness of blockchain technology for the management of PHRs, we constructed a private blockchain network and conducted verification using real patient data. The blockchain-based PHR-sharing experiment was conducted on a 64-core, 398 GB Linux CentOS 6.9 server. The blockchain network was an Ethereum version 1.8.4–based private network [26], and 301 nodes were created from one local node via the Linux screen. We connected the additional 300 nodes to one main node representing a hospital.

To investigate the effectiveness of blockchain-based PHR management, we analyzed the time taken for data transactions on the blockchain nodes and the time taken for the spread of the clinical data over the network. Since the size of the clinical data is an important variable, we used data from 100 patients for each data size assessment. All clinical data were encoded with hexadecimal codes, and the transactions were performed with hex code in the transaction data field. In the same environment, 301 nodes were created, and one of them was assumed to be a hospital. The transactions for the clinical data of 300 patients were generated from the hospital node.

The times and costs associated with propagating the transactions of the 300 patient nodes were calculated. In order to calculate the propagation time for one transaction to all nodes in the network, the time until confirmation of the block containing the last transaction was defined as one cycle. The performance of each group was measured using the time required for one cycle and a measure of the propagation speed of the blockchain network according to the amount of data transmitted. In this private blockchain network, sending clinical data from a hospital node to patient nodes was repeated 100 times, and the time and cost were calculated as the average of these 100 iterations.

Cost was calculated using the gas fee, which is a special unit used in Ethereum networks. In Ethereum networks, every operation that can be performed by a transaction or contract costs a certain amount of gas, with operations that require more computational resources costing more gas than operations that require fewer computational resources [26]. For example, a high gas fee is engendered by a costly computation or an increase in the amount of data that must be stored in the node’s state. The gas fee is calculated by repeating 100 cycles of sharing and propagating clinical data in a blockchain network, in the same way as was done for calculating time.

Data Sources

The clinical data, which forms the input to the PHRs, were collected from 300 patients randomly selected from the anonymized data warehouse of the Asan Medical Center in Seoul, South Korea [27,28]. Inclusion criteria were patients who had at least one outpatient visit record since 2014 and received at least one diagnosis and at least one laboratory test.

The purpose of this study was to evaluate the capacity of blockchain technology to manage the patients' clinical data. The patients were grouped into group A (100 patients with less than 1 year of outpatient data), group B (100 patients whose outpatient data were from between 1 and 3 years before data collection), and group C (100 patients with 3 to 5 years of outpatient visit data). We excluded all patient records that exceeded the 64 KB limit for blockchain transactions. The American Society for Testing and Materials' Continuity of Care Record, a PHR standard, was used to fit the format of each patient's clinical data [29].

Data Analysis

We used Python 3.6.4 (Python Software Foundation) to obtain and analyze data for transaction logs between nodes. The transaction logs contained each node-specific data load time, the total network spread time, and the cost per transaction. We used analyses of variance (ANOVA) to test for statistical significance in the differences between the 3 groups. All reported *P* values were 2-sided, and *P* values less than .05 were considered significant. All statistical analysis was performed using R version 3.5.0 (R Foundation for Statistical Computing).

Ethical Considerations

This study was approved by the Asan Medical Center's Institutional Review Board (No. 2018-0178). The Ethics Committee waived the need for informed consent, as all data used in this study were anonymized and anonymously managed at all stages, including during data cleaning and statistical analyses.

Results

Overall Characteristics

Of the 300 patient records, the clinical data of 74 patients exceeded the 64 KB limit for transaction records in a blockchain: 20 from group A; 16 from group B; and 38 from group C. This elimination process left 226 individual health records in the final analysis: 80 in group A, 84 in group B, and 62 in group C (Table 1). All variables except gender were significantly different between the 3 groups. The number of visits, problems, medication, results, and procedures was highest in group A and lowest in group C.

Blockchain-Based Data Transaction and Propagation

Figure 1 shows comparisons of total transaction times and total propagation times among the 3 groups. The mean data sizes were 5.7 KB, 20 KB, and 37 KB for groups A, B, and C, respectively. In terms of mean transaction times, C (128.7 seconds) had the shortest time, followed by A (132.2 seconds) and B (159.0 seconds). The mean propagation times were 1494.2 seconds, 2138.9 seconds, and 4111.4 seconds for groups A, B, and C, respectively; the mean file sizes were 5.6 KB, 18.6 KB, and 45.38 KB, respectively.

Table 1. Basic characteristics of target groups.

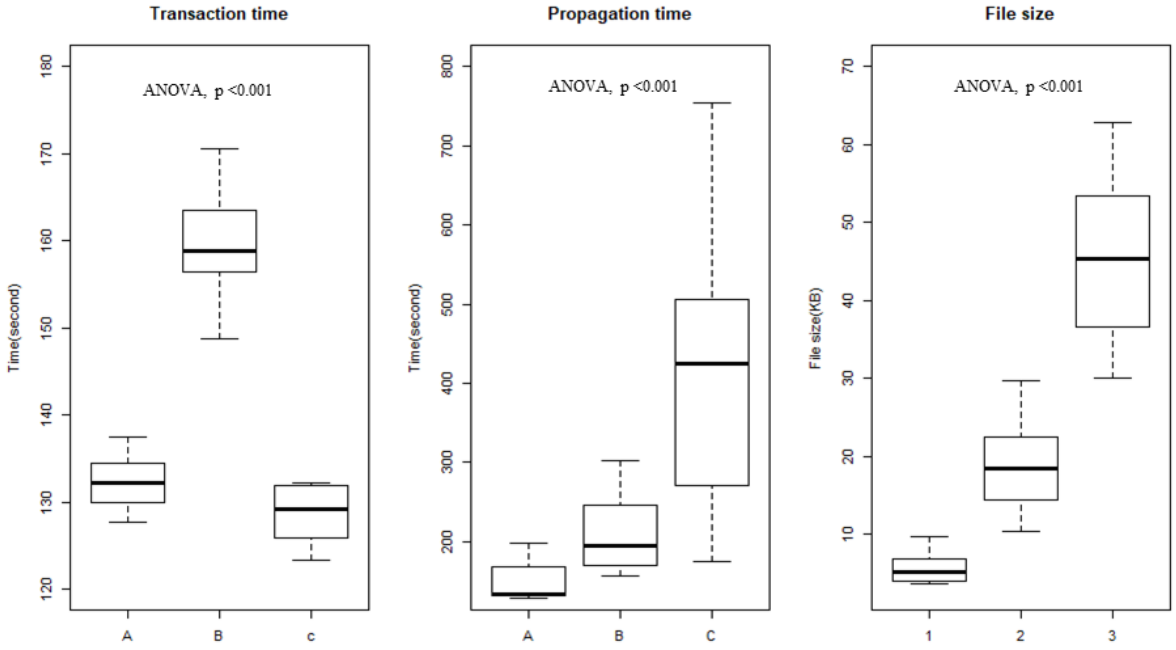
Characteristics	Group A[a] (n=80)	Group B[b] (n=84)	Group C[c] (n=62)
Gender, male, n (%)	37 (46)	32 (38)	32 (50)
Age, mean (SD)	44.27 (20.73)	45.67 (24.06)	55.89 (20.57)
Diagnosis rank, n (%)			
1st (group A joint pain; groups B and C essential [primary] hypertension)	8 (10)	19 (23)	26 (42)
2nd (group A essential [primary] hypertension; group B headache; group C chest pain, unspecified)	6 (8)	11 (13)	16 (26)
3rd (group A cough; group B chest pain, unspecified; group C Encounter for gynecological examination)	6 (8)	9 (11)	13 (21)
Visit frequency, mean (SD)	4.58 (3.91)	17.21 (9.26)	32.70 (19.94)
Problems, mean (SD)	3.68 (3.79)	18.79 (10.29)	37.9 (25.30)
Medication, mean (SD)	3.35 (4.68)	31.71 (17.14)	74.4 (47.03)
Results, mean (SD)	3.38 (4.62)	33.09 (16.86)	75.59 (46.43)
Procedure, mean (SD)	0.04 (0.25)	1.34 (1.75)	2.57 (3.04)

[a]Less than 1 year of outpatient visit data.

[b]1 to 3 years of outpatient visit data.

[c]3 to 5 years of outpatient visit data.

Figure 1. Mean transaction times, propagation times, and file sizes by group in a private blockchain network. The intergroup variance test for transaction time, propagation time, and file size was performed with analysis of variance and was found to be significant for all 3 variables.

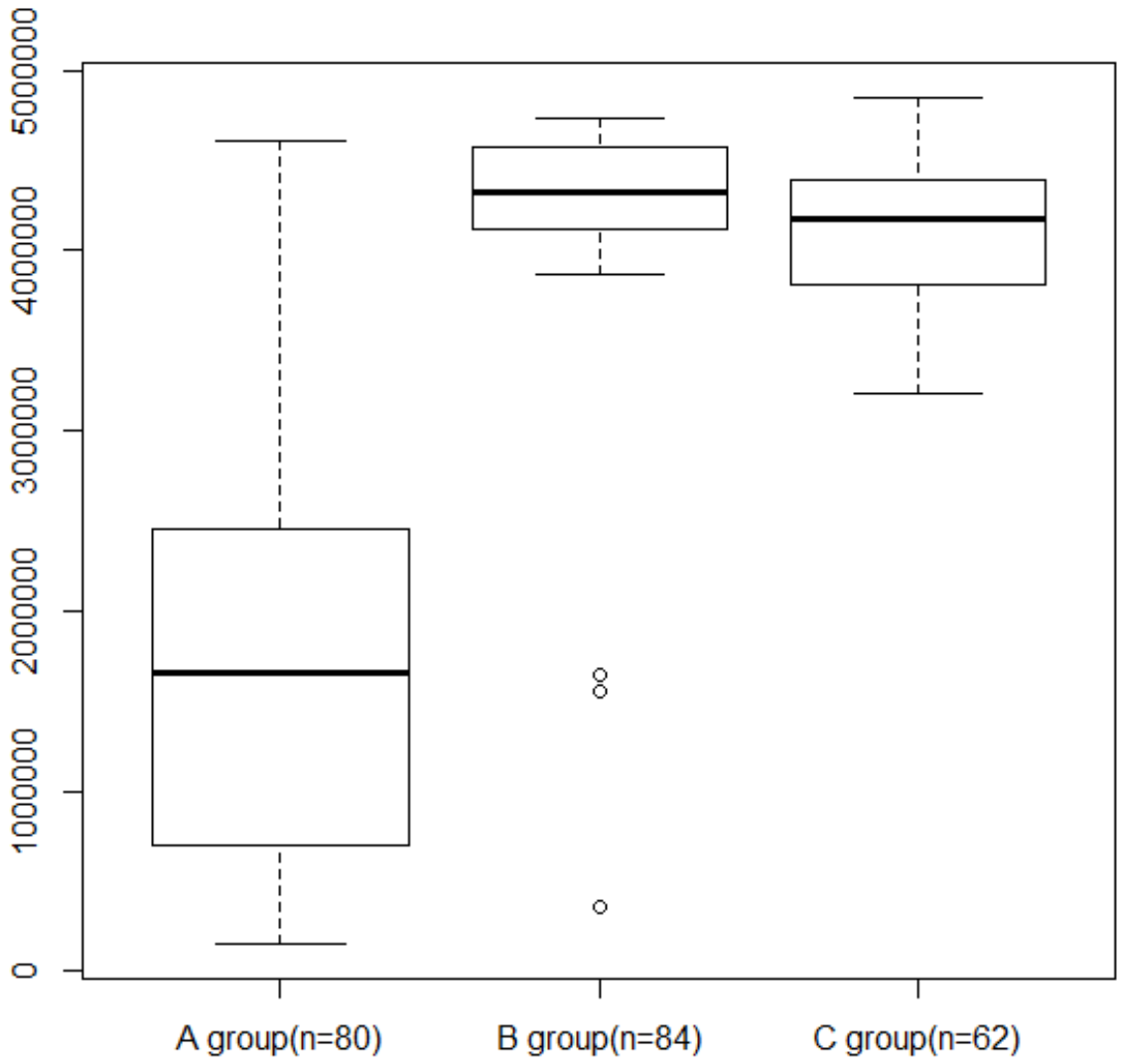


Blockchain-Based Data Operation Cost

To calculate gas fees, we divided the amount of gas required for each transaction in each group by the number of blocks actually executed. The gas consumed according to data size was found to be the same as the propagation speed. For a transaction to be propagated by each group, the blocks must be generated through a mining process. The mean number of blocks was 9.8 in group A, 9.5 in group B, and 17. 4 in group C. The larger the data size, the fewer the transactions included in one block.

Figure 2 shows the mean gas consumption per group required for the transaction and propagation of PHRs in a private blockchain network: 1,900,767 for A; 4,224,341 for B; and 4,112,784 for C. The mean data size multiplied by the gas consumption showed the same trend for propagation of the gas cost. However, B and C had similar mean data sizes, indicating that the mean number of blocks in group C was much lower than that in groups A and B.

Figure 2. Mean gas consumption per group required to transact and propagate personal health records in a private blockchain network.



Discussion

Principal Findings

To the best of our knowledge, this is the first study to investigate the possibility of transferring data between hospital- and patient-based databases using real PHRs on a blockchain network. The limitations of sharing PHRs on a private Ethereum blockchain network include, first, the fact that sharing a PHR with a transaction in a blockchain cannot save more than 64 KB. This restriction in Ethereum is to prevent distributed denial-of-service (DDoS) attacks [26]. Because of this limit, data for 74 patients were excluded from this study. A DDoS attack is a method of attacking a Web server by causing abnormal traffic to flow from multiple computers to the Web server, causing the server to fail due to excessive traffic, process progress, and excessive input/output operations. DDoS attacks on blockchains can be accomplished by creating many large blocks with dust transactions. In order to prevent such attacks, the size of a transaction is limited. This limitation of the blockchain needs to be addressed to justify the use of blockchain as a PHR platform, since patient-generated health data (PGHD), socioeconomic data, and genomic data are becoming larger, as are the standard components of PHR [1,30,31]. Second, the larger the size of the data, the greater the gas consumption on the network, resulting in unnecessary operating costs. A pattern that we identified was that the basic cost of gas increases the cost of transactions at a higher rate with increasing amounts of data. In particular, PHRs may increase in data size over time, which may increase the operational costs associated with using a blockchain network. The costs of PGHD produced in real time via the Internet of Things or wearable devices will further increase over time [1,3,4]. Third, privacy and confidentiality considerations limit the current utility of blockchain for PHR

management. In this study, data producers, owners, and content were all exposed. Exposure of a producer exposes the identity of the hospital and departments visited by a particular patient. The exposure of the transaction log means that the patient's medical history is exposed; most people consider this to be sensitive information. Last, records cannot be erased once they are recorded. This is a fundamental characteristic of a blockchain, but in the case of medical data, the data may change or may need to be removed.

Recommendations for Building a Blockchain Network for Personal Health Records

Transaction block size problems can be solved by choosing or designing blockchains that can accommodate high capacity. However, this modification could cause other problems, such as susceptibility to DDoS attacks [32], which cause data sizes to decrease. Therefore, there is a need for an alternative approach that can increase block data size without the associated security problems.

The time it takes to transact and propagate data using blockchain is a crucial variable in blockchain-based PHR. In this study, mean transaction time was longer with larger patient numbers, and mean propagation time was longer with larger file sizes but not with larger patient numbers. In this study, small patient sample groups were used for verification purposes. However, to allow for the management of and queries about large numbers of patients, it will be necessary to improve transaction and propagation times.

Increased operational costs due to large data sizes can be partially resolved by selecting a more advanced blockchain. The lack of privacy and confidentiality can also be improved by choosing a blockchain system that allows anonymization. The cryptographic technology called zk-SNARK (zero-knowledge succinct noninteractive argument of knowledge) is purported to allow users to hide both sender information and database content [33]. At present, however, anonymization techniques are imperfect and slow encryption speeds hinder performance.

Instead of writing all of the data, it is a good idea to record metadata such as data storage addresses, hash values, and timestamps in the blockchain. The actual data can then be stored elsewhere, such as on the hospital server, in patient mobile phones, or using a cloud-based storage system. This approach reduces the amount of data stored in the blockchain, regardless of the total database size, thereby freeing up storage space and minimizing costs. The approach also provides an alternative to storing personal and sensitive information on the blockchain and makes use of the advantages of both existing centralized storage technology and blockchain technology. At the same time, by allowing individuals to control storage, we can comply with the General Data Protection Regulation, which advises that "personal data shall be processed in a lawful manner, in a transparent manner in relation to the data subject" [34]. For blockchain to be optimized in a health care capacity, it should guarantee the right to health information for individuals that seek it. Recently, a model has been proposed that stores metadata in a blockchain and stores sensitive and large data in a separate storage such as a Cloud [35]. Until now, only the model has been presented, but we expect that services that implement these models will be released in the near future.

Limitations of This Study

The main limitation of this study is that the feasibility evaluation of blockchain for PHR management was done on a private rather than a public network. However, this design choice was made due to legal limitations which prevented us from uploading sensitive personal data to public networks. In the future, research using public network–based blockchains, with participant consent, will be necessary.

To calculate network latency and throughput, it is necessary to install and operate the system on multiple computers instead of configuring several nodes on a small number of computers, as was done in this study. Future studies using multiple computers will be able to test feasibility by appropriately distributing physically separated computers having researchers, services, and the Cloud in other regions.

Conclusions

Although many medical applications have been attempted using blockchain technology, studies investigating the feasibility and effectiveness of blockchain networks based on actual patient data have progressed relatively slowly. Our findings support the possibility of using blockchain technology to exchange actual patient data on a private blockchain network. Managing medical data using blockchain requires consideration of data size, operating costs, and privacy.

Acknowledgments

We would like to thank the Medical Information Office of Asan Medical Center for providing log data of mobile electronic medical records and supporting data analysis and interpretation.

Conflicts of Interest

None declared.

References

1. Kim J, Kam HJ, Park YR, Yoo S, Oh JS, Kim Y, et al. Enchanted life space: adding value to smart health by integrating human desires. Healthc Inform Res 2018 Jan;24(1):3-11 [FREE Full text] [doi: 10.4258/hir.2018.24.1.3] [Medline: 29503747]

2. Walker DM, Sieck CJ, Menser T, Huerta TR, Scheck MA. Information technology to support patient engagement: where do we stand and where can we go? J Am Med Inform Assoc 2017 Nov 01;24(6):1088-1094. [doi: 10.1093/jamia/ocx043] [Medline: 28460042]
3. Torkamani A, Andersen KG, Steinhubl SR, Topol EJ. High-definition medicine. Cell 2017 Aug 24;170(5):828-843. [doi: 10.1016/j.cell.2017.08.007] [Medline: 28841416]
4. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]
5. Singh K, Drouin K, Newmark LP, Filkins M, Silvers E, Bain PA, et al. Patient-facing mobile apps to treat high-need, high-cost populations: a scoping review. JMIR Mhealth Uhealth 2016 Dec 19;4(4):e136 [FREE Full text] [doi: 10.2196/mhealth.6445] [Medline: 27993761]
6. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing user engagement of an mhealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
7. Sosa A, Heineman N, Thomas K, Tang K, Feinstein M, Martin MY, et al. Improving patient health engagement with mobile texting: a pilot study in the head and neck postoperative setting. Head Neck 2017 May;39(5):988-995. [doi: 10.1002/hed.24718] [Medline: 28263468]
8. Park YR, Lee Y, Kim JY, Kim J, Kim HR, Kim Y, et al. Managing patient-generated health data through mobile personal health records: analysis of usage data. JMIR Mhealth Uhealth 2018 Apr 09;6(4):e89 [FREE Full text] [doi: 10.2196/mhealth.9620] [Medline: 29631989]
9. Butler J, Carter M, Hayden C, Gibson B, Weir C, Snow L. Understanding adoption of a personal health record in rural health care clinics: revealing barriers and facilitators of adoption including attributions about potential patient portal users and self-reported characteristics of early adopting users. 2013 Presented at: AMIA Annual Symposium Proceedings; 2013; Washington.
10. Tang P, Ash J, Bates D, Overhage J, Sands D. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Informat Assoc 2006;13(2):121-126. [Medline: 16357345]
11. Lober W, Zierler B, Herbaugh A, Shinstrom S, Stolyar A. Barriers to the use of a personal health record by an elderly population. 2006 Presented at: AMIA Annual Symposium Proceedings; 2006; Washington.
12. Stagnaro C. White paper: innovative blockchain uses in health care.: Freed Associates; 2017. URL: https://www.freedassociates.com/wp-content/uploads/2017/08/Blockchain_White_Paper.pdf [accessed 2019-01-10] [WebCite Cache ID 75Jcbiunt]
13. Ivan D. Moving toward a blockchain-based method for the secure storage of patient records. 2016 Presented at: ONC/NIST Use of Blockchain for Healthcare and Research Workshop; 2016; Gaithersburg.
14. MediBloc: blockchain-based healthcare information ecosystem. 2017. URL: https://medibloc-homepage.oss-us-west-1.aliyuncs.com/whitepaper/medibloc_whitepaper_en.pdf [accessed 2018-09-30]
15. Roehrs A, da Costa CA, da Rosa Righi R. OmniPHR: a distributed architecture model to integrate personal health records. J Biomed Inform 2017 Dec;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]
16. Amar D, Choudhury O. Blockchain for secure patient centered data capture and sharing. 2018. URL: https://zerista.s3.amazonaws.com/item_files/04e5/attachments/431078/original/amia_blockchain_ibmresearch.pdf [accessed 2019-01-10] [WebCite Cache ID 75JcuyP94]
17. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. 2008. URL: https://bitcoin.org/bitcoin.pdf [accessed 2019-01-10] [WebCite Cache ID 75Jd2BXh4]
18. Buterin V. Ethereum: a next-generation cryptocurrency and decentralized application platform. 2014. URL: https://bitcoinmagazine.com/articles/ethereum-next-generation-cryptocurrency-decentralized-application-platform-1390528211/ [accessed 2019-01-10] [WebCite Cache ID 75JdRivrx]
19. Underwood S. Blockchain beyond bitcoin. Communications ACM 2016;59(11):15-17. [doi: 10.1145/2994581]
20. Cohn A, West T, Parker C. Smart after all: blockchain, smart contracts, parametric insurance, and smart energy grids. Georgetown Law Technol Rev 2017;1(2):273-304.
21. Blockchain technology and regulatory investigations. 2018. URL: https://www.steptoe.com/images/content/1/7/v2/171967/LIT-FebMar18-Feature-Blockchain.pdf [accessed 2019-01-18] [WebCite Cache ID 75WzR2xvQ]
22. Wang H, Song Y. Secure cloud-based EHR system using attribute-based cryptosystem and blockchain. J Med Syst 2018 Jul 05;42(8):152. [doi: 10.1007/s10916-018-0994-6] [Medline: 29974270]
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. Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. Trials 2017 Dec 19;18(1):335 [FREE Full text] [doi: 10.1186/s13063-017-2035-z] [Medline: 28724395]
25. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: transparency and traceability of consent. F1000Res 2017 Jul 4;6:66. [doi: 10.12688/f1000research.10531.3]

26. Wood G. Ethereum: a secure decentralised generalised transaction ledger. 2014. URL: https://gavwood.com/paper.pdf [accessed 2019-01-10] [WebCite Cache ID 75Je8FcBR]
27. 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]
28. Shin S, Kim WS, Lee J. Characteristics desired in clinical data warehouse for biomedical research. Healthc Inform Res 2014 Apr;20(2):109-116 [FREE Full text] [doi: 10.4258/hir.2014.20.2.109] [Medline: 24872909]
29. Smolij K, Dun K. Patient health information management: searching for the right model. Perspect Health Inf Manag 2006 Dec 12;3:10 [FREE Full text] [Medline: 18066368]
30. Eisenstein M. Big data: the power of petabytes. Nature 2015 Nov 05;527(7576):S2-S4. [doi: 10.1038/527S2a] [Medline: 26536222]
31. Blazquez D, Domenech J. Big Data sources and methods for social and economic analyses. Technological Forecasting Soc Change 2018 May;130:99-113. [doi: 10.1016/j.techfore.2017.07.027]
32. McDowell M. Security tip (ST04-015): understanding denial-of-service attacks.: National Cyber Alert System; 2004. URL: https://www.us-cert.gov/ncas/tips/ST04-015 [accessed 2019-01-19] [WebCite Cache ID 75Wxu2y1g]
33. Sasson E, Chiesa A, Garman C, Green M, Miers I, Tromer E. Zerocash: decentralized anonymous payments from Bitcoin. 2014. URL: http://zerocash-project.org/media/pdf/zerocash-oakland2014.pdf [accessed 2019-01-10] [WebCite Cache ID 75JejuGUr]
34. Rumbold J, Pierscionek B. The effect of the general data protection regulation on medical research. J Med Internet Res 2017 Dec 24;19(2):e47 [FREE Full text] [doi: 10.2196/jmir.7108] [Medline: 28235748]
35. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-based medical records secure storage and medical service framework. J Med Syst 2018 Nov 22;43(1):5. [doi: 10.1007/s10916-018-1121-4] [Medline: 30467604]

Abbreviations

ANOVA: analysis of variance
DDoS: distributed denial of service
DLT: distributed ledger technology
EHR: electronic health record
PGHD: patient-generated health data
PHR: personal health record
zk-SNARK: zero-knowledge succinct noninteractive argument of knowledge

Edited by G Eysenbach; submitted 18.10.18; peer-reviewed by J Liu, S Sawesi; comments to author 29.11.18; revised version received 21.12.18; accepted 25.12.18; published 08.02.19.

Please cite as:
Park YR, Lee E, Na W, Park S, Lee Y, Lee JH
Is Blockchain Technology Suitable for Managing Personal Health Records? Mixed-Methods Study to Test Feasibility
J Med Internet Res 2019;21(2):e12533
URL: http://www.jmir.org/2019/2/e12533/
doi:10.2196/12533
PMID:30735142

Viewpoint

Beyond One-Off Integrations: A Commercial, Substitutable, Reusable, Standards-Based, Electronic Health Record–Connected App

Kenneth D Mandl[1,2], MD, MPH; Daniel Gottlieb[1,2], MPA; Alyssa Ellis[1], MS

[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States
[2]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Kenneth D Mandl, MD, MPH
Computational Health Informatics Program
Boston Children's Hospital
300 Longwood Avenue
Boston, MA, 02115
United States
Phone: 1 6173554145
Email: kenneth_mandl@harvard.edu

Abstract

The Substitutable Medical Apps and Reusable Technology (SMART) Health IT project launched in 2010 to facilitate the development of medical apps that are scalable and substitutable. SMART defines an open application programming interface (API) specification that enables apps to connect to electronic health record systems and data warehouses without custom integration efforts. The SMART-enabled version of the Meducation app, developed by Polyglot, has been implemented at scores of hospitals and clinics in the United States, nation-wide. After expanding their product's reach by relying on a universal, open API for integrations, the team estimates that one project manager can handle up to 20 simultaneous implementations. The app is made available through the SMART App Gallery, an open app store that supports discovery of apps and, because the apps are substitutable, market competition. This case illustrates how a universal open API for patient and clinician-facing health IT systems supported and accelerated commercial success for a start-up company. Giving end users a wide and ever-growing choice of apps that leverage data generated by the health care system and patients at home through a universal, open API is a promising and generalizable approach for rapid diffusion of innovation across health systems.

(J Med Internet Res 2019;21(2):e12902) doi:10.2196/12902

KEYWORDS

electronic medical records; application programming interfaces

Introduction

The opportunity has never been greater to build a flourishing ecosystem of modern health care software [1]. Underpinnings of interoperability for data exchange and integration of diverse software have been developed, tested, and proven, technically and commercially. We present a health IT start-up's trajectory as it transcended the common, frustrating modus operandi of *one-off* integrations of software for each new customer. Instead, the company leveraged an emerging, open, standards-based, national-scale, *app store* model. We believe that this success strongly supports an international strategy to deliver on health IT's promises of evidence-based medicine, artificial intelligence, genomics, and value-based care, while promoting the success of innovators and empowering physicians and patients with the tools they need for 21st-century health care.

The Obama administration appropriated US $34 billion for incentivizing physicians and hospitals to purchase commercial electronic health records (EHRs) [2]. Because these products evolved from billing systems, clinical functionality was underdeveloped, noninteroperable, and not designed to exchange data [3]. A novel approach addressing these gaps, proposed in 2009 [4], was to connect apps to EHRs, much the same way mobile phone users select apps from app stores. Such apps would be substitutable: readily added to or deleted from the EHR across an application programming interface (API). APIs are fundamental to software produced by Apple, Google, Microsoft, Facebook, and Amazon and are used by billions of

consumers. APIs enable software systems to interact with each other and exchange data. The Apple and Android APIs have spawned the creation of hundreds of thousands of apps without a need to devise subsequent interfaces, integrations, or agreements with Apple or Google.

To open EHRs to substitutable, third-party apps, the US Office of the National Coordinator of Health Information Technology funded the Substitutable Medical Apps and Reusable Technology (SMART) Health IT project [5-9]. The idea was that app developers, given an API specification that enabled access to the most commonly needed clinical data elements, including laboratory data, diagnoses, and medications, could readily create apps that could be added to or deleted from EHRs as easily as the (then 1-year-old) iPhone. Substitutability, a new approach to interoperability, was hoped to accelerate innovation by reducing barriers to integration and creating competition among developers; customers could select EHR-connected apps from an *app gallery*. Further, substitutability should enable customization of the end-user experience for providers using EHRs and patients accessing EHR portals.

In 2011, encouraged by the White House [10], the SMART team hosted a challenge to develop an app using the SMART API to connect to heath system data. A start-up company, Polyglot Systems, won for their app Meducation, designed to serve patients with low literacy levels, impaired vision, or language barriers. Meducation had been developed as a cloud-based medication management app, dynamically creating patient-specific medication instructions, and was designed to be intuitive and easy to read and understand in more than 20 languages. The user interface was designed through an iterative, user-centered design process with continual end-user feedback from IT teams at customer sites.

In under a week, the Polyglot team converted their existing app into a prototype SMART app that could be integrated with an EHR to pull the patient's medication list across the API. The Polyglot team struggled a bit to understand how the clinical data (eg, medications and laboratories) were represented in the SMART API. In 2014, the SMART project [5] swapped out its own data models for the emerging international standard, Health Level Seven's (HL7) Fast Health Interoperability Resources (FHIR) [11]. The contest also demonstrated the principle of reusability of apps, in that the Meducation app was able to readily run, without modification or customization, against three different clinical systems exhibiting the SMART API.

Polyglot pursued the use of the open SMART API in its strategy for integrating Meducation with EHRs and into patient and physician workflows. The company has since been acquired by First Databank, Inc. We explore implications of this approach for the rapidly evolving apps-driven health information economy [1].

One-Off Commercial Integrations

SMART was not yet widely adopted when Polyglot won the contest. Therefore, their initial implementations required time-consuming and resource-intensive custom integrations for each clinical environment. Early commercial engagements were expensive for Polyglot and, since the company did not charge separately for implementation work, they incurred large costs. Further, the Polyglot team was limited to two to three implementations at any given time and required a third-party integrator to handle project management and resolve technical issues around integration.

One of Meducation's initial implementations, prior to using SMART, was at a four-hospital system, HealthFirst, in Florida using the Allscripts Sunrise EHR. The project initiated in August 2015 and took 8 months to go live, after many integration challenges and higher-than-expected costs. There were issues in getting data to pass consistently from Allscripts Sunrise into the Meducation platform. Often the data were incomplete and missing information critical to drug safety. There were further issues in proper rendering of the user interface in the preferred workflow. System upgrades to Allscripts Sunrise often broke the interface and disrupted core functionalities, such as playing videos and rendering documents. Custom development projects came at a heavy price in time and effort, requiring substantial troubleshooting and testing.

These challenges were overcome through strong collaboration and a third-party integrator. The custom integration of a third-party app into a health system's existing clinical systems is a huge undertaking. For this project, the hospital put together a project team of clinicians, pharmacists, and informatics personnel who met regularly to oversee the project and focus on end-user experience. The hospital's IT group was brought on board early on to scope out the project and work with the Polyglot team on approaches to successful integration. The EHR vendor dedicated a team to the project for custom development work and an expert third-party integrator was brought in to deal with integration issues as they arose.

Some challenges, however, remained unresolved. For example, it would have been ideal for the Meducation document to be deposited into the EHR and included in the patient discharge instructions. This was not possible with the clinical systems and custom integration solutions available at the time, so end users needed to collate separate documents.

Subsequent implementations at clinical sites with an EHR system that Meducation had previously integrated with were more streamlined than the first-time implementation, generally taking only 3 months. The approach was to replicate the last working implementation with that EHR vendor and then troubleshoot for site-specific tailoring. Typical issues encountered during a second-time integration with a vendor include variations in the system versions, such as security settings and Internet browser type and version. Identifying and adapting to the differences across sites still required substantial involvement from the hospital's IT group.

In contrast, starting a custom integration project at a clinical site with an EHR vendor with which Meducation had not yet implemented was like starting from scratch. The only lessons that could be carried over from other implementations were around the need for trustworthy and reliable partners, as well as the collaboration and communication that needs to be established for a successful integration. Not all implementation

projects were successful. Reasons for failed integration were often related to the nature and complexity of the custom work.

Advances in the SMART Ecosystem

The 21st Century Cures Act, signed into US law in 2016, requires that certified health IT products have an API that that allows consumers to access, exchange, and use their health information "without special effort" [12,13]. In the United States, most EHR vendors are using the *SMART on FHIR* API to satisfy this requirement [14-16]. EHR vendors have collaborated over the interpretation and profiling of the FHIR standard with selected terminologies (eg, RxNorm, Systematized Nomenclature of Medicine [SNOMED], and Logical Observation Identifiers Names and Codes [LOINC]) through the Argonaut Project [16]; they have incorporated support for SMART into their products and are beginning to roll out the technology to health care institutions. Epic and Cerner have SMART-enabled hundreds of their installs. Allscripts has SMART-enabled the releases of three of their products. Other EHR companies, including Athena and eClinical Works, have also built SMART support into their products.

Apple recently launched a new version of its Health app with a health-records feature for connecting with EHR instances, to enable a patient to download a copy of her records to her phone [17]. Hundreds of hospitals and health systems offer this capability.

Technical and Business Efficiencies

Today, Medication uses the SMART on FHIR API to access and integrate with clinical systems; a single project manager can oversee 20 or more highly standardized implementation projects without a third-party integrator. While early custom implementations of Medication could not successfully integrate Medication documents into the patient discharge instructions from the EHR, with SMART on FHIR, the company is increasingly able to accomplish this within the FHIR standard.

At Polyglot, SMART on FHIR has eliminated the need for large project teams and for third-party involvement. The team was able to configure their Medication platform to the SMART on FHIR specifications by utilizing the SMART sandbox testing environment and sample patient datasets, which simulate a standards-compliant EHR.

When a health system expresses interest in Medication, there is significantly less hesitation and resource allocation required from the hospital's IT group and project management is much simpler. Integration of Medication is standardized. There is little to no custom development needed to integrate the app with each vendor's EHR system; thereby, testing and troubleshooting time is minimal.

For a new customer who is using an EHR system with which Medication has not previously been integrated, they can nonetheless use a standard project plan, checklist, and resource list. None of the project management materials need to change based on the EHR system. Therefore, the project cost and timeline can be accurately estimated from the start. By relying on standard functionality regardless of the EHR system, the costs are much lower and much more predictable.

The standardization of the Medication app on SMART has dramatically reduced the integration and implementation time by nearly eliminating the need for custom work; writing data back into EHRs still requires custom work in some cases, as the necessary components of the FHIR data model have not yet been universally implemented by vendors.

Though the app is still hosted through some legacy integrations with MEDITECH and Greenway Health, all other EHR implementations have migrated to substitutable patient-facing and provider-facing versions connecting via the SMART on FHIR API to Epic, Cerner Millennium, Allscripts Paragon, Allscripts Sunrise, eClinicalWorks, and athenahealth athenaClinicals. To date, the Medication app has been implemented in over 3000 pharmacies, 60 clinics, and 70 hospitals in the United States nationwide. Of these, about 50 of the hospitals and 15 of the clinics are integrated via SMART.

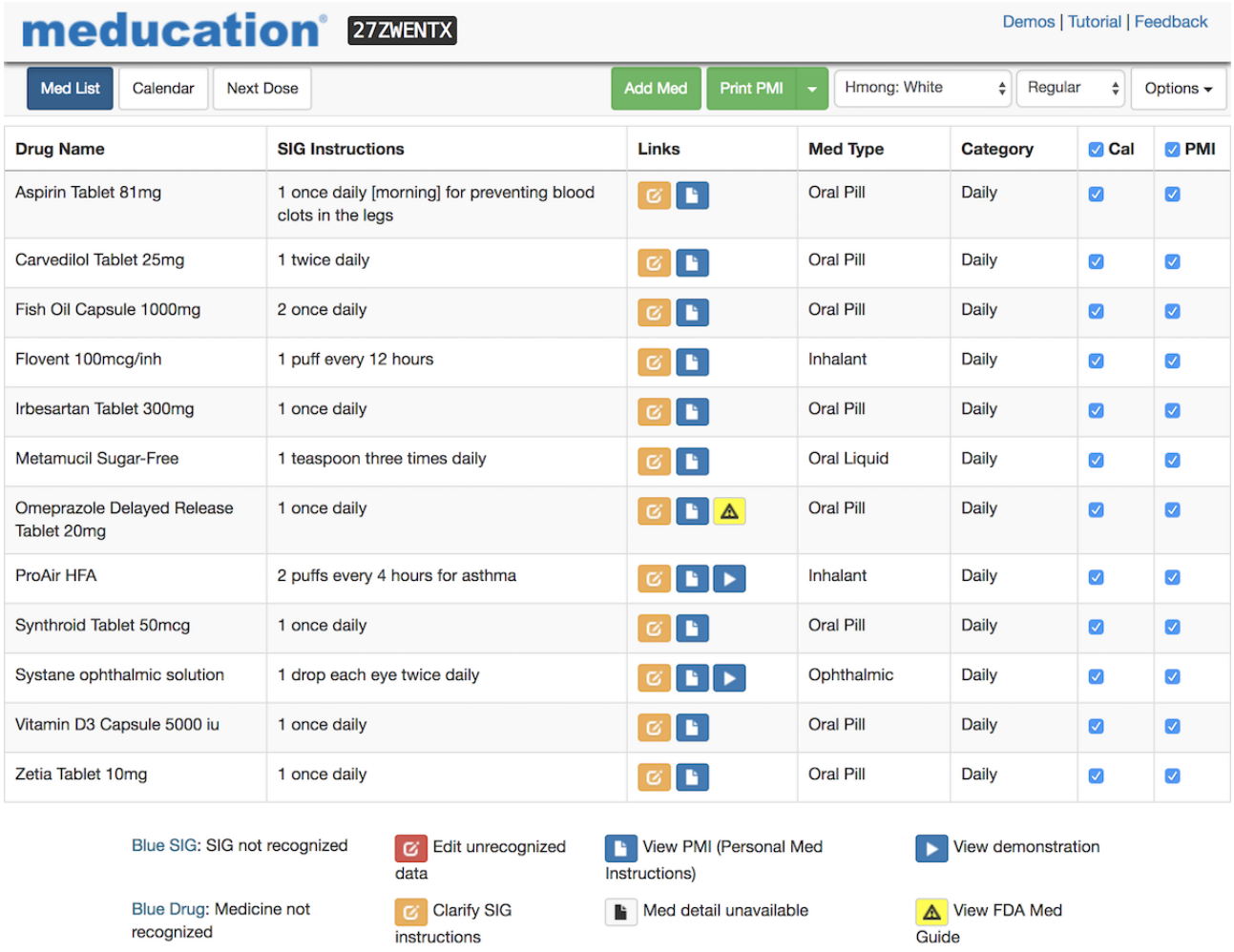
The current health apps marketplace includes the EHR vendor app stores, as well as iTunes and various Android app stores. The SMART App Gallery [18] is an open directory, showcasing SMART and FHIR apps with links to the various online stores where the app can be purchased. The Medication app is listed in the SMART App Gallery (see Figure 1). The app utilizes the gallery's demo functionality to enable users to launch the app against the sample clinical data in a sandbox environment, simulating the experience of using it in an EHR (see Figure 2). In addition, after terms were established with several EHR vendors, it has been listed in the Cerner App Gallery, Allscripts Application Store, and Athena Marketplace and will soon be listed in the Epic App Orchard.

As these emerging marketplaces for apps grow, the task of maintaining high-quality and safety levels for apps will be increasingly important. We expect this will occur through a mix of regulation [19], development-community standards [20], professional organization endorsements [21], third-party security certifications, and ongoing due diligence by health care organizations choosing to purchase and install the apps.

Innovative health app developers are using the SMART on FHIR open API specification to connect their products securely and efficiently into today's health care ecosystem. Standardizing the integration process, as the Medication team has done, eliminates the need for external, third-party services and enables efficiency and market scale for new innovations.

Figure 1. Meducation listing page on the SMART App Gallery [18].

Figure 2. SMART sandbox demonstration data. Potential users can try the app by launching it against the demonstration data.



Key Messages

APIs are ubiquitous in modern computing and enable software systems to interact with each other and exchange data. APIs have sparked ecosystems of apps for the iPhone and Android operating systems. They can similarly be used to drive an ecosystem of health care apps, essentially an *app store for health*.

In the United States, open and free API standards, developed, tested, and proven in the health care domain, have been integrated into EHR products and used to connect software to EHRs by both large (eg, Apple) and small software companies.

The experience of one commercial software company—Polyglot, recently acquired by First Databank, Inc—cogently illustrates how an open, consistent, standards-based API can foster innovation as well as commercial success.

Acknowledgments

We thank Lori McLean, Dr Daniel Fritsch, and their colleagues at First DataBank, Inc, for providing the authors with detailed information about their business model and engagements with EHR vendors and health systems. This work was funded in part by a grant from the National Institute of General Medical Sciences, National Institutes of Health (grant number R01GM104303).

Conflicts of Interest

Polyglot and First Databank, Inc, are members of the SMART Health IT Advisory Committee and contribute philanthropically to the Boston Children's Hospital Computational Health Informatics Program where the paper's authors are based.

References

1. Mandl KD, Mandel JC, Kohane IS. Driving innovation in health systems through an apps-based information economy. Cell Syst 2015 Jul;1(1):8-13 [FREE Full text] [doi: 10.1016/j.cels.2015.05.001] [Medline: 26339683]

2. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 04;362(5):382-385. [doi: 10.1056/NEJMp0912825] [Medline: 20042745]
3. Rosenbaum L. Transitional chaos or enduring harm? The EHR and the disruption of medicine. N Engl J Med 2015 Oct 22;373(17):1585-1588. [doi: 10.1056/NEJMp1509961] [Medline: 26488690]
4. Mandl KD, Kohane IS. No small change for the health information economy. N Engl J Med 2009 Mar 26;360(13):1278-1281. [doi: 10.1056/NEJMp0900411] [Medline: 19321867]
5. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Dec;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
6. Mandl KD, Mandel JC, Murphy SN, Bernstam EV, Ramoni RL, Kreda DA, et al. The SMART Platform: Early experience enabling substitutable applications for electronic health records. J Am Med Inform Assoc 2012;19(4):597-603 [FREE Full text] [doi: 10.1136/amiajnl-2011-000622] [Medline: 22427539]
7. Strategic Health IT Advanced Research Projects (SHARP) Program. Washington, DC: The Office of the National Coordinator for Health Information Technology URL: https://www.healthit.gov/sites/default/files/factsheets/get-the-facts-about-sharp-program.pdf [accessed 2019-01-22] [WebCite Cache ID 75crxMxhd]
8. SMART Health IT. Boston, MA: Computational Health Informatics Program, Boston Children's Hospital URL: https://smarthealthit.org/ [accessed 2019-01-22] [WebCite Cache ID 75csE6z8S]
9. The Office of the National Coordinator for Health Information Technology (ONC) Office of the Secretary, United States Department of Health and Human Services (HHS). 2016 Report to Congress on Health IT Progress: Examining the HITECH Era and the Future of Health IT. Washington, DC: The Office of the National Coordinator for Health Information Technology; 2016 Nov. URL: https://www.healthit.gov/sites/default/files/2016_report_to_congress_on_healthit_progress.pdf [accessed 2019-01-28] [WebCite Cache ID 75luFUlWX]
10. Chopra A. The White House, President Barack Obama. 2011 Mar 10. SMArt Prize for patients, physicians, and researchers URL: http://www.whitehouse.gov/blog/2011/03/10/smart-prize-patients-physicians-and-researchers [accessed 2019-01-22] [WebCite Cache ID 75csmEzE7]
11. FHIR. HL7 FHIR Foundation URL: http://www.fhir.org/ [accessed 2019-01-22] [WebCite Cache ID 75csvANKh]
12. 114th Congress (2015-2016). Congress.gov. 2016. H.R.34 - 21st Century Cures Act URL: https://www.congress.gov/bill/114th-congress/house-bill/34/ [accessed 2019-01-22] [WebCite Cache ID 75ct5CzLM]
13. Mandl KD, Kohane IS. A 21st-century health IT system: Creating a real-world information economy. N Engl J Med 2017 May 18;376(20):1905-1907. [doi: 10.1056/NEJMp1700235] [Medline: 28514623]
14. Halamka JD. Life as a Healthcare CIO. 2014 Dec 17. The Argonaut Project Charter URL: http://geekdoctor.blogspot.com/2014/12/the-argonaut-project-charter.html [accessed 2019-01-22] [WebCite Cache ID 75ctE2dCR]
15. Health Level Seven International. 2018. Main Page/Implementation URL: http://argonautwiki.hl7.org/index.php?title=Main_Page/Implementation [accessed 2019-01-22] [WebCite Cache ID 75ctKfUfV]
16. Ribick A. HL7 Argonaut Project Takes Major Step Forward for FHIR. Ann Arbor, MI: Health Level Seven International; 2017 Feb 15. URL: https://drive.google.com/file/d/0B6QYdUw2Ye6hRWJneTZQOGJZTFU/view?usp=drive_open [accessed 2019-01-22] [WebCite Cache ID 75ctQxLV0]
17. Mandl KD. CNBC. 2018 Jan 30. Apple will finally replace the fax machine in health care URL: https://www.cnbc.com/2018/01/30/apple-will-finally-replace-the-fax-machine-in-health-care-commentary.html [accessed 2019-01-28] [WebCite Cache ID 75lvY67jm]
18. SMART App Gallery. Boston, MA: Computational Health Informatics Program, Boston Children's Hospital URL: https://gallery.smarthealthit.org/ [accessed 2019-01-22] [WebCite Cache ID 75ctWN1H0]
19. Shuren J, Patel B, Gottlieb S. FDA regulation of mobile medical apps. JAMA 2018 Jul 24;320(4):337-338. [doi: 10.1001/jama.2018.8832] [Medline: 29971339]
20. Carin Alliance. URL: https://www.carinalliance.com/ [accessed 2019-01-22] [WebCite Cache ID 75cty7gMz]
21. O'Reilly KB. Proposed guidelines aim for safe, effective, mobile health apps. Chicago, IL: American Medical Association; 2018 Jan 22. URL: https://www.ama-assn.org/practice-management/digital/proposed-guidelines-aim-safe-effective-mobile-health-apps [accessed 2019-01-22] [WebCite Cache ID 75cuKFpTb]

Abbreviations

API: application programming interface
EHR: electronic health record
FHIR: Fast Health Interoperability Resources
HL7: Health Level Seven
LOINC: Logical Observation Identifiers Names and Codes
SMART: Substitutable Medical Apps and Reusable Technology
SNOMED: Systematized Nomenclature of Medicine

Edited by G Eysenbach; submitted 24.11.18; peer-reviewed by S Andrews; comments to author 02.01.19; revised version received 16.01.19; accepted 20.01.19; published 01.02.19.

Please cite as:
Mandl KD, Gottlieb D, Ellis A
Beyond One-Off Integrations: A Commercial, Substitutable, Reusable, Standards-Based, Electronic Health Record–Connected App
J Med Internet Res 2019;21(2):e12902
URL: http://www.jmir.org/2019/2/e12902/
doi:10.2196/12902
PMID:30707097

Original Paper

Understanding Youths' Ability to Interpret 3D-Printed Physical Activity Data and Identify Associated Intensity Levels: Mixed-Methods Study

Sam Graeme Morgan Crossley[1*], BSc, PhD; Melitta Anne McNarry[1*], BSc, PhD; Michael Rosenberg[2], BAppSc, DipEd, MPH, PhD; Zoe R Knowles[3*], HCPC, PhD; Parisa Eslambolchilar[4], BEng, MEng, PhD; Kelly Alexandra Mackintosh[1*], BSc, MSc, PhD

[1]Applied Sports Science Technology and Medicine Research Centre (A-STEM), Swansea University, Swansea, United Kingdom

[2]School of Human Sciences, The University of Western Australia, Crawley, Australia

[3]Physical Activity Exchange, School of Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom

[4]School of Computer Science and Informatics, Cardiff University, Cardiff, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Sam Graeme Morgan Crossley, BSc, PhD
Applied Sports Science Technology and Medicine Research Centre (A-STEM)
Swansea University
Bay Campus, Engineering East Building
Swansea, SA1 8EN
United Kingdom
Phone: 44 07707470931
Email: 557947@swansea.ac.uk

Abstract

Background: A significant proportion of youth in the United Kingdom fail to meet the recommended 60 minutes of moderate-to-vigorous physical activity every day. One of the major barriers encountered in achieving these physical activity recommendations is the perceived difficulty for youths to interpret physical activity intensity levels and apply them to everyday activities. Personalized physical activity feedback is an important method to educate youths about behaviors and associated outcomes. Recent advances in 3D printing have enabled novel ways of representing physical activity levels through personalized tangible feedback to enhance youths' understanding of concepts and make data more available in the everyday physical environment rather than on screen.

Objective: The purpose of this research was to elicit youths' (children and adolescents) interpretations of two age-specific 3D models displaying physical activity and to assess their ability to appropriately align activities to the respective intensity.

Methods: Twelve primary school children (9 boys; mean age 7.8 years; SD 0.4 years) and 12 secondary school adolescents (6 boys; mean age 14.1 years; SD 0.3 years) participated in individual semistructured interviews. Interview questions, in combination with two interactive tasks, focused on youths' ability to correctly identify physical activity intensities and interpret an age-specific 3D model. Interviews were transcribed verbatim, content was analyzed, and outcomes were represented via tables and diagrammatic pen profiles.

Results: Youths, irrespective of age, demonstrated a poor ability to define moderate-intensity activities. Moreover, children and adolescents demonstrated difficulty in correctly identifying light- and vigorous-intensity activities, respectively. Although youths were able to correctly interpret different components of the age-specific 3D models, children struggled to differentiate physical activity intensities represented in the models.

Conclusions: These findings support the potential use of age-specific 3D models of physical activity to enhance youths' understanding of the recommended guidelines and associated intensities.

(J Med Internet Res 2019;21(2):e11253) doi:10.2196/11253

KEYWORDS

3D printing; education; adolescent; child; comprehension; understanding; mental recall

Introduction

Regular physical activity is considered an essential part of youths' (children and adolescents) overall physiological health and psychosocial development [1-4], providing immediate and future health benefits [5-7]. Indeed, strong relationships exist between physical activity and health: Individuals who attain higher physical activity levels show a risk reduction of 30% for all-cause mortality, 20%-35% for cardiovascular diseases, 30%-40% for type 2 diabetes, and 20%-30% for cancer when compared to individuals who attain low activity [8]. Moreover, youths who frequently participate in physical activity demonstrate reduced symptoms of anxiety and depression, which subsequently leads to psychosocial benefits such as improved self-esteem and confidence [3]. Similar to physical activity, there is a dose-response relationship between increased sedentary behavior (activities in a sitting or reclining position such as watching television) and a greater risk of adverse health outcomes [9]. In the United Kingdom, youths aged 5-15 years have been reported to spend 7-8 hours per day in sedentary behavior, which accounts for 60%-65% of their day [10]. Given the pandemic rise of sedentary behavior in youth, public health sectors have produced and communicated physical activity recommendations to guide individuals toward achieving a minimum level of physical activity to reap health benefits [11]. The World Health Organization and UK Government both recommend that youths aged 5-17 years should engage in at least 60 minutes of moderate-to-vigorous physical activity every day [8,12]. Despite this, reports show that only 21% of boys and 16% of girls in the United Kingdom meet these physical activity recommendations [13,14].

Promotion of youth's physical activity relies upon our understanding of the underlying factors that influence the likelihood of achieving the desired behavior. Among the most consistently reported factors are an individual's age, sex, socioeconomic status, social and environmental support, and level of education [15-17]. However, little attention is given to individuals' knowledge regarding the recommended levels [18-21], intensities of physical activity [22-24], and, subsequently, manners in which they achieve the international physical activity guidelines. Of concern, youths most commonly cite 2 hours per week as the recommended physical activity levels [20] and demonstrate a limited ability to interpret and classify the intensities associated with daily activities [18,25-27], thereby questioning their ability to align their own activities to the recommended levels. Furthermore, youths' inability to define and understand the intensity of physical activity may, in part, explain the inconsistent reliability and validity of children's self-reported physical activity levels [28-30]. Therefore, it is important to recognize youth's lack of knowledge regarding the complexities of physical activity; content knowledge (ie, concepts) is a critical step towards youths achieving a healthy and sustainable active lifestyle that can be continued into adulthood [31]. This is particularly pertinent because adults also show a lack of knowledge of their respective physical activity targets and associated activity intensities [23]. Indeed, DiClemente et al [32] suggested that one solution to overcome youth's lack of knowledge may be the use of personalized feedback to educate an individual about a behavior and outcome. Although there is currently a paucity of literature on youths' current perceptions of physical activity intensity, it is evident that the development of personal feedback tools [33], which seek to enhance their understanding of the importance of physical activity and interpret the recommended guidelines, is warranted.

Digital mediums such as activity-tracking tools and mobile phone devices with assisted apps have allowed greater accessibility for users to visualize their personal physical activity data. Visualizations are known to enable users to understand their personal data and associations with physical activity levels, making them more comprehensible and actionable in terms of health-related aims [34]. However, on-screen visualizations are limited to visual stimulation and ignore the abundance of other senses, such as "touch," that could potentially enrich personal engagement with data [34,35]. Congruent with theories built on the notion that youth are visual and tactile learners [36-38], a number of studies support the use of tangible objects to promote youth's intellectual development [39-42]. Given that physical activity occurs in the physical world, tangible representations of physical activity that can be placed in the everyday environment have the potential to make data more available to an individual [43,44]. Indeed, Khot et al [45], investigated the use of an innovative visualization strategy involving 3D printing to create tangible physical activity data for adults, demonstrating that the visual and tactile nature of the data increased the user's awareness and reflection of their personal physical activity behaviors. Previous evidence within the educational domains suggests that tangible interfaces can play an important role in active learning among youths by increasing engagement and reflections upon a topic [46-50]. Following these developments in understanding, recent formative research on youths has demonstrated their ability to conceptualize 3D-printed objects of physical activity, with 80% of youths expressing that the models would motivate them to engage in more physical activity [24]. Moreover, youths expressed preference for 3D models, represented through abstract and graphical designs, which led to the development of two age-specific 3D-printed model prototypes. However, before introducing the age-specific 3D models into an intervention setting [51], it is important to determine their acceptability with regard to whether youths can correctly interpret the different models in terms of the amount and intensity of daily physical activity displayed [52,53]. In the absence of such formative research, researchers risk the development of 3D models and interventions that may be inappropriate or misunderstood by the target population [54]. Indeed, previous health message interventions have been limited by a lack of formative research to guide the development and delivery of messages [55]. Based on the technology design framework developed by Druin et al [56], the present study implements the role of the "tester," whereby youths are the testers of the new technology and their experiences can be observed and evaluated for impact by researchers.

The aims of this study were therefore to examine children's and adolescents' perceptions and ability to identify physical activity intensities (ie, sedentary, light, moderate, and vigorous), elicit

children's and adolescents' interpretations of the age-specific 3D model prototypes, and use the data to consolidate the design of the age-specific 3D model prototypes to inform the development of a school-based physical activity intervention.

Methods

Recruitment

Participants comprised a convenience sample taken from two primary schools and two secondary schools in South Wales, United Kingdom. In total, 12 primary school children (9 boys; mean age 7.8; SD 0.4 years) and 12 secondary school adolescents (6 boys; mean age 14.1; SD 0.3 years) participated in the study. Parents and youths provided informed written consent and assent prior to participation, respectively. All procedures were approved by the University Ethics Committee and were conducted in accordance with the Declaration of Helsinki (reference no. PG/2014/40).

Procedures

Twenty-four semistructured individual interviews were conducted with youths by the first author, either within a familiar classroom or the school library [57]. Individual interviews are a suitable method for exploratory research seeking to generate diverse and original ideas among youths [58]. Interview questions were adjusted for tone and structure to ensure age appropriateness; all interview questions and tasks were reviewed, discussed, and revised by authors SGMC, MAM, ZRK, and KAM. The interview questions (Table 1) were informed by previous formative research [24] and addressed concepts such as youths' knowledge of physical activity intensities and interpretations of the age-specific 3D models (Figures 1 and 2). Complementary to the interview questions, youths were asked to complete two interactive tasks: a physical activity and intensity-matching task and a 3D model recall and interpretation task. The first task was completed at the midpoint of the interview process and invited participants to match 20 different pictures of activities (eg, video gaming, walking, climbing stairs, and football) to the correct intensity (ie, sedentary, light, moderate, and vigorous; Table 2). Sedentary activities were based on the definition of Trost et al [59], whereas definitions for light, moderate, and vigorous activities were obtained from the youth compendium of physical activities [60,61]. After completion of the task, participants were asked to describe why they placed each activity within the specific intensity box.

Table 1. Example interview questions for children/adolescents.

Topic	Examples
Physical activity intensity	Can you tell me what you think these different levels of intensity for physical activity might be?
Physical activity intensity	What word would you use to describe the intensity of that activity (eg, climbing stairs)?
Physical activity model	What do you think the lines/bars show?
Physical activity model	Can you tell me what you think the rest of the physical activity model shows? (Prompt: how do you think this model [sun or bar chart] shows physical activity?)

Figure 1. Children's sun 3D model. PA: physical activity; MVPA: moderate-to-vigorous physical activity.

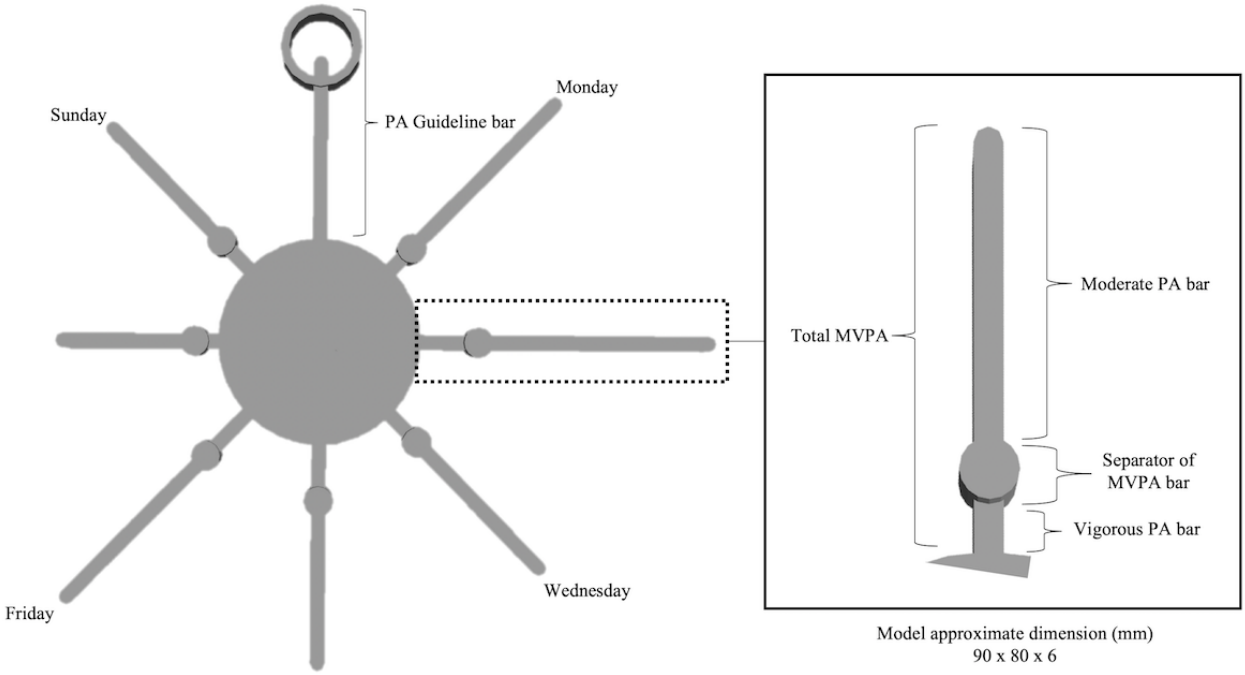


Figure 2. Adolescents' bar chart 3D model. PA: physical activity; MVPA: moderate-to-vigorous physical activity.

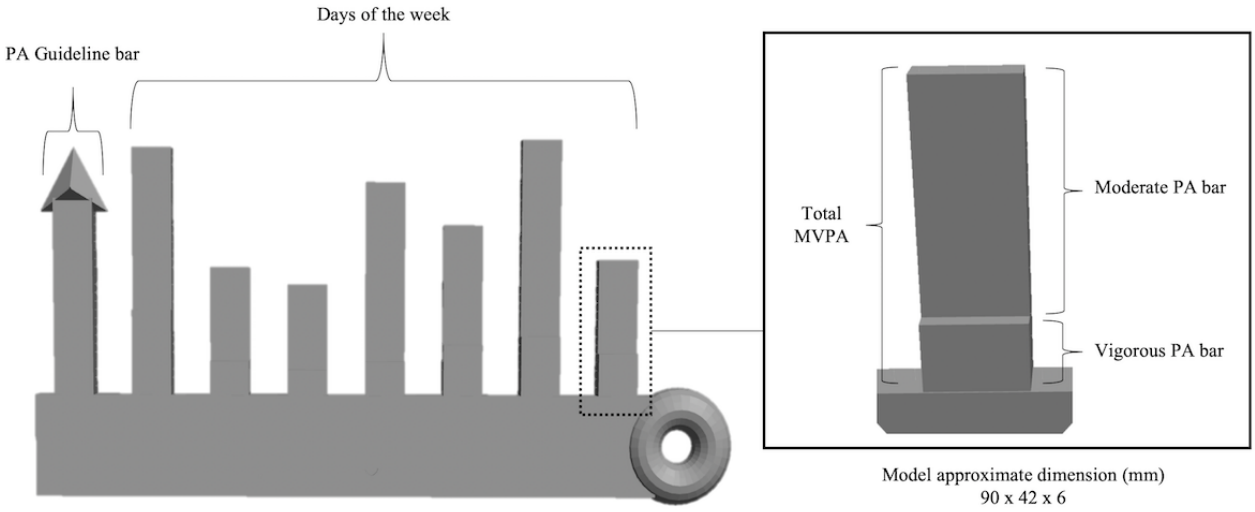


Table 2. The 20 activities and their respective intensity levels.

Intensity	Activity
Sedentary	• Eating, sitting • Reading, lying down • Mobile phone, sitting • Computer, sitting • Video games, sitting
Light (<3.0 METs[a])	• Fishing, sitting • Stretching exercises • Darts, wall • Walking, slow
Moderate (3.0-6.0 METs[a])	• Throwing, snowball • Sweeping • Mowing lawn • Climbing stairs
Vigorous (>6.0 METs[a])	• Climbing trees • Football/soccer • Tennis • Hockey, field • Running, hard effort • Swimming laps • Riding a bicycle, hard effort

[a]MET: metabolic equivalent.

The second task was completed at the end of the interview to test youths' ability to recall and interpret the different components of the age-specific 3D models. The formatively developed 3D models were designed by youths who displayed a preference for a sun (Figure 1) and adolescents (bar chart; Figure 2) by using Play-Doh as a prototype tool for the creation [24]. Both models depict example triaxial accelerometry-derived (wGT3X-BT, ActiGraph LLC, Pensacola, FL) moderate and vigorous physical activity levels achieved for each day over a week as well as a reference bar to the physical activity guidelines of 60 minutes of moderate-to-vigorous physical activity. In detail, the moderate and vigorous physical activity level achieved for each day was calculated using Evenson's child cut-points [62] on ActiLife, version 6.13.3 (ActiGraph LLC). Following this, the physical activity levels were inserted into the age-specific custom-developed 3D model code; loaded on OpenJSCAD, version 1.8.0; and subsequently 3D-printed using polylactide filament on the Ultimaker 2 Extended+ (Ultimaker, Geldermalsen, the Netherlands). All participants were asked to label a 2D diagram of the relevant model and verbally describe the model's components.

Interviews lasted for a mean of 35.8 (SD 5.3) minutes and 25.1 (SD 4.9) minutes for children and adolescents, respectively. All the interviews were digitally voice recorded (Olympus DM-520 digital voice recorder, Shinjuku, Japan), video recorded (Sony Handycam HDR-PJ540, Minato, Japan), and transcribed

verbatim. In total, 85 and 92 pages of raw transcription data (Arial font, size 12, double spaced) were produced for primary school children and secondary school adolescents, respectively. Unique identification codes were used to ensure anonymity of participants within all transcripts: B (boy) or G (girl), followed by participant number.

Data Analysis

Through the process of content analysis, transcripts were deductively analyzed through contextual 3D model themes (separator of the moderate-to-vigorous physical activity bar; physical activity guideline bar; and daily, moderate, and physical activity bars) and activity intensities (sedentary, light, moderate, and vigorous intensity). Quantitatively, we analyzed whether the classification of data was accurate (ie, activities to intensities and the different 3D model components) [63]. This mixed-methods approach allowed for a greater insight into the meaning of the data [64,65] and took into account the multiple aims of the research regarding youths' ability to identify physical activity intensities and accurately interpret the age-specific 3D models [63]. First, transcripts were thematically analyzed by the first author (SGMC) using three steps: data immersion, coding, and identifying themes [66]. The immersion of data was completed through an active way of "repeated reading" of the transcripts by searching and noting meanings and patterns within the dataset [66]. The process of coding, using a manual cut-and-paste technique, organized the data into meaningful groups that were considered pertinent to the research questions [66]. Key themes were identified by collating the relevant coded data quotes and discarding any irrelevant quotes from the analysis [66]. A frequency count of the compiled meaningful quotes was conducted to record the number of participants that noted respective points within a theme. The meaningful quotes and frequency counts were then presented diagrammatically using a pen profile approach, which is considered an appropriate method for representing diagrams of key emergent themes [67]. The last author (KAM) independently analyzed the data and discussed the outcomes with SGMC. Through the repeated process of reverse triangulation, author MAM critically cross-examined the data in reverse from the pen profiles to the transcripts until all alternative interpretations of the data were exhausted. The pen profiles were then assessed by all other authors, enabling further interpretations and adjustments before a final consensus was reached. For the activity intensity-matching task, the activities placed into certain key intensity boxes were counted (sedentary, light, moderate, and vigorous) and aligned with direct quotations (Table 3).

Statistics

An "N−1" chi-square test was conducted using SPSS Statistics 22 (IBM Corp, Chicago, IL) to determine any significant differences between boys and girls who correctly associated activities to their respective intensity, with statistical differences accepted at $P \leq .05$ [68,69].

Results

Youths' Understanding of Sedentary Behavior and Physical Activity Intensities

Youths' understanding of physical activity intensities is presented in Tables 3 and 4 (children) and Tables 5 and 6 (adolescents), with representative verbal statements for each activity reflecting youths' greatest intensity-level frequency count.

Children's Ability to Identify Sedentary Behavior and Physical Activity Intensities

Children were able to correctly align sedentary activities with the respective intensity 62% of the time, with girls demonstrating a better understanding of sedentary behavior than boys (80% vs 53%; *P*=.38). Specifically, the sedentary activities most commonly correctly identified were technology-based behaviors such as playing on a mobile phone (75%) or computer (75%) and video gaming (75%). A number of children (58%) reported that eating was a light-intensity activity: "eating's easy cause you're just like moving your arms and putting it [food] in your mouth" (PB06). Children were only able to correctly identify light-intensity activities 31% of the time, with girls showing a better understanding of light-intensity activities than boys (38% vs 28%; *P*=.75). A number of children (75%) indicated stretching as a moderate-intensity activity because "for some people stretching is really hard..." (PB06); one child stated, "when I do rugby you have to warm up and that's not hard, easy or inactive" (PB07). Furthermore, fishing was identified by five children as a sedentary behavior due to the nature of the sitting position; for example, "he's just sitting down and waiting for a fish..." (PG11). Similarly, some children struggled to define moderate-intensity activities, with only 33% of moderate activities correctly identified. Boys, as a group, fared slightly better than girls in allocating moderate-intensity activities (38% vs 25%; *P*=.66). Children perceived moderate activities such as throwing (83%), climbing stairs (75%), and sweeping (58%) as light-intensity activities. Specifically, climbing stairs was thought of as a light-intensity activity because "all you've got to do is lift a foot and put it on each step" (PB09), with sweeping noted as something that "you can relax while you're doing it " (PB02). Vigorous activities were correctly identified 68% of the time by children (boys, 73% vs girls, 57%; *P*=.58). Vigorous-intensity activities such as riding a bicycle (92%), playing hockey (92%), playing tennis (67%), swimming laps (58%), playing football (58%), running (50%), and climbing trees (50%) were all correctly classified. Children described the nature of vigorous intensity as riding a bicycle or running, which makes one "really tired" (PB09) or "a little tired" (PB01), respectively. When referring to swimming laps, children emphasized that "my swimming teacher pushes me really hard" (PB07).

Table 3. Children's ability to identify intensity of activities (n=12).

Intensity and activity item	Representative verbal statement	Frequency count according to intensity level, n			
		Sedentary	Light	Moderate	Vigorous
Sedentary					
Eating, sitting	"Eating's easy cause you're just like moving your arms and putting it [food] in your mouth" (PB06)	5	7[a]	0	0
Reading, lying down	"That one cause you're just lying there" (PB03)	5[a]	3	3	1
Mobile phone, sitting	"These [mobile phone use] are quite easy cause all you're doing is basically moving your fingers" (PB02)	9[a]	2	1	0
Computer, sitting	"Computer you just sitting down and probably typing something with mouse and this you're just going [acts out typing]..." (PB07)	9[a]	2	1	0
Video games, sitting	"They are like playing video games, this is inactive because you're not actually like moving" (PB06)	9[a]	2	1	0
Light					
Fishing, sitting	"He's just sitting down and waiting for a fish but when he winds it in he's using kind of his muscles" (PG11)	5[a]	1	4	2
Stretching exercises	"Cause when I do rugby you have to warm up and that's not hard, easy or inactive" (PB07)	0	2	9[a]	1
Darts, wall	"Throwing darts is pretty easy but not to hit the middle [of the dart board]" (PB02)	0	5[a]	4	3
Walking, slow	"Walking to school's easy, all you're doing is like moving your legs" (PB06)	1	7[a]	3	1
Moderate					
Throwing, snowball	"Throwing snowballs is quite easy because you can just throw them any way you like" (PB02)	0	10[a]	2	0
Climbing stairs	"I've put walking up steps because quite easy because all you've got to do is lift a foot and put it on each step" (PB09)	0	9[a]	3	0
Sweeping	"And sweeping because you can relax while you're doing it" (PB02)	0	7[a]	5	0
Mowing lawn	"For lawn, I've done...cause it's not easy, and it's not hard and it's not inactive so it's that one [moderate]" (PB07)	1	4	6[a]	0
Vigorous					
Climbing trees	"They're using...their tummy muscles and their arms and their legs" (PG11)	0	0	5	7[a]
Tennis	"Then tennis cause its quite active, you move a lot cos you hit and then you have to move to hit the ball again" (PB07)	0	0	4	8[a]
Swimming laps	"When I go swimming my teacher, go in the 3rd lane and my swimming teacher pushes me really hard" (PB07)	0	0	5	7[a]
Hockey, field	"Hockey's hard cause some people don't really know how to play hockey..." (PB06)	0	0	1	11[a]
Football/soccer	"...playing football's pretty hard cause...you got to get past the people who are doing skills" (PB10)	0	2	3	7[a]
Running, hard effort	"I think running because you run a long way, you get a little tired, then you get sweaty then you can't do any more" (PB01)	0	1	5	6[a]
Riding a bicycle, hard effort	"I did cycling because if you go really fast you might be really tired, and you might not want to do any more" (PB09)	0	0	1	11[a]

[a]Representative verbal statement frequency count.

Table 4. Children's ability to correctly match activities to intensity (n=12).

Correct classification	Total (%)	Boys (%)	Girls (%)
Sedentary	62	53	80
Light	31	28	38
Moderate	33	38	25
Vigorous	68	71	57

Adolescents' Ability to Identify Sedentary Behavior and Physical Activity Intensities

Adolescents correctly identified sedentary-based activities 87% of the time, with boys demonstrating a better understanding than girls (90% vs 83%; *P*=.73). Sedentary technology-based activities such as playing on a mobile phone (100%) or computer (92%) and video gaming (75%) were all correctly perceived as sedentary behaviors:

> *They're just on their electronics, playing games or watching something...they don't really have to put effort into that and they're not moving around or doing anything.* [SB01]

Light-intensity activities were correctly identified 71% of the time, with girls displaying a better understanding than boys (75% vs 67%; *P*=.77). Light-intensity activities including walking (83%), fishing (67%), playing darts (67%), and stretching (67%) were all consistently identified as light-intensity activities. Adolescents correctly identified moderate-intensity activities only 10% of the time (girls, 13% vs boys, 8%; *P*=.07). All adolescents reported that the activity of throwing (100%) was a light-intensity activity. Other moderate activities such as mowing the lawn (75%), climbing stairs (75%), and sweeping (67%) were also classified as light-intensity activities; one adolescent described moderate activities as "everyday things like mowing the lawn" (SG09). Adolescents were only able to appropriately identify vigorous-intensity activities 46% of the time, with girls demonstrating a greater ability to recognize vigorous-intensity activities than boys (62% vs 24%; *P*=.20). Adolescents correctly categorized individual fitness activities such as cycling (75%), running (67%), and swimming (50%) as vigorous-intensity activities. In contrast, organized sport activities such as football (75%), tennis (67%), and hockey (58%) were often identified as moderate-intensity activities, although they regarded football and tennis as "...quite a physical sport" (SB03) or involving "…strengths" (SB04), respectively.

Youths' Understanding of the Age-Specific 3D Models

Children's and adolescents' interpretations of the age-specific 3D models are presented in two separate pen profiles (Figures 3 and 4, respectively).

Children's Understanding and Ability to Interpret the Sun 3D Model

In total, six higher-order themes were structured around the 3D model's components: "Physical Activity Guideline Bar," "Daily Physical Activity Bars," "Moderate Physical Activity Bar," "Vigorous Physical Activity Bar," and "Separator of MVPA Bar" (Figure 3). A number of children (75%) were able to interpret the physical activity guideline bar on the 3D model as "the 60-minute time bar" (PG10). All children correctly identified that the 3D model represented a week of physical activity "Monday they did a lot [of physical activity], on Tuesday they did a tiny bit, on Wednesday they did a tiny bit less..." (PG05). The data revealed that 58% of children had some difficulty interpreting the moderate physical activity bar on the 3D model, with children describing the bar as "…the easy activity to be doing because you do easy more than hard..." (PB01). Only 42% of children were able to correctly interpret the moderate physical activity bar as "medium activity..." (PG05). Ten children (83%) correctly interpreted the vigorous physical activity bar as "how much you've done of the hard level [of physical activity]" (PG11), with only two children incorrectly interpreting the bar as the time at which the physical activity was undertaken: "the morning [of physical activity] and that might be the afternoon [of physical activity]." The circle separator along the sun's rays splitting the moderate and vigorous physical activity bars was correctly interpreted by 67% of children as "the blob splits the line up, so you know how many of the hard [physical activity] and how many of the medium [physical activities]" (PG11). Only two children expressed that they did not understand the meaning of the moderate-to-vigorous separator along the ray.

Adolescents' Understanding and Ability to Interpret the "Bar Chart" 3D Model

Four higher-order themes were identified around the 3D model components: "Physical Activity Guideline Bar," "Daily Physical Activity Bars," "Moderate Physical Activity Bar," and "Vigorous Physical Activity Bar" (Figure 4). The physical activity guideline bar was correctly interpreted by 83% of adolescents as "that's the amount [of physical activity] you need to be doing or more...sixty minutes a day" (SP12). Only two participants were unable to identify the meaning of the target bar. All adolescents had a good understanding of representation of the physical activity data as a week, and 42% of the adolescents were able to interpret the data without any previous explanation or guidance from the facilitator. The moderate-intensity physical activity bar was correctly reported by 75% of adolescents as "...the moderate activity that you [themselves] were doing" (SG01), with only three participants incorrectly defining it as "how much sport [they] have done" (SG6). All adolescents demonstrated a good understanding of the vigorous-intensity physical activity bar, stating "...this means how much hard activity [they] are doing..." (SB03).

Table 5. Adolescents' ability to identify intensity of activities (n=12).

Intensity and activity item	Representative verbal statement	Frequency count according to intensity level, n			
		Sedentary	Light	Moderate	Vigorous
Sedentary					
Eating, sitting	"Eating, maybe just a little bit of movement when you're like bringing it [the food] up to your mouth and then when you're chewing" (SB02)	10[a]	2	0	0
Reading, lying down	"Reading a book all you're doing is just flipping a page with almost nothing movement..." (SB02)	10[a]	1	1	0
Mobile phone, sitting	"They're just on their electronics...they don't really have to put effort into that and they're not moving around or doing anything" (SG01)	12[a]	0	0	0
Computer, sitting	"Yeah well obviously computer games...you're not doing much except moving your fingers maybe" (SB02)	11[a]	1	0	0
Video games, sitting	"Playing games...like some things that don't require that much movement" (SG10)	9[a]	2	0	0
Light					
Fishing, sitting	"Fishing you're just waiting in a boat and when a fish comes you have to reel it..." (SB02)	3	8[a]	2	0
Stretching exercises	"It's [stretching] not like big movement like they're not really doing much" (SB11)	2	8[a]	2	0
Darts, wall	"Darts, all you're doing is just throwing a small dart at a small target" (SB02)	3	8[a]	1	0
Walking, slow	"...walking to school you do need to walk obviously but it's not very hard..." (SB02)	3[a]	1	0	0
Moderate					
Throwing, snowball	"Throwing a snowball not much at all, all you have to do is just craft this little ball of precipitation and throw it at someone else" (SB02)	0	12[a]	0	0
Climbing stairs	"Like walking up the stairs, it's sort of easy... you can get a bit out of breath" (SB04)	0	9[a]	3	0
Sweeping	"They're just like doing something simple, like their daily life" (SG05)	4	8[a]	0	0
Mowing lawn	"Light is mostly just...everyday things like mowing the lawn" (SG09)	1	9[a]	2	0
Vigorous					
Climbing trees	"Climbing a tree cause it does take a lot of effort to climb a tree" (SG01)	0	5	6[a]	1
Tennis	"Just some like basic sports...people would think they're fairly easy...running, football and tennis" (SG10)	0	0	8[a]	4
Swimming laps	"Swimming...you have to be able to do the right streamlined technique to be able to glide through the water and then...you need to be able to breathe..." (SB02)	0	2	4	6[a]
Hockey, field	"A girl playing hockey you need to run around the pitch many times and it might get a bit tiring" (SB02)	0	0	7[a]	5
Football/soccer	"I put quite a few in medium because like football is quite a physical sport" (SB03)	0	0	9[a]	3
Running, hard effort	"These are probably the ones like make you push yourself" (SB11)	0	2	2	8[a]
Riding a bicycle, hard effort	"Like cycling when you're going up hills and stuff, it depends like how strong you are..." (SB04)	0	0	3	9[a]

[a]Representative verbal statement frequency count.

Table 6. Adolescents' ability to correctly match activities to intensity (n=12).

Correct classification	Total (%)	Boys (%)	Girls (%)
Sedentary	87	90	83
Light	71	67	75
Moderate	10	8	13
Vigorous	43	24	62

Figure 3. Children's interpretation of the sun 3D model. P: primary; B: boy; G: girl; PA: physical activity; N: frequency counts; MVPA: moderate-to-vigorous physical activity.

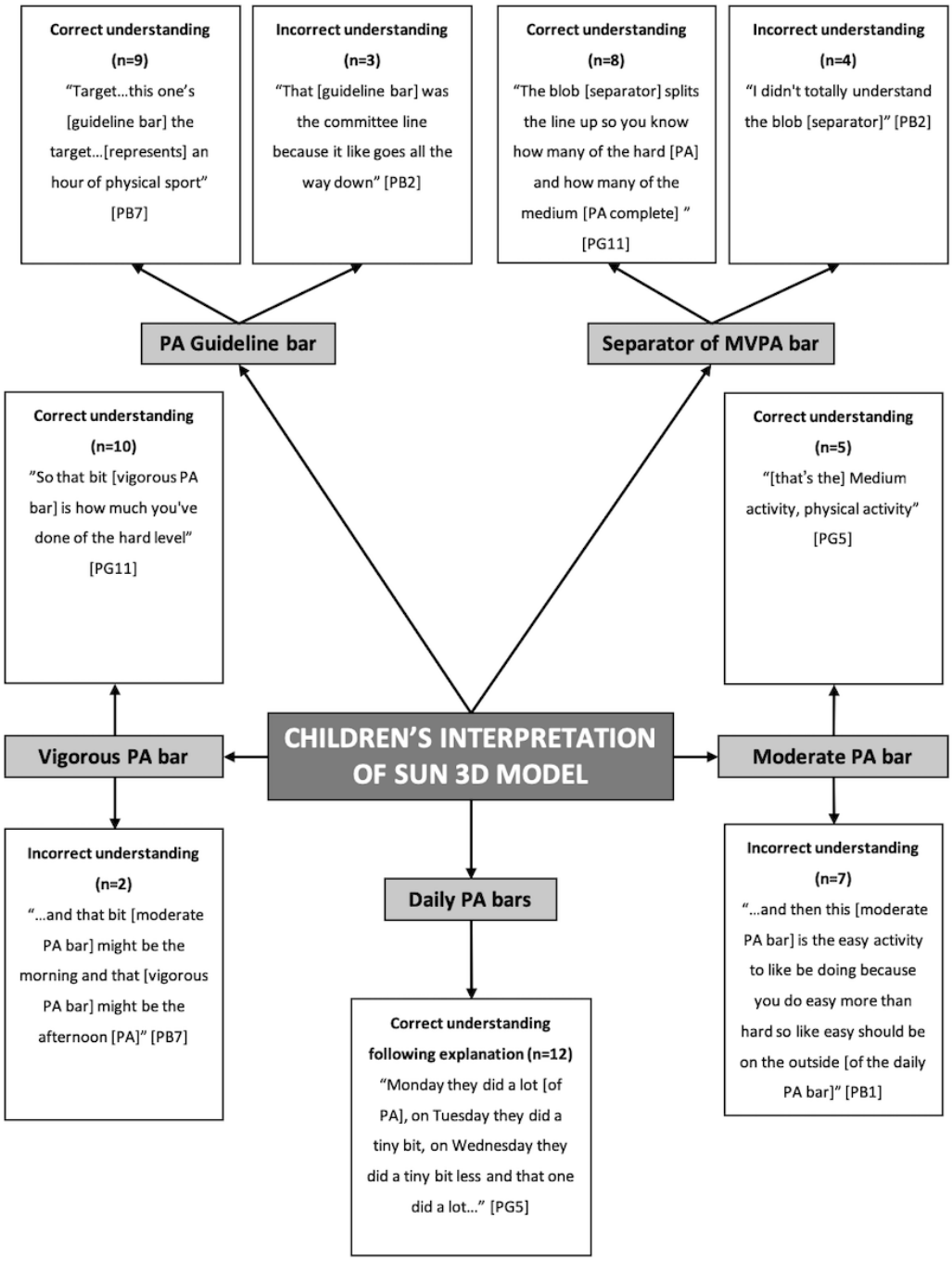
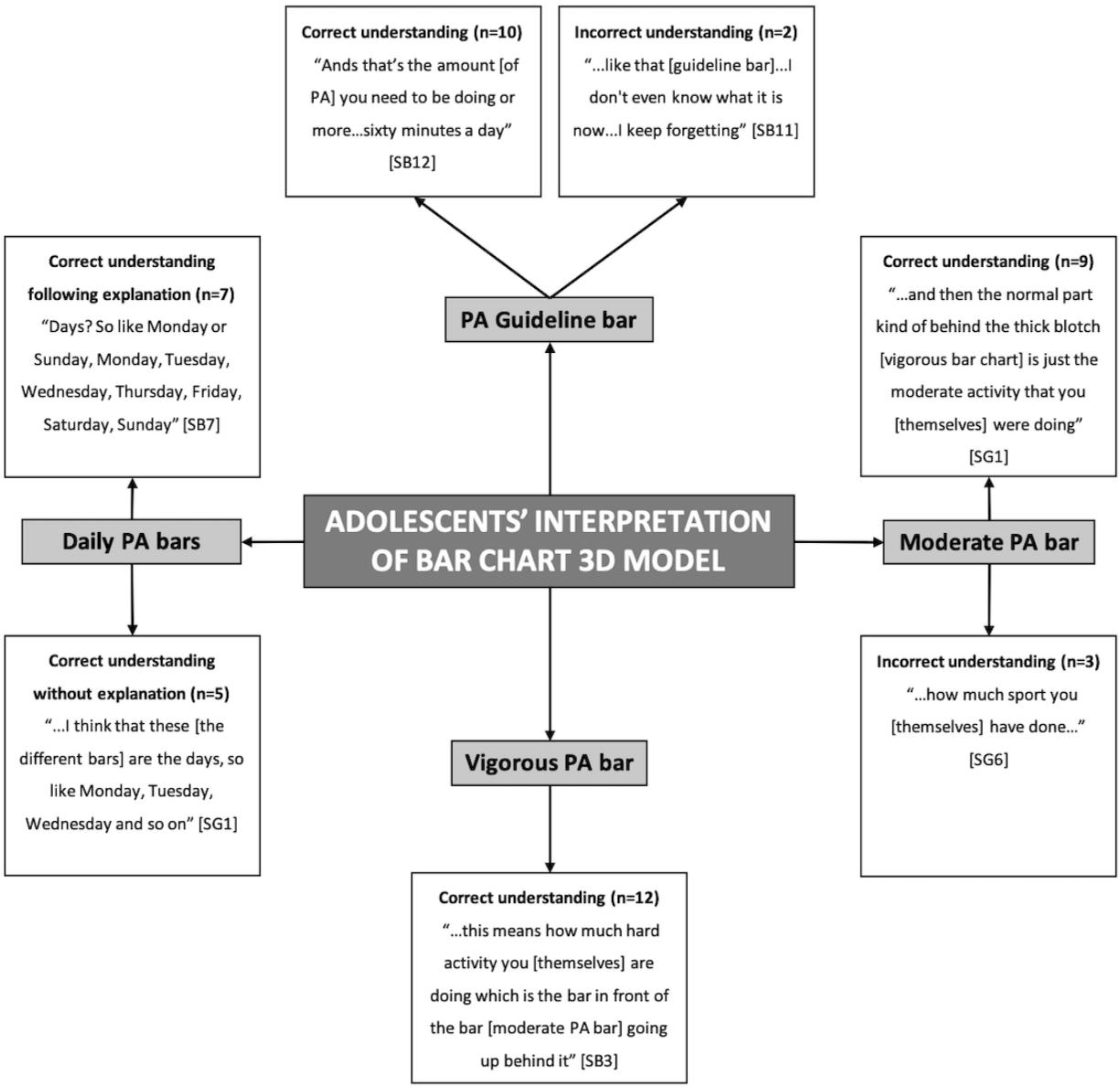


Figure 4. Adolescents' interpretations of the bar chart 3D model. S: secondary; B: boy; G: girl; PA: physical activity; N: frequency counts.



Discussion

Principal Findings

The aims of this study were to ascertain youths' understanding of the age-specific 3D model designs and to examine youths' perceptions and ability to identify activities according to their respective intensity. The study findings suggest that youths demonstrate misconceptions in defining different activity intensities. However, youths' ability to interpret the age-specific 3D models supports the use of these formatively designed tangible representations of physical activity within an intervention to aid youths' understanding and awareness of the recommended 60 minutes of moderate-to-vigorous physical activity [24].

Overview

To date, little research has explored how youths understand the meaning of the term *physical activity* [27,59,70,71]. It has previously been suggested that the terminology developed by adults used to describe physical activity is too complicated for youths, due to developmental and vocabulary differences [27]. As highlighted by Pearce et al [27], understanding how children express physical activity is a logical first step for improving overall knowledge and the development of innovative methods for enhancing physical activity. In the present study, the intensity-matching task revealed that adolescents have a greater ability to identify sedentary behaviors and light-intensity activities, whereas children showed they could more accurately identify the two extremes of intensity (ie, sedentary behaviors and vigorous-intensity activities). It could be expected that as a result of children's sporadic and explosive patterns of activity [3,72-75], moving from one extreme intensity to another, could explain, in part, the present findings demonstrating children's limited ability to identify the intermediary light- and moderate-intensity activities. Furthermore, the present study showed that only 25% (n=3) of children thought that working on the computer was a physical activity, and none of the children

characterized "sweeping" as a sedentary behavior as compared to the previous findings of 38% and 30%, respectively [59]. Although Trost et al [59] encompassed a larger sample of children aged 9-10 years, such discrepancies may be, at least in part, due to the reduced sample size and wider age range in our study. Conversely, it could be argued that the timeframe in which the research was implemented could play an influencing role on youths' understandings of physical activity. For example, since the mid-2000s, the number of campaigns with mass media components have led to an increased level of exposure to the importance of physical activity behaviors within youths, with evidence supporting this exposure-response relationship [76-78].

The majority of children tended to overestimate light-intensity activities such as stretching exercises, playing darts, and fishing. In some cases, children would associate stretching exercises with other more demanding activities such as warm-up before "...rugby..." (PB07). This type of category contamination was a recurring theme among children, with other activities such as throwing (light) and climbing trees (vigorous) being associated with "...running around" (PB01). In most cases, this category contamination led to an increase in intensity of the dominating activity (ie, going from light to moderate intensity). Furthermore, light-intensity activities such as darts and fishing were often inaccurately identified because of the perceived skill or competence required to complete the activity. Specifically, playing darts was considered a vigorous-intensity activity, as it required a certain skill to "...get [the dart] in the middle [of the dart board]" (PG05), and fishing was associated with moderate intensity because it is "...quite hard to catch fish" (PB02). Skill level was identified as a common characteristic for other activities including football, hockey, swimming, riding a bicycle, and climbing trees, with one child stating reading as a vigorous intensity because "...you have to learn how to read words" (PB06). Consistent with previous findings [27], this study emphasizes that skill in an activity, or physical competence, influenced children's perceptions of the intensity level. It could be speculated that children's perspectives of these skill- or physical competence-associated activities are likely to undergo change and refinement as a consequence of time with personal experience and maturation [70]. However, this perspective appears unlikely, as evidence suggests that adults also lack relevant knowledge in terms of determining intensities of physical activity [23]. This demonstrated inability to define intensities further highlights the importance of educating youths about different intensities of physical activity, so that as they age, their understanding of physical activity is more likely to reflect the actual intensity associated with the recommended guidelines.

It is important that youths understand the type of physical activities that form moderate- and vigorous-intensity levels to increase their chances of engaging with these type of activities and gaining the associated metabolic health benefits [79-81]. In the present study, youths demonstrated a limited ability to correctly identify moderate-intensity activities, although the degree of this inaccuracy was much greater in adolescents. It could be postulated that youths' inability to identify moderate-intensity activities could be aligned with their limited capacity to describe how a physical activity could be performed at different intensities or effort levels [60]. However, youths underestimated the intensity levels of moderate activities related to either household chores, such as sweeping and mowing the lawn, or the daily activity of climbing stairs. Adolescents described such moderate activities as "...everyday things like mowing the lawn" (SG09) and "...like it's easy" (SB04), and children suggested that when climbing stairs, "...all you've got to do is lift a foot and put it on each step" (PB09). These findings support those of Trost et al [59], as household chores and climbing stairs are not considered important contributory sources of physical activity, with the present study further highlighting that this underestimation increased with age. Nonetheless, it is perhaps pertinent to consider the applicability of some activities such as household chores, as a lack of familiarity may have led to exaggerated inaccuracies in the intensity of these activities [82]. For adolescents, the more commonly performed individual sports (swimming, running, and cycling) were correctly identified as vigorous intensity, with team sports such as football, hockey, and tennis perceived to be moderate intensity. Indeed, evidence suggests that the more the activity is considered as play or fun, the less likely youths are aware of the intensity [27]. Although there is limited evidence of this within the present findings, it could be speculated that the greater level of social interaction during team sports [83] and the perceived conception of these team activities for play or fun could function as a moderator to youths' ability to correctly assess the respective intensity [27].The present findings highlight the need to further understand how context (ie, social settings) mediates youths' ability to interpret intensities between team sports and the more individual-pursuit sports. Additionally, research is warranted to investigate the potential differences between nonathletic youths' and sports-orientated youths' understanding and ability to conceptualize intensities.

Inconsistencies in youths' ability to correctly identify and understand different activity intensities observed in this study are important, especially given that self-report physical activity questionnaires rely on youths' ability to correctly interpret activities in accordance with the intensity level [27]. The reliability and validity of data derived from measuring youth's physical activity using self-report questionnaires is problematic [28-30]. From the present findings, it could be postulated that the inconsistent reliability and validity of physical activity questionnaires are, at least in part, due to youths' misinterpretations and lack of understanding of intensity, supporting the findings of LeBlanc and Janssen [84]. Indeed, the findings support the idea that youths are not accustomed to relating their physical activity by intensity as a result of limited understanding, which makes it challenging to conduct interventions aimed at changing intensities of physical activity. Additionally, although direct comparisons between sexes failed to demonstrate any significant difference in the ability to align activities with their respective activity intensities, the present findings suggest that girls, irrespective of age, outperformed boys. For example, adolescent girls outperformed their counterparts in correctly identifying light-, moderate-, and vigorous-intensity activities. Interestingly, girls in the younger age group outperformed boys in correctly identifying sedentary and light-intensity activities, although the imbalance in the number of girls (n=3) to boys (n=9) may limit such comparisons.

Indeed, these preliminary findings could be explained by differences in cognitive development, as girls have better verbal and written language skills [85-88] and acquire vocabulary faster [89] than boys until adulthood. As a consequence, it could be postulated that the higher level of vocabulary among girls may give them an advantage in aligning activities with intensities. This is especially pertinent because previous research has shown that girls provide more reliable and valid recollections in physical activity questionnaires than boys [90]. More research is warranted to determine whether these verbal and written advantages among girls indeed play a significant role in the understanding of physical intensities, how this may impact self-report questionnaires, and how to best account for these sex differences.

One method that has the potential to develop youths' comprehension of physical activity levels and associated activity intensities is the use of personalized tangible interfaces (ie, 3D models) to aid learning [91]. The present findings support this notion, with youths demonstrating a good ability to interpret and understand the age-specific 3D models, which is an important step towards enabling a cognitive experience whereby they can start to learn about their physical activity habits [92]. Indeed, previous research has suggested that physical materials can promote playful learning in youth and might offer a more natural interaction than other types of learning interfaces (ie, digital) [46-49]. More specifically, tangible objects can enable collaborative interactions with significant others [93,94], which coincides with an individual achieving social interaction more readily than that from on-screen digital displays [95]. Moreover, evidence suggests that physical activity with the social support of significant others can significantly increase youth's motivation for physical activity [96] as well as their enjoyment [97,98], intensity [99], and overall engagement in and out of school [100], which holds promise for the 3D models. Equally important, a large proportion of youths (79%) could correctly identify and describe the current physical activity guidelines projected on the 3D models. Youths' ability to understand the physical activity guideline as a tangible representation will offer a more haptic and proprioceptive experience than visual representations alone [101], which is especially pertinent because youths are regarded as visual and tactile learners [36]. It is anticipated that the 3D models will act as a concept map, whereby youths can make connections and relationships and understand that the concepts about physical activity are not just factual, but rather ideas to increase comprehension and expand vocabulary [102]. Adolescents demonstrated a greater understanding of the age-specific 3D models and the different representations of physical intensities as compared to children, which could be explained by the Piagetian developmental theory [103]. Evidence suggests that the exploratory nature of learning through tangible interfaces such as 3D models of physical activity may offer a more supportive solution to enhancing children's understanding in identifying patterns (ie, between activities and intensities) and new concepts about physical activity than previous digital methods [91]. In this light, our findings support the use of the age-specific 3D models within a school-based intervention, whereby youths receive a personalized 3D model of their objectively assessed physical activity levels to not only enhance comprehension and understanding of the recommended guidelines and associated intensities, but also use as a unique motivational strategy to increase their physical activity.

Limitations

Although data saturation was reached and lends further credibility to the findings, this study is limited by the relatively small sample size, age range, and geographical area of data collection, which may underrepresent other socioeconomic groups and ethnic minorities. Furthermore, among children, sex difference comparisons regarding the understanding of intensities may be limited due to the small number of girls who participated in the study. Therefore, our findings on youths' understanding of the age-specific 3D models and the demonstrated ability to match activities to respective physical intensities should not be generalized but considered a stimulus for future investigation.

Conclusion

This study shows that both children and adolescents have misconceptions when identifying corresponding activity intensities. Specifically, children showed recurring-intensity classification errors such as category contamination and perceived skill or competence of an activity leading to misperceptions of intensity, with both age groups severely underestimating moderate-intensity activities. However, youths demonstrated a good ability to interpret and describe the age-specific 3D model representations of physical activity, intensity, and the recommended guideline. Therefore, this study highlights the potential utility of these age-specific 3D-printed models within an intervention to act as an educational tool to enhance youths' understanding and awareness of the recommended physical activity guidelines and associated intensities.

Acknowledgments

The authors thank all the teachers and parents for their help and the children who participated in the study.

Conflicts of Interest

None declared.

References

1. Sothern MS, Loftin M, Suskind RM, Udall JN, Blecker U. The health benefits of physical activity in children and adolescents: implications for chronic disease prevention. Eur J Pediatr 1999 Apr;158(4):271-274. [Medline: 10206121]

2. Nieman P. Psychosocial aspects of physical activity. Paediatr Child Health 2002 May;7(5):309-312 [FREE Full text] [Medline: 20046307]
3. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40. [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
4. Eime RM, Young JA, Harvey JT, Charity MJ, Payne WR. A systematic review of the psychological and social benefits of participation in sport for children and adolescents: informing development of a conceptual model of health through sport. Int J Behav Nutr Phys Act 2013 Aug 15;10:98 [FREE Full text] [doi: 10.1186/1479-5868-10-98] [Medline: 23945179]
5. Shiri R, Solovieva S, Husgafvel-Pursiainen K, Telama R, Yang X, Viikari J, et al. The role of obesity and physical activity in non-specific and radiating low back pain: the Young Finns study. Semin Arthritis Rheum 2013 Jun;42(6):640-650. [doi: 10.1016/j.semarthrit.2012.09.002] [Medline: 23270761]
6. Telama R, Yang X, Leskinen E, Kankaanpää A, Hirvensalo M, Tammelin T, et al. Tracking of physical activity from early childhood through youth into adulthood. Med Sci Sports Exerc 2014;46(5):955-962. [doi: 10.1249/MSS.0000000000000181] [Medline: 24121247]
7. Wolin K, Yan Y, Colditz GA, Lee IM. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer 2009 Feb 24;100(4):611-616 [FREE Full text] [doi: 10.1038/sj.bjc.6604917] [Medline: 19209175]
8. Department of Health, Physical Activity, Health Improvement and Protection. gov.uk. 2011. Start Active, Stay Active: A report on physical activity from the four home countries' Chief Medical Officers URL: https://www.gov.uk/government/publications/start-active-stay-active-a-report-on-physical-activity-from-the-four-home-countries-chief-medical-officers [accessed 2019-01-21] [WebCite Cache ID 75ZineZ63]
9. Tremblay M, Colley RC, Saunders TJ, Healy GN, Owen N. Physiological and health implications of a sedentary lifestyle. Appl Physiol Nutr Metab 2010 Dec;35(6):725-740. [doi: 10.1139/H10-079] [Medline: 21164543]
10. Biddle S, Cavill N, Ekelund U, Gorely T, Griffiths M, Jago R, et al. Sedentary behaviour and obesity: review of the current scientific evidence. The Sedentary Behaviour and Obesity Expert Working Group 2010:1-126 [FREE Full text]
11. Blair S, LaMonte MJ, Nichaman MZ. The evolution of physical activity recommendations: how much is enough? Am J Clin Nutr 2004 May;79(5):913S-920S. [doi: 10.1093/ajcn/79.5.913S] [Medline: 15113739]
12. World Health Organization. 2011. Global Recommendations on Physical Activity for Health: 5-17 years old URL: https://www.who.int/dietphysicalactivity/factsheet_recommendations/en/ [accessed 2019-01-20] [WebCite Cache ID 75ZNydkbH]
13. Hills A, Andersen LB, Byrne NM. Physical activity and obesity in children. Br J Sports Med 2011 Sep;45(11):866-870. [doi: 10.1136/bjsports-2011-090199] [Medline: 21836171]
14. Townsend N, Wickramasinghe K, Williams J, Bhatnagar P, Rayner M. British Heart Foundation.: Nuffield Department of Population Health, University of Oxford British Heart Foundation: London; 2015. Physical Activity Statistics 2015 URL: https://www.bhf.org.uk/informationsupport/publications/statistics/physical-activity-statistics-2015 [accessed 2019-01-25]
15. Bauman A, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW, Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? Lancet 2012 Jul 21;380(9838):258-271. [doi: 10.1016/S0140-6736(12)60735-1] [Medline: 22818938]
16. Carver A, Timperio A, Crawford D. Playing it safe: the influence of neighbourhood safety on children's physical activity. A review. Health Place 2008 Jun;14(2):217-227. [doi: 10.1016/j.healthplace.2007.06.004] [Medline: 17662638]
17. Trost S, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in physical activity: review and update. Med Sci Sports Exerc 2002 Dec;34(12):1996-2001. [doi: 10.1249/01.MSS.0000038974.76900.92] [Medline: 12471307]
18. Snethen J, Broome ME. Weight, exercise, and health: children's perceptions. Clin Nurs Res 2007 May;16(2):138-152. [doi: 10.1177/1054773806298508] [Medline: 17452432]
19. Sleap M, Wormald H. Perceptions of Physical Activity Among Young Women aged 16 and 17 Years. European Journal of Physical Education 2006 Aug 10;6(1):26-37. [doi: 10.1080/1740898010060104]
20. Harris J, Cale L, Duncombe R, Musson H. Young people's knowledge and understanding of health, fitness and physical activity: issues, divides and dilemmas. Sport, Education and Society 2016 Sep 03;23(5):407-420. [doi: 10.1080/13573322.2016.1228047]
21. Roth M, Stamatakis E. Linking young people's knowledge of public health guidelines to physical activity levels in England. Pediatr Exerc Sci 2010 Aug;22(3):467-476. [Medline: 20814041]
22. Grewal TK. Electronic Thesis and Dissertation Repository, The University of Western Ontario. 2013. Awareness of Physical Activity Levels and Sedentary Behaviour: An Assessment of Awareness of Physical Activity Levels and Sedentary Behaviour Among Parents and Children URL: https://ir.lib.uwo.ca/etd/1155/ [accessed 2019-01-25]
23. Knox E, Esliger DW, Biddle SJH, Sherar LB. Lack of knowledge of physical activity guidelines: can physical activity promotion campaigns do better? BMJ Open 2013 Dec 05;3(12):e003633 [FREE Full text] [doi: 10.1136/bmjopen-2013-003633] [Medline: 24319271]
24. Crossley SGM, McNarry MA, Hudson J, Eslambolchilar P, Knowles ZR, Mackintosh KA. Perceptions of Visualizing Physical Activity as a 3D-Printed Object: Formative Study. J Med Internet Res 2019 Jan 30;21(1):e12064 [FREE Full text] [doi: 10.2196/12064] [Medline: 30698532]

25. Cowden R, Plowman SA. The Self-Regulation and Perception of Exercise Intensity in Children in a Field Setting. Pediatric Exercise Science 1999 Feb;11(1):32-43. [doi: 10.1123/pes.11.1.32]
26. Prochaska J, Sallis JF, Long B. A physical activity screening measure for use with adolescents in primary care. Arch Pediatr Adolesc Med 2001 May;155(5):554-559. [Medline: 11343497]
27. Pearce PF, Harrell JS, McMurray RG. Middle-School Children's Understanding of Physical Activity: "If You're Moving, You're Doing Physical Activity". Journal of Pediatric Nursing 2008;23(3):169-182. [doi: 10.1016/j.pedn.2007.09.003]
28. Aggio D, Fairclough S, Knowles Z, Graves L. Validity and reliability of a modified english version of the physical activity questionnaire for adolescents. Arch Public Health 2016;74:3 [FREE Full text] [doi: 10.1186/s13690-016-0115-2] [Medline: 26807217]
29. Chinapaw M, Mokkink LB, van Poppel MNM, van Mechelen W, Terwee CB. Physical activity questionnaires for youth: a systematic review of measurement properties. Sports Med 2010 Jul 01;40(7):539-563. [doi: 10.2165/11530770-000000000-00000] [Medline: 20545380]
30. Martínez-Gómez D, Martínez-de-Haro V, Pozo T, Welk GJ, Villagra A, Calle ME, et al. [Reliability and validity of the PAQ-A questionnaire to assess physical activity in Spanish adolescents]. Rev Esp Salud Publica 2009;83(3):427-439 [FREE Full text] [Medline: 19701574]
31. Brusseau T, Kuilinna PH, Cothran DJ. Health and physical activity content knowledge of Pima children. Physical Educator 2011;68(2):66.
32. DiClemente C, Marinilli AS, Singh M, Bellino LE. The role of feedback in the process of health behavior change. Am J Health Behav 2001;25(3):217-227. [Medline: 11322620]
33. Kremers S, Dijkman MA, de Meij JS, Jurg ME, Brug J. Awareness and habit: important factors in physical activity in children. Health Education 2008 Oct 17;108(6):475-488. [doi: 10.1108/09654280810910881]
34. Khot RA, Hjorth L, Mueller FF. Understanding material representations of physical activity. CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2016:3835-3844. [doi: 10.1145/2556288.2557144]
35. Ullmer B, Ishii H. Emerging frameworks for tangible user interfaces. IBM Syst. J 2000;39(3.4):915-931. [doi: 10.1147/sj.393.0915]
36. Dunn RS, Dunn KJ. Learning Styles/Teaching Styles: Should They ... Can They ... Be Matched? Educational Leadership 1979:238-244.
37. Fleming ND, Mills C. Not Another Inventory, Rather a Catalyst for Reflection. To Improve the Academy 1992;11(1):137-155. [doi: 10.1002/j.2334-4822.1992.tb00213.x]
38. Cole M, Wertsch JV. Beyond the Individual-Social Antinomy in Discussions of Piaget and Vygotsky. Human Development 1996;39(5):250-256. [doi: 10.1159/000278475]
39. Price S, Rogers Y, Scaife M, Stanton D, Neale H. Using 'tangibles' to promote novel forms of playful learning. Interacting with Computers 2003 Apr;15(2):169-185. [doi: 10.1016/S0953-5438(03)00006-7]
40. Marshall P. Do tangible interfaces enhance learning?. URL: https://dl.acm.org/citation.cfm?doid=1226969.1227004 [accessed 2019-01-25]
41. Rogers Y, Scaife M, Gabrielli S, Smith H, Harris E. A conceptual framework for mixed reality environments: designing novel learning activities for young children. URL: https://dl.acm.org/citation.cfm?id=641632 [accessed 2019-01-25]
42. Bara F, Gentaz E, Colé P, Sprenger-Charolles L. The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. Cognitive Development 2004 Jul;19(3):433-449. [doi: 10.1016/j.cogdev.2004.05.003]
43. Khot R, Mueller FF. Sweat-atoms: turning physical exercise into physical objects. 2013 Presented at: CHI '13 Extended Abstracts on Human Factors in Computing Systems; 2013; Paris, France p. 3075-3078. [doi: 10.1145/2468356.2479614]
44. Sauvé K, Houben S, Marquardt S, Bakker S, Hengeveld B, Gallacher S, et al. LOOP: A physical artifact to facilitate seamless interaction with personal data in everyday life. 2017 Presented at: DIS '17 Companion Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems; 2017; Edinburgh, United Kingdom p. A. [doi: 10.1145/3064857.3079175]
45. Rohit AK, Floyd MF, Larissa H. SweatAtoms: materializing physical activity. 2013 Presented at: in Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death; September 30 - October 01, 2013; Melbourne, Australia.
46. Dourish P. Where the Action Is: The Foundations of Embodied Interaction. Cambridge, Massachusetts, United States: The MIT Press; 2001.
47. Jacob RJK, Ishii H, Pangaro G, Patten J. A tangible interface for organizing information using a grid. 2002 Presented at: in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. . ACM; April 20 - 25, 2002; Minneapolis, Minnesota, USA p. 339-346. [doi: 10.1145/503376.503437]
48. Triona L, Klahr D. CiteSeerX. 2005. Point and click or build by hand: comparing the effects of physical vs. virtual materials on middle school students' ability to optimize an engineering design URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.133.2590&rank=2 [accessed 2019-01-25]
49. Price S, Rogers Y, Scaife M, Stanton D, Neale H. Using 'tangibles' to promote novel forms of playful learning. Interacting with Computers 2003 Apr;15(2):169-185. [doi: 10.1016/S0953-5438(03)00006-7]

50. Rogers Y, Scaife M, Gabrielli S, Smith H, Harris E. A Conceptual Framework for Mixed Reality Environments: Designing Novel Learning Activities for Young Children. Presence: Teleoperators and Virtual Environments 2002 Dec;11(6):677-686. [doi: 10.1162/105474602321050776]
51. Crossley SGM, McNarry MA, Eslambolchilar P, Knowles ZR, Mackintosh KA. The Tangibility of Personalised 3D Printed Feedback May Enhance Youths Physical Activity Awareness, Goal-Setting and Motivation. J Med Internet Res (forthcoming) 2019:1-18 (forthcoming). [doi: 10.2196/preprints.12067]
52. Gittelsohn J, Steckler A, Johnson CC, Pratt C, Grieser M, Pickrel J, et al. Formative research in school and community-based health programs and studies:. Health Educ Behav 2006 Feb;33(1):25-39 [FREE Full text] [doi: 10.1177/1090198105282412] [Medline: 16397157]
53. Meade C, Calvo A, Cuthbertson D. Impact of culturally, linguistically, and literacy relevant cancer information among Hispanic farmworker women. J Cancer Educ 2002;17(1):50-54. [doi: 10.1080/08858190209528793] [Medline: 12000108]
54. Bopp M, Fallon E. Community-based interventions to promote increased physical activity: a primer. Appl Health Econ Health Policy 2008;6(4):173-187. [doi: 10.2165/00148365-200806040-00001] [Medline: 19382818]
55. Martinez J, Latimer AE, Rivers SE, Salovey P. Formative research for a community-based message-framing intervention. Am J Health Behav 2012 Mar;36(3):335-347 [FREE Full text] [doi: 10.5993/AJHB.36.3.5] [Medline: 22370435]
56. Druin A. The role of children in the design of new technology. Behaviour & Information Technology 2002 Jan;21(1):1-25. [doi: 10.1080/01449290110108659]
57. Kennedy C, Kools S, Krueger R. Methodological considerations in children's focus groups. Nurs Res 2001;50(3):184-187. [Medline: 11393641]
58. Heary C, Hennessy E. Focus Groups Versus Individual Interviews with Children: A Comparison of Data. The Irish Journal of Psychology 2006 Jan;27(1-2):58-68. [doi: 10.1080/03033910.2006.10446228]
59. Trost SG, Morgan AM, Saunders RP, Felton GA, Ward DS. Children 's Understanding of the Concept of Physical Activity. Pediatric Exercise Science 2000;12(3):293-299.
60. Ridley K, Ainsworth BE, Olds TS. Development of a compendium of energy expenditures for youth. Int J Behav Nutr Phys Act 2008 Sep 10;5:45 [FREE Full text] [doi: 10.1186/1479-5868-5-45] [Medline: 18782458]
61. Butte N, Watson KB, Ridley K, Zakeri IF, McMurray RG, Pfeiffer KA, et al. A Youth Compendium of Physical Activities: Activity Codes and Metabolic Intensities. Med Sci Sports Exerc 2018 Dec;50(2):246-256 [FREE Full text] [doi: 10.1249/MSS.0000000000001430] [Medline: 28938248]
62. Evenson K, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sports Sci 2008 Dec;26(14):1557-1565. [doi: 10.1080/02640410802334196] [Medline: 18949660]
63. Hamad E, Savundranayagam MY, Holmes JD, Kinsella EA, Johnson AM. Toward a Mixed-Methods Research Approach to Content Analysis in The Digital Age: The Combined Content-Analysis Model and its Applications to Health Care Twitter Feeds. J Med Internet Res 2016 Mar 08;18(3):e60 [FREE Full text] [doi: 10.2196/jmir.5391] [Medline: 26957477]
64. Holsti OR. Content analysis for the social sciences and humanities. Boston, USA: Addison-Wesley Pub. Co; 1969:235.
65. Pool IS. Trends in Content Analysis: A Review Article. Journal of Folklore Research 1959:31-38.
66. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
67. Boddy L, Knowles ZR, Davies IG, Warburton GL, Mackintosh KA, Houghton L, et al. Using formative research to develop the healthy eating component of the CHANGE! school-based curriculum intervention. BMC Public Health 2012 Aug 29;12:710 [FREE Full text] [doi: 10.1186/1471-2458-12-710] [Medline: 22931457]
68. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med 2007 Aug 30;26(19):3661-3675. [doi: 10.1002/sim.2832] [Medline: 17315184]
69. Richardson J. The analysis of 2 × 2 contingency tables--yet again. Stat Med 2011 Apr 15;30(8):890; author reply 891-890; author reply 892. [doi: 10.1002/sim.4116] [Medline: 21432882]
70. Brustad R. Children's perspectives on exercise and physical activity: measurement issues and concerns. J Sch Health 1991 May;61(5):228-230. [Medline: 1943049]
71. Cardinal B, Engels H, Zhu W. Application of the Transtheoretical Model of Behavior Change to Preadolescents' Physical Activity and Exercise Behavior. Pediatric Exercise Science 1998 Feb;10(1):69-80. [doi: 10.1123/pes.10.1.69]
72. Sleap M, Warburton P. Physical activity levels of 5-11-year-old children in England: cumulative evidence from three direct observation studies. Int J Sports Med 1996 May;17(4):248-253. [doi: 10.1055/s-2007-972841] [Medline: 8814504]
73. Welk G, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. Res Q Exerc Sport 2000 Jun;71(2 Suppl):S59-S73. [Medline: 10925827]
74. Baquet G, Stratton G, Van Praagh E, Berthoin S. Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: a methodological issue. Prev Med 2007 Feb;44(2):143-147. [doi: 10.1016/j.ypmed.2006.10.004] [Medline: 17157370]
75. Adamo K, Prince SA, Tricco AC, Connor-Gorber S, Tremblay M. A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: a systematic review. Int J Pediatr Obes 2009;4(1):2-27. [doi: 10.1080/17477160802315010] [Medline: 18720173]

76. Economos C, Hyatt RR, Goldberg JP, Must A, Naumova EN, Collins JJ, et al. A community intervention reduces BMI z-score in children: Shape Up Somerville first year results. Obesity (Silver Spring) 2007 May;15(5):1325-1336 [FREE Full text] [doi: 10.1038/oby.2007.155] [Medline: 17495210]
77. Huhman M, Potter LD, Duke JC, Judkins DR, Heitzler CD, Wong FL. Evaluation of a national physical activity intervention for children: VERB campaign, 2002-2004. Am J Prev Med 2007 Jan;32(1):38-43. [doi: 10.1016/j.amepre.2006.08.030] [Medline: 17218189]
78. Sanigorski A, Bell AC, Kremer PJ, Cuttler R, Swinburn BA. Reducing unhealthy weight gain in children through community capacity-building: results of a quasi-experimental intervention program, Be Active Eat Well. Int J Obes (Lond) 2008 Jul;32(7):1060-1067. [doi: 10.1038/ijo.2008.79] [Medline: 18542082]
79. Andersen L, Harro M, Sardinha LB, Froberg K, Ekelund U, Brage S, et al. Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). Lancet 2006 Jul 22;368(9532):299-304. [doi: 10.1016/S0140-6736(06)69075-2] [Medline: 16860699]
80. Ekelund U, Anderssen SA, Froberg K, Sardinha LB, Andersen LB, Brage S, European Youth Heart Study Group. Independent associations of physical activity and cardiorespiratory fitness with metabolic risk factors in children: the European youth heart study. Diabetologia 2007 Sep;50(9):1832-1840. [doi: 10.1007/s00125-007-0762-5] [Medline: 17641870]
81. Rizzo N, Ruiz JR, Hurtig-Wennlöf A, Ortega FB, Sjöström M. Relationship of physical activity, fitness, and fatness with clustered metabolic risk in children and adolescents: the European youth heart study. J Pediatr 2007 Apr;150(4):388-394. [doi: 10.1016/j.jpeds.2006.12.039] [Medline: 17382116]
82. Li S. The Conversation. 2016 Oct 06. 'It's all about me, me, me!' Why children are spending less time doing household chores URL: https://theconversation.com/its-all-about-me-me-me-why-children-are-spending-less-time-doing-household-chores-66134 [accessed 2019-01-18] [WebCite Cache ID 75VxlEeOM]
83. Brettschneider W. Effects of sport club activities on adolescent development in Germany. European Journal of Sport Science 2001 Jun;1(2):1-11. [doi: 10.1080/17461390100071201]
84. LeBlanc A, Janssen I. Difference between self-reported and accelerometer measured moderate-to-vigorous physical activity in youth. Pediatr Exerc Sci 2010 Nov;22(4):523-534. [Medline: 21242602]
85. Lynn R. Sex Differences on the Differential Aptitude Test in British and American Adolescents. Educational Psychology 1992 Jan;12(2):101-102. [doi: 10.1080/0144341920120201]
86. Mann V, Sasanuma S, Sakuma N, Masaki S. Sex differences in cognitive abilities: a cross-cultural perspective. Neuropsychologia 1990;28(10):1063-1077. [Medline: 2267058]
87. Martin D, Hoover HD. Sex Differences in Educational Achievement: A Longitudinal Study. The Journal of Early Adolescence 2016 Sep 15;7(1):65-83. [doi: 10.1177/0272431687071007]
88. Undheim J, Nordvik H. Socio - economic Factors and Sex Differences in an Egalitarian Educational System: academic achievement in 16 - year - old Norwegian students. Scandinavian Journal of Educational Research 1992 Jan;36(2):87-98. [doi: 10.1080/0031383920360201]
89. Roulstone S, Loader S, Northstone K, Beveridge M. The Speech and Language of Children Aged 25 Months: Descriptive Data from the Avon Longitudinal Study of Parents and Children. Early Child Development and Care 2002 Jun;172(3):259-268. [doi: 10.1080/03004430212126]
90. Rangul V, Holmen TL, Kurtze N, Cuypers K, Midthjell K. Reliability and validity of two frequently used self-administered physical activity questionnaires in adolescents. BMC Med Res Methodol 2008 Jul 15;8:47 [FREE Full text] [doi: 10.1186/1471-2288-8-47] [Medline: 18627632]
91. Marshall P. Do tangible interfaces enhance learning? 2007 Presented at: in Proceedings of the 1st international conference on Tangible and embedded interaction; February 15-17, 2007; Baton Rouge, Louisiana p. 163-170. [doi: 10.1145/1226969.1227004]
92. Forlizzi J, Battarbee K. Understanding experience in interactive systems. 2004 Presented at: in Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods,techniques; August 01-04, 2004; Cambridge, MA, USA p. 261-268. [doi: 10.1145/1013115.1013152]
93. Fernaeus Y, Tholander J. "Looking At the Computer but Doing It On Land": Children's Interactions in a Tangible Programming Space. 2006 Presented at: People and Computers XIX — The Bigger Picture; 2005; Edinburgh, Scotland p. 3-18. [doi: 10.1007/1-84628-249-7_1]
94. Suzuki H, Kato H. Interaction-level support for collaborative learning: AlgoBlock—an open programming language. 1995 Presented at: in The first international conference on Computer support for collaborative learning; 1995; Indiana Univ., Bloomington, Indiana, USA p. 349-355. [doi: 10.3115/222020.222828]
95. Svendsen G. The influence of interface style on problem solving. International Journal of Man-Machine Studies 1991 Sep;35(3):379-397. [doi: 10.1016/S0020-7373(05)80134-8]
96. Salvy S, Roemmich JN, Bowker JC, Romero ND, Stadler PJ, Epstein LH. Effect of peers and friends on youth physical activity and motivation to be physically active. J Pediatr Psychol 2009 Mar;34(2):217-225 [FREE Full text] [doi: 10.1093/jpepsy/jsn071] [Medline: 18617572]

97. Jago R, Page AS, Cooper AR. Friends and physical activity during the transition from primary to secondary school. Med Sci Sports Exerc 2012 Jan;44(1):111-117. [doi: 10.1249/MSS.0b013e318229df6e] [Medline: 21697746]
98. Salvy S, de la Haye K, Bowker JC, Hermans RCJ. Influence of peers and friends on children's and adolescents' eating and activity behaviors. Physiol Behav 2012 Jun 06;106(3):369-378 [FREE Full text] [doi: 10.1016/j.physbeh.2012.03.022] [Medline: 22480733]
99. Barkley J, Salvy SJ, Sanders GJ, Dey S, Von Carlowitz KP, Williamson ML. Peer influence and physical activity behavior in young children: an experimental study. J Phys Act Health 2014 Feb;11(2):404-409. [doi: 10.1123/jpah.2011-0376] [Medline: 23359051]
100. Pearce M, Page AS, Griffin TP, Cooper AR. Who children spend time with after school: associations with objectively recorded indoor and outdoor physical activity. Int J Behav Nutr Phys Act 2014 Mar 30;11(1):45 [FREE Full text] [doi: 10.1186/1479-5868-11-45] [Medline: 24679149]
101. Gillet A, Sanner M, Stoffler D, Olson A. Tangible interfaces for structural molecular biology. Structure 2005 Mar;13(3):483-491 [FREE Full text] [doi: 10.1016/j.str.2005.01.009] [Medline: 15766549]
102. Butzow C, Butzow JW. Science through Children's Literature: An Integrated Approach. Science Activities: Classroom Projects and Curriculum Ideas 1990 Sep;27(3):29-38. [doi: 10.1080/00368121.1990.9956738]
103. Piaget J, Cook M. The origins of intelligence in children. Madison, Connecticut, United States: International Universities Press; 1952.

Abbreviations

MET: metabolic equivalent

Edited by G Eysenbach; submitted 07.06.18; peer-reviewed by J Griffin, J Arsenijevic; comments to author 08.10.18; revised version received 15.11.18; accepted 10.12.18; published 22.02.19.

Please cite as:
Crossley SGM, McNarry MA, Rosenberg M, Knowles ZR, Eslambolchilar P, Mackintosh KA
Understanding Youths' Ability to Interpret 3D-Printed Physical Activity Data and Identify Associated Intensity Levels: Mixed-Methods Study
J Med Internet Res 2019;21(2):e11253
URL: http://www.jmir.org/2019/2/e11253/
doi:10.2196/11253
PMID:30794204

Original Paper

Implementing Blockchains for Efficient Health Care: Systematic Review

Anuraag A Vazirani[1]; Odhran O'Donoghue[1]; David Brindley[1], MEng, MSc, DPhil, FRSA; Edward Meinert[1,2], MA, MSc, MBA, MPA, CEng FBCS

[1]Healthcare Translation Research Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom

[2]Global Digital Health Unit, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Edward Meinert, MA, MSc, MBA, MPA, CEng FBCS
Healthcare Translation Research Group
Department of Paediatrics
University of Oxford
John Radcliffe Hospital
Oxford,
United Kingdom
Phone: 44 7824446808
Email: edward.meinert@paediatrics.ox.ac.uk

Abstract

Background: The decentralized nature of sensitive health information can bring about situations where timely information is unavailable, worsening health outcomes. Furthermore, as patient involvement in health care increases, there is a growing need for patients to access and control their data. Blockchain is a secure, decentralized online ledger that could be used to manage electronic health records (EHRs) efficiently, therefore with the potential to improve health outcomes by creating a conduit for interoperability.

Objective: This study aimed to perform a systematic review to assess the feasibility of blockchain as a method of managing health care records efficiently.

Methods: Reviewers identified studies via systematic searches of databases including PubMed, MEDLINE, Scopus, EMBASE, ProQuest, and Cochrane Library. Suitability for inclusion of each was assessed independently.

Results: Of the 71 included studies, the majority discuss potential benefits and limitations without evaluation of their effectiveness, although some systems were tested on live data.

Conclusions: Blockchain could create a mechanism to manage access to EHRs stored on the cloud. Using a blockchain can increase interoperability while maintaining privacy and security of data. It contains inherent integrity and conforms to strict legal regulations. Increased interoperability would be beneficial for health outcomes. Although this technology is currently unfamiliar to most, investments into creating a sufficiently user-friendly interface and educating users on how best to take advantage of it would lead to improved health outcomes.

International Registered Report Identifier (IRRID): RR2-10.2196/10994

(J Med Internet Res 2019;21(2):e12439) doi:10.2196/12439

KEYWORDS

blockchain; electronic health records; efficiency; interoperability; health; information science; computers

Introduction

Medical records in Britain comprise of legacy paper records and numerous disconnected electronic systems. Despite the advancement of other health care fields such as oncology and neurology in their use of technology [1,2], there remains a lack of interoperability in health care systems, arising from the nonuniform record storage methods, that restricts doctors in their capacity to provide appropriate care [3]. Furthermore, the lack of (correct) information has been considered the primary cause of problems in health care, leading to medical errors and adverse events [4]. Not only does this include clinical errors, but administrative ones, such as the recent failure by the National Health Service (NHS) to invite almost 50,000 women

for their cervical screening tests [5,6]. Patients must recount their history multiple times, which may be done incompletely. They too appreciate that interoperability would be beneficial in alleviating these adverse events [7,8]. The NHS planned to mend the situation by computerizing all records by 2018, however this target was delayed first to 2020 and again to 2023 [9]. In the United States, almost 90% of physicians already use a computerized system [10], although these are not all interoperable. Efficient health information technology (IT) systems are especially important at a time when the NHS faces shortages of more than 100,000 doctors and nurses [11,12], a total that is expected to increase with increasing demands on the service. Blockchain could solve the problem of interoperability by allowing doctors to gather information about a patient from multiple independent systems.

A blockchain is a decentralized online ledger (database), first implemented to store an ever-increasing record of all transactions using the cryptocurrency, Bitcoin [13]. It works by replacing trusted third-party signatories of a transaction (in a financial context, typically a “middle-man” payment provider such as Visa) with computational (cryptographic) proof to validate transactions. This validation is carried out by a network of users (“full nodes”) who collectively adhere to previously agreed rules, which are implemented by the software. This method saves both the cost of mediation, as a blockchain involves no mediator, and the cost associated with reversing transactions when disputes arise, as blockchain transactions are essentially irreversible. The transaction records are grouped into blocks, each of which is locked to the next with a cryptographic hash. Once recorded, data in any given block cannot be modified without altering all subsequent blocks (as each block’s hash depends on the last), nor without the agreement of a majority of the members of the network. As well as in financial services, distributed ledger technology has also been applied in the manufacturing industry to track goods within a supply chain [14,15], in governments for voting and public records [16,17], and in retail for unmediated goods trading and to allow more sophisticated loyalty packages [18,19].

The system is also flexible enough to allow the addition of arbitrary logic to process, validate, and access the data. This is implemented via components of business logic known as smart contracts, which reside on the blockchain and are synchronized across all nodes. A smart contract is a string of computer code that executes whenever certain conditions are met, ensuring security and authorized access [20]. The ability to create smart contracts makes blockchain suitable for health care, where strict regulations govern how sensitive data can be used [21,22]. Information exchange using smart contracts is transparent and conflict-free and eliminates the need for a middleman, as the blockchain executes the data sharing based on the preagreed conditions of the contract [23,24].

Ownership and privacy of data are important issues that blockchain could solve. It is currently debated whether the health care provider or the patient owns health care data relating to a patient (although patients have a definite right to access the data [25]). In addition to ownership issues, with the introduction of the General Data Protection Regulation (GDPR) in the European Union, it is important for patients to know how their personal information is being handled [21,26]. Smart contracts implemented by a blockchain would simplify the consent process for data access by doctors. The current consent process is not standardized or personalized, which makes it difficult for a patient to express clearly via an access control policy, which may, for example, involve allowing selected access to particular specialists.

Another concern with medical records is the cost currently associated with transferring records between locations [27]. Repeated imaging studies carried out because of unavailability of prior results can be dangerous in terms of delayed treatment as well as financially costly. Sending data via email is considered a security risk [28,29], and there is clear inefficiency inherent in transcribing a digital asset onto optical media which is commonly read only once at the receiving site [22,30]. A system integrating patient consent as well as access to authorized individuals would save on these costs.

Medical information is no longer limited to written reports, imaging studies, and blood tests. Genomic data and that collected by wearable devices, such as bracelets and watches embedded with sensors, are increasingly accumulated. If exploited effectively, the availability of these new forms of data may lead to improved treatment options and outcomes and may also be examined by health insurance companies offering discounts for “healthy” behavior. Further benefits arise in the realm of artificial intelligence. When given the appropriate data, this can infer trends from the data that are then used to generate population-level insight, and so achieve population health as a whole. These new data formats, however, will require careful integration to allow appropriate analysis while maintaining patient privacy and security against hackers.

Although digitization of health records has been in place in the general practitioner (GP) sector for over 30 years (albeit lacking essential data sharing and exchanging capabilities), secondary care has not yet successfully achieved this de facto standard. Distributed ledger technology, initiated and exemplified by the bitcoin blockchain, is having a growing impact on IT environments in which conformation to legislative regulations and maintenance of public trust are increasingly paramount [31], and it may be used in realizing NHS Digital’s target. The aim of this review was to summarize the evidence relating to the implementation of blockchains to manage electronic health records (EHRs), and to discuss whether this would improve efficiency of record management.

Methods

The review protocol [32] was used with the following modifications: (1) The research question was narrowed to focus on efficiency, as the issues of privacy and scalability would broaden the review excessively, (2) A total of 5 additional search databases were included to account for the potential lack of published research on the topic, and (3) The population was extended to anyone whose health data are stored in a blockchain.

Research Question and Definitions

RQ: What strategies have been proposed or trialed to implement a blockchain or blockchains for the

management of electronic medical records, and how do they improve efficiency compared to currently employed medical record management methods?

1. *Medical record*: [any] record consisting of information about the physical or mental health or condition of an identifiable individual made by or on behalf of a health professional in connection with the care of that individual [33].
2. *Efficiency*: Either improved interoperability or cost-effectiveness, or improved health outcomes as a result of these.
3. *Current methods*: These may consist of traditional paper-based methods or more advanced technology adopted to provide more coordinated health care.
4. *Interoperability*: Automatic and seamless exchange of health information across health information systems [34]. (This corresponds to syntactic interoperability.)

Ethics and Dissemination

As data collection was executed via published literature, ethical approval was not be required for this review.

Search Strategy and Study Selection

We searched PubMed, Scopus, EMBASE, MEDLINE, ProQuest, CINAHL, AMED, Global Health, Books@Ovid, and Cochrane Library for all relevant literature. We also searched for other systematic reviews on the topic on the PROSPERO [35] database. We used the MeSH database [34] to derive keywords and search term combinations (Multimedia Appendix 1). As blockchains applied to the health care sector remain a novel approach, we did not place restrictions on the study type, dates of publication, or geographic locations. However, only studies in English were included.

Results of the search strings (Multimedia Appendix 2) were imported into EndNote X8.0.1 (Clarivate Analytics), which was used to remove duplicate articles. Remaining duplicates were deleted manually. We used an iterative approach of title, abstract, and full text searches (Multimedia Appendix 3) and exported the results into Microsoft Excel. Titles and abstracts were then independently screened by 2 reviewers, based on the following inclusion criteria:

1. Articles must discuss the use of blockchain to manage medical records in some manner, and
2. Articles must describe the benefits and/or disadvantages of using this technology (and compare this to currently used methods of managing medical data).

Where the second point may not be determined from the abstract alone, the study should be taken to full text screening. Studies may not identify a comparator, and these may be included provided the remaining inclusion criteria are met.

A total of 2 reviewers (AV and OO) resolved discrepancies through discussion (Multimedia Appendix 4), and no adjudication from a third reviewer was required. The full texts of the remaining articles were subsequently assessed for their eligibility, based on the same eligibility criteria. This selection process is demonstrated using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram (Figure 1).

Data Extraction

A template was designed to collect information required to address the research question. Basic metadata were collected automatically by EndNote, and the remaining data items (Multimedia Appendix 5) were gathered after reading the papers in full.

Outcomes

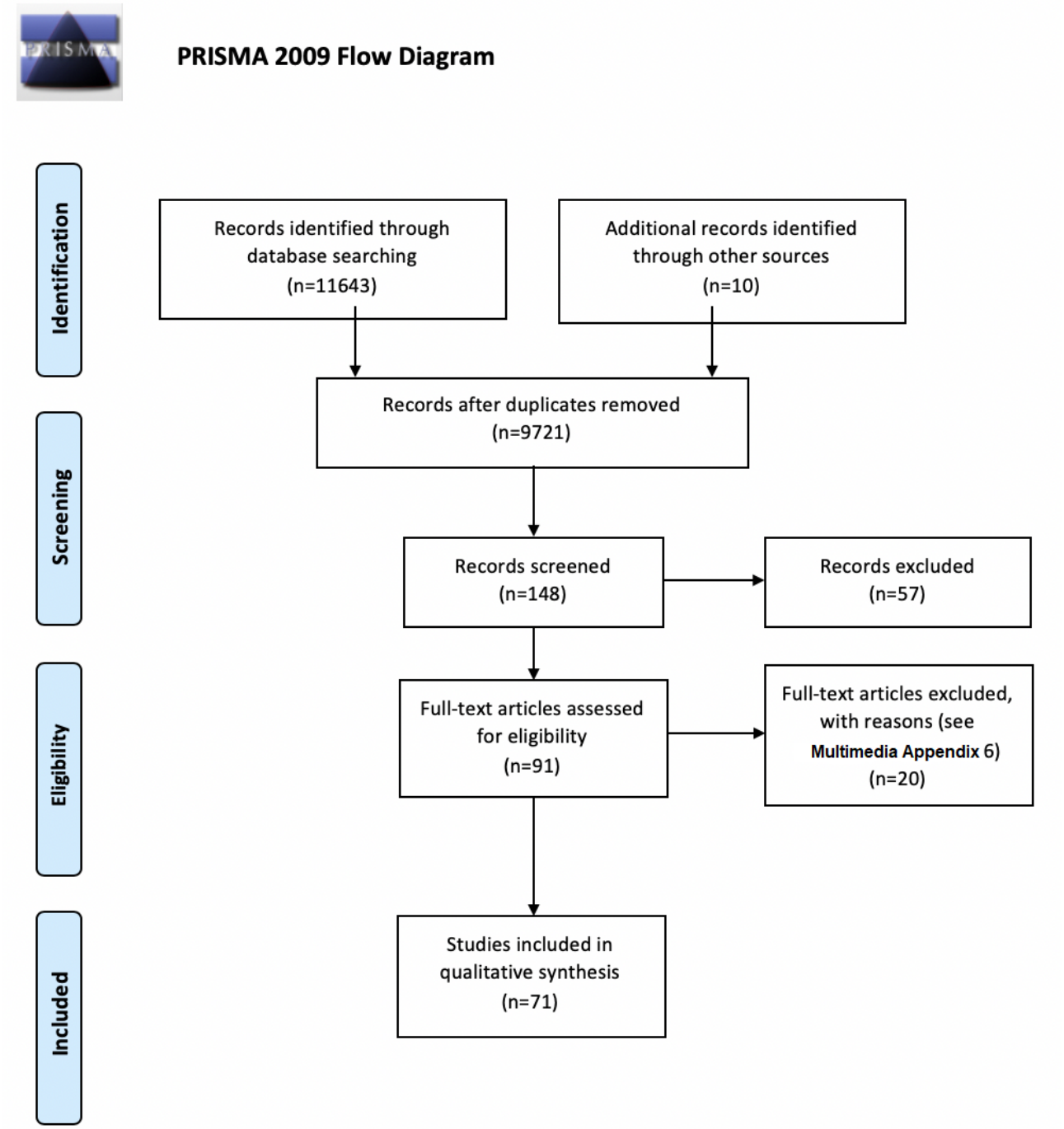
The primary outcome measures were interoperability and cost-effectiveness (our definition of efficiency). The secondary outcome measure was improved health outcomes, although it was noted that it might be difficult to determine a quantitative measure of this with respect to blockchains.

Strength of Evidence and Data Synthesis

Studies of interventions involving randomized and nonrandomized methods were assessed for risk of bias using the Cochrane Collaboration Risk of Bias Tool and the Risk of Bias in Nonrandomized Studies - of Interventions tools, respectively.

The extracted data were subsequently summarized qualitatively. No meta-analysis was performed, because application of blockchains in health care remains a novel method and articles with sufficient numerical data were not found. In addition, the heterogeneity of studies prevented a meta-analysis.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Results

Description and Characteristics of Included Studies

After the initial literature search, removal of duplicates, eligibility (Multimedia Appendix 6), and full-text screening, 61 articles were included in the study. An additional 10 articles were added via snowballing (review of the references from the included articles) of the full texts screened (Table 1).

Few studies described the implementation of a blockchain system to real-world medical data, highlighting the novelty of this technique. Of these, 1 [36] used smart contracts to manage access to medical data that were stored on the cloud, whereas the others stored medical data directly on the blockchain. Of the largest group of articles which proposed a framework (without testing it on live data), the majority advocated cloud-based data storage with blockchain-based access control. In addition to the primary outcome of interoperability, issues considered in these articles included those of privacy and data security, scalability, and administrative affairs. There are also a number of companies currently implementing blockchains in health care, which were referred to throughout the literature, and many of which have not published any academic literature. These are collated in Multimedia Appendix 7.

Table 1. Summary themes of included articles.

Article type	Number of articles
Implementation of EHR[a] blockchain system (tested on real data)	4 [36-39]
Proposal of EHR blockchain framework (not tested on real data)	18 [21-23,31,40-52]
Discussion of benefits and drawbacks of the use of blockchain for EHRs	10 [8,24,28,53-59]
Description of company in the field	20 (Multimedia Appendix 7)
Newspaper, magazine, and columns (opinion articles)	19 in searched databases

[a]EHR: electronic health record.

The majority of the information comparing blockchain's potential versus current methods of managing records was found in opinion articles, which were set more broadly in the context of developments in health care technology; many described the disarray of current health record management. Some used the successes of blockchain in other fields than health care and finance to demonstrate its versatility.

Outcome Measures

Interoperability, one of our primary outcome measures, was seen as feasible using a blockchain approach to store an index of records and to manage access to cloud-based records [36,51]. This approach saves the administrative costs involved in transporting records, as well as the medical costs associated with waiting for their arrival. Logistical difficulties and costs may arise in collating legacy data, although we expect that these would be accounted for in savings from improved health outcomes in the long term.

Discussion

Summary of Evidence

A blockchain can allow improved interoperability as data across multiple systems can be exchanged and accessed simultaneously and immediately. The interfacing of different systems would also save costs, as no expensive data mediators or escrow services are required to broker trust [8,45]. Costs are also saved in the process of data transfer, which becomes immediate [27]. This is especially useful in the management of patients suffering from chronic diseases, facilitating the delivery of care by connecting both patients and providers with medical data. These factors, therefore, reduce administrative delays, as does the use of smart contracts to execute patients' consent preferences immediately [23,24]. An off blockchain data repository ("data lake") is scalable and can store a variety of data types, as well as being a tool for research. This is in contrast to current health care data systems, which are not synchronized to allow intersystem communication or equipped to manage emerging data formats [9]. Blockchain is interactive and supports high throughput data analysis and machine learning, while being encrypted and digitally signed to ensure data privacy and authenticity [51]. These are particularly important factors in a field that is both dominated by evidence-based research but bound by strict data regulations [21]. The disjointed systems currently in use do not lend themselves toward such modern research techniques, nor are they stored in a secure or immutable manner. The collaboration between patients, doctors, and researchers arising from a blockchain-based system allows for a greater degree of exchange and comparison, leading to specific and personalized care pathways [31,45]. The Office of the National Coordinator for Health Information Technology (ONC) has recently described several features critical to the development of an interoperable health system, which are addressed by blockchain [60], such as establishing a directory of resource locations that can be easily referenced to locate information and creating an economic environment in which interoperability is a sound business decision[22].

Interoperability

Health data are dynamic and expansive, and seamless exchange of health data across health information systems would be advantageous. As it would not be practical in terms of speed, storage capacity, or sustainability to replicate all health records on every computer in the blockchain network [54], we instead advocate blockchain as a method to manage access control (and for smart contract management) by systematically storing a catalogue of all users' health records and related metadata. Each time data are added to the EHR by a doctor or patient (from a mobile app or wearable sensor), a metadata-containing pointer to this is added to the blockchain, while the data are stored securely on the cloud [51]. A full index of a particular patient's records is stored in a single location along with related metadata, regardless of the whereabouts of the medical data. The blockchain, with this secured index of records, then directs authorized individuals to the cloud-based data, thus allowing the immediate exchange of information between approved professionals, while also keeping an immutable ledger of those readers. The fact that blockchain is based on open-source software also has potential benefits, as health trusts can use the open application programming interface to integrate data as they wish, giving them timely access to accurate information in a format useable by them. Striving for interoperability is also a key feature of the Health Information Technology for Economic and Clinical Health Act, which has meant that since 2011; American health care providers have been given financial incentives to demonstrate meaningful use of EHRs [61].

Health

Fast access to a comprehensive set of patient data would allow doctors to treat patients without the need to wait for the arrival of previous results. The availability not only of prompt but more frequent data would allow physicians to create specialized treatment plans on the basis of outcomes and treatment efficacy. Daily health data would also engage a patient more in their own health care, and improve patient compliance [51], a historic

challenge in the field [62]. The capability of personalized medicine would, therefore, be improved with this interoperability, as a single access point for all real-time health data is created for each patient. Data gathered from wearable sensors and mobile apps would contribute information on the risks and benefits of treatments, and on patient-reported outcome measures.

Integrity

The immutability of a blockchain that stems from linking the hashes of subsequent blocks, carries with it inherent integrity as blocks cannot be rewritten without collaboration of a majority of nodes. This is key to maintaining a true record of patient-provider interactions as well as data originating from devices, both of which could influence not only medical decisions but also those involving insurance. This property was exemplified by RadBit at the 2017 Yale Healthcare Hackathon [46], a system which allows patients to keep possession of their medical images, along with an immutable chain of custody. Temporary keys ("tokens") can be created by users of the blockchain and passed onto those such as health care providers and insurance companies, providing them temporary access. The token is independent of the data, containing only authorization commands, and is verified and validated (by recording them on the chain) before the required reports are dispatched. Integrity may also be maintained by the use of external auditors, who may verify the accuracy of the system in real time and retrospectively [21]. Potential ways to improve the integrity are to use blind signatures, which reinforce protection from tampering as well as confirming the sender's and viewer's identities [53], or to use signatures from multiple authorities [42].

The Office of the National Coordinator for Health Information Technology Blockchain Challenge

In 2016, the ONC organized the "Use of Blockchain in Health IT and Health-Related Research" challenge, seeking ideas to address the difficulties of managing health records [20,63,64]. Winning papers described innovative ways to securely empower patients through interoperability [57,65]. MedRec, one of the winning entries, is now being implemented in Boston. This proposal from Massachusetts Institute of Technology's media lab involves associating a medical record with viewing permissions and data retrieval instructions for execution on external databases, thus using the blockchain to record patient-provider interactions via smart contracts. Once a doctor creates a record, it is verified, and its viewing permissions are authorized by the patient. The party receiving new information receives an automated notification [3] and a hashed pointer to the new medical record, and its permissions are stored on the chain. This system allows patients to be empowered through access and control of their data, options which have until recently not been available outside a trial setting [57], and which we hope will soon become a standard of care. So far, their system has been successful with medications, blood tests, vaccination histories, and other therapeutic interventions [66].

Large-Scale Implementation

Blockchain has recently been adopted on a large scale by the Estonian government, in collaboration with Guardtime, where it now secures millions of records. Other companies involved in introducing blockchain to everyday health care include Medicalchain, which allows users to sign up and use the interface to interact with their GP, and Patientory that connects doctors, health providers, and consumers, and others listed in Multimedia Appendix 7.

Challenges and Limitations

Data Ownership and Privacy

Achieving interoperability depends on patients taking control of their data and deciding on how it is to be used, features integral to blockchain. Although shifting data ownership from the government and companies to patients would require extensive reengineering of legacy systems to introduce a blockchain, it would incentivize patients to become active agents in their own care by contributing data to get the best possible treatment [8,67,68]. The blockchain would also give patients the sole power to authorize data access to various providers at their discretion [41], eliminating delays associated with the current bureaucracy [52,60,69], and ensuring patient privacy [21]. These benefits bestowed by blockchain stand to empower patients with control over their data, a new expectation in a time when mutualistic and consumerist patient-doctor relationships are becoming the norm. Patients could also selectively share data with researchers using blockchain, either for the greater scientific good or to enable studies on their unique condition [70]. The system would guarantee their consent, a factor key to establishing autonomy and patient-centered care. A recent example of patient autonomy over health data is 23andMe, 80% of whose users chose to make their genomic data available to researchers. This demonstrates that patients will be happy to share data for research should they stand to benefit. Enabling direct patient involvement in controlling the use of their records in this open and secure manner enabled by blockchain will enhance the uptake of such platforms and potentially lead to improved health outcomes [31]. Sharing data may be accomplished with a trusted system, but as an additional incentive, there is the idea of "rewarding" patients for healthy behavior, such as with lower insurance premiums [21,56]. This and other similar ideas will need to be piloted to determine how effectively they will overcome this challenge.

Legal

Under GDPR (Article 17), the Organization of Economic Cooperation and Development privacy guideline, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, and others [58], individuals may request for their data to be erased. This is possible when the records themselves are not stored on the blockchain. However, a record of the data's previous existence may still be maintained within the chain, even if the data were to be deleted. The legal question arising from this relates to whether metadata of personal data classifies as personal data [54]. Regardless, the fact that data in the system would in any case be controlled by the patient is a definite sign of progress. A potential limitation here arises with respect to

the use of a private or consortium-led blockchain, as these data privacy concerns would need to be addressed to make the services compliant and vendor neutral. This may be achieved by having a governing authority enforce open standards. Although these sorts of regulatory constraints are necessary to ensure such appropriate use of information, they may slow development in the field. HIPAA, for example, requires that an institutional review board approve the use of data [21]. That any such delay would be substantial is however unlikely, and so compliance with these regulations is paramount to the success of blockchain.

Security

Sensitive data must be kept safe from eavesdroppers and intruders [41,71]. Breaches have a negative impact on the public perception of the health care field and threaten to hinder future research through more stringent regulatory restrictions [72,73]. The WannaCry attack of May 2017 infected many thousands of computers worldwide, including those of the NHS [64,74,75]. One earlier attack in Los Angeles targeted EHRs in particular, demanding thousands of dollars in ransom [76]. A blockchain is more secure than legacy methods, which would issue patients with credentials [23]. It achieves this property by the use of public-key cryptography (as opposed to symmetric-key cryptography, the method typically used for encryption). This involves generating a public and private key for each user using a one-way encryption function, known as a hash. There is no way for anyone but the recipient to see information sent over the blockchain, as it is secured by their private key. The only security flaw that may arise with a blockchain is unrelated to encryption and comes about if a public blockchain is used: hackers could collude in a "51% attack," resulting in the rewriting of the chain structure [77]. Thus, to realize the advantages of a decentralized system, patients must have some trust that at least 50% of mining nodes would not want to violate the immutability of the blockchain. The public blockchain also leads to the possibility of patient identification, which would need to be avoided by pseudonymizing data to protect patients' identities [22,56]. If a private or consortium blockchain was to be used, however, mining nodes would be limited to hospitals and other trusted health providers, eliminating these security flaws.

Other Concerns

Although the major concerns with blockchain are those of security, privacy, and legal restrictions, for which various workarounds have been developed, there remain some further challenges to consider. First, consolidating data from legacy systems will involve removing data that are duplicated in different parts of the system [8] and converting outdated file formats. This introduces an implementation cost, in excess of the basic cost to introduce a blank system [78], which a government may not be willing to spend [79]. Second, as with any system, it is necessary for users to input good quality information: the trustworthiness arising from blockchain's immutability and decentralization concedes to the input of low-quality (incorrect) information [8,40]. Finally, the issue of currency, used in blockchain to incentivize users to mine blocks in the new network. An initial coin offering [45,57,80] could initiate this process by valuing the new token as funds are raised. However, extremes of price could deter miners, and so mining may need to be restricted to health care providers to avoid this. Another view is to remove all currency, as data are owned by the patient and are not inherently an exchangeable currency [23]. On the basis of this, we may assume that providers already have an incentive to secure patients' medical information, and so there would be no need to incentivize mining beyond the simple use of the system. Finally, the reliance of a blockchain on essentially arbitrary computation could be seen to introduce administrative inefficiency [31]. Transactions are therefore energy-intensive, as each must be computationally verified and validated by the whole network [45]. However, such a mechanism is still beneficial, as rather than providing economic value it demonstrates proof-of-participation, which would be required for ongoing use of the system.

Conclusions

The storage and sharing of medical data (developing interoperability) are vital for improved health outcomes. Respecting privacy of sensitive information while doing this remains a big challenge in health care. The literature shows that with the appropriate regulatory guidelines and use standards, blockchain can act as a vehicle to manage consented access to EHRs. This will increase interoperability without compromising security, while also protecting patient privacy. These issues would most effectively be tackled by the use of a private or consortium-led blockchain; however, this would need to be regulated to ensure appropriate use of data. The improved interoperability and reduced long-term administrative costs would lead to improved health outcomes.

Blockchain represents a new form of technology in which the current literature is lacking in this application context, and no usage feedback or statistical comparisons with traditional systems exist. There are costs associated with transferring to a new system, and in educating health professionals and patients on how best to take advantage of it for improved health. Blockchain involves concepts unfamiliar to the vast majority of the population, such as cryptographic signature and key management. Investments into new systems would, however, be outweighed through returns. In the primary stages of implementation, the practical usefulness of the proposed system will likely depend on the end-user experience—the complexities underlying the blockchain will need to be hidden behind a sufficiently user-friendly interface such as an online or mobile app to be adopted successfully. Short-term trials will outline the most effective ways to implement such a user-friendly experience, which may be expanded thereafter.

Acknowledgments

This work was supported by the Sir David Cooksey Fellowship in Healthcare Translation (EM), the SENS Research Foundation (EM), the Oxford Academic Health Sciences Centre (EM), and the Final Honor School of Medical Sciences, Cell & Systems, Biology and Neuroscience at the University of Oxford (AV).

Authors' Contributions

EM conceived the study topic and oversaw the project. AV investigated and executed bench research and completed manuscript drafting on his own. OD completed peer review of papers for analysis and gave helpful discussion on content. EM gave feedback on the completed manuscript to AV who incorporated all feedback. DB also provided feedback on iterations. AV responded to peer review facilitated by EM. The final manuscript was approved by all authors. EM is the guarantor.

Conflicts of Interest

DB is a stockholder in Translation Ventures Ltd (Charlbury, UK), IP Asset Ventures Ltd, and Biolacuna Ltd (Oxford, UK), companies that, among other services, provide biomanufacturing, regulatory, and financial advice to pharmaceutical clients. DB is also subject to the CFA Institute's codes, standards, and guidelines, so he must stress that this piece is provided for academic interest only and must not be construed in any way as an investment recommendation. Additionally, at time of publication, DB and the organizations with which he is affiliated may or may not have agreed to and/or pending funding commitments from the organizations named herein.

Multimedia Appendix 1

Search terms and keywords.

[DOCX File, 12KB - jmir_v21i2e12439_app1.docx]

Multimedia Appendix 2

Search strings.

[DOCX File, 13KB - jmir_v21i2e12439_app2.docx]

Multimedia Appendix 3

Electronic screening.

[DOCX File, 13KB - jmir_v21i2e12439_app3.docx]

Multimedia Appendix 4

Title and abstract screening.

[DOCX File, 12KB - jmir_v21i2e12439_app4.docx]

Multimedia Appendix 5

Data extraction items.

[DOCX File, 12KB - jmir_v21i2e12439_app5.docx]

Multimedia Appendix 6

Eligibility stage search exclusions.

[DOCX File, 12KB - jmir_v21i2e12439_app6.docx]

Multimedia Appendix 7

Blockchain and company system names.

[DOCX File, 12KB - jmir_v21i2e12439_app7.docx]

References

1. Sample I. The Guardian. 'It's going to create a revolution': how AI is transforming the NHS URL: https://www.theguardian.com/technology/2018/jul/04/its-going-create-revolution-how-ai-transforming-nhs [accessed 2018-11-14] [WebCite Cache ID 749i1L3v9]
2. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, et al. A deep learning model to predict a diagnosis of alzheimer disease by using F-FDG PET of the brain. Radiology 2018 Nov 6:180958. [doi: 10.1148/radiol.2018180958] [Medline: 30398430]
3. Brock J. Healthcare Innovation. 2017 May 25. Healing revenue cycle management URL: https://www.hcinnovationgroup.com/finance-revenue-cycle/article/13008199/healing-revenue-cycle-management [accessed 2019-02-02] [WebCite Cache ID 75tD2Hagj]
4. CRICO. 2015. Malpractice Risks in Communication Failures URL: https://www.rmf.harvard.edu/Malpractice-Data/Annual-Benchmark-Reports/Risks-in-Communication-Failures
5. Capita. Capita statement on cervical screening correspondence delay URL: https://www.capita.com/news/news/2018/capita-statement-on-cervical-screening-correspondence-delay/ [accessed 2019-02-02] [WebCite Cache ID 75tDEQHcT]
6. Iacobucci G. Cervical screening: GP leaders slam Capita over failure to send up to 48 500 letters. Br Med J 2018 Nov 14:k4832. [doi: 10.1136/bmj.k4832]
7. PwC UK. Majority of public have never accessed their own health records URL: https://www.pwc.co.uk/press-room/press-releases/majority-of-public-have-never-accessed-their-own-health-records.html [accessed 2019-02-02] [WebCite Cache ID 75tDJNqKK]
8. Engelhardt MA. Hitching healthcare to the chain: an introduction to blockchain technology in the healthcare sector. TIM Rev 2017 Oct 27;7(10):22-34. [doi: 10.22215/timreview/1111]
9. Department of Health & Social Care. Making IT work: harnessing the power of health information technology to improve care in England URL: https://www.gov.uk/government/publications/using-information-technology-to-improve-the-nhs/making-it-work-harnessing-the-power-of-health-information-technology-to-improve-care-in-england [accessed 2018-07-13] [WebCite Cache ID 73gDg6PI5]
10. Jamoom E, Yang N. Centers for Disease Control and Prevention. 2016. National Electronic Health Records Survey: 2015 State and National Electronic Health Record Adoption Summary Tables URL: https://www.cdc.gov/nchs/data/ahcd/nehrs/2015_nehrs_web_table.pdf [accessed 2018-09-11] [WebCite Cache ID 73gDyiV4m]
11. The Nuffield Trust. The health care workforce in England URL: https://www.nuffieldtrust.org.uk/files/2018-11/health-foundation-king-s-fund-and-nuffield-trust-the-health-care-workforce-in-england.pdf [accessed 2019-02-02] [WebCite Cache ID 75tDcPHRK]
12. Iacobucci G. NHS staff shortages could reach 250 000 by 2030 without urgent action, experts warn. Br Med J 2018 Nov 14;363:k4831. [doi: 10.1136/bmj.k4831] [Medline: 30442691]
13. Nakamoto S. Bitcoin. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System URL: https://bitcoin.org/bitcoin.pdf [accessed 2018-11-04] [WebCite Cache ID 73gEKsOBp]
14. Skuchain. URL: http://www.skuchain.com/ [accessed 2018-11-14] [WebCite Cache ID 73vJXmYwP]
15. Wayback Machine. Blockchain: the solution for transparency in product supply chains URL: http://web.archive.org/web/20180920095159/https://www.provenance.org/whitepaper [accessed 2018-11-14] [WebCite Cache ID 73vJRTl2X]
16. Aki J. Bitcoin Magazine. Thailand Uses Blockchain-Supported Electronic Voting System in Primaries URL: https://bitcoinmagazine.com/articles/thailand-uses-blockchain-supported-electronic-voting-system-primaries/ [accessed 2018-11-14] [WebCite Cache ID 73vK241Jm]
17. Ruubel M. Guardtime. Blockchain-Enabled Cloud: Estonian Government selects Ericsson, Apcera and Guardtime URL: https://guardtime.com/blog/blockchain-enabled-cloud-estonian-government-selects-ericsson-apcera-and-guardtime [accessed 2018-11-14] [WebCite Cache ID 73vJmtlDt]
18. OpenBazaar. URL: https://openbazaar.org/ [accessed 2018-11-14] [WebCite Cache ID 73vK9DRMy]
19. Loyyal. URL: https://www.loyyal.com/ [accessed 2018-11-14] [WebCite Cache ID 73vKSBkb3]
20. Wright G, Wright A, Landman A. NEJM Catalyst. 2017 Feb 9. Blockchain in Health Care: Decoding the Hype URL: https://catalyst.nejm.org/decoding-blockchain-technology-health/ [accessed 2018-11-04] [WebCite Cache ID 73gERTHFx]
21. Mamoshina P, Ojomoko L, Yanovich Y, Ostrovski A, Botezatu A, Prikhodko P, et al. Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget 2018 Jan 19;9(5):5665-5690 [FREE Full text] [doi: 10.18632/oncotarget.22345] [Medline: 29464026]
22. Patel V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health Informatics J 2018 Apr 1:1460458218769699. [doi: 10.1177/1460458218769699] [Medline: 29692204]
23. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustain Cities Soc 2018 May;39:283-297. [doi: 10.1016/j.scs.2018.02.014]
24. Alhadhrami Z, Alghfeli S, Alghfeli M, Abedlla JA, Shuaib K. Introducing Blockchains for Healthcare. 2017 Nov 23 Presented at: 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA); November 21-23, 2017; Ras Al Khaimah, United Arab Emirates. [doi: 10.1109/ICECTA.2017.8252043]

25. Nottinghamshire Local Medical Committee Ltd. 2018 Jun. Access to health records: Updated to reflect the General Data Protection Regulation URL: https://www.bma.org.uk/-/media/files/pdfs/employment%20advice/ethics/access%20to%20health%20records%2027%20june%202018.pdf [accessed 2019-02-02] [WebCite Cache ID 75tDxfUIG]
26. Health Information Privacy. Your Rights Under HIPAA URL: https://www.hhs.gov/hipaa/for-individuals/guidance-materials-for-consumers/index.html [accessed 2018-11-04] [WebCite Cache ID 73gElGR1a]
27. Patel V, Barker W, Siminerio E. Office of the National Coordinator for Health Information Technology. Trends in Consumer Access and Use of Electronic Health Information URL: https://www.healthit.gov/sites/default/files/briefs/oncdatabrief30_accesstrends_.pdf [accessed 2018-07-13] [WebCite Cache ID 73gFcnBMo]
28. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and trustable electronic medical records sharing using blockchain. AMIA Annu Symp Proc 2017;2017:650-659 [FREE Full text] [Medline: 29854130]
29. Health Information Privacy. HIPAA Privacy Rule URL: https://www.hhs.gov/hipaa/for-professionals/privacy/index.html [accessed 2018-11-04] [WebCite Cache ID 73gFk9wju]
30. Erickson BJ. Experience with importation of electronic images into the medical record from physical media. J Digit Imaging 2011 Aug;24(4):694-699 [FREE Full text] [doi: 10.1007/s10278-011-9364-3] [Medline: 21286776]
31. Cunningham J, Ainsworth J. Enabling patient control of personal electronic health records through distributed ledger technology. Stud Health Technol Inform 2017;245:45-48. [doi: 10.3233/978-1-61499-830-3-45] [Medline: 29295049]
32. Meinert E, Alturkistani A, Foley KA, Osama T, Car J, Majeed A, et al. Blockchain Implementation in Health Care: Protocol for a Systematic Review. JMIR Res Protoc 2019 Feb 08;8(2):e10994 [FREE Full text] [doi: 10.2196/10994] [Medline: 30735146]
33. National Health Service. Guidance for Access to Health Records Requests URL: https://www.igt.hscic.gov.uk/WhatsNewDocuments/Access%20to%20Health%20Records%20Feb%202010.pdf [accessed 2018-11-04] [WebCite Cache ID 73gGTGgOc]
34. MeSH Browser. Medical Subject Headings URL: https://meshb.nlm.nih.gov/search [accessed 2019-02-02] [WebCite Cache ID 75tEgmIg4]
35. PROSPERO. URL: https://www.crd.york.ac.uk/prospero/ [accessed 2019-02-02] [WebCite Cache ID 75tEiPdRt]
36. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: Using Blockchain for Medical Data Access and Permission Management. 2016 Aug Presented at: 2nd International Conference on Open and Big Data (OBD); August 22-24, 2016; Vienna, Austria. [doi: 10.1109/OBD.2016.11]
37. Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
38. Lee SH, Yang CS. Fingernail analysis management system using microscopy sensor and blockchain technology. Int J Distrib Sens Netw 2018 Mar 28;14(3):155014771876704. [doi: 10.1177/1550147718767044]
39. e-Estonia. eHealth authority partners with Guardtime to accelerate transparency and auditability in health care URL: https://e-estonia.com/ehealth-authority-partners-with-guardtime-to-accelerate-transparency-and-auditability-in-health-care/ [accessed 2018-11-04] [WebCite Cache ID 73gGdQOAX]
40. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. How Blockchain Could Empower eHealth: An Application for Radiation Oncology. 2017 Presented at: Data Management and Analytics for Medicine and Healthcare - Third International Workshop, DMAH 2017; September 1, 2017; Munich, Germany p. 0494. [doi: 10.1007/978-3-319-67186-4_1]
41. Omar A, Rahman M, Basu A, Kiyomoto S. MediBchain: A Blockchain Based Privacy Preserving Platform for Healthcare Data. 2017 Dec 9 Presented at: 10th International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage (SpaCCS 2017); December 12-15, 2017; Guangzhou, China p. 534-543. [doi: 10.1007/978-3-319-72395-2_49]
42. Guo R, Shi H, Zhao Q, Zheng D. Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems. IEEE Access 2018;6:11676-11686 [FREE Full text] [doi: 10.1109/ACCESS.2018.2801266]
43. Liang X, Shetty S, Zhao J, Bowden D, Li D. ResearchGate. Towards Decentralized AccountabilitySelf-sovereignty in Healthcare Systems URL: https://www.researchgate.net/publication/321137917_Towards_Decentralized_Accountability_and_Self-Sovereignty_in_Healthcare_Systems
44. Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating Blockchain for Data Sharing and Collaboration in Mobile Healthcare Applications. 2017 Oct 8 Presented at: 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC); October 8-13, 2017; Bologna, Italy p. 1-5. [doi: 10.1109/PIMRC.2017.8292361]
45. Mannaro K, Baralla G, Pinna A, Ibba S. A blockchain approach applied to a teledermatology platform in the Sardinian region (Italy). Information 2018 Feb 23;9(2):44 [FREE Full text] [doi: 10.3390/info9020044]
46. Nichol PB. CIO. 2017. Top 3 practical innovations from Yale healthcare hackathon URL: https://www.cio.com/article/3161886/innovation/top-3-practical-innovations-from-yale-healthcare-hackathon.html [accessed 2018-07-13] [WebCite Cache ID 73gGohT1l]
47. Raju S, Rajesh V, Deogun JS. The Case for a Data Bank: an Institution to Govern Healthcare and Education. In: Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance. 2017 Mar 7 Presented at: ICEGOV'17; March 7-9, 2017; New Delhi, India p. 538-539. [doi: 10.1145/3047273.3047275]

48. Siddiqi M, All S, Sivaraman V. Secure Lightweight Context-driven Data Logging for Bodyworn Sensing Devices. 2017 Apr 26 Presented at: 5th International Symposium on Digital Forensic and Security (ISDFS 2017); April 26-28, 2017; Tirgu Mures, Romania. [doi: 10.1109/ISDFS.2017.7916500]
49. Zhang J, Xue N, Huang X. A secure system for pervasive social network-based healthcare. IEEE Access 2016;4:9239-9250 [FREE Full text] [doi: 10.1109/ACCESS.2016.2645904]
50. Zhao Z, Zhang Y, Peng Y, Xu R. Lightweight backup and efficient recovery scheme for health blockchain keys. 2017 Mar 22 Presented at: IEEE 13th International Symposium on Autonomous Decentralized Systems; March 22-24, 2017; Bangkok, Thailand. [doi: 10.1109/ISADS.2017.22]
51. Linn LA, Koo MB. Office of the National Coordinator for Health Information Technology. Blockchain For Health Data and Its Potential Use in Health IT and Health Care Related Research URL: https://www.healthit.gov/sites/default/files/11-74-ablockchainforhealthcare.pdf [accessed 2018-11-05] [WebCite Cache ID 73hP7Ddos]
52. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218 [FREE Full text] [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
53. Liu P. Medical record system using blockchain, big data tokenization. 2016 Nov 25 Presented at: International Conference on Information and Communications Security (ICICS 2016); November 29–December 2, 2016; Singapore p. 254-261. [doi: 10.1007/978-3-319-50011-9_20]
54. Esposito C, de Santis A, Tortora G, Chang H, Choo KR. Blockchain: a panacea for healthcare cloud-based data security and privacy? IEEE Cloud Comput 2018 Jan;5(1):31-37 [FREE Full text] [doi: 10.1109/MCC.2018.011791712]
55. Mertz L. (Block) chain reaction: a blockchain revolution sweeps into health care, offering the possibility for a much-needed data solution. IEEE Pulse 2018;9(3):4-7. [doi: 10.1109/MPUL.2018.2814879] [Medline: 29757744]
56. Mertz L. Hospital CIO explains blockchain potential: an interview with Beth Israel Deaconess Medical Center's John Halamka. IEEE Pulse 2018 May;9(3):8-9. [doi: 10.1109/MPUL.2018.2814878] [Medline: 29757745]
57. Pauwels E, Grevatt N. Wilson Center. 2017. The Social Benefits of Blockchain for Health Data: Securing Patient Privacy & Control URL: https://www.wilsoncenter.org/sites/default/files/blockchain_for_health_data_securing_patient_privacy_control.pdf [accessed 2018-11-05] [WebCite Cache ID 73hPNSZel]
58. Weiss M, Botha A, Herselman M, Loots G. Blockchain as an enabler for public mHealth solutions in South Africa. : 30 May-2 June, 2017; 2017 May 30 Presented at: IST-Africa Week Conference (IST-Africa); 2017; Windhoek, Namibia. [doi: 10.23919/ISTAFRICA.2017.8102404]
59. Wong MC, Yee K, Nøhr C. Socio-technical considerations for the use of blockchain technology in healthcare. Stud Health Technol Inform 2018;247:636-640 [FREE Full text] [Medline: 29678038]
60. Office of the National Coordinator for Health Information Technology. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap URL: https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf [accessed 2018-11-05] [WebCite Cache ID 73hPb3aIE]
61. Office of the National Coordinator for Health Information Technology. Index for Excerpts from the American Recovery and Reinvestment Act of 2009 (ARRA) URL: https://www.healthit.gov/sites/default/files/hitech_act_excerpt_from_arra_with_index.pdf [accessed 2018-09-18] [WebCite Cache ID 73hPkxfxS]
62. Scambler G. Sociology as Applied to Medicine Sixth Edition. London, UK: Saunders Ltd; 2008.
63. Office of the National Coordinator for Health Information Technology. Announcing the Blockchain Challenge URL: https://www.healthit.gov/topic/grants-contracts/announcing-blockchain-challenge [accessed 2018-11-05] [WebCite Cache ID 73hPshgqZ]
64. Slabodkin G. Blockchain remains a work in progress for use in healthcare. Health Data Manage 2017;25(3):37-39 https://www.healthdatamanagement.com/news/blockchain-remains-a-work-in-progress-for-use-in-healthcare.
65. ID2020 Alliance. URL: https://id2020.org/ [accessed 2018-11-05] [WebCite Cache ID 73hPwddbe]
66. Journal of Ahima. New Administration Gets to Know 'Blockchain' URL: http://www.ahimajournal-digital.com/ahimajournal/april_2017?pg=11#pg11 [accessed 2019-02-02] [WebCite Cache ID 75tGNXr5x]
67. Kitson A, Marshall A, Bassett K, Zeitz K. What are the core elements of patient-centred care? A narrative review and synthesis of the literature from health policy, medicine and nursing. J Adv Nurs 2013 Jan;69(1):4-15. [doi: 10.1111/j.1365-2648.2012.06064.x] [Medline: 22709336]
68. Stewart M. Towards a global definition of patient centred care. Br Med J 2001 Feb 24;322(7284):444-445 [FREE Full text] [doi: 10.1136/bmj.322.7284.444] [Medline: 11222407]
69. Potts J. The big cure. IPA Review 2016;68(4):- [FREE Full text]
70. Prajna S. CommsMEA. 2017. Unleashing the power of the ledger URL: https://www.commsmea.com/16927-unleashing-the-power-of-the-ledger [accessed 2018-11-05] [WebCite Cache ID 73hQDGaxT]
71. Health Information Privacy. HIPAA Security Rule URL: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html [accessed 2018-11-05] [WebCite Cache ID 73hQPLMHD]
72. Patil HK, Seshadri R. Big Data Security and Privacy Issues in Healthcare. 2014 Oct 27 Presented at: IEEE International Congress on Big Data; June 27-July 2, 2014; Washington, DC. [doi: 10.1109/BigData.Congress.2014.112]

73. Padmanabhan P. CIO. How AI and blockchain are driving precision medicine in 2017 URL: https://www.cio.com/article/3157477/healthcare/how-ai-and-blockchain-are-driving-precision-medicine-in-2017.html [accessed 2018-07-13] [WebCite Cache ID 73hQZnu0o]
74. Fox News. 'Ransomware' cyberattack cripples hospitals across England URL: https://www.foxnews.com/health/ransomware-cyberattack-cripples-hospitals-across-england [accessed 2018-11-05] [WebCite Cache ID 73hQg0X4g]
75. Recode. US hospitals have been hit by the global ransomware attack URL: https://www.recode.net/2017/6/27/15881666/global-eu-cyber-attack-ushackers-nsa-hospitals [accessed 2018-09-18] [WebCite Cache ID 73hQjc8JH]
76. Conn J. Modern Healthcare. Could blockchain help cure health IT's security woes? URL: https://www.modernhealthcare.com/article/20161105/MAGAZINE/311059966 [accessed 2018-11-05] [WebCite Cache ID 73hQu0K6H]
77. Wood G. Gavin Wood. Ethereum: A secure decentralized transaction ledger URL: http://gavwood.com/paper.pdf [accessed 2019-02-02] [WebCite Cache ID 75tGmqUkv]
78. Moss J, Smith C, Davies J. Healthcare Dive. 2017 Sep 1. Blockchain shows promise in healthcare URL: https://www.healthcaredive.com/news/blockchain-shows-promise-in-improving-interoperability-idc-report-says/528105/
79. Padmanabhan P. CIO. 2017. Technology-led innovation in digital health: the law of inverse relationships URL: https://www.cio.com/article/3211596/innovation/technology-led-innovation-in-digital-health-the-law-of-inverse-relationships.html [accessed 2018-11-05] [WebCite Cache ID 73hRC9v8Y]
80. Damiani J. Forbes. 2017. SimplyVital Health Is Using Blockchain To Revolutionize Healthcare URL: https://www.forbes.com/sites/jessedamiani/2017/11/06/simplyvital-health-blockchain-revolutionize-healthcare [accessed 2018-07-13] [WebCite Cache ID 73hPChtda]

Abbreviations

EHR: electronic health record
GDPR: General Data Protection Regulation
GP: general practitioner
HIPAA: Health Insurance Portability and Accountability Act
IT: information technology
NHS: National Health Service
ONC: Office of the National Coordinator for Health Information Technology

Edited by G Eysenbach, K Clauson, P Zhang; submitted 07.10.18; peer-reviewed by B Schreiweis, S Zheng; comments to author 25.10.18; revised version received 27.11.18; accepted 29.12.18; published 12.02.19.

Please cite as:
Vazirani AA, O'Donoghue O, Brindley D, Meinert E
Implementing Blockchains for Efficient Health Care: Systematic Review
J Med Internet Res 2019;21(2):e12439
URL: http://www.jmir.org/2019/2/e12439/
doi:10.2196/12439
PMID:30747714

Review

Mobile Digital Education for Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration

Gerard Dunleavy[1], MSc; Charoula Konstantia Nikolaou[2], PhD; Sokratis Nifakos[3], MSc; Rifat Atun[4,5], MBBS, FRCP; Gloria Chun Yi Law[1], PhD; Lorainne Tudor Car[6,7], MD, MSc, PhD

[1]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[2]Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[3]Health Informatics Centre, Karolinska Institutet, Stockholm, Sweden
[4]Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States
[5]Harvard Medical School, Harvard University, Boston, MA, United States
[6]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[7]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Lorainne Tudor Car, MD, MSc, PhD
Family Medicine and Primary Care
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
Level 18, Clinical Science Building
11 Mandalay Road
Singapore, 308232
Singapore
Phone: 65 6904 1258
Email: lorainne.tudor.car@ntu.edu.sg

Abstract

Background: There is a pressing need to implement efficient and cost-effective training to address the worldwide shortage of health professionals. Mobile digital education (mLearning) has been mooted as a potential solution to increase the delivery of health professions education as it offers the opportunity for wide access at low cost and flexibility with the portability of mobile devices. To better inform policy making, we need to determine the effectiveness of mLearning.

Objective: The primary objective of this review was to evaluate the effectiveness of mLearning interventions for delivering health professions education in terms of learners' knowledge, skills, attitudes, and satisfaction.

Methods: We performed a systematic review of the effectiveness of mLearning in health professions education using standard Cochrane methodology. We searched 7 major bibliographic databases from January 1990 to August 2017 and included randomized controlled trials (RCTs) or cluster RCTs.

Results: A total of 29 studies, including 3175 learners, met the inclusion criteria. A total of 25 studies were RCTs and 4 were cluster RCTs. Interventions comprised tablet or smartphone apps, personal digital assistants, basic mobile phones, iPods, and Moving Picture Experts Group-1 audio layer 3 player devices to deliver learning content. A total of 20 studies assessed knowledge (n=2469) and compared mLearning or blended learning to traditional learning or another form of digital education. The pooled estimate of studies favored mLearning over traditional learning for knowledge (standardized mean difference [SMD]=0.43, 95% CI 0.05-0.80, N=11 studies, low-quality evidence). There was no difference between blended learning and traditional learning for knowledge (SMD=0.20, 95% CI –0.47 to 0.86, N=6 studies, low-quality evidence). A total of 14 studies assessed skills (n=1097) and compared mLearning or blended learning to traditional learning or another form of digital education. The pooled estimate of studies favored mLearning (SMD=1.12, 95% CI 0.56-1.69, N=5 studies, moderate quality evidence) and blended learning (SMD=1.06, 95% CI 0.09-2.03, N=7 studies, low-quality evidence) over traditional learning for skills. A total of 5 and 4 studies assessed attitudes (n=440) and satisfaction (n=327), respectively, with inconclusive findings reported for each outcome. The risk of bias was judged as high in 16 studies.

Conclusions: The evidence base suggests that mLearning is as effective as traditional learning or possibly more so. Although acknowledging the heterogeneity among the studies, this synthesis provides encouraging early evidence to strengthen efforts aimed at expanding health professions education using mobile devices in order to help tackle the global shortage of health professionals.

(J Med Internet Res 2019;21(2):e12937) doi:10.2196/12937

KEYWORDS

mLearning; digital education; health workforce; systematic review; meta-analysis

Introduction

In 2013, the World Health Organization estimated that there was a shortage of 17.4 million health care workers worldwide: around 2.6 million doctors, and approximately 9 million nurses and midwives [1]. This shortage is more apparent in certain regions like Africa where there is an average of 1.9 health workers per 1000 population when 4.5 are needed to reach the health-related sustainable development goals [2]. This situation is further exacerbated with the migration of both students and fully qualified workers, either from rural to urban areas within a country or migration outside the country [3]. This dearth and disproportionate distribution of health workers worldwide [4] may be aggravated by the inadequacy of training programs (in terms of content, organization, and delivery) to provide trainees with the necessary skills, competencies, and experience for the context in which they will work [5]. Therefore, focused effort and resources are needed to develop and implement strategies aimed at increasing both the number of health professionals and the quality and relevance of their training [2,6]. The deployment of information and communication technologies for educational purposes (ie, digital education) has been recognized as a strategic platform to build robust health professions education and training systems [7].

Digital education is a broad construct describing a wide range of teaching and learning strategies that are exclusively based on the use of electronic media and devices as training, communication, and interaction tools [8]. The construct covers aspects that may pertain to educational approaches, concepts, methods and technologies [9]. Digital education facilitates distant learning, which may help address the shortage of health professionals and educators in settings with limited resources by reducing the constraints of time and geographic barriers to training. When digital education is used alongside traditional educational strategies such as classroom-based, face-to-face teaching, this method of education can be considered *blended learning*.

Digital education can entail various types of interventions that can be characterized in different ways: according to delivery tools, content, learning objectives, pedagogical approaches or settings of delivery. We categorized digital education according to the mode of delivery of digital education intervention and the pedagogical methods. Digital education includes, but is not limited to, offline and online computer-based education, serious gaming and gamification, massive open online courses, virtual reality environments, augmented reality, virtual patient simulations, psychomotor skills trainers and mobile digital education (mLearning) among others [9]. Each of these types of digital education has its own specificities, advantages, limitations, and challenges. This review is part of a series of reviews [10-14] evaluating the efficacy of different types of digital education in improving skills, knowledge, attitudes, and ultimately clinical competencies of pre and postregistration health professionals. This review focuses on mLearning for pre and postregistration health professions education [14].

There is no uniformly accepted definition of mLearning. This lack of consensus not only arises from the rapid evolution of the field but also from ambiguity of the term “mobile.” Earlier definitions of mLearning were technocentric and only focused on the types of devices used, (eg, through a smartphone or tablet), or the situational context in which learning takes place (eg, on the way back home) [15], whereas more recent definitions of mLearning give more weight to the learner and the context in which the learning takes place. In the Handbook of Mobile Learning, mLearning was defined as “learning across multiple contexts, through social and content interactions, using personal electronic devices” [16]. However, the latter definition creates ambiguity around the type of devices, particularly given the number of personal consumer devices, such as laptops, that are currently available in the market. To avoid such ambiguity, we considered mLearning in health professions education as any intervention using handheld, mobile devices connected through wireless connections to deliver educational content to pre and postregistration health professionals in order to extend the reach of learning and teaching beyond physical space and distance.

mLearning is increasingly used in health professions education before (preregistration) and after qualification (postregistration), for example, as part of specialty training, continuous medical education or continuous personal development. In this review, we present the evidence collated on the use of mLearning in pre and postregistration health professions education. We considered eligible studies on candidates for, and holders of, the qualifications listed in the Health Field of Education and Training of the International Standard Classification of Education. We combine both the technocentric and the learner-centered approaches by defining handheld, mobile devices as being “small, autonomous, and unobtrusive enough to accompany us in every moment of our every-day life” [17]. Arguably, considering the recent advances in the capabilities of modern handheld devices, many if not all of the digital education interventions could foreseeably be delivered via mLearning.

Past reviews have underlined the potential of mLearning interventions but also stressed upon the need for further research and reviews on the topic [18-21]. Considering the rapidly

evolving nature of mLearning technologies, up-to-date evidence is essential to evaluate the effectiveness of mLearning for health professions education. The most recently published of these reviews was in 2015 with a search strategy that was applied in 2012 [19]. However, the technology and field of mLearning have advanced rapidly since. The past reviews had methodological flaws, which as a result garnered less evidence, with some reviews focusing singularly on 1 medium of mLearning rather than being inclusive across a range of mLearning devices [20]. With a more robust and systematic methodology, this review collates new evidence published since these reviews were performed [18-21], providing a more comprehensive, focused and up-to-date evaluation of mLearning in health professions education.

The primary objective of this review is to evaluate the effectiveness of mLearning educational interventions for delivering preregistration and postregistration health professions education.

Methods

We adhered to the published protocol [14] and followed Cochrane methodology in every step of the review [22]. For a more detailed description of the methodology, please refer to the methodology paper by Car J et al [9].

Search Strategy and Data Sources

Electronic Searches

We developed a comprehensive search strategy for Medical Literature Analysis and Retrieval System Online (Ovid), EMBASE (Elsevier), Cochrane Central Register of Controlled Trials (Wiley), PsycINFO (Ovid), Educational Research Information Centre (Ovid), Cumulative Index to Nursing and Allied Health Literature (Ebsco) and Web of Science Core Collection (Thomson Reuters). Multimedia Appendix 1 contains the Medical Literature Analysis and Retrieval System Online [Ovid] search strategy used. Databases were searched from January 1990 to August 2017. The reason for selecting 1990 as the starting year for our search is because before this year, the use of mobile devices for education was limited to very basic tasks. There were no language restrictions.

We searched reference lists of all the studies that we deemed eligible for inclusion in our review and relevant systematic reviews. We also searched the International Clinical Trials Registry Platform Search Portal and metaRegister of Controlled Trials to identify unpublished trials from and including 1990 to August 16, 2017.

Inclusion Criteria

We included RCTs and cluster RCTs. We excluded crossover trials because of a high likelihood of a carry-over effect. We included studies with students enrolled in either preregistration or postregistration health professions educational programs. We defined preregistration, undergraduate education or basic vocational training as any type of study leading to a qualification that (1) is recognized by the relevant governmental or professional bodies of the country where the studies were conducted and (2) entitles the qualification-holder to apply for entry-level positions in the health care workforce. Postregistration health profession educational programs were defined as any type of study after a qualification, which is recognized by the relevant governmental or professional bodies that enable the qualification holder entry into or continuation of work in the health care workforce in a more independent or senior role. Participants were not excluded based on age, gender, or any other sociodemographic characteristic.

We included studies in which mLearning interventions were used to deliver the learning content of the course. This includes studies where mLearning methods were the sole means by which the intervention was delivered or where mLearning methods were used in combination with traditional learning (ie, blended learning), as long as the contribution of the mLearning component to overall learning has been assessed.

mLearning interventions were defined as any teaching, learning and/or training intervention that was delivered through handheld mobile devices using wireless transmissions: third generation of mobile telecommunications technology, fourth generation of mobile telecommunications technology, global system for mobile communications, originally groupe spécial mobile (GSM), general packet radio services (GPRS), enhanced data rates for GSM evolution (EDGE or EGPRS), multimedia messaging service, short message service, universal mobile telecommunications system, wireless networking (Wi-Fi or any other wireless local area network) or long term evolution standard. Handheld mobile devices include but are not limited to mobile phones, smartphones, personal digital assistants (PDAs), tablets and Moving Picture Experts Group-1 audio layer 3 (MP3) players.

To be eligible for inclusion, studies have to report at least one of the following primary or secondary outcomes. The primary outcomes (measured using any validated or nonvalidated instruments) were the following: (1) students' knowledge postintervention scores, (2) students' skills postintervention, (3) students' attitudes postintervention toward the mLearning intervention, education, or new clinical knowledge, and (4) learners' satisfaction postintervention with the mLearning intervention.

Secondary outcomes were patient-related outcomes, change in clinical practices, economic aspects of the mLearning interventions (eg, cost-effectiveness, implementation cost, return on investment), changes in accessibility and/or availability of education and any adverse and/or unintended effects of mLearning interventions.

Data Collection and Analysis

Selection of Studies

The search results from different electronic databases were combined in a single library and duplicates were removed. A total of 2 authors (GD and CKN) independently screened titles and abstracts of all records to identify potentially relevant studies. We retrieved full-text copies of those articles deemed potentially relevant. Finally, the same 2 reviewers independently assessed the full-text versions of the retrieved articles against the eligibility criteria. Any disagreements were resolved through

discussion between the 2 screeners with a third review author (LTC) acting as an arbiter.

Data Extraction and Management

A total of 2 reviewers out of 4 (GD, CKN, LTC, and SN) independently extracted relevant characteristics related to participants, intervention, comparators, outcome measures, and results from all the included studies using a standard data collection form (see Multimedia Appendix 2). Any disagreements were resolved by discussion. We contacted study authors for any missing information.

Assessment of Risk of Bias in Included Studies

A total of 2 authors out of 3 (GD, CKN, and LTC) independently assessed the risk of bias of RCTs and cluster-RCTs using the Cochrane risk of bias assessment tool [22]. We piloted the risk of bias assessment among the reviewers and contacted study authors in case of any unclear or missing information. We assessed risk of bias in included RCTs using the following domains: random sequence generation, allocation sequence concealment, blinding (participants, personnel), blinding (outcome assessment), completeness of outcome data, selective outcome reporting, and other sources of bias (eg, baseline imbalance, inappropriate administration of an intervention, and contamination).

For cluster RCTs, we also assessed the risk of these additional biases: recruitment bias, baseline imbalance, loss of clusters, incorrect analysis, and comparability with individually randomized trials. Judgments concerning the risk of bias for each study will be classified using “yes,” “no,” or “unclear,” indicating high, low, or unclear risk of bias, respectively. We incorporated the results of the risk of bias assessment into the review using a graph and a narrative summary.

Measures of Treatment Effect

We were unable to identify a clinically meaningful interpretation of effect size in the literature for digital education interventions. Therefore, in line with other studies in the field, we present outcomes using postintervention standardized mean difference (SMD) and interpret the effect size using Cohen’s ‘rule of thumb’ (ie, with 0.2 representing a small effect, 0.5 a moderate effect, and 0.8 a large effect) [22,23]. This type of effect size interpretation has been used in previous studies [23]. For papers that reported median and range for the various outcomes, we converted this to mean and SD via the method mentioned by Wan et al [24], and then recalculated these values to provide an SMD for each outcome measure. We used the standard way to convert the results as recommended by Cochrane [22].

Data Synthesis

We aimed to present uniform postintervention data (ie, SMDs for continuous outcomes with their respective confidence intervals) to ensure consistency and comparability of data. For the meta-analysis, we used a random-effects model as different scales were used in different studies. We used the generic inverse variance method to combine cluster and noncluster RCTs of continuous outcomes. The effect estimates with corresponding 95% CIs for each study as well as a pooled effect size with 95% CI were displayed in the forest plots. We performed meta-analysis using Review Manager 5.3 (Cochrane Library Software, Oxford, UK) [25]. We adhered to the statistical approach described in the Cochrane Handbook [22].

We developed a preliminary synthesis by grouping the included studies by the type of interventions and comparators used:

1. mLearning versus traditional learning
2. Blended learning versus traditional learning
3. mLearning versus other forms of digital education

We prepared a *Summary of Findings* table to present a summary of the results and a judgment on the quality of the evidence, on the basis of the methods described in chapter 11 of the Cochrane Handbook for Systematic Reviews of Interventions [26]. Moreover, 2 authors used the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) criteria to rank the quality of the evidence using the GRADE profiler (GRADEpro) software [26].

Results

Results of the Search

Our search strategy retrieved 30,532 unique references, of these, 29 studies fulfilled inclusion criteria [27-55] (see Figure 1).

Included Studies

We included 29 studies involving 3175 participants [27-55] (see Multimedia Appendix 3 for characteristics of included studies).

A total of 25 out of 29 studies were RCTs, and the remaining 4 studies were cluster RCTs [31,32,43,47]. A total of 26 studies randomized participants into 2 groups [28-39,41-45,47-55]. Furthermore, 1 study randomized participants into 3 groups [46] and 2 studies randomized participants into 4 groups [27,40]. Participants included preregistration and postregistration health professionals. A total of 15 studies included preregistration participants, 9 involving medical students [27,29,34,38, 42,44,45,48,50], 4 studies involved nursing students [33, 40,41,52], 1 study each, involved dental [37] and physiotherapy students [36]. A total of 13 studies included postregistration health professionals [27,31,32,35,39,43,46,47,49,51,53-55]. The postregistration health professional participants included registered nurses, physicians, internal medicine residents, family medicine residents, neurosurgeon trainees, midwives, health extension workers, and trauma and critical care fellows. One study involved pre and postregistration health professionals [30] (ie, medical students and gastroenterology residents and fellows). A total of 24 studies were conducted in high-income countries [27,29-31,33-41,44-49,51-55], 13 of which were conducted in the United States. A total of 4 studies were conducted in middle-income countries, including China [32], Iran [28], Kenya [50], and Turkey [42]. Only 1 study was conducted in a low-income country [43], namely Ethiopia. No included study was published before 2006. A total of 8 studies were published between 2006 and 2013 [33,34,38,40, 44,46,53,55], whereas the remaining 21 studies (72%) were published between 2014 and 2017 (see Figures 2 and 3). For the intervention groups, 18 studies used a tablet or smartphone device to deliver the intervention [27,29-31,34,36,

37,39,41-43,45,48-52,54]; 5 studies used basic mobile phones [28,32,33,44,47]; 3 studies used iPods [38,40,53]; 2 studies used a personal digital assistant [46,55], and 1 study used MP3 players [35] (see Figure 2). Only 6 studies directly mentioned the use of learning theories in their instructional design for mobile learning [33,37,38,40,42,51], each of which used theories that are of cognitive-behaviorist pedagogy [56]. A total of 2 studies adopted the cognitive theory of multimedia to improve clinical skills [38] or knowledge [42]; 3 studies adopted cognitive learning theories such as information processing theory [33], dual coding theory [37], and adult learning theory [51], and 1 study combined cognitive theory (ie, Bloom's Taxonomy) with social constructivism [40]. The remaining 23 studies did not mention any learning theories explicitly in their reports. Most of the studies only described the teaching or instructional practices in mLearning, which lead to the change of knowledge, skills, attitudes, or satisfaction. For the control groups in the included studies, 26 studies used traditional forms of learning (eg, didactic lectures, conference, small group teaching, paper-based, standard box trainer, clinical placement, or usual learning) [28,30-45,47-55]; 2 studies used a different form of mLearning intervention (eg, limited functions) [27,46]; 1 study used another form of digital education (eg, video access to a lecture) [29]. The duration of the interventions ranged from 20 min [39] to 12 months [43,51]. One study did not report the duration of the intervention [40]. There were no studies reporting repeated interventions.

Figure 1. Study flow diagram. RCT: randomized controlled trial.

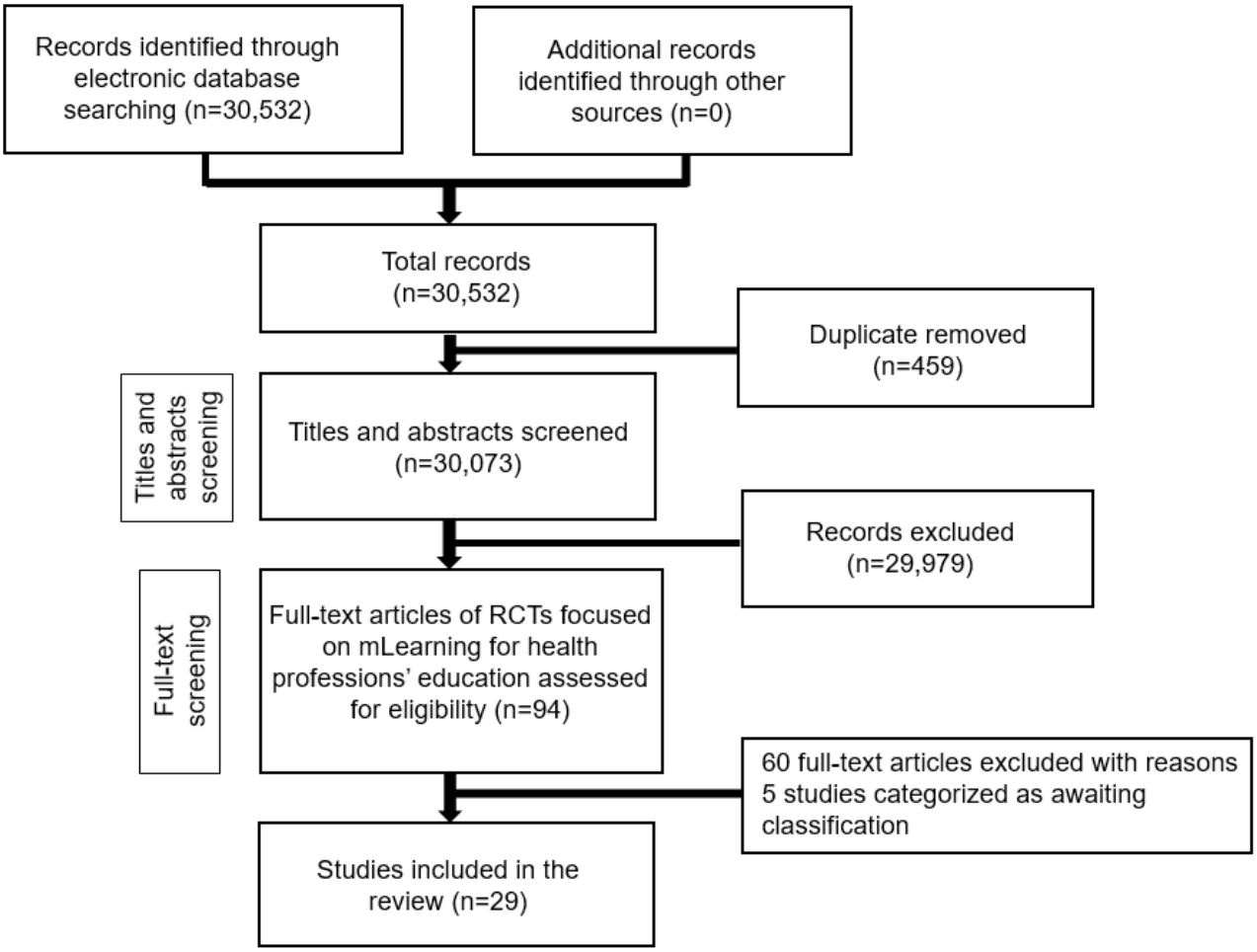


Figure 2. Number of studies by year of publication and mLearning device. PDA: personal digital assistant.

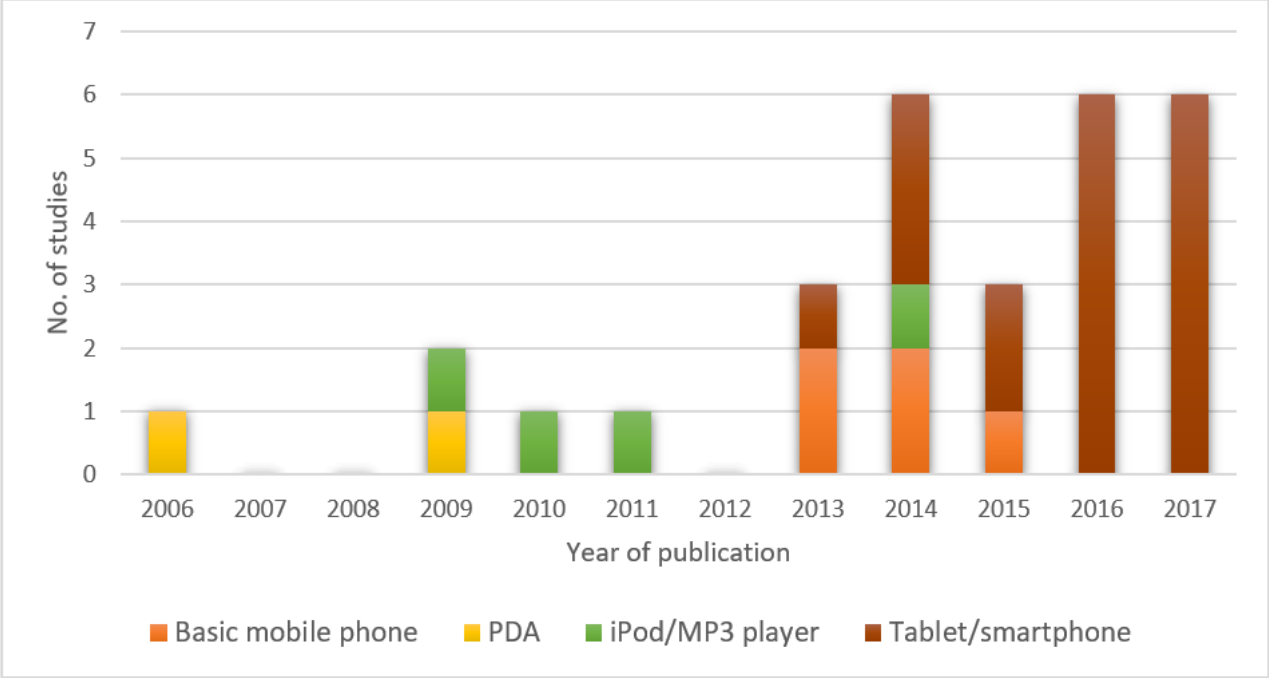


Figure 3. Country of origin of included studies for low- and middle-income and high-income countries separately. WHO: World Health Organization.

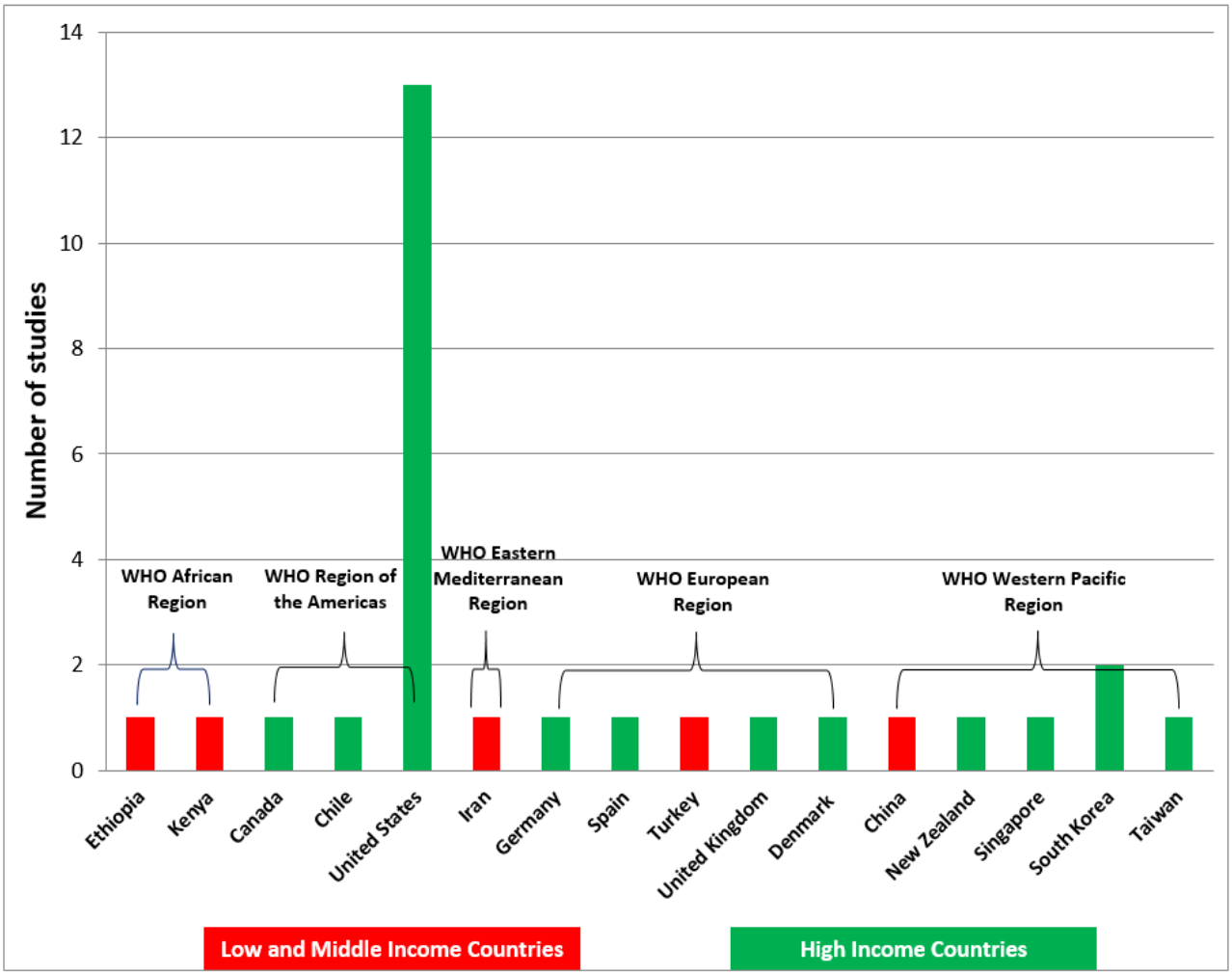
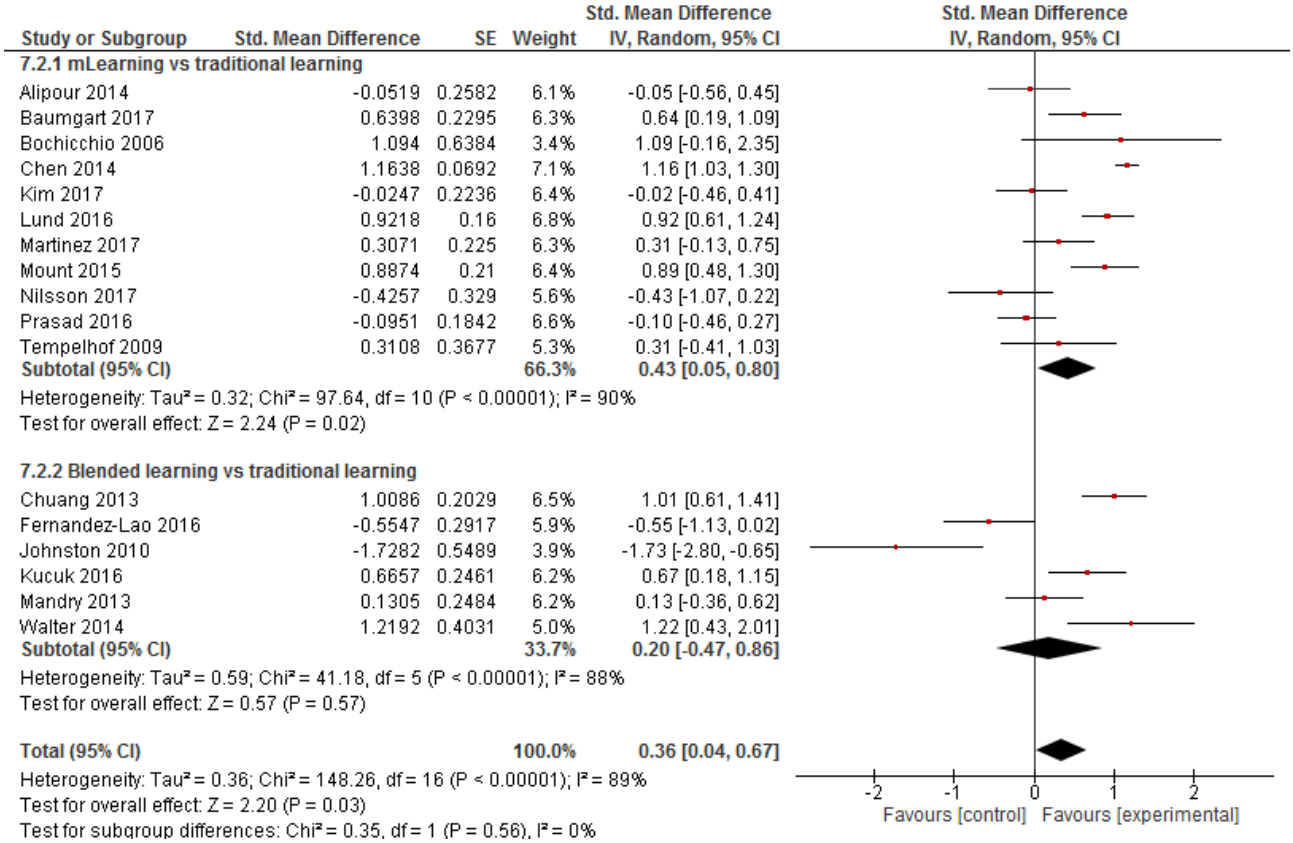


Figure 4. Forest plot for the knowledge outcome (postintervention).



Primary Outcomes

Knowledge

A total of 20 studies (n=2469) assessed knowledge posttest scores as a primary outcome [27-30,32,33,36,40-47,49,51,53-55] with 75% of studies using multiple choice questionnaires (MCQs) as their outcome measure. A total of 13 studies assessed knowledge using nonvalidated instruments [27-29,32,36,40,43,45,49,51,53-55]. A total of 5 studies stated measures that were performed to validate their measurement instruments [33,41,42,46,47], whereas 2 studies stated using standardized tests that are regularly used in medical education [30,44]. A total of 17 studies assessed knowledge posttest scores as a primary outcome immediately postintervention [27-30,32,36,40,42-46,49,51,53-55]; 1 study assessed knowledge 1 week postintervention [33]; 1 study assessed knowledge 1 month postintervention [41]; 1 study assessed knowledge 6 weeks postintervention [47]. A total of 10 studies focused on postregistration health professions education [28,32,43,46,47,49,51,53-55]. A total of 9 studies focused on preregistration health professions education [27,29,33,36,40-42,44,45], whereas the remaining 1 study included both pre and postregistration health professions education [30].

mLearning Versus Traditional Learning

A total of 11 studies compared mLearning methods versus traditional learning, assessing knowledge gain postintervention [28,30,32,41,43,45,47,49,51,53,55] (n=1828). For a summary of the effects of these comparisons on knowledge scores, see Multimedia Appendix 4.

The pooled estimate of the studies favored mLearning over traditional learning in terms of postintervention knowledge scores (SMD=0.43, 95% CI 0.05-0.80, N=11 studies, low-quality evidence; see Figure 4). There was a substantial amount of heterogeneity in the pooled analyses (I^2=90%).

Blended Learning Versus Traditional Learning

A total of 6 studies compared blended learning methods (mLearning plus traditional learning) with traditional learning to assess knowledge gain postintervention [33,36,40,42,44,54] (n=345). For a summary of the effects of these comparisons on knowledge scores, see Multimedia Appendix 5.

There was no difference between blended learning and traditional learning groups in terms of postintervention knowledge scores (SMD=0.20, 95% CI –0.47, 0.86, N=6 studies, low-quality evidence; see Figure 4). There was a substantial amount of heterogeneity in the pooled analyses (I^2=88%).

mLearning Versus mLearning

A total of 2 studies compared one form of mLearning with another form of mLearning to assess knowledge gain postintervention [27,46]. Of these, 1 study (63 participants) included 4 groups receiving varying forms of an mLearning intervention, viewing an iPad with a podcast that was either a narrated presentation for group 1, a narration with video demonstration of skills for group 2, a narrated presentation with guided mental practice for group 3, or a narrated presentation with video demonstration of skill and guided mental practice for group 4 [27]. Knowledge gain was significantly higher for both group 2 and group 3 compared with group 1 (*P*=.01; *P*=.01, respectively); knowledge gain was also significantly higher for

group 4 compared with all other groups. Furthermore, 1 study (72 participants) included 3 groups; however, only 2 of these were randomized [46]. The 2 randomized groups (38 participants) received either a basic PDA or a PDA with an additional software, a Geriatric Assessment Tool (GAT) program added. The authors reported that the PDA with additional GAT software may have little or no difference in knowledge gain postintervention compared with the basic PDA group (SMD=0.03, 95% CI –0.61 to 0.67, small effect size) [46].

mLearning Versus Another Form of Digital Education

One study (100 participants) compared mLearning with another form of digital education to assess knowledge gain postintervention [29]. The study compared an mLearning group who used the module Carpal Tunnel Surgery on the Touch Surgery app 3 times with a group who watched an audio-dubbed slide show lecture 3 times [29]. Compared with another form of digital education, mLearning was reported to improve postintervention knowledge (SMD=1.82, 95% CI 1.35-2.29, large effect size) [29].

Skills

A total of 14 studies assessed skill acquisition of mLearning interventions compared with various controls and included a total of 1097 participants [27,31,34-39,41,43,48-50,52]. A total of 11 studies used direct observation assessments to assess skills [27,31,34,36-39,41,43,49,50], 1 study used a timed quiz [48], 1 study used a survey [52], and 1 study used an MCQ to assess skills [35]. A total of 8 studies assessed cognitive skills [27,35,36,41,43,48-50], while 4 studies assessed psychomotor skills [34,37-39] and further 2 studies assessed nontechnical skills [31,52]. All 14 studies that assessed skills assessed the outcome immediately postintervention. A total of 9 studies focused on preregistration health professionals [27,34,36-38,41,48,50,52]. A total of 5 studies focused on postregistration health professionals [31,35,39,43,49].

mLearning Versus Traditional Learning

A total of 5 studies compared mLearning methods with traditional learning, assessing skill acquisition postintervention [31,35,41,43,49] (n=529). For a summary of the effects of these comparisons on skill acquisition scores, see Multimedia Appendix 4.

The pooled estimate of the studies favored mLearning over traditional learning in terms of postintervention skill acquisition (SMD=1.12, 95% CI 0.56 to 1.69, N=5 studies, moderate quality evidence; see Figure 5). There was a substantial amount of heterogeneity in the pooled analyses (I^2=87%).

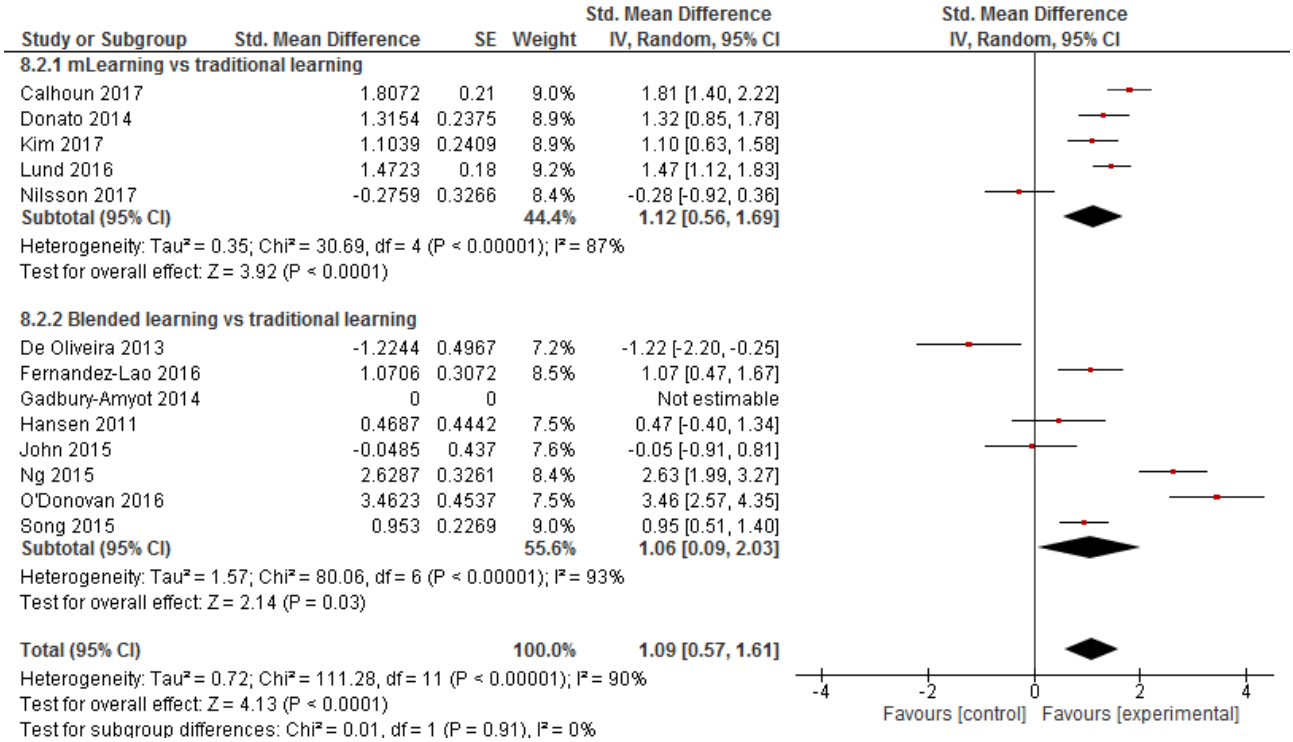
Blended Learning Versus Traditional Learning

A total of 8 studies compared blended learning methods versus traditional learning, assessing skill acquisition postintervention [34,36-39,48,50,52] (n=504). For a summary of the effects of these comparisons on knowledge scores, see Multimedia Appendix 5.

The pooled estimate of the studies favored blended learning over traditional learning in terms of postintervention skill acquisition scores (SMD=1.06, 95% CI 0.09-2.03, N=7 studies, low-quality evidence; see Figure 5). There was a substantial amount of heterogeneity in the pooled analyses (I^2=93%).

We are uncertain about the effect of 1 study (183 participants) because of incomparable outcome data [37]. However, the authors reported that blended learning may have little or no difference in dental procedural skill acquisition at postintervention compared with traditional learning [37].

Figure 5. Forest plot for the skills outcome (postintervention).



mLearning Versus mLearning

One study (63 participants) included 4 groups receiving varying forms of an mLearning intervention, as was described earlier in the review [27]. Skill acquisition was assessed using a key-elements scale, critical error checklist, and the Ottawa global rating scale (GRS), as students had to manage a manikin-based simulated airway crisis. Group 1 significantly underperformed in comparison with all other groups on the key-events scale, the critical error checklist, and the Ottawa GRS. Group 4 reported greater improvement on the key events checklist compared with group 2 and 3, but there was no difference in terms of the critical error checklist and the Ottawa GRS between the groups.

Attitude

A total of 5 studies assessed participants' attitudes following the mLearning or blended learning intervention and included a total of 440 participants [35,37,38,41,48]. All studies assessing attitude used self-report surveys. All studies assessing attitude, assessed attitude in terms of the participants' self-confidence as a result of the intervention. A total of 4 studies focused on preregistration health professionals [37,38,41,48], whereas 1 study focused on postregistration health professionals [35].

mLearning Versus Traditional Learning

A total of 2 studies comparing mLearning methods versus traditional learning assessed participants' attitude postintervention [35,41] (n=167). For a summary of the effects of these comparisons on attitudes, see Multimedia Appendix 4.

The pooled estimate of the studies favored mLearning over traditional learning in terms of postintervention attitudes (SMD=0.51, 95% CI 0.20-0.81, N=2 studies, low-quality evidence, I^2=0%).

Blended Learning Versus Traditional Learning

One study (72 participants) reported that participants in the blended learning group felt more confident in their ability to identify the anatomical structures postintervention compared with traditional learning [48]. Furthermore, 1 study (21 participants) assessed learners' postintervention attitude and reported inconclusive findings in terms of postintervention self-confidence for both male and female catheterization [38]. A further study (183 participants) assessed participants' post-intervention attitudes toward their intervention, but this was only conducted among the blended learning intervention group; therefore, we were unable to judge the effect of the interventions [37].

Satisfaction

A total of 4 studies [33,35,36,41] assessed participants' satisfaction following mLearning or blended learning interventions compared with various controls and included a total of 327 participants. A total of 2 studies assessed satisfaction with the learning method in both the intervention and control groups [35,41], whereas the remaining 2 studies only assessed satisfaction with the learning method in the intervention group [33,36]. A total of 3 studies focused on preregistration health professionals [33,36,41], whereas 1 study focused on postregistration health professionals [35].

mLearning Versus Traditional Learning

A total of 2 studies assessed participants' postintervention satisfaction scores in the mLearning interventions compared with traditional learning [35,41] (n=167).

There was no difference between mLearning and traditional learning groups in terms of postintervention satisfaction (SMD=0.39, 95% CI –0.29, 1.06, N=2 studies, very low-quality evidence). There was a substantial amount of heterogeneity in the pooled analyses (I^2=79%).

Blended Learning Versus Traditional Learning

No study assessed participants' postintervention satisfaction scores in both the blended learning intervention group and the traditional learning group. A total of 2 studies [33,36] assessed participants' post-intervention satisfaction in the blended learning intervention group only; therefore, we were unable to judge the effect of the interventions because of missing or incomparable outcome data.

Secondary Outcomes

Cost-Effectiveness

A total of 2 studies [32,49] performed an economic analysis of mLearning interventions compared with traditional learning interventions. One study performed a more thorough and comprehensive economic comparison [49]. The Programme Effectiveness and Cost Generalization model for conducting cost-effectiveness analyses [57] was used to compare the mLearning group using a mobile app with the traditional learning group using textbooks. An incremental cost-effectiveness ratio of –861.967 (95 % CI –1071.7 to –3.2) US $/pct. point change in Objective Structured Assessment of Ultrasound Skills scale score was reported indicating that traditional learning was significantly more cost-effective than the mLearning [49]. In contrast, Chen et al 2014 reported that mLearning was more cost-effective than traditional learning [32]. Short message service (SMS) text messages over 6 weeks for the intervention group cost less than 2 Yuan (US $0.32) per health worker compared with 560 Yuan (US $89.96) per health worker for the 1-day training for the control group. An additional study reported on the cost of the mLearning intervention device used, namely the "connecTAB" [50]. Each "connecTAB," which came preloaded with the intervention groups instructional videos, reportedly cost US $50.

Patient-Related Outcomes

One study reported on patient-related outcomes [43]. The primary outcome in the study was perinatal death, which was defined as a composite of a stillbirth or an early neonatal death. The mLearning intervention group, which included midwives and health extension workers, received a smartphone with the "Safe Delivery App" downloaded. The app included information and animated videos around the topic of perinatal survival. The control group engaged in standard care and did not receive an active intervention. A lower perinatal mortality of 14 per 1000 births was reported in the intervention clusters compared with 23 per 1000 births in control clusters; however, this difference was not significant. Similarly, the intervention group reported a lower stillbirth rate of 10 per 1000 births compared with 16

per 1000 births in control clusters, this difference was not statistically significant.

Changes in Clinical Practices/Behaviors

A total of 2 studies reported on changes in clinical practices/behaviors [32,55]. One study (n=479) reported on changes in antibiotic and steroid prescriptions comparing an mLearning group who received SMS text messages over a 6-week period with a traditional learning group who received standard continuous medical education [32]. In the mLearning group, there was no change in the prescription of antibiotics, whereas prescriptions for steroids fell by 5%. In contrast, for the traditional learning group, prescriptions for antibiotics and steroids increased by 17 and 11 percentage points, respectively. Antibiotic decision appropriateness was assessed in 1 study (n=12) but was only performed in the mLearning group; therefore, no comparison with the control group was possible [55]. The authors reported an improvement in antibiotic decision accuracy from 66% during the first 3 months to 86.6% during the second 3-month period. No other studies assessed secondary outcomes.

Adverse and/or Unintended Effects

None of the included studies reported any reported any adverse and/or unintended effects of the mLearning interventions.

Changes in Accessibility and/or Availability of Education

We were unable to assess the changes in accessibility and/or availability of education because of limited information in the included studies.

Sensitivity Analyses

There was not sufficient data to allow sensitivity analyses to be conducted.

Assessment of Publication Bias

There were not enough comparisons to carry out a formal assessment of publication bias.

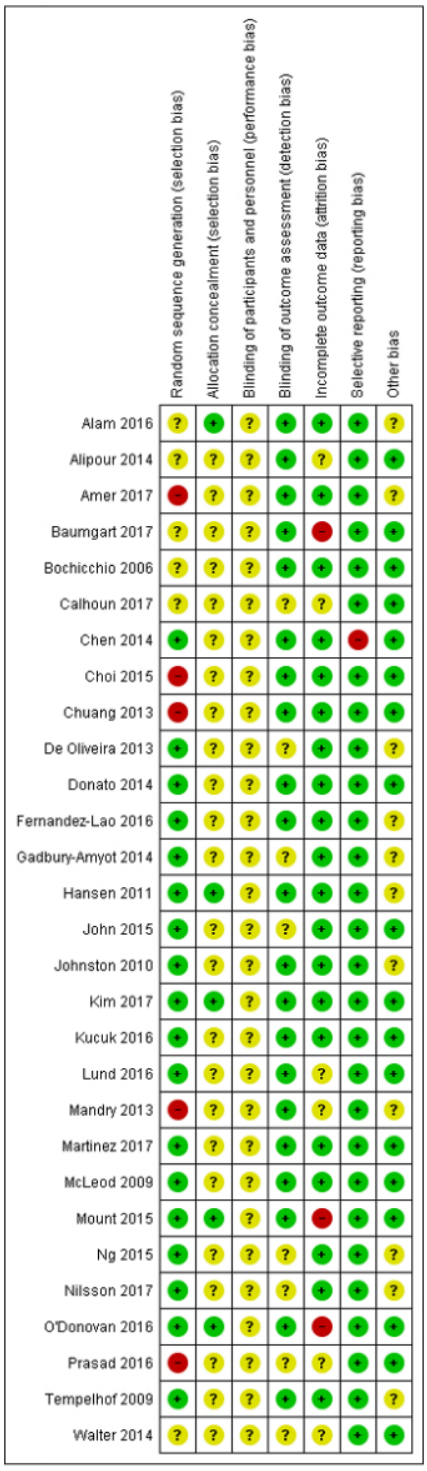
Risk of Bias in Included Studies

As presented in the risk of bias summary (Figure 6), the risk of bias was mostly judged to be high or unclear because of a lack of relevant information in the included studies (see Multimedia Appendix 6 for the risk of bias graph). We judged that the overall risk of bias was high in 16 studies (55%) as the studies had an unclear risk of bias in at least 4 out of 7 domains or a high risk in at least one domain [28-34,37,44,47-52,54]. We judged that the risk of bias was low in 2 studies (7%) because of the 2 components of selection bias being graded as low plus at least 3 of the remaining 5 domains [38,41].

More than four-fifth of studies (86%) did not provide information on the method of randomization and sequence allocation. The majority of studies (72%) reported the use of outcome measures to blind assessors or used self-report questionnaires or MCQs in the outcome assessment, which we believed did not require blinding, and thus these studies were judged to be of a low risk of bias. The remaining studies (28%) were judged to be of an unclear risk of bias because of a lack of information. A total of 3 studies (10%) were judged to be of a high risk of attrition bias as these studies had a high dropout rate (35%-73%) and/or no reasons for missing data were reported and/or lacked intention to treat analysis for the missing data, a further 6 studies (20%) were judged as unclear because of a lack of information. One study (3%) was judged to be of a high risk of bias for selective reporting, as an outcome stated in the methods section was not reported in the results section, the rest of the studies were judged to be of a low risk of bias for selective reporting. A total of 11 studies (38%) did not provide any information on a baseline assessment and were judged to be of an unclear risk of other bias.

In the clustered RCTs, only 1 study accounted for clustering reporting both individual level and cluster levels results [32], whereas there was no evidence of attrition of clusters in the studies. Additional analyses of the risk of bias for the cluster RCT are presented in Multimedia Appendix 7.

Figure 6. Risk of bias summary: review authors' judgments about each risk of bias item across all included studies.



Discussion

Principal Findings

Learning is changing and is increasingly becoming mobile. Our findings suggest that mLearning interventions are equivalent or possibly superior to traditional learning methods for improving knowledge and skill in pre and postregistration health professions education (see the summary of findings Multimedia Appendices 4 and 5). Reporting that mLearning is as effective as traditional learning has important policy implications, and we do so with caution and consideration when interpreting these findings, acknowledging the high heterogeneity among the included studies. With more than 5 billion people owning a mobile phone worldwide [58], the global reach of mLearning is evident. mLearning's potential to extend the reach of health professions training and education via mobile devices is significant.

With 21 of the 29 included studies (72%) published between 2014 and 2017, it's clear that mLearning is an emerging educational strategy. The remaining 8 studies were published

between 2006 and 2013, with no studies published before 2006, further highlighting the modern nature of this approach to health professions education and its relevance. The novel nature of mLearning, coupled with the continuing rapid advancements in mobile technology will likely see mLearning continue to evolve, as it has with the studies included in this review. This evolution is graphically illustrated in Figure 3, with PDAs being used more in the initial years, whereas for 14 out of the 15 studies from 2015 onwards, the main mode of delivery of mLearning was via smartphone/tablet devices. We identified a lack of research on mLearning in low- and middle-income countries (LMICs), with only 5 of the 29 included studies (17%) conducted in LMICs, reducing the applicability of evidence to more resource-constrained settings where the shortage of health professionals is greatest. Of note, all 5 of the studies conducted in LMICs were published between 2014 and 2016, suggesting that the field of mLearning is developing in these countries. The studies included in this review covered a variety of areas within the medical, nursing, physiotherapy, and dental field; however, the diversity in the subjects taught, outcomes measured, and the inconsistent measurement tools used in the assessments, also need consideration when interpreting our findings.

The majority of studies focused on preregistration medical and nursing students and residents undertaking specialty training. A smaller number of studies focused on practicing physicians [32,35], practicing nurses [28], physiotherapy students [36] dental students [37], midwives and health extension workers [43]. There were no studies included involving pharmacists or occupational therapists and there was a lack of studies in general among allied health professionals indicating that mLearning interventions may not be implemented as commonly outside the field of medical and nursing education. More research is needed, especially in these fields, to assess whether mLearning is applicable more broadly across the health professions field.

As mLearning may involve the use of new technology, which can entail significant financial investment, the cost of such an introduction, especially in low-income settings, is a key factor when deciding to adopt new mediums for education. Unfortunately, only 2 studies, with conflicting results, performed an economic analysis comparing mLearning with traditional learning methods. As a result, there is limited evidence to draw any definitive conclusions on the costs and cost-effectiveness of mLearning. Similarly, there was little information on patient-related outcomes or changes in clinical behaviors/practices. A further aspect of mLearning that needs addressing is the impact of mLearning on the adverse/unintended effects of mLearning on patients as well as the learner.

For the majority of the studies, the intervention was introduced as supplementary learning to their standard education, and therefore questions remain as to how effective mLearning would be when introduced to a formal curriculum. In addition, the interventions were often introduced on an optional basis, where the use and intensity of the intervention were dependent on the individual and exact exposure and effect of the intervention are unknown.

Only 6 studies mentioned the use of learning theories to inform their instructional design. This indicates a major limitation of mLearning within health professions education. There is a lack of theoretical frameworks guiding effective instructional design so that learning pathways using adopted technologies could be delineated clearly [59,60]. To meet the future needs of the evolving landscape of health professions education, emphasis should be put on training health professionals who can fulfill these needs using mobile technologies. Therefore, the first step in these efforts is to carefully investigate how to use existing pedagogical frameworks to inform the design and development of mobile learning interventions that aim to achieve desired learning outcomes. In designing mLearning interventions, considerations of feasibility in real-life contexts, scalability and, sustainability over time are important for long term success [60].

Strengths and Limitations of the Review

This review provides the most up-to-date evidence on the effectiveness of mLearning in health professions education and is supported by a comprehensive search strategy and the robust methodology that was applied at each stage of the screening, data extraction, and assessment of the evidence. This is illustrated by the larger body of evidence gathered in this review compared with previous reviews [18-21]. However, several biases may have been introduced in the review process.

Common biases include study eligibility criteria, identification and selection of studies, data extraction, and study appraisal. We tried to minimize or eliminate these biases in this review by adopting a variety of quality checks. We prespecified the eligibility criteria of the studies to be included in the review, and those were clearly defined in the protocol published before carrying out the review. This measure ensured that decisions on which studies were to be included were consistent and not based on characteristics of potentially eligible studies. The search strategy was devised and conducted by experienced librarians including all appropriate databases. As the search strategy was devised for a general project on digital education, it included Medical Subject Headings terms for several different digital education modalities. The search strategy retrieved a very large number of references. Titles and abstracts were screened independently by a team of reviewers and full-text inclusion assessment involved at least 2 reviewers. Furthermore, lead authors of the reviews of the other digital education modalities transferred studies from their reviews to other more appropriate reviews if necessary. Data availability bias may occur if some data are unavailable in the included studies and their unavailability is related to the study results. As with publication bias, this situation may lead to unrepresentative data and toward a false favorable effect. We contacted authors of studies with missing data or no data on specific outcomes to ask for those or to ask clarifications. Overall the risk of bias for most studies was judged to be high (because of a lack of information), with some instances of a high risk of bias for sequence generation, attrition, and reporting bias identified.

Reasons for downgrading the evidence included inconsistency, that is, high heterogeneity/differences in the direction of effect, with high the I^2 values reported for each of the knowledge and

skill comparisons. We did not identify a sufficient number of studies within the review comparisons to allow for the performance of subgroup analyses, which were prespecified in the protocol. With the number of studies of mLearning in health professions education continuing to rise over time, future reviews will be able to perform more focused subgroup analyses.

Future Research

The review identified gaps in evidence, which if addressed, would provide more conclusive evidence on the effectiveness and cost-effectiveness of mLearning. Further research should do the following: assume validated and standardized outcome measures, use adequately powered trials, ensure that participants are adequately trained and empowered to use the mLearning devices, be adequately reported to allow independent replications, shift toward competency-based assessments, include theoretical underpinning in instructional design, include participants from other health professions such as pharmacy and occupational therapy; be conducted in more LMICs, incorporate a more in-depth study of the various aspects of mLearning (eg, interactivity, feedback) and how each specific component affects study outcomes, provide information about the effects of mLearning on patient outcomes, provide information on cost and cost-effectiveness of mLearning, provide information on potential unintended effects of mLearning, and include both short-term and retention (follow up) outcome data.

Conclusions

mLearning is a novel educational strategy that is rapidly developing in the field of health professions education. The synthesis of data in this review shows that mobile learning is at least as or potentially more effective than traditional learning. However, the effectiveness of mLearning in health professions education is not certain because of the lack of validated and standardized outcome measures, and heterogeneity between both interventions and outcome assessments. Furthermore, there is a need for research to expand to the realm of cost-effectiveness, to fully understand the value of mLearning in health professions education. Further research is necessary to conclusively evaluate the effectiveness and cost-effectiveness of mLearning.

Acknowledgments

This review was conducted in collaboration with the Health Workforce Department at the World Health Organization. We would like to thank Mr Carl Gornitzki, Ms GunBrit Knutssön and Mr Klas Moberg from the University Library, Karolinska Institutet, Sweden, for developing the search strategy, and the peer reviewers for their comments. We also would like to thank Dr Ram Chandra Bajpai for his suggestions in data analysis. The authors gratefully acknowledge funding from the Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, eLearning for health professionals education grant, and from National Healthcare Group, Singapore.

Authors' Contributions

LTC conceived the idea for the review. GD and CN wrote the review. LTC provided methodological guidance, drafted some of the methodology-related sections and critically revised the review. SN, RA, GCYL provided comments on and edited the review.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE (Ovid) search strategy.

[DOCX File, 19KB - jmir_v21i2e12937_app1.docx]

Multimedia Appendix 2

Data extraction form.

[DOCX File, 17KB - jmir_v21i2e12937_app2.docx]

Multimedia Appendix 3

Characteristics of included studies.

[DOCX File, 23KB - jmir_v21i2e12937_app3.docx]

Multimedia Appendix 4

Summary of findings table for mLearning versus traditional learning.

[DOCX File, 15KB - jmir_v21i2e12937_app4.docx]

Multimedia Appendix 5

Summary of findings table for blended learning vs traditional learning.

[DOCX File, 15KB - jmir_v21i2e12937_app5.docx]

Multimedia Appendix 6

Risk of bias graph.

[PNG File, 9KB - jmir_v21i2e12937_app6.png]

Multimedia Appendix 7

Risk of bias for cluster randomized controlled trials.

[DOCX File, 15KB - jmir_v21i2e12937_app7.docx]

References

1. Horton R, Araujo ES, Bhorat H, Bruysten S, Jacinto CG, McPake B, et al. World Health Organization. 2016 Mar 2. Final report of the expert group to the High-Level Commission on Health Employment and Economic Growth URL: https://www.who.int/hrh/com-heeg/reports/report-expert-group/en/ [accessed 2018-10-01] [WebCite Cache ID 757WXrJwP]
2. World Health Organization. 2014. Health Workforce 2030: A global strategy on human resources for health URL: http://www.who.int/workforcealliance/knowledge/resources/strategy_brochure9-20-14.pdf [accessed 2017-12-11] [WebCite Cache ID 748Tlsq0e]
3. Nair M, Webster P. Health professionals' migration in emerging market economies: patterns, causes and possible solutions. J Public Health (Oxf) 2013 Mar;35(1):157-163 [FREE Full text] [doi: 10.1093/pubmed/fds087] [Medline: 23097260]
4. Chen LC. Striking the right balance: health workforce retention in remote and rural areas. Bull World Health Organ 2010 May;88(5):323 [FREE Full text] [doi: 10.2471/BLT.10.078477] [Medline: 20461215]
5. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. The Lancet 2010 Dec 4;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
6. World Health Organization. 2011. Transformative scale up of health professional education: an effort to increase the numbers of health professionals and to strengthen their impact on population health URL: http://apps.who.int/iris/bitstream/handle/10665/70573/WHO_HSS_HRH_HEP2011.01_eng.pdf?sequence=1&isAllowed=y [accessed 2018-11-23] [WebCite Cache ID 75kzQdnoy]
7. Crisp N, Gawanas B, Sharp I, Task Force for Scaling Up Education Training for Health Workers. Training the health workforce: scaling up, saving lives. The Lancet 2008 Feb 23;371(9613):689-691. [doi: 10.1016/S0140-6736(08)60309-8] [Medline: 18295028]
8. Sangrà A, Vlachopoulos D, Cabrera N. Building an inclusive definition of e-learning: An approach to the conceptual framework. Int Rev Res Open Dis Learn 2012 Apr 13;13(2):145-159 [FREE Full text] [doi: 10.19173/irrodl.v13i2.1161]
9. Car J, Carlstedt-Duke J, Car LC, Posadzki P, Whiting P, Zary N, et al. Digital education for health professions: methods for overarching evidence syntheses. J Med Internet Res 2019:/ (forthcoming) [FREE Full text] [doi: 10.2196/12913]
10. Gentry S, Gauthier A, Alvarez J, Wortley D, van Rijswijk J, Car J, et al. Serious gaming and gamification interventions for health professional education. Cochrane Database Syst Rev 2016 Jun 3:- [FREE Full text] [doi: 10.1002/14651858.CD012209.pub2]
11. Hervatis V, Kyaw BM, Semwal M, Dunleavy G, Tudor Car L, Zary N, et al. PROSPERO. 2016 Apr 15. Offline and computer-based eLearning interventions for medical students' education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045679 [accessed 2018-11-27] [WebCite Cache ID 75r2hfXwX]
12. Kononowicz AA, Woodham L, Georg C, Edelbring S, Stathakarou N, Davies D, et al. PROSPERO. 2016 May 19. Virtual patient simulations for health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045912 [accessed 2018-11-27] [WebCite Cache ID 75r2aDPSX]
13. Paul P, Toon E, Hadadgar A, Jirwe M, Saxena N, Lim KT, et al. Online- and local area network (LAN)-based eLearning interventions for medical doctors' education. Cochrane Database Syst Rev 2016;2016(3):CD012108 [FREE Full text] [doi: 10.1002/14651858.CD012108]
14. Tudor Car L, Riboli-Sasco EF, Marcano Belisario JS, Nikolaou CK, Majeed A, Zary N, et al. PROSPERO. 2015 Sep 15. Mobile learning for delivering health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42015029786 [accessed 2018-11-27] [WebCite Cache ID 75r28PGRf]
15. Hashemi M, Azizinezhad M, Najafi V, Nesari AJ. ScienceDirect. 2011. What is Mobile Learning? Challenges and Capabilities URL: https://ac.els-cdn.com/S1877042811023081/1-s2.0-S1877042811023081-main.pdf?_tid=3e3fe140-2b4b-4d48-b11d-4614e327322c&acdnat=1549364460_b8ebd2c926c4e2c4a0f08cf3abb34302 [accessed 2017-11-27]

16. Crompton H. A historical overview of mobile learning: toward learner-centered education. In: Berge ZL, Muilenburg L, editors. Handbook of Mobile Learning. London: Routledge; 2013:3-14.
17. Ronchetti M, Trifonova A. Where is mobile learning going? San Diego, CA: Association for the Advancement of Computing in Education (AACE); 2003 Presented at: Proceedings of E-World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education; 2003; Phoenix, Arizona p. 1794-1801.
18. 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 Jan 15;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
19. Guo P, Watts K, Wharrad H. An integrative review of the impact of mobile technologies used by healthcare professionals to support education and practice. Nurs Open 2015 Nov 27;3(2):66-78 [FREE Full text] [doi: 10.1002/nop2.37] [Medline: 27708817]
20. Kho A, Henderson LE, Dressler DD, Kripalani S. Use of handheld computers in medical education. A systematic review. J Gen Intern Med 2006 May;21(5):531-537 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00444.x] [Medline: 16704405]
21. Siddiqui ZS, Jonas-Dwyer DR. Mobile learning in health professions education: a systematic review. In: Pedagogical Applications Social Effects of Mobile Technology Integration. Pennsylvania: IGI Global; 2013:193-204.
22. Higgins JP, Green S. The Cochrane Collaboration. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] URL: http://crtha.iums.ac.ir/files/crtha/files/cochrane.pdf [accessed 2019-01-29] [WebCite Cache ID 75mWykNDL]
23. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. J Am Med Assoc 2011 Sep 7;306(9):978-988. [doi: 10.1001/jama.2011.1234] [Medline: 21900138]
24. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
25. Cochrane Community. 2014. Review Manager (RevMan) Version 5.3 URL: https://community.cochrane.org/help/tools-and-software/revman-5 [accessed 2019-02-05] [WebCite Cache ID 75xTpZsGT]
26. Schünemann HJ, Oxman AD, Higgins JP, Vist GE, Glasziou P, Guyatt GH. Cochrane Handbook for Systematic Reviews of Interventions.: The Cochrane Collaboration; 2011 Mar. Chapter 11: Presenting Results and 'Summary of Findings' Tables URL: https://handbook-5-1.cochrane.org/chapter_11/11_presenting_results_and_summary_of_findings_tables.htm [accessed 2018-11-01] [WebCite Cache ID 75eVmp3hx]
27. Alam F, Boet S, Piquette D, Lai A, Perkes CP, LeBlanc VR. E-learning optimization: the relative and combined effects of mental practice and modeling on enhanced podcast-based learning-a randomized controlled trial. Adv Health Sci Educ Theory Pract 2016 Oct;21(4):789-802. [doi: 10.1007/s10459-016-9666-9] [Medline: 26846221]
28. Alipour S, Jannat F, Hosseini L. Teaching breast cancer screening via text messages as part of continuing education for working nurses: a case-control study. Asian Pac J Cancer Prev 2014 Jul 30;15(14):5607-5609. [doi: 10.7314/APJCP.2014.15.14.5607] [Medline: 25081673]
29. Amer KM, Mur T, Amer K, Ilyas AM. A mobile-based surgical simulation application: a comparative analysis of efficacy using a carpal tunnel release module. J Hand Surg Am 2017 May;42(5):389e1-389e9. [doi: 10.1016/j.jhsa.2017.02.008] [Medline: 28318742]
30. Baumgart DC, Wende I, Grittner U. Tablet computer enhanced training improves internal medicine exam performance. PLoS One 2017 Apr;12(4):e0172827 [FREE Full text] [doi: 10.1371/journal.pone.0172827] [Medline: 28369063]
31. Calhoun AW, Sutton ER, Barbee AP, McClure B, Bohnert C, Forest R, et al. Compassionate Options for Pediatric EMS (COPE): addressing communication skills. Prehosp Emerg Care 2017 Jan 19;21(3):334-343. [doi: 10.1080/10903127.2016.1263370] [Medline: 28103120]
32. Chen Y, Yang K, Jing T, Tian J, Shen X, Xie C, et al. Use of text messages to communicate clinical recommendations to health workers in rural China: a cluster-randomized trial. Bull World Health Organ 2014 Jul 1;92(7):474-481 [FREE Full text] [doi: 10.2471/BLT.13.127076] [Medline: 25110372]
33. Chuang YH, Tsao CW. Enhancing nursing students' medication knowledge: the effect of learning materials delivered by short message service. Comput Educ 2013 Feb;61:168-175. [doi: 10.1016/j.compedu.2012.09.013]
34. de Oliveira GS, Glassenberg R, Chang R, Fitzgerald P, McCarthy RJ. Virtual airway simulation to improve dexterity among novices performing fibreoptic intubation. Anaesthesia 2013 Oct;68(10):1053-1058 [FREE Full text] [doi: 10.1111/anae.12379] [Medline: 23952805]
35. Donato AA, Kaliyadan AG, Wasser T. Self-directed study using MP3 players to improve auscultation proficiency of physicians: a randomized, controlled trial. J Contin Educ Health Prof 2014;34(2):131-138. [doi: 10.1002/chp.21220] [Medline: 24939355]
36. Fernández-Lao C, Cantarero-Villanueva I, Galiano-Castillo N, Caro-Morán E, Díaz-Rodríguez L, Arroyo-Morales M. The effectiveness of a mobile application for the development of palpation and ultrasound imaging skills to supplement the traditional learning of physiotherapy students. BMC Med Educ 2016 Oct 19;16(1):274 [FREE Full text] [doi: 10.1186/s12909-016-0775-1] [Medline: 27756288]

37. Gadbury-Amyot CC, Purk JH, Williams BJ, van Ness CJ. Using tablet technology and instructional videos to enhance preclinical dental laboratory learning. J Dent Educ 2014 Feb;78(2):250-258 [FREE Full text] [Medline: 24489032]
38. Hansen M, Oosthuizen G, Windsor J, Doherty I, Greig S, McHardy K, et al. Enhancement of medical interns' levels of clinical skills competence and self-confidence levels via video iPods: pilot randomized controlled trial. J Med Internet Res 2011 Mar 1;13(1):e29 [FREE Full text] [doi: 10.2196/jmir.1596] [Medline: 21447471]
39. John NW, Phillips NI, Cenydd LA, Coope D, Carleton-Bland N, Kamaly-Asl I, et al. A tablet-based virtual environment for neurosurgery training. Presence (Camb) 2015;24(2):155-162. [doi: 10.1162/PRES_a_00224]
40. Johnston R, Hepworth J, Goldsmith M, Lacasse C. Use of iPod™ technology in medical-surgical nursing courses: effect on grades. Int J Nurs Educ Scholarsh 2010 Nov 30;7(1):e43. [doi: 10.2202/1548-923X.2092] [Medline: 21126230]
41. Kim SJ, Shin H, Lee J, Kang S, Bartlett R. A smartphone application to educate undergraduate nursing students about providing care for infant airway obstruction. Nurse Educ Today 2017 Jan;48:145-152. [doi: 10.1016/j.nedt.2016.10.006] [Medline: 27810633]
42. Küçük S, Kapakin S, Göktaş Y. Learning anatomy via mobile augmented reality: effects on achievement and cognitive load. Anat Sci Educ 2016 Oct;9(5):411-421. [doi: 10.1002/ase.1603] [Medline: 26950521]
43. 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]
44. Mandry C, White ML, Hayes A, Hoonpongsimanont W, Moreno Walton L, Schiller JG, et al. Text messaging enhanced curriculum for medical student emergency medicine clerkship. 2013 Presented at: Annual meeting of the Society for Academic Emergency Medicine (SAEM); May 14-18, 2013; Atlanta, Georgia, US p. S208.
45. Martínez F, Tobar C, Taramasco C. Implementation of a smartphone application in medical education: a randomised trial (iSTART). BMC Med Educ 2017 Sep 18;17(1):168 [FREE Full text] [doi: 10.1186/s12909-017-1010-4] [Medline: 28923048]
46. Mcleod TG, McNaughton DA, Hanson GJ, Cha SS. Educational effectiveness of a personal digital assistant-based geriatric assessment tool. Med Teach 2009 May;31(5):409-414. [doi: 10.1080/01421590802216241] [Medline: 18937094]
47. Mount H, Zakrajsek T, Huffman M, Deffenbacher B, Gallagher K, Skinker B, et al. Text messaging to improve resident knowledge: a randomized controlled trial. Fam Med 2015 Jan;47(1):37-42 [FREE Full text] [Medline: 25646876]
48. Ng CL, Liu X, Chee SC, Ngo RY. An innovative 3-dimensional model of the epitympanum for teaching of middle ear anatomy. Otolaryngol Head Neck Surg 2015 Nov;153(5):832-837. [doi: 10.1177/0194599815584600] [Medline: 25994233]
49. Nilsson PM, Todsen T, Subhi Y, Graumann O, Nolsøe CP, Tolsgaard MG. Cost-effectiveness of mobile app-guided training in extended focused assessment with sonography for trauma (eFAST): a randomized trial. Ultraschall Med 2017 Dec;38(6):642-647. [doi: 10.1055/s-0043-119354] [Medline: 28950381]
50. O'Donovan J, Ahn R, Nelson BD, Kagan C, Burke TF. Using low-cost Android tablets and instructional videos to teach clinical skills to medical students in Kenya: a prospective study. JRSM Open 2016 Aug;7(8):1-7 [FREE Full text] [doi: 10.1177/2054270416645044] [Medline: 27540487]
51. Prasad M, Waterbury T, Kebed K, Bois J, Killu A, Klarich K, et al. A novel, mobile, electronic curriculum based on adult learning theory is noninferior to a traditional didactic lecture-based curriculum. 2016 Aug 29 Presented at: ESC Congress 2016; August 27-31, 2016; Rome, Italy p. 1265. [doi: 10.1093/eurheartj/ehw434]
52. Choi Y, Song E, Oh E. Effects of teaching communication skills using a video clip on a smart phone on communication competence and emotional intelligence in nursing students. Arch Psychiatr Nurs 2015 Apr;29(2):90-95. [doi: 10.1016/j.apnu.2014.11.003] [Medline: 25858200]
53. Tempelhof MW, Garman KS, Langman MK, Adams MB. Leveraging time and learning style, iPod vs realtime attendance at a series of medicine residents conferences: a randomised controlled trial. Inform Prim Care 2009;17(2):87-94 [FREE Full text] [doi: 10.14236/jhi.v17i2.720] [Medline: 19807950]
54. Walter RJ, Mallia AK, Massoumi A, Hartzell J. Mobile-centered spaced learning: A novel approach to graduate medical education in the digital era. In: J Gen Intern Med. 2014 Presented at: 37th Annual Meeting of the Society of General Internal Medicine; April 23-26, 2014; San Diego, California p. S526-S527 URL: http://paperity.org/p/53794699/abstracts-from-the-37th-annual-meeting-of-the-society-of-general-internal-medicine
55. Bochicchio GV, Smit PA, Moore R, Bochicchio K, Auwaerter P, Johnson SB, POC-IT Group. Pilot study of a web-based antibiotic decision management guide. J Am Coll Surg 2006 Mar;202(3):459-467. [doi: 10.1016/j.jamcollsurg.2005.11.010] [Medline: 16500251]
56. Anderson T, Dron J. Three generations of distance education pedagogy. Int Rev Res Open Dist Learn 2011 Mar 25;12(3):80-97 [FREE Full text] [doi: 10.19173/irrodl.v12i3.890]
57. Tolsgaard MG, Tabor A, Madsen ME, Wulff CB, Dyre L, Ringsted C, et al. Linking quality of care and training costs: cost-effectiveness in health professions education. Med Educ 2015 Dec 1;49(12):1263-1271 [FREE Full text] [doi: 10.1111/medu.12882] [Medline: 26611191]
58. GSM Association. 2018. The Mobile Economy 2018 URL: https://www.gsma.com/mobileeconomy/wp-content/uploads/2018/05/The-Mobile-Economy-2018.pdf [accessed 2018-11-23] [WebCite Cache ID 748UD5Sey]

59. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
60. Park Y. A pedagogical framework for mobile learning: categorizing educational applications of mobile technologies into four types. Int Rev Res Open Dist Learn 2011 Feb 28;12(2):78-102 [FREE Full text] [doi: 10.19173/irrodl.v12i2.791]

Abbreviations

GAT: Geriatric Assessment Tool
GPRS: general packet radio services
GRS: global rating scale
GSM: global system for mobile communications
LMIC: low- and middle-income countries
MP3: Moving Picture Experts Group-1 audio layer 3
MCQ: multiple choice questionnaire
PDA: personal digital assistant
RCT: randomized controlled trial
SMD: standardized mean difference
SMS: short message service

Edited by A Marusic; submitted 29.11.18; peer-reviewed by O Polasek, M Mahmic Kaknjo, S Peters; comments to author 20.12.18; revised version received 15.01.19; accepted 17.01.19; published 12.02.19.

Please cite as:
Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GCY, Tudor Car L
Mobile Digital Education for Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration
J Med Internet Res 2019;21(2):e12937
URL: http://www.jmir.org/2019/2/e12937/
doi:10.2196/12937
PMID:30747711

Viewpoint

Digital Education in Health Professions: The Need for Overarching Evidence Synthesis

Josip Car[1], MD, PhD, FRCPE, FFPH; Jan Carlstedt-Duke[2], MD, PhD; Lorainne Tudor Car[3,4], MD, PhD; Pawel Posadzki[5], PhD; Penny Whiting[6], PhD; Nabil Zary[7,8,9], MD, PhD; Rifat Atun[10,11], MBBS, FRCP; Azeem Majeed[12], MBBS, MD; James Campbell[13], MPH, MSc; Digital Health Education Collaboration

[1]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[2]Nanyang Institute of Technology in Health and Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[3]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[4]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom
[5]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[6]Bristol Medical School, University of Bristol, Bristol, United Kingdom
[7]Medical Education Research and Scholarship Unit, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[8]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden
[9]International Medical Simulation Centre, Mohammed VI University of Health Sciences, Casablanca, Morocco
[10]Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States
[11]Harvard Medical School, Harvard University, Boston, MA, United States
[12]Department of Primary Care & Public Health, School of Public Health, Imperial College London, London, United Kingdom
[13]Health Workforce Department, World Health Organization, Geneva, Switzerland

Corresponding Author:
Josip Car, MD, PhD, FRCPE, FFPH
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road
Singapore,
Singapore
Phone: 65 6340 2480
Email: josip.car@ntu.edu.sg

Abstract

Synthesizing evidence from randomized controlled trials of digital health education poses some challenges. These include a lack of clear categorization of digital health education in the literature; constantly evolving concepts, pedagogies, or theories; and a multitude of methods, features, technologies, or delivery settings. The Digital Health Education Collaboration was established to evaluate the evidence on digital education in health professions; inform policymakers, educators, and students; and ultimately, change the way in which these professionals learn and are taught. The aim of this paper is to present the overarching methodology that we use to synthesize evidence across our digital health education reviews and to discuss challenges related to the process. For our research, we followed Cochrane recommendations for the conduct of systematic reviews; all reviews are reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidance. This included assembling experts in various digital health education fields; identifying gaps in the evidence base; formulating focused research questions, aims, and outcome measures; choosing appropriate search terms and databases; defining inclusion and exclusion criteria; running the searches jointly with librarians and information specialists; managing abstracts; retrieving full-text versions of papers; extracting and storing large datasets, critically appraising the quality of studies; analyzing data; discussing findings; drawing meaningful conclusions; and drafting research papers. The approach used for synthesizing evidence from digital health education trials is commonly regarded as the most rigorous benchmark for conducting systematic reviews. Although we acknowledge the presence of certain biases ingrained in the process, we have clearly highlighted and minimized those biases by strictly adhering to scientific rigor, methodological integrity, and standard operating procedures. This paper will be a valuable asset for researchers and methodologists undertaking systematic reviews in digital health education.

(J Med Internet Res 2019;21(2):e12913) doi:10.2196/12913

KEYWORDS

methods; education, medical; systematic reviews; evidence-based; education, distance; education, professional

Background

A global shortfall of 18 million health workers has been estimated by 2030 [1]. This shortage is recognized as an important obstacle to the achievement of universal health coverage, which ensures that all people have access to health services of high quality without the risk of financial hardship [2]. Digital health education has been identified as one of the potential means of addressing these growing challenges and is seeing an increasing adoption at all levels from primary, secondary, and pre- and postgraduate university education to life-long learning and continuous professional development. This perpetuating trend, partially driven by advances in science and technology and rising consumer demand is also seen in education for health professionals [3], where digital technologies are ubiquitous and diverse. In line with the growing adoption and innovations in digital health education, literature on the effectiveness of digital health educational interventions has seen a rapid growth over the last two decades [4,5]. This literature encompasses a wide range of digital education interventions delivered in a variety of settings. Furthermore, it includes diverse health professionals, a multitude of comparisons, several interventions, and a range of different outcome measures. Despite a growing number of trials and systematic reviews in the area of digital health education, there is a lack of conceptual clarity and robust evidence-based recommendations for many of the existing digital health education modalities.

The Digital Health Education Collaboration has been established as an international initiative in a quest for effective digital education interventions for health professionals. It is a response to a growing shortage of health professionals worldwide [6] and aims at providing robust evidence to support the transformation of education for health professionals [7,8] through the use of digital technology. The initiative is driven by a global need for reliable recommendations for health professions education.

We aim to address an important gap in the evidence by undertaking a series of focused, high-quality, methodologically robust systematic reviews on the effectiveness of digital interventions in health professions education, focusing primarily on evidence from randomized controlled trials (RCTs). This approach is complemented by evidence syntheses of studies deploying other designs such as qualitative research. We built upon the Cochrane systematic review methodology to develop a tailored, tried-and tested approach to systematic reviews of digital health education literature [9].

Given the relatively recent emergence of and continuous innovation in digital health education, the literature in this academic field is fraught with many challenges. The objective of this study is to provide a detailed description of the methodology developed to tackle these challenges, which we have applied across a number of systematic reviews [10-25]. We present our comprehensive search strategies; explain eligibility criteria in terms of populations, interventions, comparator groups, outcome measures, and study designs; discuss our literature-screening processes, data extraction and management, and risk of bias assessment; outline our approach to data synthesis, analysis, and visualization; and highlight some of the challenges we faced.

Digital Health Education Collaboration for Evidence Synthesis

Our collaboration involves a range of experts including educationalists, content experts, digital technology experts, methodological experts, information specialists, and statisticians. It also includes many authors of systematic reviews. We report our reviews in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines [9]. Correspondingly, we first developed and published or registered protocols and then embarked on systematic reviews [13,15-19,23]. We present our methodological approach below.

Defining the Scope

Since there are a number of, and at times, disorganized, definitions of digital health education in the literature [26], we attempted to develop our own definitions and conceptual framework by scoping the literature; creating standard operating procedures; and performing multiple consultations, discussions, and meetings with the aforementioned field experts. Given the nature of our approach to classifying the modalities and the rapid evolution of evidence in this field, we recognize that our and others' classifications will evolve as new evidence becomes available. We present the definitions of digital, traditional, and blended education in Textbox 1. Potential advantages of digital education may include ubiquitous delivery and flexible access to learning content, personalization of learning experience, better sensation of content, deeper information processing, adaptability, enhanced collaboration capacities, increased motivation and enjoyment of learning, cost-effectiveness, scalability, and equity. On the other hand, there are potential disadvantages such as implementation restrictions caused by digital divide (requirement of information technology infrastructure and digital literacy); additional development and set-up costs; and untoward effects of digital education such as anxiety, dizziness, and isolation. We were unable to find a robust framework that would capture and describe the variety of digital education applications in health, particularly in relation to the employed technology, which we termed "modalities." Through discussion with educational experts within our teams and review of relevant literature, we identified the following modalities: mobile learning (m-learning) or mobile digital education [13,24], virtual reality [19,22], virtual patient [16], serious gaming and gamification [15,27], offline or online digital education [17,18], massive open online courses [28], digital psychomotor skills

trainers [29], and virtual learning environment [30]. Operational definitions for these modalities are presented in Table 1.

In line with this classification, we performed a series of systematic reviews focusing on each modality (except massive open online courses, digital psychomotor skills trainers, and virtual learning environment [11,13,15-19,23,25,27]). In addition, we adopted alternative approaches to literature synthesis, keeping in mind the needs of various stakeholders in health professions education [34]. We have focused our evidence synthesis on digital education for a specific topic or condition (eg, smoking cessation [20], diabetes [21], domestic violence, and antibiotic management), discipline (eg, dermatology [10], pediatrics, geriatrics, leadership, and management [14]), roles of health professionals (eg, medical doctors [18], medical students [17], and preregistration [undergraduate] and postregistration [postgraduate] health professionals [11]), pedagogical foundations with relevant technology applications (eg, digital problem-based learning [35]), and type of outcome (eg, cognitive skills such as communication or diagnostic skills) [12].

Textbox 1. Definitions of digital, traditional, and blended education.

- *Digital education* (also known as electronic learning or digital learning) is the act of teaching and learning by means of digital technologies. It is an overarching term for an evolving multitude of educational approaches, concepts, methods, and technologies. Digital education can be further characterized by specific pedagogies and instructional methods, contexts of provision, and technical affordances of hardware and software. Modalities of digital education range from the basic conversion of content into a digital format (eg, a book into a PDF or HTML format) to complex deployment of digital technologies (eg, mobile education, serious games, virtual patients, and virtual reality).
- *Traditional education* is the act of any teaching and learning based on nondigital educational material (eg, textbook or model) or in-person human interaction (eg, teacher or other learner). Traditional education in form of in-person human interaction can also include nondigital and digital educational aids such as images, charts, maps, objects, boards, and videos.
- *Blended education* is the act of teaching and learning, which integrates aspects of traditional and digital education. Blended education can take on diverse formats depending on the type and share of digital and traditional education employed in the blended educational approach. The digital component of blended learning includes online learning as well as the use of other digital education modalities. Nonetheless, education delivered via in-person human interaction supported by digital educational aids (eg, images, charts, maps, objects, and boards) is considered traditional education and not blended education.

Table 1. Description of the digital modalities.

Digital education modality	Working definitions/description
Offline computer-based digital education (offline digital education)	An intervention that requires no internet or local area network connection and can be delivered through media including CD-ROM, external hard disc, and universal serial bus stick [17].
Online computer-based digital education (online digital education)	An intervention that requires the use of a "Transmission Control Protocol" and an "Internet Protocol" as standards for learning activities. Alternatively, these may be referred to as being "online," "Web-based," or "on a network" [18].
Serious gaming and gamification interventions	A competitive activity in which students set educational goals intended to promote knowledge acquisition. The games may either be designed to promote learning or the development of cognitive skills, or take the form of simulations that allow learners to practice their skills in a virtual environment [31].
Massive open online course	An online course that is designed for participation of large numbers of geographically dispersed students [28].
Virtual learning environment	An environment that is based on a certain pedagogical model, incorporates or implies one or more didactic objectives, provides users with experiences they would otherwise not be able to experience in the physical world, and rebounds specific learning outcomes [30].
Virtual reality	A computer-generated representation of a real or artificial environment that can be interacted with by external entities, allowing for a first-person active-learning experience through immersion [19].
Virtual patient	"Interactive computer simulations of real-life clinical scenarios for the purpose of medical training, education, or assessment" [16].
Digital psychomotor skills trainers	An intervention in which digital technologies are utilized to train skills belonging to the psychomotor domain; mental and motor activities are required to execute a manual task [29]. Examples include high-fidelity mannequins; virtual reality using probes; and laparoscopy, otolaryngoscopy, endoscopy, ureteroscopy, cystoscopy simulators, or robotic surgery [32].
Mobile digital education (m-learning)	"Learning across multiple contexts, through social and content interactions, using personal electronic devices" [33].

Inclusion Criteria

For all reviews, we adopted the following general inclusion criteria in addition to review-specific criteria, as appropriate.

Types of Studies

We aimed to evaluate the highest-quality evidence on the effectiveness of digital education interventions in education for health professionals [36]. Therefore, we only considered individually, cluster-, or quasi-randomized controlled trials eligible for inclusion in these reviews [10-25]. We excluded crossover trials including those with a stepped-wedge design due to the high likelihood of carry-over effect.

Types of Participants

When defining eligible health professionals, we used the Education and Training (091) criteria from the Health Field of the International Standard Classification of Education [37]. We considered eligible candidates for, and holders of, the qualifications listed with the exclusion of students and practitioners of traditional, alternative, and complementary medicine. We therefore included students from the following disciplines: all health professionals working in health care settings (hospice, hospitals, clinics, and community health centers) in medicine, nursing and midwifery, dental studies, medical diagnostic and treatment technology, therapy and rehabilitation, and pharmacy. More specifically, we included physio/occupational therapists, pharmacists, radiographers, radiotherapists, paramedics, environmental and occupational health and hygiene professionals, audiologists, speech therapists, nutritionists/dieticians, medical/nuclear medicine technologists, optometrists/opticians, public health staff, community health agents, and any health care educators/counsellors. Studies were considered eligible if participants were enrolled in any of the following programs: (1) A preregistration, undergraduate, health-related university degree or a basic, health-related vocational training program defined as any type of study leading to a qualification that is recognized by the relevant governmental or professional bodies of the country where the studies were conducted and entitles the qualification holder to apply for entry-level positions in the health care workforce or have direct contact with patients. For this reason, graduate medical education courses will be included in this category. (2) A postregistration health professional educational program, defined as any type of study that enables the qualification holder entry into or continuation of work in the health care workforce in a more independent or senior role. Continued professional development and continued medical education [38] are essential for postregistration health professionals to stay up-to-date with the latest advancements and therefore were considered eligible. We defined continued medical education as "all educational activities that serve to maintain, develop, or increase the knowledge, skills, and professional performance and relationships that a physician uses to provide services for patients, the public, or the profession" [39] and continued professional development as "a range of learning activities through which health and care professionals maintain and develop throughout their career to ensure that they retain their capacity to practice safely, effectively and legally within their evolving scope of practice" [40].

Participants were not excluded based on age, sex, or any other sociodemographic characteristic. However, in some reviews, we focused on a certain group of participants such as medical doctors [18], medical students [17], or allied health professionals [25]. This was due to the unmanageable number of RCTs identified during the initial scoping searches of the literature.

Types of Interventions

We included studies in which digital education was used to deliver the learning content of a course related to health education. Studies of blended learning, which represents a continual convergence between traditional and digital education, were eligible. Studies that use digital education in patients, consumers, or lay health workers were excluded. The operational definitions of individual digital education modalities are presented in Table 1.

We included studies that made any of the following intervention comparisons: (1) digital education or blended learning compared to traditional learning (eg, face-to-face learning, one-to-one learning, classroom-based learning, or self-directed learning), (2) digital education or blended learning intervention compared to another form of digital education, or (3) digital health education or blended learning intervention compared to no intervention.

Types of Outcome Measures

The most appropriate outcomes and tools to measure those outcomes in educational digital health education trials are currently under debate [41]. The selection and classification of outcomes in our reviews were aligned with the Miller classification of clinical competence, which differentiates the following levels of clinical competence: "knows," "knows how," "shows how," and "does" [42]. The different types of tests for health professionals' knowledge and skills were grouped and analyzed together. For example, multiple-choice questions assessing knowledge (ie, "knows") were analyzed together and essay questions assessing competence (ie, "knows how") were analyzed together. This framework also specifies the type of measurement tools used to assess these different outcomes. We focused on the testing method rather than the delivery method (ie, if skills were assessed by a knowledge test, we categorized them as "knowledge"). Correspondingly, we mapped these levels of competencies to outcomes reported in the included studies. We decided on the following primary outcomes and their definitions: learners' postintervention knowledge, defined as learners' factual or conceptual understanding; learners' postintervention skills, defined as learners' ability to demonstrate a procedure or technique in an educational setting; learners' postintervention attitudes toward the digital education intervention, defined as an unobservable psychological construct, which can manifest itself in relevant beliefs, feelings, and behavioral components; learners' postintervention satisfaction with the digital health education intervention, defined as the level of approval when comparing perceived performance in digital health education with one's expectations; learner's postintervention change in behavior or clinical practice

(eg, reduced prescription of antibiotics, improved diagnosis, and improved quality of care); and patient-related outcomes (only for interventions delivered to postregistration learners), defined as the results of a clinical intervention obtained by the patient.

We believe that knowledge, skills, and attitudes combined together ultimately form professional competencies. We also included the following secondary outcomes: economic outcomes (eg, cost and cost-effectiveness of the interventions), adverse or unintended effects of digital health education (on both patients and learners; eg, patient mortality, patient morbidity, medical errors, addiction, and dizziness), and self-efficacy measured as self-rated competence of health professionals in delivering a treatment or therapy.

We included outcome data for all specified outcomes measured using both validated and nonvalidated instruments. If multiple measures of the same outcome were reported, we selected the primary outcome as defined by the authors. In case this was not specified in the study, we used the measurement that was the most consistent with outcomes reported in other studies. Another alternative was to calculate the mean value of all measures. For papers that reported median and range of the outcomes, we converted these values to mean and SD using the methods described by Wan [43]. If a study did not report SD but provided confidence intervals, we estimated SD from studies using a previously described method [9].

Search Methods for Identification of Studies

Our aim was to develop a highly sensitive search strategy that would capture all relevant studies. An experienced team of librarians/information specialists from the Karolinska Institutet developed and piloted the search strategy. We performed regular, yearly updates of our searches.

We searched the following databases: MEDLINE (Ovid), Embase (Elsevier), The Cochrane Central Register of Controlled Trials (CENTRAL; Wiley), PsychINFO (Ovid), Educational Resource Information Centre (ERIC; Ovid), Cumulative Index to Nursing and Allied Health Literature (CINAHL; Ebsco), and Web of Science Core Collection (Thomson Reuters). We used the MEDLINE strategy and keywords presented in the appendix of studies, which were adapted to search the other databases. Databases were searched from January 1990 to August 2017. We selected 1990 as the starting year for our search because prior to this year, the use of the computers was limited to very basic tasks. We searched for and included papers in any language. Our searches were focused around three major topics: effectiveness, digital technologies, and educational outcomes of health professionals. We also searched two trial registers: International Clinical Trials Registry Platform and metaRegister of Controlled Trials. We screened the reference lists of all eligible studies and relevant systematic reviews to identify additional relevant studies.

We implemented the search strategy and imported all identified references into the reference-management software (EndNote, Version X8, Clarivate Analytics, Philadelphia, PA). The search results from different electronic databases were combined, and duplicate records of the same studies were removed.

Selection of Studies

We developed and piloted a decision tree with the main inclusion criteria and operational definitions to assist with our screening process. We screened references in two steps to ensure maximum sensitivity and specificity. Two reviewers independently screened titles and abstracts for eligibility. We retrieved the full texts of all articles that appeared eligible for inclusion. Two reviewers independently assessed the full text of the retrieved articles for compliance with our inclusion and exclusion criteria. Any disagreements were resolved through discussion between the two authors. If no agreement could be reached, we consulted a third author. Studies that appeared to be relevant but were excluded at this stage were listed in the "characteristics of excluded studies" tables with the reasons for exclusion (as per the Cochrane standards) [9]. Two reviewers verified the final list of included studies. We presented the results of the literature search and screening process using the PRISMA flow diagram.

Data Extraction and Management

All relevant data were extracted using a structured, piloted form in Microsoft Excel or Covidence (Veritas Health Innovation, Melbourne, Australia) by different teams of reviewers. These forms were piloted on five studies by the authors and amended according to the received feedback. For each review, two researchers independently extracted and managed the data for each of the included studies. We extracted standard data on study design and setting, participants, interventions, controls, and outcomes. We extracted specific data in relation to the factors including intervention type, mode of delivery, field of study, duration, frequency, and interactivity. We also collected data on the type and validity of outcome-measurement instruments and study funding. Disagreements between review authors were resolved by discussion and consensus. A third review author acted as an arbiter in cases where disagreements could not be resolved.

Assessment of Risk of Bias in Included Studies

Two reviewers independently assessed the risk of bias for RCTs using the Cochrane "Risk of Bias" tool, with any disagreements resolved by discussion and consensus [9]. We piloted the risk-of-bias assessments to investigate agreement among the reviewers. RCTs were assessed for risk of bias using the following domains: random sequence generation, allocation concealment, blinding of outcome assessors, completeness of outcome data, and selective outcome reporting (eg, the presence or absence of a published protocol). We also assessed other sources of bias such as baseline imbalance and inappropriate administration of an intervention. For cluster RCTs, we assessed the risk of the following additional domains: recruitment bias, baseline imbalance, loss of clusters, incorrect analysis, and comparability with individually randomized trials as previously

recommended [44]. Judgements concerning the risk of bias for each study were classified as high, low, or unclear risk of bias, supported by a quote from the study report and a justification for our judgement for each item presented in a "Risk of bias" table. We incorporated the results of the risk of bias assessment into the review using risk of bias tables, summary of findings tables, a graph, and a narrative summary. For objectively reported outcomes, we did not judge studies to be at a high risk of bias due to a lack of participant blinding, as the nature of the intervention precluded this type of blinding.

Measures of Treatment Effect

For continuous outcomes, we presented the data in the form of standardized mean difference (SMD) along with 95% CIs, as the outcomes were measured with a range of different outcome-measurement tools. The majority of studies presented postintervention data instead of mean change scores. As SMD does not allow for pooling of both change and postintervention scores, we decided to use postintervention mean scores for all reviews [10-12,14,16,20-22,24,25,27,35]. For dichotomous data, we calculated odds ratios, risk ratios, or hazard ratios along with 95% CIs and *P* values.

We were unable to identify a clinically meaningful difference in effect size in the literature on digital health education. Therefore, in line with other studies in the field, we presented outcomes using postintervention SMD and interpreted the effect size using the Cohen rule of thumb (ie, with 0.2 representing a small effect, 0.5 representing a moderate effect, and 0.8 representing a large effect) [9,45]. If studies had multiple arms, we compared the intervention arm to the least-active control arm and assessed differences in the postintervention outcomes. This type of effect-size interpretation has been used in previous studies [46].

Management of Missing Data

We contacted the original investigators for clarification or to request missing information. If we were unable to obtain this information, we used data available from the published studies and deemed the risk of bias in respective domains as unclear. We did not impute any missing data; complete case analysis was used for data analyses. We conducted analysis on an intention-to-treat basis, where possible, including all participants who were randomized to either the digital health education group or comparator group, regardless of losses to follow-up and withdrawals [9]. We reported data on the loss to follow-up and assessed this as a potential source of bias. When data were unavailable for an intention-to-treat analysis, we analyzed data as reported.

Data Synthesis

Data were extracted and entered into tables grouped by study design and type of intervention to create a descriptive synthesis using Review Manager, version 5.3 (The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, Denmark). Where feasible, we pooled the results quantitatively and presented findings in forest plots to provide effect estimates and 95% CIs for each individual study as well as a pooled effect estimate and 95% CIs. For meta-analysis of dichotomous outcomes, we planned to use the Mantel-Haenszel random-effects model [9]. For cluster RCTs and where appropriate, we planned to use meta-analysis for the data using a generic inverse-variance method, which accounts for clustering of data. However, the dichotomous data were mostly limited and data from cluster RCTs were either limited or already adjusted for clustering; therefore, they were analyzed together with the data originating from RCTs.

Assessment of Heterogeneity

We performed a qualitative assessment of clinical heterogeneity across the included studies by determining whether the included studies were similar enough (in terms of their population, intervention characteristics, and reported outcomes) to yield meaningful conclusions and by visually inspecting the overlap of confidence intervals on forest plots. If a meta-analysis of the included studies was appropriate, we assessed statistical heterogeneity by calculating the I^2 statistic [9]. In case of a high degree of heterogeneity (I^2 >50%), we explored the reasons for variability by conducting subgroup analyses. We attempted to explore possible clinical or methodological reasons for this variation by performing prespecified subgroup analysis. In most reviews, high heterogeneity precluded statistical pooling [10-12,20,21], and prespecified subgroup analyses did not provide an explanation for the observed heterogeneity. In such cases, we performed a narrative synthesis of findings. We presented the findings from the studies, organized by interventions, outcomes, or comparisons, in line with the objectives and research questions in each review. We analyzed the direction, magnitude, and heterogeneity of the effect of the intervention as well as the quality of the included evidence. We took note of consistencies/inconsistencies and outliers in the data. At times, we further analyzed and visualized the outcomes using albatross plots [47]. This is a novel approach, which allows an approximate examination of underlying effect sizes and additional exploration of sources of heterogeneity across studies. This is achieved by drawing contours that show the range of effect sizes that might lead to each *P* value for the given sample sizes under simple study designs.

Assessment of Reporting Biases

We assessed reporting bias qualitatively based on the characteristics of the included studies (eg, if only small studies that indicate positive findings were identified for inclusion). When at least 10 studies were quantitatively pooled, we planned to construct a funnel plot to investigate publication bias. However, in some reviews [22,24,25,35], the number of included studies in any of the pooled analyses did not allow for a formal assessment of the reporting bias [48].

Subgroup Analysis and Sensitivity Analysis

We considered stratifying the following variables, where appropriate: countries' income status (low- and middle-income

countries versus high-income countries), registration stage (pre- and postregistration interventions), discipline (ie, dental studies, medicine, nursing and midwifery, medical diagnostic and treatment technology, therapy and rehabilitation, and pharmacy), type of digital education intervention, and inclusion in formal institutional curriculum.

We also considered performing sensitivity analyses by removing, for example, the high-risk-of-bias studies or studies with a small sample size to investigate their impact on the outcomes. However, performing the prespecified subgroup and sensitivity analyses was unfeasible for all reviews [10-12,14,20-22,24,25,27,35] because of the unequal distribution, limited information, and insufficient number of studies in respective subgroups, comparisons, and outcomes.

Grading of Recommendations, Assessment, Development and Evaluations Assessment and Summary of Findings Tables

In most reviews [10-12,20-22,24,25,27,35], we performed Grading of Recommendations, Assessment, Development and Evaluations (GRADE) assessment to determine the quality of the included evidence and prepared a "Summary of findings" table to present a summary of the results and a judgement on the quality of evidence, based on meta-analysis or narrative synthesis. We followed the Cochrane Handbook for Systematic Reviews of Interventions guidelines [49]. Two authors used the GRADE criteria to rank the quality of evidence. We applied the following downgrading criteria in our assessment of the quality of evidence: study limitations, inconsistency, indirectness, imprecision, and publication bias. For study limitations, we summarized the risk-of-bias assessment for each outcome across studies and incorporated it into judgements about the GRADE assessment and "Summary of findings" tables. Although we have taken into consideration the GRADE upgrading criteria (ie, large effect, dose-response relations, and direction of residual confounding and biases), they were not applicable to our systematic reviews. As we were unable to pool studies in many reviews, we presented the findings in the tables narratively in line with the approach presented previously [50]. We presented the findings for each of the major primary outcomes as defined in the "Types of outcome measures" section.

Discussion

In this paper, we present the methodology that we developed and employed in a series of systematic reviews on the use of digital technology in health professions education [10-25,27,35]. Digital health education is an evolving research field that has undergone tremendous development over the last 20 years. The aim of our collaboration is to provide a robust evidence base to make evidence-based recommendations on the use of digital technology in health professions education.

When performing these systematic reviews [10-12,14,20-22,24,25], we faced numerous challenges and observed important limitations of the evidence. We found a lack of conceptual frameworks and unclear classifications and definitions to guide evidence syntheses in this area. We also observed high clinical and methodological heterogeneity across studies, poor reporting, lack of information on participants' randomization and baseline assessments, and small sample sizes. In our methodological approach, we aimed to tackle these challenges and provide a strong, comprehensive, thorough analysis and synthesis of the data. We have classified, conceptualized, and defined digital health education. In addition, we have undertaken a comprehensive, expert-informed literature search; transparent screening; data extraction; risk-of-bias assessment; and robust data analysis and synthesis. We also delineated gaps in the literature and provided clear recommendations for future research.

In addition to providing reliable evidence-based recommendations, our collaboration strives to further advance the field of digital health education. Our projects include development of reporting standards for digital health education studies and mapping the type of outcomes and tools used to measure outcomes. We recognize that our list of digital modalities may not be exhaustive and that our classification will need to evolve with the innovation and progress of digital education as further evidence becomes available. We are constantly improving and adapting our methodology in line with new insights. For example, we are now exploring the impact of various types of traditional learning as a control intervention, both passive (eg, reading a textbook) and active forms (eg, small, interactive workshop), on outcomes. We are also delineating the most active components and essential features of blended learning, establishing a possible relationship between the level of complexity and effectiveness and investigating the length of the causal pathway between the intervention and educational outcomes. Further, we aim to evaluate the effect of various digital education features such as interactivity, feedback, immersion, or spaced education on the educational outcomes. Digital education is an evolving field, permeating all aspects of education, with a potentially huge impact on health professional training and ultimately, quality of patient care. We aim to keep up with these changes by providing cutting-edge, strong evidence that will guide judicious and successful adoption of digital education in health professions education.

Acknowledgments

We gratefully acknowledge funding (eLearning for health professionals' education grant) from the Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore. We also want to acknowledge the multi-disciplinary and international team of people who have been participating and contributing to the development and activities of the Digital Health Education Collaboration. The contributors are: Ram Bajpai, Shweta Bajpai, Gerard Dunleavy, Andy Hau Yan Ho, Krister Järbrink,

Geronimo Jimenez, Bhone Myint Kyaw, Laura Martinengo, Nuraini Nazeha, Yuri Rykov, and Monika Semwal from the Centre for Population Health Sciences (CePHaS), Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore; Onyema Ajuebor from the World Health Organization (WHO), Geneva, Switzerland; Andrzej Kononowicz from the Jagiellonian University Medical College, Krakow, Poland; Vasilis Hervatis, Anneliese Lilienthal, Italo Masiello, Natalia Stathakarou, and Luke Woodham from the Karolinska Institutet, Stockholm, Sweden; Hayfaa Wahabi, Khawater Bahkali, and Titi Maher from the King Saud University, Riyadh, Saudi Arabia; Sarah Gentry from the University of East Anglia, Norwich, UK; Serena Brusamento, Nikolas Mastellos and Edward Meinert from Imperial College London, London, UK; Priya Lall from the Queen Mary University of London, London, UK; Rebecca Rees from the University College London, London, UK; Pradeep Paul George and Nakul Saxena from the National Healthcare Group, Singapore, Singapore; Ziva Cotic from the University of Ljubljana, Ljubljana, Slovenia; David Davies from the University of Warwick, Coventry, UK; Samuel Edelbring from the Örebro University, Örebro, Sweden; Sandra Kemp from the Curtin University, Perth, Australia; Magdalena Koperny from the Jagiellonian University Medical College, Krakow, Poland; Craig Lockwood from The University of Adelaide, Adelaide, Australia; Neil Smart from the University of New England, Armidale, Australia; Jitka Vseteckova from The Open University, Milton Keynes, UK; David Wortley, nonaffiliated, from the United Kingdom.

Authors' Contributions

JoC conceived the project and wrote the manuscript with input from PP and LTC. JCD, PW, NZ, RA, AM, and JaC provided critical feedback to the manuscript. All other authors contributed to the development of the methods and reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Dublin Declaration on Human Resources for Health: Building the Health Workforce of the Future. Dublin: Fourth Global Forum on Human Resources for Health, World Health Organization; 2017 Nov 17. URL: https://www.who.int/hrh/events/Dublin_Declaration-on-HumanResources-for-Health.pdf?ua=1 [accessed 2019-01-02] [WebCite Cache ID 757V0GjMj]
2. Norheim OF. Ethical priority setting for universal health coverage: challenges in deciding upon fair distribution of health services. BMC Med 2016 May 11;14:75 [FREE Full text] [doi: 10.1186/s12916-016-0624-4] [Medline: 27170046]
3. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
4. Choules AP. The use of elearning in medical education: a review of the current situation. Postgrad Med J 2007 Apr;83(978):212-216 [FREE Full text] [doi: 10.1136/pgmj.2006.054189] [Medline: 17403945]
5. Wutoh R, Boren SA, Balas EA. eLearning: a review of Internet-based continuing medical education. J Contin Educ Health Prof 2004;24(1):20-30. [doi: 10.1002/chp.1340240105] [Medline: 15069909]
6. World Health Organization. 2016. Global strategy on human resources for health: workforce 2030 URL: https://www.who.int/hrh/resources/pub_globstrathrh-2030/en/ [accessed 2018-11-15] [WebCite Cache ID 757WPzGb6]
7. High-Level Commission on Health Employment and Economic Growth. Geneva, Switzerland: World Health Organization; 2016 Sep. URL: https://www.who.int/hrh/com-heeg/reports/report-expert-group/en/ [accessed 2019-01-02] [WebCite Cache ID 757WXrJwP]
8. Human resources for health and implementation of the outcomes of the United Nations' High-Level Commission on Health Employment and Economic Growth. Geneva, Switzerland: World Health Organization; 2018 May 17. URL: http://apps.who.int/gb/ebwha/pdf_files/WHA70/A70_18-en.pdf [accessed 2018-11-15] [WebCite Cache ID 757WikbRq]
9. Cochrane Handbook for Systematic Reviews of Interventions Update September 2018 - Version 6. 2018. URL: https://training.cochrane.org/handbook [accessed 2019-01-28] [WebCite Cache ID 75lhPBv7G]
10. Xu X, Posadzki PP, Lee GE, Car J, Smith HE. Digital Education for Health Professions in the Field of Dermatology: A Systematic Review by Digital Health Education Collaboration. Acta Derm Venereol 2018 Oct 15 [FREE Full text] [doi: 10.2340/00015555-3068] [Medline: 30320871]
11. Wahabi HA, Esmaeil SA, Bahkali KH, Titi MA, Amer YS, Fayed AA, et al. Medical doctors' offline-computer-assisted digital education: a systematic review by the Digital Health Education collaboration. J Med Internet Res 2019 Preprint available [FREE Full text] [doi: 10.2196/preprints.12998]
12. Posadzki PP, Paddock S, Campbell J, Car J. Medical students' digital education of communication skills: a systematic review by the Digital Health Education collaboration. J Med Internet Res 2018 Preprint available. [doi: 10.2196/preprints.12967]
13. Tudor Car L, Riboli-Sasco EF, Marcano Belisario JS, Nikolaou CK, Majeed A, Zary N, et al. PROSPERO CRD42015029786. 2015. Mobile learning for delivering health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42015029786 [accessed 2019-02-01] [WebCite Cache ID 75r28PGRf]

14. Tudor Car L, Kyaw BM, Atun R. The role of eLearning in health management and leadership capacity building in health system: a systematic review. Hum Resour Health 2018 Sep 03;16(1):44 [FREE Full text] [doi: 10.1186/s12960-018-0305-9] [Medline: 30176899]
15. Gentry S, L'Estrade Ehrstrom B, Gauthier A, Alvarez J, Wortley D, van Rijswijk J, et al. PROSPERO CRD42016045997. 2016. Serious Gaming and Gamification interventions for health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045997 [accessed 2019-02-01] [WebCite Cache ID 75r2UFjRj]
16. Kononowicz AA, Woodham L, Georg C, Edelbring S, Stathakarou N, Davies D, et al. PROSPERO CRD42016045912. 2016. Virtual patient simulations for health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045912 [accessed 2019-02-01] [WebCite Cache ID 75r2aDPSX]
17. Hervatis V, Kyaw BM, Semwal M, Dunleavy G, Tudor Car L, Zary N. PROSPERO CRD42016045679. 2016. Offline and computer-based eLearning interventions for medical students' education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045679 [accessed 2019-02-01] [WebCite Cache ID 75r2hfXwX]
18. George PP, Toon E, Hadadgar A, Jirwe M, Saxena N, Lim KTK, et al. PROSPERO CRD42016045571. 2016. Online- and local area network (LAN)-based eLearning interventions for medical doctors' education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045571 [accessed 2019-02-01] [WebCite Cache ID 75r2li0AE]
19. Saxena N, Kyaw BM, Vseteckova J, Dev P, George PP, Lim KTK, et al. PROSPERO CRD42017083206. 2017. Virtual reality environments for health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045470 [accessed 2019-02-01] [WebCite Cache ID 75r2u3Isx]
20. Semwal M, Whiting P, Bajpai R, Bajpai S, Kyaw BM, Tudor Car L. Digital education for health professions on smoking cessation management: a systematic review by the Digital Health Education collaboration. J Med Internet Res 2019 (preprint) [FREE Full text] [doi: 10.2196/preprints.13000]
21. Huang Z, Semwal M, Lee SY, Tee WKM, Ong W, Tan WS, et al. Digital Health Professions Education on Diabetes Management: Systematic Review by the Digital Health Education Collaboration. J Med Internet Res (forthcoming) 2019 [FREE Full text] [doi: 10.2196/12997]
22. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou C, George PP, et al. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
23. Semwal M, Wahabi HA, Posadzki P, Divakar U, Lim KTK, Audouard-Marzin Y, et al. PROSPERO CRD42017083206. 2017. Offline and computer-based eLearning interventions for medical doctors' education [Cochrane protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017083206 [accessed 2019-02-01] [WebCite Cache ID 75r34ZaiS]
24. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GYC, Tudor Car L. Mobile Digital Education for Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Preprint available [FREE Full text] [doi: 10.2196/12937]
25. Posadzki P, Bala M, Kyaw BM, Semwal M, Divakar U, Koperny M, et al. Offline digital education for post-registration health professions: a systematic review by the Digital Health Education collaboration. J Med Internet Res 2019 Preprint available [FREE Full text] [doi: 10.2196/preprints.12968]
26. Peek R. Web-based digital education opportunities. The Journal of Academic Librarianship 2000 Jan 25;26(1):49-52. [doi: 10.1016/S0099-1333(99)00123-8]
27. Gentry S, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious gaming and gamification education in health professions: a systematic review by the Digital Health Education collaboration. J Med Internet Res 2019:a Preprint available [FREE Full text] [doi: 10.2196/preprints.12994]
28. Stuchlíková L, Kósa A. Massive open online courses - Challenges and solutions in engineering education. 2013 Oct 25 Presented at: 11th IEEE International Conference on Emerging eLearning Technologies and Applications (ICETA); 2013; Slovakia p. 24-25 URL: http://www-old.cnl.sk/~jakab/2013%20ICETA%20FLASH-FILMS/alliceta2013papers.pdf
29. Nicholls D, Sweet L, Hyett J. Psychomotor Skills in Medical Ultrasound Imaging An Analysis of the Core Skill Set. J Ultras Med 2014 Aug 01;33(8):1349-1352. [doi: 10.7863/ultra.33.8.1349] [Medline: 25063399]
30. Mikropoulos TA, Natsis A. Educational virtual environments: A ten-year review of empirical research (1999–2009). Computers & Education 2011 Apr 04;56(3):769-780. [doi: 10.1016/j.compedu.2010.10.020]
31. Erhel S, Jamet E. Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. Computers & Education 2013 Sep;67:156-167 [FREE Full text] [doi: 10.1016/j.compedu.2013.02.019]
32. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. Med Educ 2010 Jan;44(1):50-63. [doi: 10.1111/j.1365-2923.2009.03547.x] [Medline: 20078756]
33. Crompton H. A historical overview of mobile learning: Toward learner-centered education. In: Berge ZL, Muilenburg LY, editors. Handbook of Mobile Learning. London: Routledge; 2013:3-14.

34. Lall P, Rees R, Law GCY, Dunleavy G, Cotic Z, Car J. What influences the Implementation of mLearning for Medical and Nursing Education? A qualitative systematic review by the Digital Health Education collaboration. J Med Internet Res 2019 Preprint available [FREE Full text] [doi: 10.2196/12895]
35. Tudor Car L, Kyaw BM, Dunleavy G, Smart N, Semwal M, Rotgans J, et al. Health professions digital problem-based learning: a systematic review and meta-analysis by the Digital Health Education collaboration. J Med Internet Res 2019:a Preprint available [FREE Full text] [doi: 10.2196/preprints.12945]
36. Petrisor B, Bhandari M. The hierarchy of evidence: Levels and grades of recommendation. Indian J Orthop 2007 Jan;41(1):11-15 [FREE Full text] [doi: 10.4103/0019-5413.30519] [Medline: 21124676]
37. UN Statistics Division. Montréal: UNESCO; 2013. International Standard Classification of Education: Fields of Education Training 2013 (ISCED-F 2013) URL: https://unstats.un.org/unsd/classifications/Family/Detail/1046 [accessed 2018-11-23] [WebCite Cache ID 75dvwBEeA]
38. Health and Care Professions Council. London, UK; 2012. Our standards for CPD URL: https://www.hcpc-uk.org/cpd/your-cpd/our-standards-for-cpd/ [accessed 2019-01-02] [WebCite Cache ID 75dvJ0D71]
39. Accreditation Council for Continuing Medical Education. URL: https://www.accme.org/ [accessed 2019-01-02] [WebCite Cache ID 757XyDPcC]
40. Health and Care Professions Council. 2018. What is CPD? URL: https://www.hcpc-uk.org/cpd/what-is-cpd [accessed 2019-01-02] [WebCite Cache ID 757YDOeZx]
41. Schuwirth L, Cantillon P. The need for outcome measures in medical education. BMJ 2005 Oct 29;331(7523):977-978 [FREE Full text] [doi: 10.1136/bmj.331.7523.977] [Medline: 16254276]
42. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [Medline: 2400509]
43. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
44. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. BMJ 2003 Oct 02;327(7418):785-789 [FREE Full text] [doi: 10.1136/bmj.327.7418.785] [Medline: 14525877]
45. Cochrane Handbook For Systematic Reviews Of Interventions. 2008. Assessing Risk of Bias in Included Studies URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470712184.ch8 [accessed 2019-01-28] [WebCite Cache ID 75lhBcfw9]
46. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA 2011 Sep 07;306(9):978-988. [doi: 10.1001/jama.2011.1234] [Medline: 21900138]
47. Harrison S, Jones HE, Martin RM, Lewis SJ, Higgins JPT. The albatross plot: A novel graphical tool for presenting results of diversely reported studies in a systematic review. Res Synth Methods 2017 Sep;8(3):281-289 [FREE Full text] [doi: 10.1002/jrsm.1239] [Medline: 28453179]
48. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [Medline: 9310563]
49. Schünemann JH, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011), The Cochrane Collaboration. 2011 Mar. Presenting Results and 'Summary of Findings' Tables URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470712184.ch11 [accessed 2019-01-28] [WebCite Cache ID 75lh1GAud]
50. Chan R, Webster J, Marquart L. Information interventions for orienting patients and their carers to cancer care facilities. Cochrane Database Syst Rev 2011 Dec 07(12):CD008273. [doi: 10.1002/14651858.CD008273.pub2] [Medline: 22161425]

Abbreviations

GRADE: Grading of Recommendations, Assessment, Development and Evaluations
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by A Marusic; submitted 28.11.18; peer-reviewed by I Buljan, S Gallagher, S O'Connor, J Jenkinson; comments to author 20.12.18; revised version received 13.01.19; accepted 15.01.19; published 14.02.19.

Please cite as:
Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Atun R, Majeed A, Campbell J, Digital Health Education Collaboration
Digital Education in Health Professions: The Need for Overarching Evidence Synthesis
J Med Internet Res 2019;21(2):e12913
URL: http://www.jmir.org/2019/2/e12913/
doi:10.2196/12913
PMID:30762583

Review

Digital Health Professions Education on Diabetes Management: Systematic Review by the Digital Health Education Collaboration

Zhilian Huang[1,2], MPH; Monika Semwal[1], PhD; Shuen Yee Lee[3], BSc; Mervin Tee[4], BSc; William Ong[2], BEng; Woan Shin Tan[1,2,5], PhD; Ram Bajpai[1], PhD; Lorainne Tudor Car[6], MD, PhD

[1]Centre for Population Health Sciences (CePHaS), Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[2]Institute for Health Technologies (HealthTech NTU), Interdisciplinary Graduate School, Nanyang Technological University Singapore, Singapore, Singapore

[3]Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

[4]School of Mechanical Aerospace and Engineering, Nanyang Technological University Singapore, Singapore, Singapore

[5]Health Services and Outcomes Research Department, National Healthcare Group, Singapore, Singapore

[6]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Lorainne Tudor Car, MD, PhD
Family Medicine and Primary Care
Lee Kong Chian School of Medicine
Nanyang Technological University
11 Mandalay Road, Level 18
Clinical Sciences Building
Singapore, 308232
Singapore
Phone: 65 69041258
Email: lorainne.tudor.car@ntu.edu.sg

Abstract

Background: There is a shortage of health care professionals competent in diabetes management worldwide. Digital education is increasingly used in educating health professionals on diabetes. Digital diabetes self-management education for patients has been shown to improve patients' knowledge and outcomes. However, the effectiveness of digital education on diabetes management for health care professionals is still unknown.

Objective: The objective of this study was to assess the effectiveness and economic impact of digital education in improving health care professionals' knowledge, skills, attitudes, satisfaction, and competencies. We also assessed its impact on patient outcomes and health care professionals' behavior.

Methods: We included randomized controlled trials evaluating the impact of digitalized diabetes management education for health care professionals pre- and postregistration. Publications from 1990 to 2017 were searched in MEDLINE, EMBASE, Cochrane Library, PsycINFO, CINAHL, ERIC, and Web of Science. Screening, data extraction and risk of bias assessment were conducted independently by 2 authors.

Results: A total of 12 studies met the inclusion criteria. Studies were heterogeneous in terms of digital education modality, comparators, outcome measures, and intervention duration. Most studies comparing digital or blended education to traditional education reported significantly higher knowledge and skills scores in the intervention group. There was little or no between-group difference in patient outcomes or economic impact. Most studies were judged at a high or unclear risk of bias.

Conclusions: Digital education seems to be more effective than traditional education in improving diabetes management–related knowledge and skills. The paucity and low quality of the available evidence call for urgent and well-designed studies focusing on important outcomes such as health care professionals' behavior, patient outcomes, and cost-effectiveness as well as its impact in diverse settings, including developing countries.

(J Med Internet Res 2019;21(2):e12997) doi:10.2196/12997

KEYWORDS

evidence-based practice; health personnel; learning; systematic review; diabetes mellitus

Introduction

Diabetes is one of the biggest global public health concerns affecting an estimated 425 million adults worldwide, and this number is expected to rise to 629 million by 2045 [1]. This is coupled with a shortage of health care professionals competent in delivering high-quality diabetes care [2,3]. Enhancing both the size and competencies of health care professionals is a priority and improving health professions education is seen as one of the key strategies to this end [4]. Digital education, broadly defined as the use of digital technology in education, has been recognized as having the potential to improve health professions education by making it scalable, interactive, personalized, global, and cost-effective [5-7].

Past systematic reviews on digital education have focused mainly on diabetes self-management education for patients, showing an improvement in patients' knowledge and outcomes [8-10]. The effectiveness of digital education interventions for health care professionals on diabetes management is still unknown [11]. To address this gap, we performed a systematic review to evaluate the effect of digital education on diabetes management on health care professionals' knowledge, skills, attitudes, competencies, and behaviors, as well as its impact on patient outcomes.

Methods

Systematic Review Guidance

We followed the Cochrane Handbook of Systematic Reviews for our methodology [12] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement for reporting [13]. For a detailed description of the methodology, please refer to the study by Car et al [14].

Data Sources and Searches

This review is part of an evidence-synthesis initiative on digital health professions education, where an extensive search strategy was developed for a series of systematic reviews on different modalities of digital health education for health care professionals (see Multimedia Appendix 1) [15]. The following databases were searched from January 1990 to August 2017:

1. The Cochrane Central Register of Controlled Trials (The Cochrane Library,)
2. MEDLINE (Ovid)
3. EMBASE (Elsevier)
4. PsycINFO (Ovid)
5. Educational Resource Information Centre (ERIC; Ovid)
6. Cumulative Index to Nursing and Allied Health Literature (CINAHL; EBSCO)
7. Web of Science Core Collection (Clarivate analytics).

We included studies in all languages and at all stages of publication. Our search strategy included gray literature sources such as Google scholar, trial registries, theses, dissertations, and academic reports. The citations retrieved from different sources were combined into a single library and screened by 2 authors independently. We also screened references of included papers for potentially eligible studies. Discrepancies and disagreements were resolved through discussion until a consensus was reached.

Study Selection

We included randomized controlled trials (RCTs), cluster RCTs, and quasi-RCTs and excluded cross-over trials due to high likelihood of a carry-over effect in this type of studies [12]. Studies on pre- or postregistration health care professionals taking part in digital education interventions on diabetes management were considered eligible. We defined health care professionals in line with the Health Field of Education and Training (091) in the International Standard Classification of Education [16]. Studies on digital education on both type 1 and type 2 diabetes at all educational levels were included.

We defined digital education as any teaching and learning that occurs by means of digital technologies. We considered eligible all digital education modalities, including offline and online education, Serious Gaming and Gamification, Massive Open Online Courses, Virtual Reality Environments, Virtual Patient Simulations, Psychomotor Skills Trainers, and mobile learning. Eligible comparisons were traditional, blended, or another form of digital education intervention on diabetes management. Traditional education was defined as any teaching and learning taking place via nondigital educational material (eg, textbooks) or in-person human interaction (eg, lecture or seminar). Traditional education also included usual learning, for example, usual revisions as well as on-the-job learning without a specific intervention in postregistration health care professionals. Blended education was defined as the act of teaching and learning that combines aspects of traditional and digital education. Eligible primary outcomes measured using any validated and non-validated instruments were knowledge, skills, competencies, attitudes, and satisfaction. Eligible attitudes-related outcomes comprised all attitudes toward patients, new clinical knowledge, skills, and changes to clinical practice.

Eligible secondary outcomes included patient outcomes in studies on postregistration health care professionals (eg, patients' blood pressure, blood glucose, and blood lipid levels), change in health care professional's behavior (ie, treatment intensification, defined as an intensity or dose increase of an existing treatment or the addition of a new treatment/class of medication), and economic impact of the intervention.

Data Extraction

In this study, 2 authors independently extracted data from studies using a structured and piloted data extraction form. We extracted information on study design, participants' demographics, type, content and delivery of digital education, and information pertinent to the intervention. Study authors were contacted in case of unclear or missing information.

Risk of Bias and Quality of Evidence Assessment

The methodological quality of included RCTs was independently assessed by 2 authors using the Cochrane Risk of Bias Tool [12]. The risk of bias assessment was piloted between the reviewers, and we contacted study authors in case of any unclear or missing information. We assessed the risk of bias in included

RCTs for the following domains: (1) random sequence generation; (2) allocation concealment; (3) blinding of participants to the intervention; (4) blinding of outcome assessment; (5) attrition; (6) selective reporting; and (7) other sources of bias [17]. Cluster RCTs were assessed using 5 additional domains: (1) recruitment bias; (2) baseline imbalance; (3) loss of clusters; (4) incorrect analysis; and (5) comparability with individually randomized trials [12].

Data Synthesis and Analysis

In line with Miller's classification, a learning model for assessment of clinical competence [18], we classified outcomes based on the type of outcome measurement instruments used in the study. For example, multiple-choice questionnaires were classified as assessing knowledge and objective structured clinical examinations as assessing participants' skills.

Although some studies reported change scores, we presented only postintervention data as those were more commonly reported and to ensure consistency and comparability of findings. Continuous outcomes are presented using mean difference (for outcomes measured using the same measurement tool), standardized mean difference (SMD; for outcomes measured using diverse measurement tools), and 95% CIs. Dichotomous outcomes are presented using risk ratios (RRs) and 95% CIs. As we were unable to identify a clinically meaningful interpretation of effect size in the literature for digital education interventions, we interpreted the effect size using Cohen *rule of thumb* with SMD greater than or equal to 0.2 representing a small effect, SMD greater than or equal to 0.5 a moderate effect, and SMD greater than or equal to 0.8 a large effect [19,20]. In studies that reported more than one measure for each outcome, the primary measure, as defined by the primary study authors, was considered.

Heterogeneity and Subgroup Analyses

Heterogeneity was assessed qualitatively using information relating to participants, interventions, controls, and outcomes as well as statistically using the I^2 statistic for outcomes allowing for pooled analysis [17]. Due to substantial methodological, clinical, and statistical heterogeneity (I^2>50%), we conducted a narrative synthesis according to type of comparison, that is, (1) digital education versus traditional education, (2) digital education versus blended education, and (3) one digital education type versus another digital education type. Subgroup analyses were not feasible owing to the small number of studies and limited information. We presented the study findings in a forest plot using the random effects model and standardized mean difference as the measurement scales were different and without the pooled estimates.

Results

Included Studies

Our search strategy for a series of systematic reviews focusing on different digital health professions education modalities yielded 30,532 unique references. We removed 459 duplicates, and upon screening of titles and abstracts, the screening excluded 30,050 citations. We identified 23 potentially eligible studies for which we retrieved and screened full texts. Of these, we included 12 studies: 9 RCTs and 3 cluster RCTs, all published in English (Figure 1). Moreover, 1 study was reported by 3 journal papers [21-23]. Although presented as a cluster RCT, this study included randomization at the individual, physician level and was therefore considered an RCT. A total of 9 studies were excluded due to ineligible study design (n=3), missing data (n=5), and ineligible participants (n=1; Figure 1).

Participant Characteristics

There were 2263 health care professionals in 12 included studies [23-34]. A third of the studies included less than 50 participants. The study with 3 published reports had 1182 patient records as a measure of clinical outcomes [21-23]. Only 1 study targeted pediatric patients with type 1 diabetes [28]. All other studies reporting patient outcomes focused on adult patients with type 2 diabetes. A total of 8 studies focused on doctors [23,24,26,27,29,32-34]. Moreover, 1 study each focused on medical students [30], pharmacy students [25], nurses [31], and jointly on doctors, nurses, and dietitians [28].

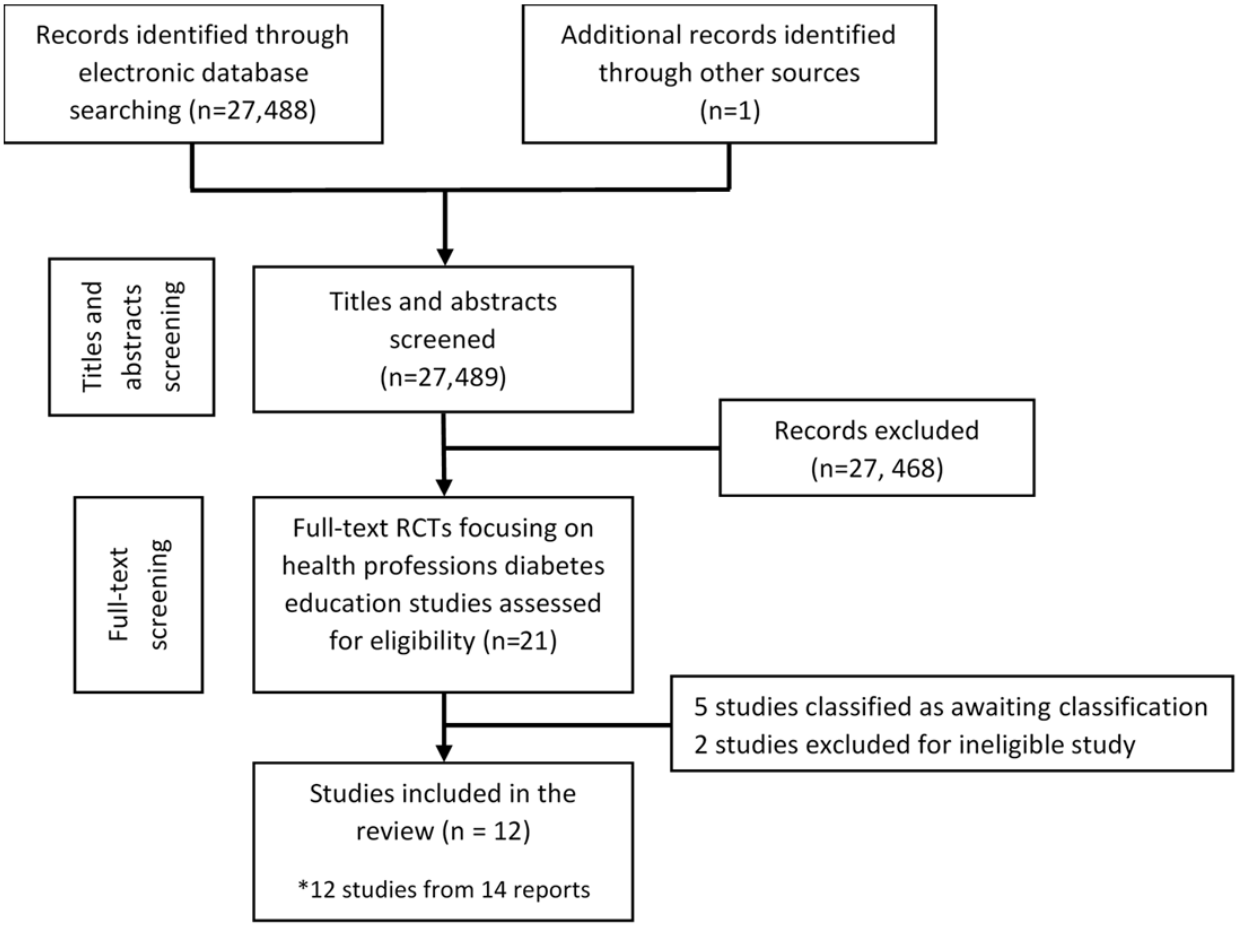
Study Characteristics

A total of 10 studies were conducted in high-income countries including Australia [30], the United States [23,24,26,29,31-33], and the United Kingdom [27,28]. A total of 2 studies were conducted in middle-income countries such as Thailand [25] and Brazil [34] each.

A total of 6 studies compared digital education with traditional education [25,26,29,30,32,34]. A total of 3 studies compared 2 different methods of digital education interventions [23,27,31], 2 compared blended education with usual education [28,33], and 1 study with 3 arms compared usual, blended, and digital education [24]. Only 4 studies reported duration of the intervention lasting from an hour to 2 weeks [25,26,30,34].

Various types of modalities were used to deliver the digital education interventions. A total of 3 studies used a Web-based or online portal [23,27,28]; 3 used a scenario-based simulation software [24,26,32]; 1 study each assessed high-fidelity mannequins [31]; an online game app on the computer [34]; periodic email reminders on the lecture content [33]; personal digital assistant–delivered learning materials [29]; and a computer-based diabetes management program [25].

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of included studies.



All studies except 3 employed clinical scenarios in the digital education intervention [21,24-27,30-32,34]. The remaining 3 studies used text-based learning [29,33,28]. Feedback was provided to participants in the intervention group in 7 studies [21,24,26-28,32,34]. A total of 2 studies comparing different forms of digital education reminded participants to log into the system [21,27], whereas one employed an email reminder to consolidate learned knowledge [33]. Half of the studies evaluated interactive digital education interventions [21,24,26,31,32,34].

Comparison interventions were also varied; 3 studies utilized a Web-based system (online portal) for the control group [23,27,31]; 4 compared the digital education intervention with face-to-face education [25,29,33,34]; 1 provided hard copy materials [29]; 1 reported *revision as usual* where participants could access relevant materials available to them [30]. A total of 4 studies focusing on postregistration education did not include any control intervention [24,26,28,32].

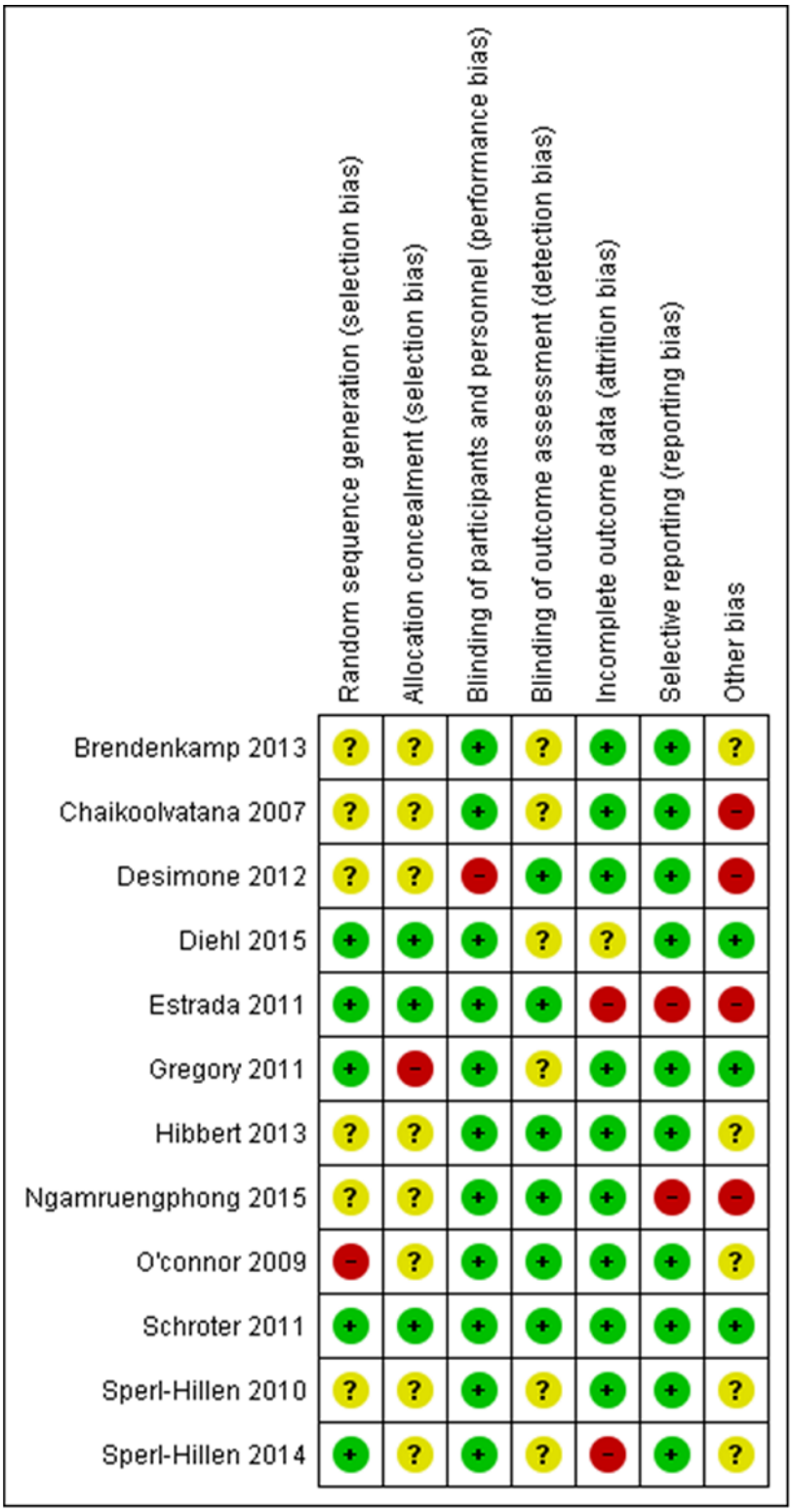
A total of 11 studies measured primary outcomes; 6 assessed knowledge with questionnaires [25,27,29,31-33]; 5 assessed skills and competency (measured as a combination of knowledge and skill) [25,28,30,32,34]; 2 assessed learners' attitude [27,34]; and 4 assessed learners' satisfaction [24,26,30,34]. A total of 5 studies measured secondary outcomes; 2 assessed the cost of the intervention [26,28]; 4 assessed patient outcomes (ie, patients meeting glycated hemoglobin [23,24,26,28]; low-density lipoprotein [23,24,26]; and blood pressure control [23,26] goals); and 2 assessed treatment intensification (intensifying the treatment regimen as required) [21,24].

Participant type and content of diabetes education across the studies varied widely and included diabetes management skills for primary care physicians (PCPs) [23,24,26,27,34]; diabetes clinical care for primary care, family, and internal medicine residents [29,32]; communication skills for pediatric doctors, nurses, and dieticians managing type 1 diabetes patients [28]; clinical endocrinology skills for medical students [30]; primary care residents' training on Hepatitis B vaccination for diabetes patients [33]; nursing care for hypoglycemic patients [31]; and diabetes management knowledge, communication, and patient note writing skills for pharmacy students [25].

Risk of Bias in the Included Studies

Of 12 included studies, 7 were judged at a high risk of bias and three studies had an unclear risk of bias for at least three domains. Of three cluster RCTs, two were judged at a high risk due to baseline imbalance (Figure 2, Multimedia Appendix 2).

Figure 2. Risk of bias summary: review authors' judgement about each risk of bias item for each included study. The symbol "+" indicates a low risk of bias, "?" indicates unclear risk of bias and "-" indicates a high risk of bias. The methodology of 2 studies (Crenshaw 2010 and Billue 2012) are duplicated with (Estrada 2011) and not presented in this figure.



Digital Education Versus Traditional Education

A total of 4 RCTs [25,29,30,34] and 2 cluster RCTs [26,32] compared digital education with traditional education, including no intervention (ie, knowledge acquisition as usual or usual on-the-job training), face-to-face lectures or hard copy printouts (Table 1). A total of 3 studies measured knowledge outcome. Of these studies, 2 compared online virtual simulation and computer-based learning intervention with no intervention, respectively, and reported moderate-to-large postintervention knowledge gain in the digital education group compared with the control group (Multimedia Appendix 3 and Multimedia Appendix 4). The final study compared learning materials either printed or displayed on a mobile electronic device for medical

residents, where no between-group difference in postintervention knowledge scores was found [29].

Skills were assessed in 4 studies [25,30,32,34], which largely reported higher effectiveness of digital education interventions (Multimedia Appendix 3 and Multimedia Appendix 4). Moreover, 1 study comparing a training video and no intervention (usual revision) for medical students reported significant improvement in lower limb examination (RR: 2.29; 95% CI 1.05-4.99) and diabetes history taking skills (RR: 4.17; 95% CI 1.18-14.77), and no difference in thyroid disease examination. Another study comparing computer-based and face-to-face learning for final year pharmacy students found large improvements in subjective, objective, assessment, and plan note writing skills (SMD 0.78; 95% CI 0.33-1.22) in the digital education group and no difference in patient history taking skills between the groups [25]. The third study, comparing an online virtual case-based simulation with no intervention for medical residents, reported higher proportion of patients meeting safe treatment goals in 3 out of 4 hypothetical simulation cases [32] (Multimedia Appendix 4). The final study compared an online game with face-to-face learning for PCPs and assessed their competency, that is, a combination of factual knowledge and problem-solving skills of PCPs on insulin therapy for diabetes reported small improvements in the digital education group (SMD: 0.4; 95% CI 0.09-0.70]) [34].

Only 1 study comparing a simulated physician learning software and no intervention measured patient clinical outcomes and cost [24]. The study reported the mean pre- and postintervention change (95% CI) in glycated hemoglobin, systolic blood pressure, diastolic blood pressure, and low-density lipoprotein levels of patients under the care of participating physicians. Improvements were observed for all measures from baseline to postintervention in both intervention and control groups. However, when comparing the groups, results were mixed (Multimedia Appendix 4) [24]. Cost savings of US $71 per patient was reported for the intervention group compared with the control group from the health plan perspective, but the difference was not statistically significant.

Learner's satisfaction was assessed with self-reported surveys in 3 studies. Only 1 study focusing on online games evaluated satisfaction for both intervention and control groups, but the use of different questionnaires did not allow between-group comparisons [32]. The same study found significantly better diabetes management and insulin-related attitudes and beliefs toward the digital education intervention in the intervention group [32]. The remaining 2 studies assessed satisfaction only in the intervention group, and more than 80% of participants were satisfied with the digital intervention [24,28].

Blended Education Versus Traditional Education

One cluster RCT [26] and one RCT [31] compared blended education with traditional education and evaluated knowledge, skills, patients' glycated hemoglobin levels, and economic impact. The blended education within the RCT comprised the standard education and an additional 30-min didactic lecture, a pocket card, and monthly email reminders on lecture content. The study reported large improvement in postintervention knowledge score in the blended education group compared with the control group (SMD: 1.98; 95% CI 1.21-2.74) [31].

The blended learning program of the cluster RCT included Web-based training and practical workshops for behavioral change in pediatric patients with type 1 diabetes, whereas the control group received no intervention [26]. The blended education group had a large improvement in the postintervention communication skills score (SMD: 1.58; 95% CI 0.99-2.17) and a higher proportion of tasks done or partially done in shared agenda setting (RR: 7.49; 95% CI 1.88-29.9) compared with the control group (Multimedia Appendix 3). Cost differences in the mean total National Health Service cost (direct costs: training; indirect costs: clinic visits) were not statistically significant; although, the blended education intervention incurred an additional mean cost of £183.96 per patient. There was no statistically significant difference between the groups in patient outcomes (ie, glycated hemoglobin levels) and patients' quality of life (Multimedia Appendix 4).

Digital Education Versus Blended Education Versus Traditional Education

One RCT study compared digital education, blended education, and traditional education to improve the safety and quality of diabetes care delivered by PCPs [24]. The digital education group received online case-based simulation, and the blended education group also received feedback in the form of additional face-to-face physician opinion. Learners' satisfaction and patient clinical outcomes (ie, mean change in glycated hemoglobin, blood pressure, and low-density lipoprotein level and treatment intensification) were assessed. Over 97% of PCPs who completed the education intervention rated their satisfaction with the digital education and blended interventions as excellent or very good after completing the simulated cases. The mean glycated hemoglobin level significantly improved in the digital education group compared with blended or traditional education (Multimedia Appendix 4). There was no statistically significant difference across the intervention groups in the remaining patient outcomes (Multimedia Appendix 4).

Table 1. Characteristics of the included studies.

Study, design, and country	Learning modality	Type of participants	Number of sites and participants	Intervention duration	Type of outcome
Digital education versus traditional education					
Chaikoolvatana 2007 [25]; RCT[a]; Thailand	I[b]: Computer based learning (CBL); C[c]: face-to-face lectures	Final year pharmacy students	I: 43, C: 40	I: 2 hours; C: 2 3-hour sessions; (over 2 months)	(1) Knowledge; (2) skills
Desimone 2012 [29]; RCT; United States	I: PDA[d] version education materials; C: Printed materials	Internal medicine residents	I: 11, C: 11	Over 1 month	Knowledge
Diehl 2017 [34]; RCT; Brazil	I: Online game; C: Face-to-face lectures and activities	Primary Care Physicians	I: 94, C: 76	4 hours (over 3 months)	Skills
Hibbert 2013 [30]; RCT; Australia	I: Training Video; C: No intervention (usual revision)	Second year medical students	I: 12, C: 10	Over 2 Weeks	Skills
Sperl-Hillen 2010 [26]; cRCT[e]; United States	I: Simulation Software; C: No intervention	Primary Care Physicians and their patients	I: 20 sites, (1847 patients), C: 21 sites, (1570 patients)	5.5 days; (over 6 months)	Patient outcomes; Economic impact
Sperl-Hillen 2014 [32]; cRCT; United States	I: Simulation software; C: No intervention (Not assigned learning cases)	Family/ internal medicine residents	I: 10 sites (177 residents), C: 9 sites (164 residents)	Over 6 months	Knowledge; Skills
Blended learning versus traditional education					
Gregory 2011 [28]; cRCT; United Kingdom	I: Web-based intervention and practical workshops; C: No intervention	Paediatric doctors, nurses, psychologists, dieticians, and their patients	I: 13 sites (356 patients), C: 13 sites (333 patients)	Over 12 months	Skills; Patient outcomes; Economic impact
Ngamruengphong 2011 [33]; RCT; United States	I: Standard education+30 min didactic lecture, a pocket card, and monthly e-mail reminders that consisted of the lecture content; C: Standard residency education	Primary care residents	I: 20, C: 19	Over 2 months	Knowledge
Digital education versus digital education					
Billue 2012 [21]; RCT United States; Estrada 2011[23]; RCT United States; Crenshaw 2010[22]; RCT; United States	I: Web-based intervention with feedback; C: Web-based intervention without feedback	Family/ general/ internal medicine physicians	I: 48 physicians (479 patients), C: 47 physicians (466 patients)	Over 2 years	Patient outcomes
Brendenkamp 2013 [31]; RCT; United States	I: Simulation (High fidelity Mannequin); C: Web-based intervention	Staff nurses	I: 47, C: 49	Not reported	Knowledge
Schroter 2011 [27]; RCT; United Kingdom	I: Web-based learning + Diabetes Needs assessment tool (DNAT); C: Web-based learning without DNAT	Diabetes doctors and nurses	I: 499, C: 498	Over 4 months	Knowledge
Blended learning versus digital education versus traditional education					
O'Connor 2009 [24]; RCT; United States	Group A: No intervention; Group B: Simulated web-based learning; Group C: simulated case-based physician learning + physician opinion leader feedback	Primary care physicians and their patients	Group A: 100 physicians, 691 patients; Group B: 100 physicians, 725 patients; Group C: 99 physicians, 604 patients	Not reported	Patient outcomes

[a]RCT: randomized controlled trial.

[b]I: intervention group.

[c]C: control group.

[d]PDA: Personal Digital Assistance.

[e]cRCT: cluster RCT.

Digital Education Versus Digital Education

A total of 3 RCT studies compared 2 different digital education modalities [21,25,29]. Moreover, 1 study compared a high-fidelity simulation mannequin with an online learning system [29]. The other 2 studies, employing the same Web-based (online) system in both the groups, evaluated the addition of an interactive learning needs assessment tool or feedback to the intervention group, respectively [21,25].

Studies reported no significant difference in terms of knowledge, attitudes, and patient outcomes. The study evaluating the use of feedback as part of the digital education intervention reported higher study engagement in the intervention group as reflected by the total number of pages viewed (SMD: 1.40; 95% CI 0.95-1.85), total number of visits (SMD: 1.38; 95% CI 0.93-1.83]), duration of Web access in min (SMD: 1.07; 95% CI 0.64-1.50), and the number of components viewed (SMD: 1.14; 95% CI 0.70-1.57) [18].

Discussion

Principal Findings

We found 12 studies evaluating the effectiveness of digital health professions education on diabetes management. Although evidence is limited, heterogeneous, and of low quality, our findings suggest that digital and blended education may improve health care professionals' knowledge and skills compared with traditional education. However, an improvement in knowledge and skills does not seem to translate into improvements in diabetes care as reflected by little or no difference in sparsely reported patient outcomes in the included studies. Although simulated learning seems to be more effective in improving patient outcomes compared with the other strategies assessed, studies comparing different forms of digital education reported no statistically significant difference between groups.

The inconsistency between the effect on health care professionals' and patients' outcomes observed in our review is in line with the existing literature, where knowledge and skill gains outweigh improvements in patient outcomes [33]. Yet patient outcomes were only reported in 4 diverse studies in this review. The lack of patient-related data is common in digital education studies, possibly owing to difficulty in measuring patient outcomes, especially in preregistration health care professionals. Furthermore, patient outcomes are potentially affected by contextual factors unrelated to health care professionals' competence, such as patients' health beliefs and financial barriers [34]. Finally, a lack of difference between the groups observed in the included studies may be merely due to their insufficient statistical power to evaluate patient outcomes.

Although digital education has been present in health professions' education for the last 2 decades, its technological development and adoption has been expedited in recent years [35], particularly in high-income countries. Likewise, most studies in our review were published since 2010 and are from high-income countries. Widespread access to digital technology in high-income countries may diminish the effects of digital education interventions in RCTs, given that blinding is not possible, and the control group participants may interact or have alternative electronic access to information. Studies on the use of digital education in low- and middle-income countries would provide a more comprehensive assessment as the technological setup and learning infrastructure is more limited [36-38]. Although there is a universal need for scalable and high-quality education to build health care professionals' competencies in diabetes management and care, this is especially important for developing countries facing severe workforce shortages and increasing burden of chronic disease [5,39,40].

Digital education interventions in this review, although diverse in terms of the mode of delivery, mostly employed clinical scenarios for presentation of educational content. Furthermore, the included digital education interventions were mainly asynchronous and aimed at postregistration health care professionals. Although this digital education format may indeed be optimal for busy clinicians as part of their continuing professional development, there is scope for more research on other digital education formats as well as preregistration health care professionals [38,41,42].

Limitations

There are limitations to the evidence included in this review. First, studies were too heterogeneous to be pooled. Second, many studies were at a high risk of biases such as selection and attrition bias. Third, satisfaction with digital education interventions may be overestimated by a heavy reliance on self-reported measures and a disproportionate focus on only the intervention group. Satisfaction is important in ensuring the success of digital education interventions as it impacts the user's intention to sustain learning through digital means [33]. Therefore, alternative methods should be used to explore satisfaction with digital education interventions such as the actual time spent on digital learning or in-depth qualitative analyses on the perceptions of digital education. Finally, studies, in general, did not refer to a learning theory in the intervention design. Digital education presents a new model of learning where technological and Web-based learning expands and changes the paradigm of usual learning. Furthermore, the complexity of diabetes management may warrant a unique learning pedagogy. The use of technological or adult learning theories in the development of digital education interventions may improve the quality, reporting, and ingenuity of the digital education research if grounded in existing theoretical frameworks [34].

Future Research

Digital education is rapidly transforming health professions training and is expected to gain even more prominence in the coming years. It is critical for digital education adoption and implementation to be guided by a robust evidence base. There is a need for more high-quality and standardized studies from a range of settings, including developing countries, which would focus on all aspects of diabetes management. Future research should also aim to assess the economic impact to inform planning, development, and adoption of digital health professions education interventions on diabetes management.

Conclusions

Digital education holds the promise of a scalable and affordable approach to health professions education, with particular relevance to developing countries tackling severe shortage of skilled health care staff. In this review, we aimed to determine the effectiveness and cost-effectiveness of digital education for health professions education on diabetes management. We identified 12 studies showing that digital education is well-received and seems to improve knowledge and skills scores in health care professionals compared with traditional or usual education. Although digital education seems to be more effective, or not inferior to other forms of education on diabetes management, the paucity and low quality of data prevent us from making recommendations about its adoption. Future studies should focus on a range of outcomes using validated and standardized outcome measurements in different settings to improve the quality and credibility of evidence.

Acknowledgments

This review was conducted in collaboration with the World Health Organization, Health Workforce Department. We gratefully acknowledge funding from the Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, eLearning for health professionals' education grant. WST would like to acknowledge funding from the Singapore National Medical Research Council Research Training Fellowship and the Singapore National Healthcare Group PhD in Population Health Scheme for funding. ZH, SYL, MT, and WO were funded by the NTU research scholarship.

Authors' Contributions

LTC conceived the idea for the review. ZH and MS conducted the study selection, data extraction, and wrote and revised the review. LTC provided methodological guidance and critically revised the review. RB provided statistical guidance and comments on the review. SYL, MT, WST, and WO assisted in studies selection and data extraction and provided comments on the review. All authors commented on the review and made revisions following the first draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE (Ovid) search strategy.

[PDF File (Adobe PDF File), 72KB - jmir_v21i2e12997_app1.pdf]

Multimedia Appendix 2

Forest plots for knowledge and skill outcomes.

[PDF File (Adobe PDF File), 113KB - jmir_v21i2e12997_app2.pdf]

Multimedia Appendix 3

Risk of bias assessments.

[PDF File (Adobe PDF File), 172KB - jmir_v21i2e12997_app3.pdf]

Multimedia Appendix 4

Detailed characteristics of the included studies.

[PDF File (Adobe PDF File), 135KB - jmir_v21i2e12997_app4.pdf]

References

1. International Diabetes Federation. 2017. IDF Diabetes Atlas Eighth Edition URL: https://www.idf.org/e-library/epidemiology-research/diabetes-atlas/134-idf-diabetes-atlas-8th-edition.html [accessed 2019-02-05] [WebCite Cache ID 75xNABN2R]
2. Byrne JL, Davies MJ, Willaing I, Holt RIG, Carey ME, Daly H, et al. Deficiencies in postgraduate training for healthcare professionals who provide diabetes education and support: results from the Diabetes Attitudes, Wishes and Needs (DAWN2) study. Diabet Med 2017 Feb 13;34(8):1074-1083. [doi: 10.1111/dme.13334] [Medline: 28195662]
3. Childs BP. Core competencies in diabetes care: educating health professional students. Diabetes Spectr 2005 Apr 1;18(2):67-68. [doi: 10.2337/diaspect.18.2.67]

4. Groupe, Livres. World Health Organization. 2014. A universal truth: no health without a workforce URL: https://www.who.int/workforcealliance/knowledge/resources/GHWA_AUniversalTruthReport.pdf [accessed 2019-02-05] [WebCite Cache ID 75xPH0o69]
5. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]
6. Rider B, Lier S, Johnson T, Hu D. Interactive web-based learning: translating health policy into improved diabetes care. Am J Prev Med 2016;50(1):122-128. [doi: 10.1016/j.amepre.2015.07.038] [Medline: 26602941]
7. Wood EH. MEDLINE: the options for health professionals. J Am Med Inform Assoc 1994;1(5):372-380. [Medline: 7850561]
8. Årsand E, Frøisland DH, Skrøvseth SO, Chomutare T, Tatara N, Hartvigsen G, et al. Mobile health applications to assist patients with diabetes: lessons learned and design implications. J Diabetes Sci Technol 2012 Sep 1;6(5):1197-1206 [FREE Full text] [doi: 10.1177/193229681200600525] [Medline: 23063047]
9. Cooper H, Cooper J, Milton B. Technology-based approaches to patient education for young people living with diabetes: a systematic literature review. Pediatr Diabetes 2009;10(7):a. [doi: 10.1111/j.1399-5448.2009.00509.x] [Medline: 19490492]
10. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
11. Jackson CL, Bolen S, Brancati FL, Batts-Turner ML, Gary TL. A systematic review of interactive computer-assisted technology in diabetes care. Interactive information technology in diabetes care. J Gen Intern Med 2006 Mar 8;21(2):105-110 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00310.x] [Medline: 16390512]
12. Higgins JP, Green S. Iran University of Medical Sciences. 2011 Oct 18. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 URL: http://crtha.iums.ac.ir/files/crtha/files/cochrane.pdf [accessed 2019-01-30] [WebCite Cache ID 75oXZS37B]
13. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Br Med J 2015 Jan 2;349:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
14. Car J, Carlstedt-Duke J, Car LT, Posadzki P, Whiting P, Zary N, et al. Digital education for health professions: methods for overarching evidence syntheses. J Med Internet Res 2019:1-36 Accepted but not published at the point of citation(forthcoming). [doi: 10.2196/12913]
15. Posadzki P, Car L, Jarbrink K, Bajpai R, Semwal M, Kyaw B. The EQUATOR Network. 2017. Protocol for development of the Standards for Reporting of Digital Health Education Intervention Trials (STEDI) statement URL: http://www.equator-network.org/wp-content/uploads/2018/01/STEDI-Statement-Protocol-updated-Jan-2018.pdf [accessed 2018-11-29] [WebCite Cache ID 74Huf6Nas]
16. Statistics U. National Center for Educational Quality Enhancement. 2014. ISCED Fields of Education and Training 2013 (ISCED-F 2013): manual to accompany the International Standard Classification of Education URL: https://eqe.ge/res/docs/228085e.pdf [accessed 2018-11-29] [WebCite Cache ID 74HuiKfy8]
17. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
18. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]
19. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. J Clin Epidemiol 2013 Feb;66(2):173-183. [doi: 10.1016/j.jclinepi.2012.08.001] [Medline: 23116689]
20. Leucht S, Kissling W, Davis JM. How to read and understand and use systematic reviews and meta-analyses. Acta Psychiatr Scand 2009 Jun;119(6):443-450. [doi: 10.1111/j.1600-0447.2009.01388.x] [Medline: 19469725]
21. Billue K, Safford M, Salanitro A, Houston T, Curry W, Kim Y, et al. Medication intensification in diabetes in rural primary care: a cluster-randomised effectiveness trial. BMJ Open 2012;2(5):e000959 [FREE Full text] [doi: 10.1136/bmjopen-2012-000959] [Medline: 22991217]
22. Crenshaw K, Curry W, Salanitro A, Safford M, Houston T, Allison J. Is physician engagement with Web-based CME associated with patients' baseline hemoglobin A1c levels? The Rural Diabetes Online Care study. Acad Med 2010 Sep;85(9):1511-1517. [doi: 10.1097/ACM.0b013e3181eac036] [Medline: 20736679]
23. Estrada CA, Safford MM, Salanitro AH, Houston TK, Curry W, Williams JH, et al. A web-based diabetes intervention for physician: a cluster-randomized effectiveness trial. Int J Qual Health Care 2011 Dec;23(6):682-689 [FREE Full text] [doi: 10.1093/intqhc/mzr053] [Medline: 21831967]
24. O'Connor P, Sperl-Hillen J, Johnson P, Rush W, Asche S, Dutta P, et al. Simulated physician learning intervention to improve safety and quality of diabetes care: a randomized trial. Diabetes Care 2009 Apr;32(4):585-590 [FREE Full text] [doi: 10.2337/dc08-0944] [Medline: 19171723]

25. Chaikoolvatana A, Haddawy P. Evaluation of the effectiveness of a computer-based learning (CBL) program in diabetes management. J Med Assoc Thai 2007 Jul;90(7):1430-1434. [Medline: 17710988]
26. Sperl-Hillen JM, O'Connor PJ, Rush WA, Johnson PE, Gilmer T, Biltz G, et al. Simulated physician learning program improves glucose control in adults with diabetes. Diabetes Care 2010 Aug;33(8):1727-1733 [FREE Full text] [doi: 10.2337/dc10-0439] [Medline: 20668151]
27. Schroter S, Jenkins RD, Playle RA, Walsh KM, Probert C, Kellner T, et al. Evaluation of an online interactive Diabetes Needs Assessment Tool (DNAT) versus online self-directed learning: a randomised controlled trial. BMC Med Educ 2011 Jun 16;11:35 [FREE Full text] [doi: 10.1186/1472-6920-11-35] [Medline: 21679446]
28. Gregory J, Robling M, Bennert K, Channon S, Cohen D, Crowne E. Development and evaluation by a cluster randomised trial of a psychosocial intervention in children and teenagers experiencing diabetes: the DEPICTED study. Health Technol Assess 2011 Aug;15(29):1-202. [doi: 10.3310/hta15290] [Medline: 21851764]
29. Desimone ME, Blank GE, Virji M, Donihi A, DiNardo M, Simak DM, et al. Effect of an educational Inpatient Diabetes Management Program on medical resident knowledge and measures of glycemic control: a randomized controlled trial. Endocr Pract 2012 Mar;18(2):238-249. [doi: 10.4158/EP11277.OR] [Medline: 22440993]
30. Hibbert EJ, Lambert T, Carter JN, Learoyd D, Twigg S, Clarke S. A randomized controlled pilot trial comparing the impact of access to clinical endocrinology video demonstrations with access to usual revision resources on medical student performance of clinical endocrinology skills. BMC Med Educ 2013 Oct 3;13:135. [doi: 10.1186/1472-6920-13-135] [Medline: 24090039]
31. Bredenkamp N. Simulation versus online learning: effects on knowledge acquisition, knowledge retention, and perceived effectiveness. In: University of Northern Colorado PhD Dissertation. Denver: University of Northern Colorado; May 2013:E.
32. Sperl-Hillen J, O'Connor PJ, Ekstrom HL, Rush WA, Asche SE, Fernandes OD, et al. Educating resident physicians using virtual case-based simulation improves diabetes management: a randomized controlled trial. Acad Med 2014 Dec;89(12):1664-1673 [FREE Full text] [doi: 10.1097/ACM.0000000000000406] [Medline: 25006707]
33. Ngamruengphong S, Horsley-Silva JL, Hines SL, Pungpapong S, Patel TC, Keaveny AP. Educational intervention in primary care residents' knowledge and performance of hepatitis B vaccination in patients with diabetes mellitus. South Med J 2015 Sep 9;108(9):510-515. [doi: 10.14423/SMJ.0000000000000334] [Medline: 26332473]
34. Diehl LA, Souza RM, Gordan PA, Esteves RZ, Coelho IC. InsuOnline, An electronic game for medical education on insulin therapy: a randomized controlled trial with primary care physicians. J Med Internet Res 2017 Dec 9;19(3):e72 [FREE Full text] [doi: 10.2196/jmir.6944] [Medline: 28279950]
35. Lee MC. Explaining and predicting users' continuance intention toward e-learning: an extension of the expectation–confirmation model. Comput Educ 2010 Feb;54(2):506-516. [doi: 10.1016/j.compedu.2009.09.002]
36. Andrade J, Ares J, García R, Rodríguez S, Seoane M, Suárez S. Guidelines for the development of e-learning systems by means of proactive questions. Comput Educ 2008 Dec;51(4):1510-1522. [doi: 10.1016/j.compedu.2008.02.002]
37. Walsh K. The future of e-learning in healthcare professional education: some possible directions. Commentary. Ann Ist Super Sanita 2014;50(4):309-310 [FREE Full text] [Medline: 25522068]
38. Scott R, Mars M, Hebert M. How global is 'e-Health' and 'Knowledge Translation'? In: Ho K, Jarvis-Selinger S, Novak Lauscher H, Cordeiro J, Scott R, editors. Technology Enabled Knowledge Translation for eHealth. New York, NY: Springer New York; 2012:339-357.
39. Al-Shorbaji N, Atun R, Car J, Majeed A, Wheeler E. World Health Organization. 2015. eLearning for undergraduate health professional education: a systematic review informing a radical transformation of health workforce development URL: https://www.who.int/hrh/documents/14126-eLearningReport.pdf [accessed 2019-01-30] [WebCite Cache ID 75oA808uE]
40. Hernández-Torrano D, Ali S. First year medical students' learning style preferences and their correlation with performance in different subjects within the medical course. BMC Med Educ 2017 Aug;17(1):131. [doi: 10.1186/s12909-017-0965-5] [Medline: 28789690]
41. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. Acad Med 2010 May;85(5):909-922. [doi: 10.1097/ACM.0b013e3181d6c319] [Medline: 20520049]
42. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. J Am Med Assoc 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]

Abbreviations

PCP: primary care physician
RCT: randomized controlled trial
RR: risk ratio
SMD: standardized mean difference

Edited by A Marusic; submitted 05.12.18; peer-reviewed by P O'Connor, M Mahmic Kaknjo, J Brož, H Zamanian; comments to author 29.12.18; revised version received 12.01.19; accepted 12.01.19; published 21.02.19.

Please cite as:
Huang Z, Semwal M, Lee SY, Tee M, Ong W, Tan WS, Bajpai R, Tudor Car L
Digital Health Professions Education on Diabetes Management: Systematic Review by the Digital Health Education Collaboration
J Med Internet Res 2019;21(2):e12997
URL: http://www.jmir.org/2019/2/e12997/
doi:10.2196/12997
PMID:30789348

Review

Online Digital Education for Postregistration Training of Medical Doctors: Systematic Review by the Digital Health Education Collaboration

Pradeep Paul George[1,2,3], BSMS, MSc, PhD; Olena Zhabenko[4], MD, PhD; Bhone Myint Kyaw[5], MSc, MBBS, PhD; Panagiotis Antoniou[6], PhD; Pawel Posadzki[4], PhD; Nakul Saxena[7], PhD; Monika Semwal[4], PhD; Lorainne Tudor Car[5], MSc, PhD, MD; Nabil Zary[8,9,10], PhD; Craig Lockwood[2], PhD; Josip Car[4,11], MD, PhD, DIC, MSc, FFPH, FRCP (Edin)

[1]Health Services and Outcomes Research, National Healthcare Group, Singapore, Singapore

[2]Joanna Briggs Institute, University of Adelaide, Adelaide, Australia

[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[4]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[5]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[6]Laboratory of Medical Physics, Aristotle University of Thessaloniki, Thessaloníki, Greece

[7]Ophthalmology Team, Novartis, Singapore, Singapore

[8]Medical Education Research and Scholarship Unit, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[9]Department of Learning, Informative, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[10]10I Emerging Technologies Lab, Mohammed VI University of Health Sciences, Casablanca, Morocco

[11]Department of Family Medicine, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

Corresponding Author:
Josip Car, MD, PhD, DIC, MSc, FFPH, FRCP (Edin)
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University
Experimental Medicine Building
59 Nanyang Drive
Singapore,
Singapore
Phone: 65 6592 3960
Email: josip.car@ntu.edu.sg

Abstract

Background: Globally, online and local area network–based (LAN) digital education (ODE) has grown in popularity. Blended learning is used by ODE along with traditional learning. Studies have shown the increasing potential of these technologies in training medical doctors; however, the evidence for its effectiveness and cost-effectiveness is unclear.

Objective: This systematic review evaluated the effectiveness of online and LAN-based ODE in improving practicing medical doctors' knowledge, skills, attitude, satisfaction (primary outcomes), practice or behavior change, patient outcomes, and cost-effectiveness (secondary outcomes).

Methods: We searched seven electronic databased for randomized controlled trials, cluster-randomized trials, and quasi-randomized trials from January 1990 to March 2017. Two review authors independently extracted data and assessed the risk of bias. We have presented the findings narratively. We mainly compared ODE with self-directed/face-to-face learning and blended learning with self-directed/face-to-face learning.

Results: A total of 93 studies (N=16,895) were included, of which 76 compared ODE (including blended) and self-directed/face-to-face learning. Overall, the effect of ODE (including blended) on postintervention knowledge, skills, attitude, satisfaction, practice or behavior change, and patient outcomes was inconsistent and ranged mostly from no difference between the groups to higher postintervention score in the intervention group (small to large effect size, very low to low quality evidence). Twenty-one studies reported higher knowledge scores (small to large effect size and very low quality) for the intervention, while

20 studies reported no difference in knowledge between the groups. Seven studies reported higher skill score in the intervention (large effect size and low quality), while 13 studies reported no difference in the skill scores between the groups. One study reported a higher attitude score for the intervention (very low quality), while four studies reported no difference in the attitude score between the groups. Four studies reported higher postintervention physician satisfaction with the intervention (large effect size and low quality), while six studies reported no difference in satisfaction between the groups. Eight studies reported higher postintervention practice or behavior change for the ODE group (small to moderate effect size and low quality), while five studies reported no difference in practice or behavior change between the groups. One study reported higher improvement in patient outcome, while three others reported no difference in patient outcome between the groups. None of the included studies reported any unintended/adverse effects or cost-effectiveness of the interventions.

Conclusions: Empiric evidence showed that ODE and blended learning may be equivalent to self-directed/face-to-face learning for training practicing physicians. Few other studies demonstrated that ODE and blended learning may significantly improve learning outcomes compared to self-directed/face-to-face learning. The quality of the evidence in these studies was found to be very low for knowledge. Further high-quality randomized controlled trials are required to confirm these findings.

(J Med Internet Res 2019;21(2):e13269) doi:10.2196/13269

KEYWORDS

randomized controlled trials; effectiveness; systematic review; medical education; internet

Introduction

Information communication technology (ICT) has transformed the way information is exchanged and shared around the world [1-4]. In medical education, ICT facilitated a paradigm shift from traditional learning to a dynamic system, moving away from the instructor- or student-focused presentation session to a student-centered process, where students can learn anywhere, anytime, and at their own pace. It also provides unique opportunities for interactive communication and networking [5].

Online and local area network–based (LAN) digital education (ODE) encompasses a variety of interventions characterized by their tools, content, learning objectives, pedagogical approaches, and delivery settings. ODE also varies widely in its configuration (eg, tutorial, asynchronous discussion, and live conferencing), instructional methods (eg, practice exercises and cognitive interactivity), and presentation [6]. ODE uses a full electronic approach, which is entirely driven by technology, or a mix of traditional learning and digital technology (ie, blended learning). Blended learning may be more suitable for health care training, which commonly needs to combine hands-on skill-based training at a practical level and self-directed learning such as ODE [7-9].

ODE has been used widely in undergraduate medical and other health professionals' education [10] and is now gaining popularity in postregistration medical education for lifelong learning (ie, continuing education), evidenced by the growing number of studies. Continuing medical education (CME) is defined as "all educational activities which serve to maintain, develop, or increase the knowledge, skills, and professional performance and relationships that a physician used to provide services for patients, the public, or the profession" [11] and continuing professional development (CPD) is defined as "a range of learning activities through which medical professionals maintained and developed throughout their career to ensure that they retain their capacity to practice safely, effectively and legally within their evolving scope of practice" [12]. Recently, nearly all medical schools in the United States and Canada moved to providing some form of online learning material as part of their CME for physicians [6].

Research shows that learning is influenced more by the content and instructional strategy than by the type of technology used to deliver the content [13]; in other words, the design of a course determines its effectiveness in learning [14]. There is a significant methodological, educational, and clinical heterogeneity amongst the studies [15-37], which highlighted the need for a review on ODE that focused specifically on the education of medical doctors with more homogenous learning technologies. The *a priori* protocol reported here has also been published in the Cochrane library [38].

The primary objective of this review was to evaluate the effectiveness of ODE in improving doctors' knowledge, skills, attitude, and satisfaction. The secondary objectives were to assess changes in clinical practices or behaviors, patient outcomes, costs and cost-effectiveness of the intervention, and unintended/adverse effects on patients and physicians.

Methods

Search Strategy and Data Sources

A search strategy was developed in accordance with the Cochrane Handbook of Systematic Reviews of Interventions [39] to search the Cochrane CENTRAL, MEDLINE (Ovid), Embase (Elsevier), PsycINFO (Ovid), ERIC (Ovid), CINAHL (Ebsco), Web of Science Core Collection (Thomson Reuters), and International Clinical Trials Platform (World Health Organization) databases. The detailed search strategy is presented in Multimedia Appendix 1, and a detailed description of the methodology has been published elsewhere [40]. Databases were searched from January 1, 1990, to March 9, 2017. We selected 1990 as the starting year for our search because prior to this, the use of ICT for education was limited. We identified additional studies both by scanning relevant systematic reviews and meta-analyses and hand searching reference lists of all included studies.

Selection Criteria

Only randomized controlled trials (RCTs), cluster RCTs (cRCTs), and quasi-randomized controlled trials of postregistration education for medical doctors using ODE (standalone or blended) with any type of controls measuring knowledge, cognitive skills, attitudes, satisfaction (primary outcome), changes in practice or behavior, patient outcomes, costs, or adverse effects (secondary outcome) outcomes were eligible for inclusion in this review (Multimedia Appendix 2). Participants were not excluded on the basis of age, gender, or any other sociodemographic variables. We used the gold-standard systematic review methods recommended by the Cochrane Collaboration and strictly adhered to the published protocol [38].

Data Extraction

Three reviewers (PG, OZ, and BK) independently screened the titles and abstracts and full-text versions of the eligible studies and performed the data extraction. We extracted the relevant data on participants, intervention and control, outcome measures, and study designs. We contacted the study authors in cases of any missing information. A fourth review author (PP) acted as an arbiter in cases of disagreement.

Risk of Bias Assessment

Two reviewers independently assessed the risk of bias for RCTs using the Cochrane Collaboration's "risk of bias" tool [39]. For RCTs, we did so across the domains of random sequence generation, allocation concealment, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other bias including the comparability of intervention and control group; characteristics at baseline; validity and reliability of outcome assessment tools; and protection against contamination. Blinding of participants and personnel was not assessed, as the nature of the intervention precludes blinding.

We assessed the risk of bias for cRCTs across the domains of recruitment bias [41], baseline imbalances, loss of clusters incorrect analysis, and comparability with individual randomized trials [39]. For each study, two reviewers independently categorized each domain as low, high, or unclear risk of bias.

Assessment of Heterogeneity

Clinical heterogeneity was assessed to check if the included studies were similar in terms of their population, intervention characteristics, and reported outcomes and to ascertain the possibility of pooling the measures of effect. The extracted data were analyzed using RevMan 5.3 software (The Nordic Cochrane Centre, Copenhagen, Denmark). Statistical heterogeneity was assessed using the Chi-square and I^2 tests [39]. We found significant heterogeneity (clinical and statistical) among the included studies; hence, meta-analysis was not suitable for analysis.

Data Synthesis

The results from individual RCTs were reported as the standardized mean difference (SMD) for continuous variables and risk ratios (RR) for dichotomous variables. Where studies reported more than one measure for each outcome, the primary measure, as defined by the study authors, was used in the analysis. If studies had multiple arms, we compared the intervention arm to the least active control arm and assessed the difference in postintervention outcomes. Similarly, when multiple domains of the same outcome were measured, only the primary domains identified and agreed upon by the review authors were reported. Meta-analyses were not possible because there was significant clinical and methodological heterogeneity across the included studies.

Summary of Findings

Summary of findings tables (Tables 1-3) were prepared based on the methods described in Chapter 11 of the Cochrane Handbook for Systematic Reviews of Interventions [39]. Two review authors (PG and BK) independently used the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) criteria to rank the quality of the evidence using the GRADE profiler (GRADEpro) software [42]. In the main review, we only compared ODE with self-directed/face-to-face learning and blended learning with self-directed/face-to-face learning; the rest of the comparisons are presented in Multimedia Appendix 3.

Table 1. Summary of findings for online and local area network–based digital education as compared to self-directed learning. patient or population: postregistration medical doctors; setting: universities, hospitals, and primary care; intervention: online and local area network–based digital education; comparison: self-directed learning.

Outcomes	Number of participants (number of RCTs[a])	Quality of evidence (GRADE[b])	Direction of effects
Knowledge assessed with multiple-choice questions. Follow-up ranged from posttest to 1 year	3067 (29)	Very low[c,d,e,f]	Seventeen studies [43-60] reported that ODE[g] was significantly more effective than self-directed learning (very low certainty evidence). Two studies [61,62] reported mixed results (very low certainty evidence). Ten studies [63-72] reported that ODE was as effective as self-directed learning (very low certainty evidence).
Skills assessed with OSCE[h], diagnostic assessment, examination, questionnaires, and surveys. Follow-up ranged from posttest to 4 years	829 (8)	Low[c,d,i]	Five studies [65,73-76] reported that ODE was significantly more effective than self-directed learning (low certainty evidence). Two studies [77,78] reported that ODE was as effective as self-directed learning (low certainty evidence). One study [54] reported self-directed learning was more effective than ODE (low certainty evidence).
Attitude assessed with questionnaires. Follow-up ranged from posttest to 136 days	392 (4)	Low[c,d]	One study [47] reported that ODE was significantly more effective than self-directed learning (low certainty evidence). Another [66] reported that ODE was as effective as self-directed learning (low certainty evidence). Two studies [44,58] reported mixed results (low certainty evidence).
Satisfaction assessed with questionnaires. Follow-up ranged from posttest to 6 months	934 (6)	Low[c,d]	Two studies [67,79] reported that ODE was significantly more effective (low certainty evidence). Three studies [54,58,80] reported that ODE was as effective as self-directed learning (low-certainty evidence). One study [61] reported mixed results (low certainty evidence).

[a]RCT: randomized controlled trial.

[b]GRADE: Grading of Recommendations, Assessment, Development and Evaluations.

[c]Rated down by one level for study limitations. Most studies were considered to be at an unclear or high risk of bias. Overall, the risk of bias for most studies was unclear due to a lack of information reported.

[d]Rated down by one level for inconsistency. There was variation in effect size (ie, very large and very small effects were observed).

[e]Rated down by one level for publication bias. The effect estimates were asymmetrical, suggesting possible publication bias.

[f]Very low quality (+ – – –): We have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect.

[g]ODE: online and local area network–based digital education.

[h]OSCE: objective structured clinical examination.

[i]Low quality (+ + – –): Our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect

Table 2. Summary of findings for online digital education as compared to face-to-face learning. patient or population: postregistration medical doctors; setting: universities, hospitals, and primary care; intervention: online and local area network–based digital education; comparison: face-to-face learning.

Outcomes	Number of participants (number of RCTs[a])	Quality of evidence (GRADE[b])	Direction of effects
Knowledge assessed with multiple-choice questions. Follow-up ranged from posttest to 18 months	1202 (9)	Very low[c,d,e,f]	Two studies [81,82] reported that ODE[g] was significantly more effective in improving physicians' knowledge scores than face-to-face learning (very low certainty evidence). Six studies [83-88] found that ODE was as effective as face-to-face learning in improving physicians' knowledge scores (very low certainty evidence). One study [89] reported that face-to-face learning was significantly more effective than ODE in improving physicians' knowledge scores.
Skills assessed with OSCE[h], diagnostic assessment, examination, questionnaires, and surveys. Follow-up ranged from posttest to 12 months	291 (7)	Low[c,d,i]	Six studies [84,87,90-93] reported ODE was as effective as face-to-face learning in improving physicians' skills (low certainty evidence). In one study [94], data were missing.
Attitude assessed with questionnaires. Follow-up ranged from posttest to 18 months	220 (2)	Low[c,d]	Two studies [82,95] reported that ODE was as effective as face-to-face learning in improving physicians' attitude (low certainty evidence).
Satisfaction assessed with questionnaires. Follow-up ranged from posttest to 12 weeks	260 (4)	Low[c,d]	Two studies [83,87] reported that ODE was significantly more effective than face-to-face learning for improving physicians' satisfaction (low certainty evidence). Two studies [81,84] reported that ODE was as effective as face-to-face learning in improving physicians' satisfaction (low certainty evidence).

[a]RCT: randomized controlled trial.

[b]GRADE: Grading of Recommendations, Assessment, Development and Evaluations.

[c]Rated down by one level for study limitations. Most studies were considered to be at an unclear or high risk of bias. Overall, the risk of bias for most studies was unclear due to a lack of information reported.

[d]Rated down by one level for inconsistency. There was variation in effect size (ie, very large and very small effects were observed).

[e]Rated down by one level for publication bias. The effect estimates were asymmetrical, suggesting possible publication bias.

[f]Very low quality (+ – – –): We have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect.

[g]ODE: online and local area network–based digital education.

[h]OSCE: objective structured clinical examination.

[i]Low quality (+ + – –): Our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect

Table 3. Summary of findings for blended learning as compared to self-directed/face-to-face learning. patient or population: postregistration medical doctors; setting: universities, hospitals, and primary care; intervention: blended learning; comparison: self-directed/face-to-face learning.

Outcomes	Number of participants (number of studies)	Quality of evidence (GRADE[a])	Direction of effects
Knowledge assessed with multiple-choice questions. Follow-up ranged from posttest to 26 months	4413 (7 RCTs[b])	Very low[c,d,e,f]	Two studies [96,97] reported that blended learning was significantly more effective in improving physicians' knowledge than self-directed/face-to-face learning (very low certainty evidence). Five studies assessed together [96,98-101] reported that blended learning was as effective as self-directed/face-to-face learning (very low certainty evidence).
Skills assessed with OSCE[g], diagnostic assessment, examination, questionnaires, and surveys. Follow-up ranged from posttest to 26 months.	4131 (6 RCTs)	Low[c,d,h]	Two studies [96,102] reported that blended learning may significantly improve physicians' skills, and four studies [98,99,103,104] reported that blended learning may be as effective as face-to-face learning in improving skills (low certainty evidence).
Attitude assessed with a questionnaire. Follow-up assessed posttest	61 (1 cRCT[i])	Low[c,d]	Kulier et al [105] compared a blended learning course on EBM[j] to a face-to-face EBM course and reported that the intervention may be as effective as the controls for improving physicians' attitude.
Satisfaction assessed with questionnaires on a Likert scale. Follow-up ranged from posttest to 6 months	166 (3 RCTs)	Low[c,d]	Ali et al [98] compared ATLS[k] delivered through blended learning to a standard ATLS course and reported no difference in satisfaction between the groups (low certainty evidence). Kronick et al [106] compared 3 hours of online training to no training (self-directed training) and found that the intervention slightly improved satisfaction (low certainty evidence). Platz et al [100] compared basic ultrasound principles and extended focused assessment with sonography for trauma using blended learning as compared to face-to-face training and reported mixed results (low certainty evidence).

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

[b]RCT: randomized controlled trial.

[c]Rated down by one level for study limitations. Most studies were considered to be at an unclear or high risk of bias. Overall, the risk of bias for most studies was unclear due to a lack of information reported.

[d]Rated down by one level for inconsistency. There was variation in effect size (ie, very large and very small effects were observed).

[e]Rated down by one level for publication bias. The effect estimates were asymmetrical, suggesting possible publication bias.

[f]Very low quality (+ – – –): We have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect.

[g]OSCE: objective structured clinical examination.

[h]Low quality (+ + – –): Our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect.

[i]cRCT: cluster-randomized trial.

[j]EBM: evidence-based medicine.

[k]ATLS: Advanced Trauma Life Support

Results

Search Results

Our searches yielded a total of 27,488 citations and 93 studies (Figure 1). Of those, 74 studies were RCTs including 12,537 participants and 19 were cRCTs including 1262 clusters, 3727 physicians, and 7690 patients. Sixty-four studies were published between 2010 and 2017, and the remaining 29 studies were published between 1999 and 2009.

Participants, Settings, and Countries of Origin

A total of 29 (31.1%) studies were conducted among primary care practitioners (general practitioners, family medicine practitioners/residents, and occupational physicians), 12 among surgeons (12.9%), 11 among general and internal medicine practitioners (11.8%), and 8 among pediatricians (8.6%; Figure 2 and Multimedia Appendices 4-10). Only 2 (2.2%) [83,96] of the 93 studies were conducted in low- to middle-income countries; all the remaining studies were conducted in high-income countries with majority in the United States (53.8%), Canada (10.8%), and Germany (5.4%; Multimedia Appendix 11). Fifty studies were carried out in hospital settings, 31 studies were conducted in university settings, 11 studies were conducted in primary care settings, and one study was conducted in a mixed hospital and university setting.

Figure 1. Modified Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart of the search results and study-selection process.

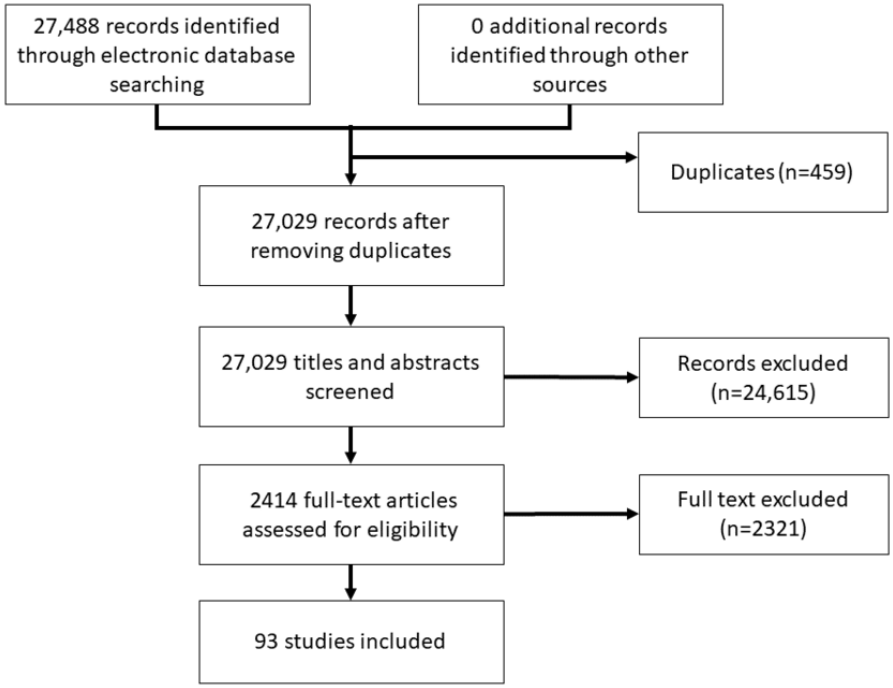
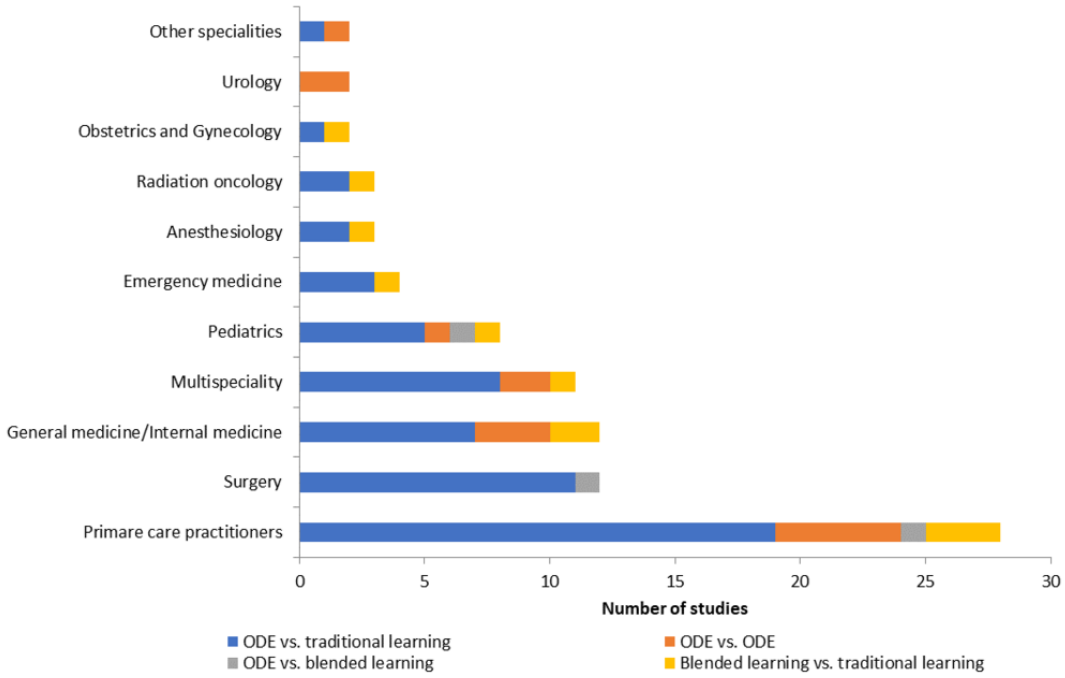


Figure 2. Number of ODE studies by specialty and type of learning. ODE: online and local area network–based digital education.



Interventions and Comparisons

A total of 61 studies compared ODE and self-directed/face-to-face learning. Self-directed learning was defined as self-learning through books and journals, and face-to-face learning was defined as learning through didactic classroom lectures and courses. Fourteen studies compared ODE and other types of ODE, 3 studies compared ODE and blended learning, and 15 studies compared blended learning and self-directed/face-to-face learning. Two studies used synchronous learning technology (video-conferencing systems) for training, and 39 studies used asynchronous learning technologies such as Web-based libraries/repositories of video modules, CD-ROM, emails, and online discussion groups to deliver the intervention. In the main review, we only compared ODE with self-directed/face-to-face learning and blended learning with self-directed/face-to-face learning; the rest of the comparisons are presented in Multimedia Appendix 3.

Risk of Bias in Randomized Controlled Trials and Cluster Randomized Controlled Trials

A total of 51 studies were considered to be at high risk of bias for at least one of the risk of bias domains (Figures 3 and 4). Six studies were rated as having a high risk of selection bias, 31 studies were rated as having a high risk of attrition bias due to a high drop-out rate (>20%), and 3 studies [103,107,108] had a high risk of reporting bias. Epstein et al reported a high risk of detection bias [109]. Further, 25 studies had a high risk of "other biases." Similarly, among the cRCTs, 12 studies had a high risk of bias for baseline imbalance, 8 studies had a high risk of bias for loss of clusters, and 3 studies had a high risk of bias for incorrect analyses. Risk of bias is described in detail in Multimedia Appendices 4 and 12.

Effects of Interventions by Outcomes

The characteristics of included studies categorized by participants' specialty, outcomes, comparisons, and intervention types are presented in Multimedia Appendices 5-13. The educational content was heterogeneous among the included studies. Studies that compared ODE/blended ODE with self-directed or face-to-face learning are presented in the manuscript; for other comparisons, see Multimedia Appendix 3.

Primary Outcomes

Knowledge

A total of 54 studies assessed knowledge: 20 studies used questionnaires (open ended), 28 studies used multiple-choice questions, and 6 studies did not specify the type of instrument used to measure knowledge.

Online and Local Area Network–Based Digital Education Versus Self-Directed Learning

A total of 29 studies compared ODE and self-directed learning; of these studies, only 18 studies reported numerical data in a format that could be used (Figures 5 and 6). Eleven studies [43-50,61,63,64] presented incomplete data (missing means, SDs, or CIs), which could not be included in the data analysis. Seventeen studies (n=2107) [43-60] reported that the ODE was significantly more effective than self-directed learning (small to large effect size, very low certainty evidence). Nine studies (n=796) reported that ODE was as effective as self-directed learning (very low certainty evidence). Two studies [61,62] reported mixed results (very low certainty evidence).

Figure 3. Risk-of-bias summary for each included study.

Figure 4. Risk-of-bias item results presented as percentages across all included studies.

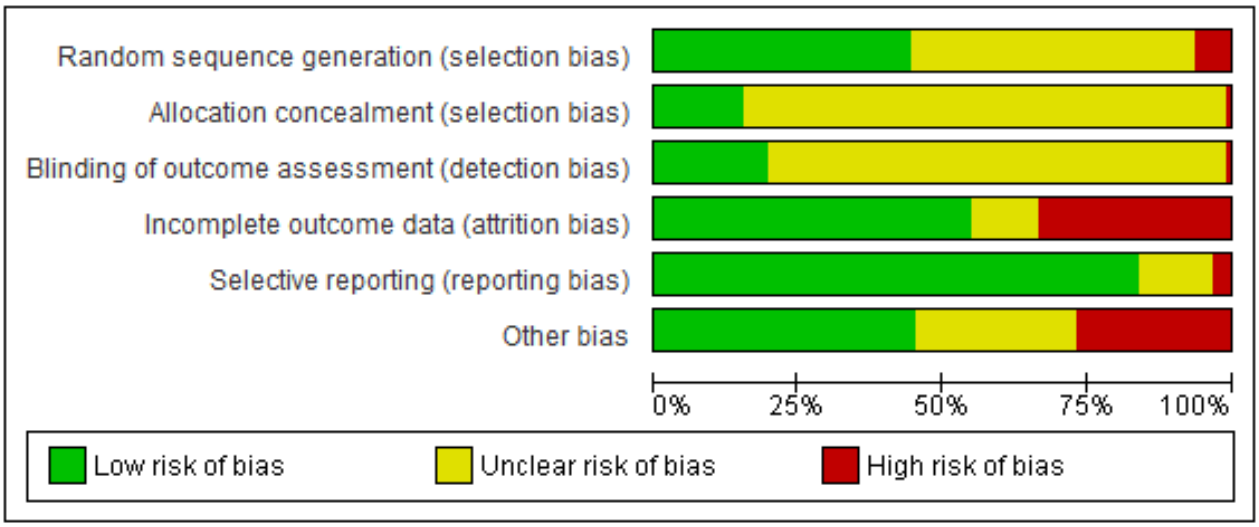
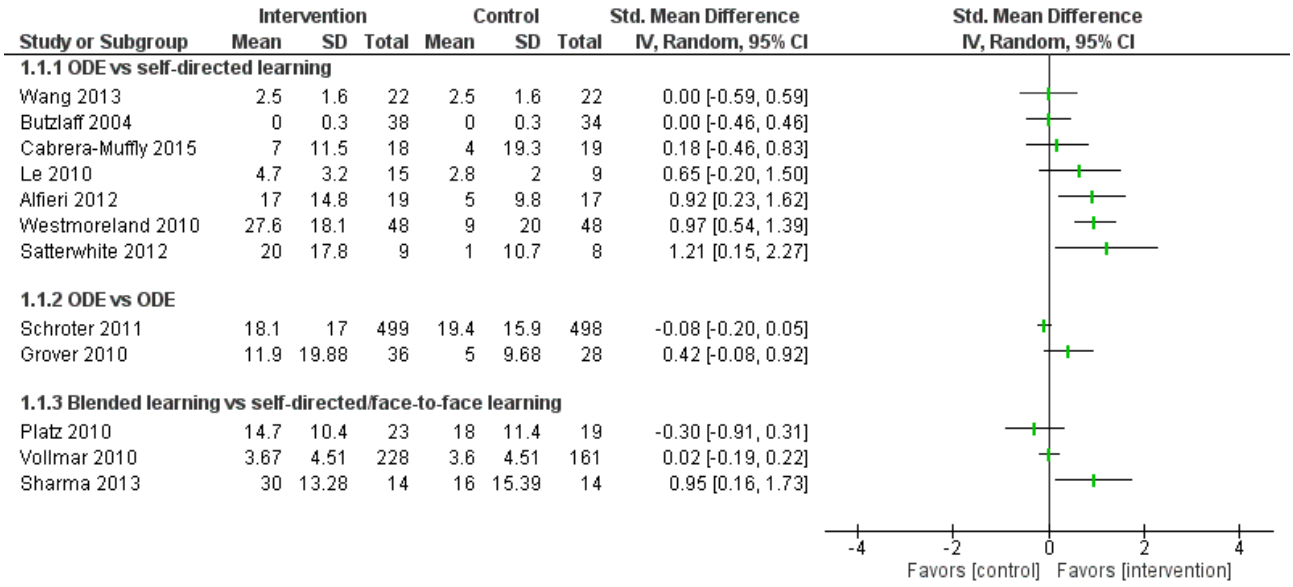


Figure 5. Comparison of change in knowledge scores (postintervention). ODE: online and local area network–based digital education; IV: inverse variance.



Online and Local Area Network–Based Digital Education Versus Face-to-Face Learning

Nine studies compared ODE with face-to-face learning; of these, only four studies [83-86] reported numerical data in a useable format (Figure 6). ODE improved physicians' knowledge compared with face-to-face learning in studies by Fordis et al [81] and Pelayo-Alvarez et al [82]. Six studies [83-88] reported that ODE may be equally effective as face-to-face learning for improving knowledge scores (n=489). McLeod et al [89] reported that knowledge scores were higher with face-to-face learning than with ODE. Overall, empirical evidence suggests that ODE may have little effect on knowledge as compared with face-to-face learning. However, since the evidence was of very low quality, we are uncertain about the true estimate.

Blended Learning Versus Self-Directed/Face-to-Face Learning

Seven studies assessed learners' knowledge, of which two studies [96,97] reported that blended learning was significantly more effective in improving physicians' knowledge than self-directed/face-to-face learning (n=232; large effect size, very low certainty evidence; Figures 5 and 6). Five studies assessed together [96,98-101] reported that blended learning was as effective as self-directed/face-to-face learning for improving physicians' knowledge scores (very low certainty evidence).

Skills

Twenty-one studies assessed participants' skills: five studies used an objective structured clinical examination, five studies used questionnaires, four studies used practical skills test/exam, three studies used checklists, and the remaining used other methods to assess skills (Multimedia Appendix 6).

Online and Local Area Network–Based Digital Education Versus Self-Directed Learning

Eight studies [54,65,73-78] compared ODE with self-directed learning, which included no intervention or text-based learning. Of these, only four studies [65,74,76,78] had numerical information in a presentable format (Figures 7-9). Five studies [65,73-76] reported that ODE was significantly more effective than self-directed learning (low certainty evidence). Two studies [77,78] reported that ODE was as effective as self-directed learning (low certainty evidence). One study [54] reported that self-directed learning was more effective than ODE (low certainty evidence). Overall, empirical evidence from five studies [65,73-76] indicated that ODE interventions can improve physicians' skills as compared to self-directed learning. Similarly, evidence from two [54,77] of the eight studies indicated that ODE may be as effective as self-directed learning for improving physicians' postintervention skills scores.

Figure 6. Comparison of postintervention knowledge scores. ODE: online and local area network–based digital education; IV: inverse variance.

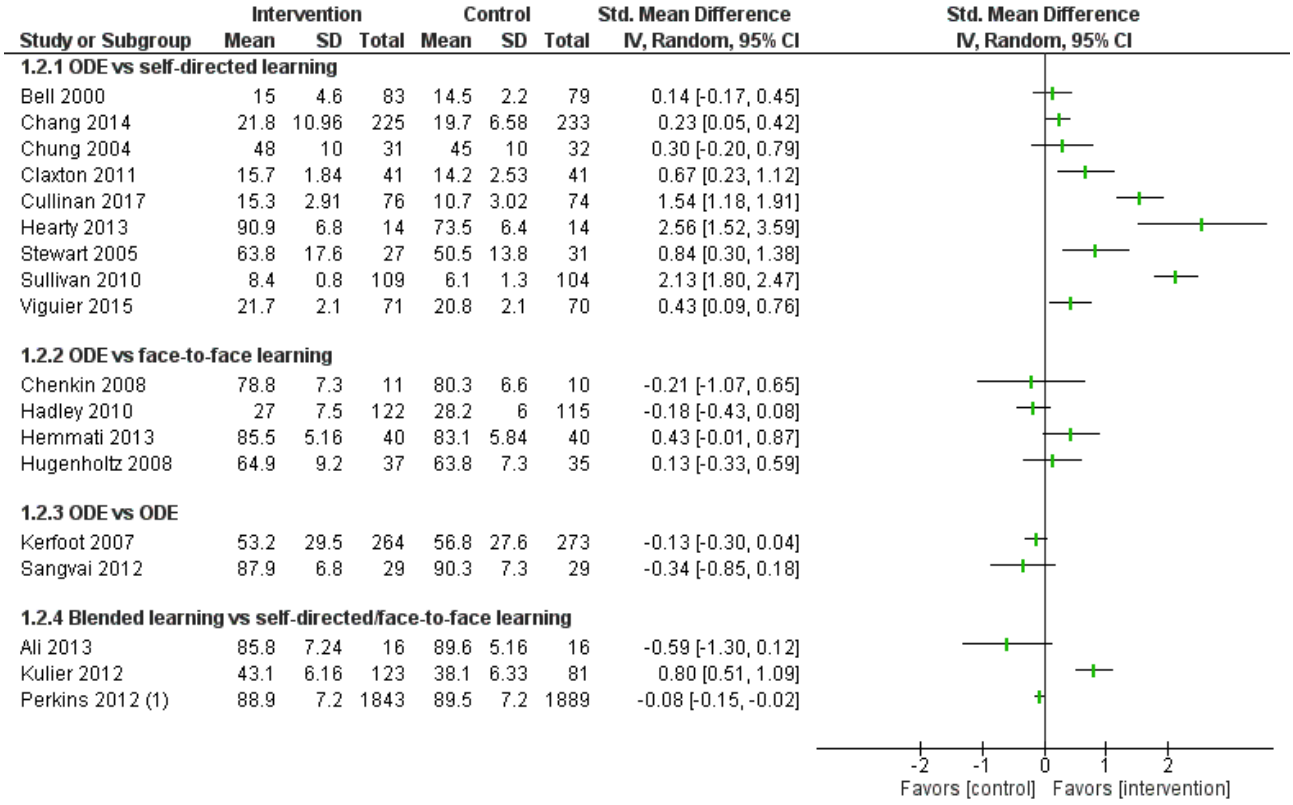


Figure 7. Comparison of change in skills scores (postintervention). ODE: online and local area network–based digital education; IV: inverse variance.

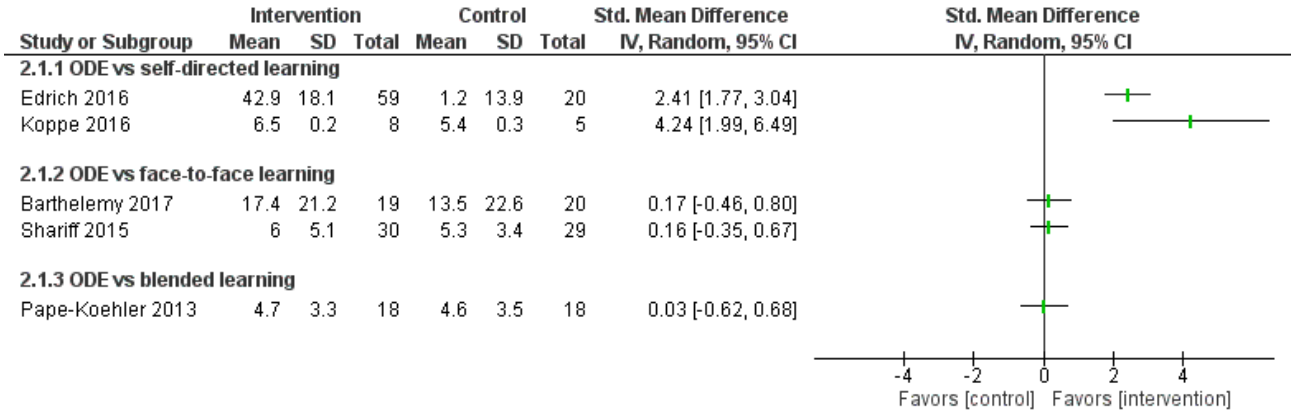


Figure 8. Comparison of postintervention skills scores. ODE: online and local area network–based digital education; IV: inverse variance.

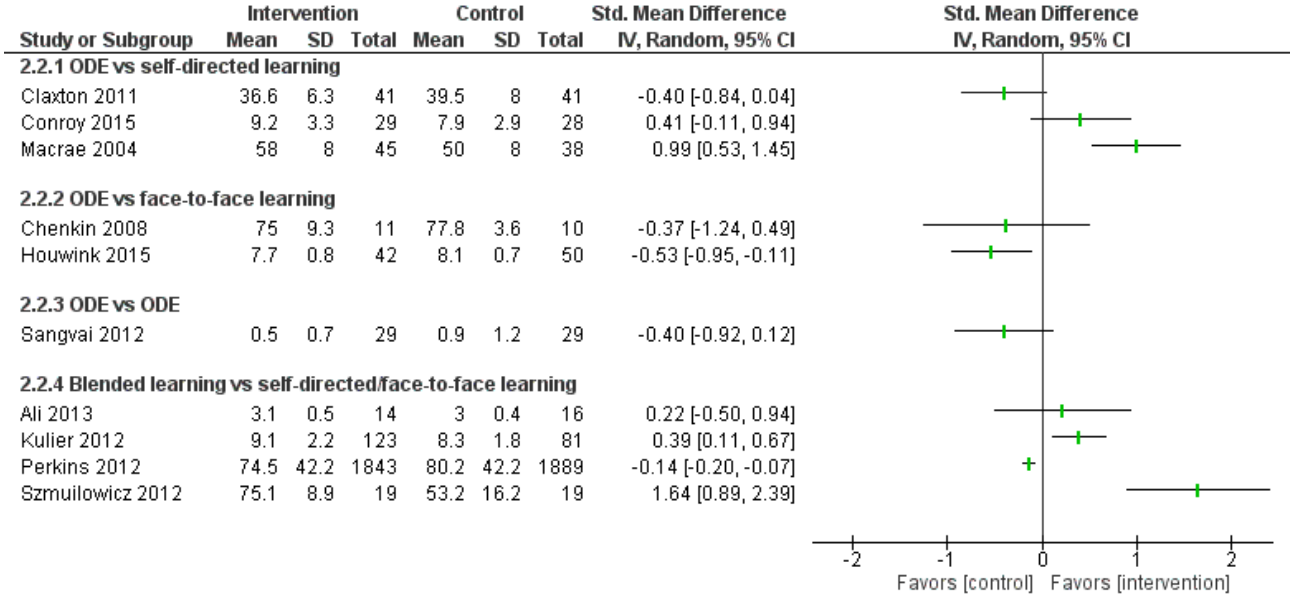
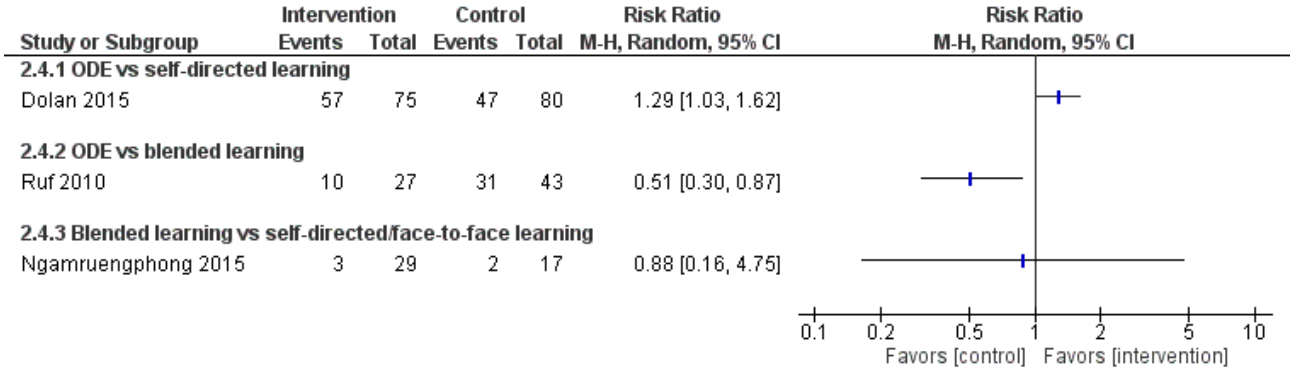


Figure 9. Comparison of postintervention skill scores (dichotomous). ODE: online and local area network–based digital education; M-H: Mantel-Haenszel.



Online and Local Area Network–Based Digital Education Versus Self-Directed Learning

Seven studies [84,87,90-94] compared ODE and face-to-face learning (classroom didactic lecture–based learning), of which six studies [84,87,90-93] reported that ODE was as effective as face-to-face learning in improving physicians' skills (low certainty evidence). In one study [94], data were missing (Figures 7-9). Overall, empirical evidence from six studies suggests that ODE may be as effective as face-to-face learning in improving physicians' skills.

Blended Learning Versus Self-Directed/Face-to-Face Learning

Six studies assessed skills [96,98,99,102-104]: Kulier et al [96] and Szmuilowicz et al [102] reported that blended learning may improve physicians' skills compared with self-directed / face-to-face learning (Figure 8). Overall, empirical evidence from two of the six studies suggests that blended learning may be as effective as self-directed/face-to-face learning in improving physicians' skills (moderate to large effect size; high certainty evidence). Evidence from four [98,99,103,104] of the six studies suggests that blended learning may be as effective as self-directed/face-to-face in learning to improve physicians' skills.

Attitude

Eight studies assessed participants' attitude: six studies used questionnaires and two studies [66,105] used Likert scales.

Online and Local Area Network–Based Digital Education Versus Self-Directed Learning

Four studies [44,47,58,66] compared ODE with self-directed learning, of which only two studies [58,66] reported numerical data. Le et al [66] reported a change in mean attitude scores (SMD 0.46, 95% CI –0.38 to 1.30 [low certainty]) and Sullivan et al [58] assessed attitude posttest (SMD –0.01, 95% CI –0.28 to 0.26 [low certainty]). Harris et al [47] reported that ODE interventions may improve physicians' attitude compared to self-directed learning. Le et al [66] reported that ODE may be as effective as self-directed learning for improving physicians' attitude. Connolly et al [44] and Sullivan et al [58] reported mixed results. Overall, empirical evidence from the four studies reported mixed results for this outcome.

Online and Local Area Network–Based Digital Education Versus Face-to-Face Learning

Two studies [82,95] compared ODE with face-to-face learning (classroom didactic lecture–based learning). Only Putnam et al [95] reported a change in learners' attitude as a dichotomous

outcome (RR 0.94, 95% CI 0.72-1.22 [small effect size, low quality]); the other study did not provide data for inclusion in the analysis. Overall, empirical evidence from the two studies reported mixed results for this outcome.

Blended Learning Versus Self-Directed/Face-to-Face Learning

Kulier et al [105] compared an integrated ODE course with face-to-face training on evidence-based medicine among obstetrics and gynecology residents; another study by Kulier et al [96] compared an integrated ODE course with a self-directed course on evidence-based medicine and assessed attitude scores at baseline only. Kulier et al [105] reported that blended learning may be as effective as face-to-face training for improving physicians' attitude (RR 3.00, 95% CI 0.76-11.88).

Satisfaction

Sixteen studies assessed participants' satisfaction, of which 10 studies used questionnaires and 6 studies used Likert scales.

Online and Local Area Network–Based Digital Education Versus Self-Directed Learning

Six studies [54,58,61,67,79,80] compared ODE with self-directed learning, which included no intervention or text-based learning. Of these, only three studies [58,67,79] reported numerical data in a useable format (Figures 10 and 11). Two studies [58,67] assessed mean satisfaction scores posttest (Figure 10). Bell et al [67] reported higher satisfaction for the ODE group than the self-directed learning group (SMD 0.68, 95% CI 0.36-0.99 [moderate effect size, low quality]). Sullivan et al [58] reported no difference in satisfaction scores between the groups (SMD 0.18, 95% CI –0.09 to 0.45). Similarly, Matzie et al [79] assessed posttest satisfaction as a dichotomous outcome (Figure 11) and reported higher satisfaction in four of the five domains (RR 1.13, 95% CI 1.03-1.23 [small effect size, low certainty evidence]). Overall, empirical evidence from two studies [67,79] suggests that ODE may be as effective as self-directed learning on physicians' satisfaction (moderate to large effect size, low quality). Similarly, evidence from four [54,58,79,80] of the six studies suggests that ODE may be as effective as self-directed learning in improving physicians' satisfaction. Gold et al [61] reported mixed results.

Online and Local Area Network–Based Digital Education Versus Face-to-Face Learning

Four studies [81,83,84,87] compared ODE with face-to-face learning. Of these, two studies [83,84] reported numerical data in a useable format (Figures 10 and 11). Overall, empirical evidence from two [83,87] of the four studies suggested that ODE may be effective compared with face-to-face learning in improving physicians' satisfaction (large effect size, low certainty evidence). Similarly, evidence from two [81,84] of the four studies suggested that ODE may be as effective as face-to-face learning in improving physicians' satisfaction.

Blended Learning Versus Self-Directed/Face-to-Face Learning

Three studies [98,100,106] assessed satisfaction (Figures 10 and 11). Ali et al [98] reported no difference in satisfaction between the groups, while Kronick et al [106] reported mixed effects. Platz et al [100] found higher satisfaction with face-to-face learning compared to blended learning (moderate effect size, very certainty evidence). Overall, empirical evidence from three [98,100,106] studies suggests that the interventions have mixed effects on learners' satisfaction.

Secondary Outcomes

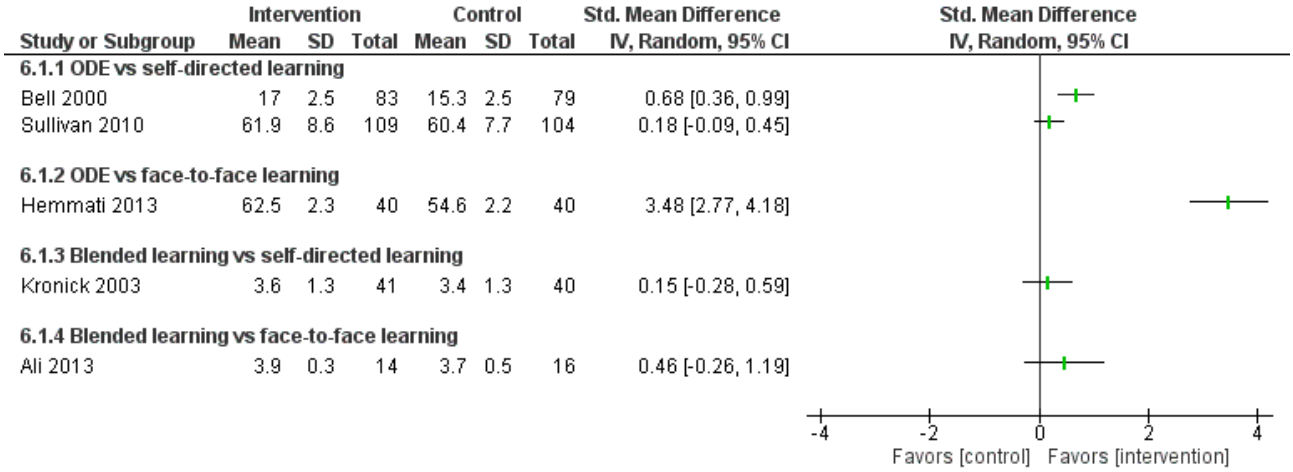
Practice or Behavior Change

Fourteen studies assessed practice or behavior change: four studies used questionnaires; four studies used hospital chart audits or case note reviews; and four other studies used a Likert scale [66], an Intimate Partner Violence Survey scale [64], and patient data from an administrative database and a scenario-based decision-support system [110]. Two studies [111,112] did not state the assessment tools used to measure practice or behavior change.

Online and Local Area Network–Based Digital Education Versus Self-Directed Learning

Fourteen studies [43,45,62,64,66,110-118] compared ODE with self-directed learning, which included no intervention or text-based learning. Of these, only nine studies [62,66,110-113,115,117,118] reported numerical data in a useable format (Figures 12-14).

Figure 10. Comparison of postintervention satisfaction scores. ODE: online and local area network–based digital education; IV: inverse variance.



Overall, empirical evidence from 7 [43,45,110,112,114,116,117] of the 14 studies suggests that ODE may be more effective than self-directed learning in improving physicians' practice or behavior change (moderate to large effect size, very low certainty evidence). Evidence from 4 [62,66,113,115] of the 14 studies suggests that ODE may be as effective as self-directed learning in improving physicians' practice or behavior change. Three studies [64,111,113] reported mixed results (Figures 12-14).

Fordis et al [81] compared ODE on cholesterol management with face-to-face learning for physicians. The study reported that ODE (online CME) may be as effective as face-to-face learning (live CME) for improving physicians' practice or behavior change (RR 0.58, 95% CI –0.06 to 1.21; Figures 12 and 13).

Figure 11. Comparison of postintervention satisfaction scores (dichotomous). ODE: online and local area network–based digital education; M-H, Mantel-Haenszel.

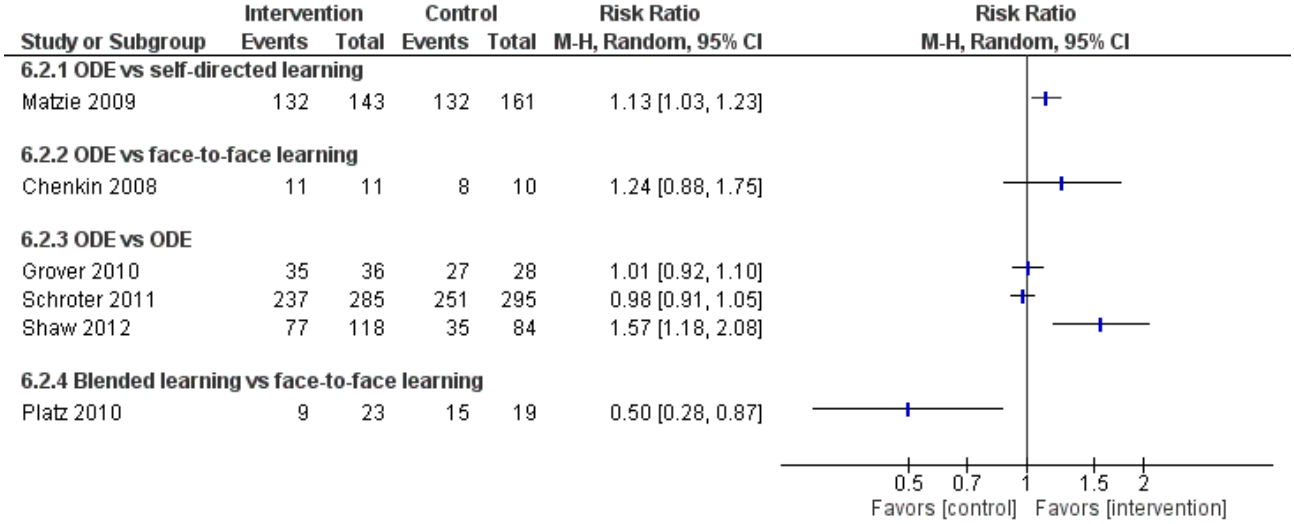


Figure 12. Comparison of practice or behavior change scores (pre-post intervention). ODE: online and local area network–based digital education; IV: inverse variance.

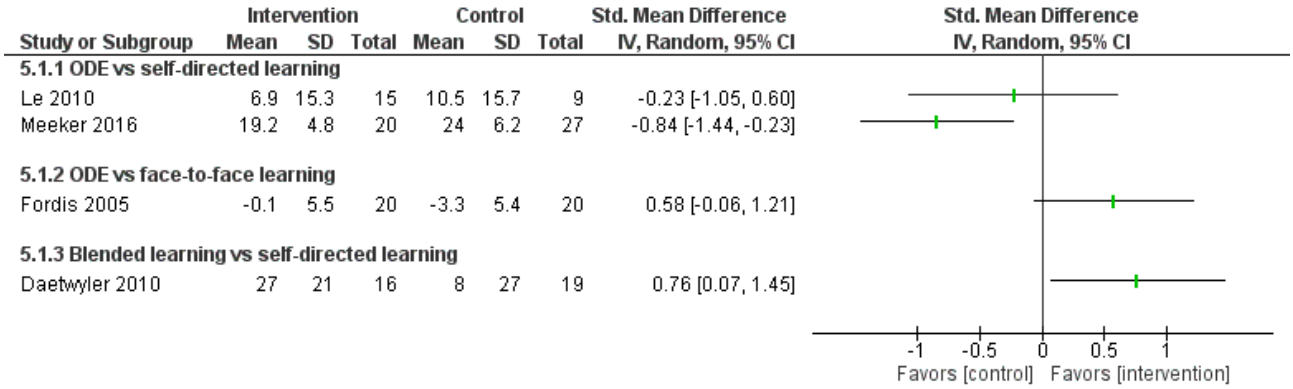


Figure 13. Comparison of postintervention practice or behavior-change scores. ODE: online and local area network–based digital education; IV: inverse variance.

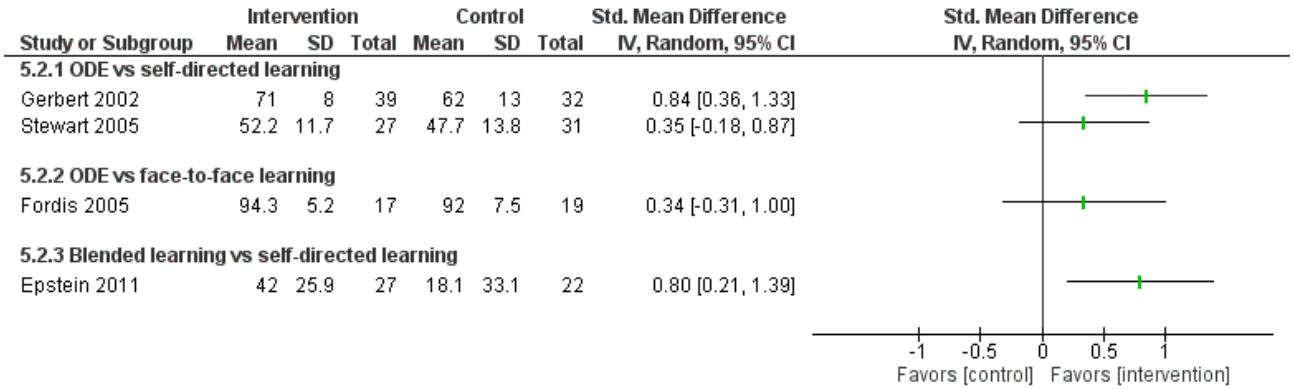
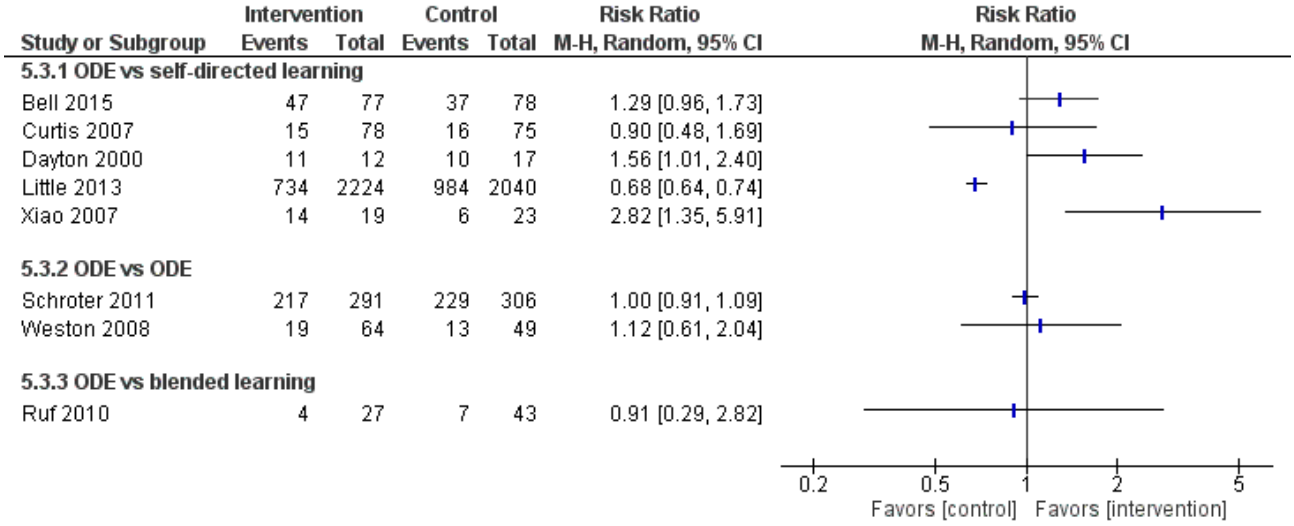


Figure 14. Comparison of postintervention practice or behavior change (dichotomous). ODE: online and local area network–based digital education; M-H: Mantel-Haenszel.



Blended Learning Versus Self-Directed/Face-to-Face Learning

Of the three studies [104,107,109] that assessed practice or behavior change, only two studies [107,109] reported numerical data; these studies reported mixed results for several behavior change outcomes (Figures 12 and 13). Midmer [104] reported that blended learning may be as effective as face-to-face learning in improving physicians' practice or behavior change. Overall, empirical evidence from three studies suggests that blended learning may be as effective as self-directed/face-to-face learning.

Patient Outcomes for All Comparisons

Four studies assessed patient outcomes: These studies used hospital chart audits [65,114,119] and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30.

Four studies [65,114,119,120] compared ODE or blended learning with self-directed/face-to-face learning. Butler [114] reported that ODE may be as effective as self-directed learning for improving patient outcomes. Dolan et al [65] reported that ODE may be more effective than self-directed learning for improving patient outcomes (RR 1.07, 95% CI 1.01-1.13). Girgis et al [120] reported that blended learning may be as effective as face-to-face learning for improving patient outcomes (SMD 0.00, 95% CI –0.20 to 0.20). However, Legare et al [119] found that blended learning may be more effective than self-directed learning for improving patient outcomes (RR 2.80, 95% CI 1.44-5.44). Overall, empirical evidence from the two studies [65,114,120] suggests that ODE may be as effective as self-directed/face-to-face learning in improving patient outcomes. In contrast, empirical evidence from two studies [65,119] suggests that ODE/blended learning may be more effective than self-directed learning.

Cost

Three studies reported the cost of the ODE interventions [43,99,114]. Braido et al [43] performed inter- or intragroup comparisons and a cost-minimization analysis and reported pharmaceutical cost containment of 29% in the ODE group compared to the self-directed learning group; however, the spending on diagnostic investigations increased by 13.4% in the ODE group and reduced by 24.4% in the control group. Butler et al [114] presented cost information for the Stemming the Tide of Antibiotic Resistance educational program and reported greater reduction in costs postintervention in the ODE group compared to self-directed learning (intervention: £120.76; control: £2.21 per 1000 patients). Perkins et al [99] presented the cost of the Advanced Life Support training program and reported that faculty, catering, and facility costs were 47% lower for the blended learning group than for the conventional training group.

No studies reported on the adverse or unintended effects of ODE or blended interventions.

Discussion

Principal Findings

The review identified a variety of ODE interventions used for postregistration training of medical doctors; these interventions were used among diverse medical specialties, and a range of outcomes and comparators were reported. Because of the high heterogeneity, we were unable to pool the data quantitatively.

A total of 93 RCTs were included, of which 76 studies compared ODE/blended learning with self-directed/face-to-face learning (N=12,424). The results are presented in the main review, and the rest of the comparisons are presented in the Multimedia Appendices 1-13. Among these, 21 studies with 2611 participants reported higher postintervention knowledge scores (small to large effect size, very low quality) for the intervention group, while 20 studies with 5496 participants reported no difference in knowledge scores between the groups. Seven studies with 794 participants reported higher postintervention skill scores in the intervention group (large effect size, low quality), while 13 studies with 4447 participants reported no difference in skill scores between the groups. One study with 99 participants reported higher postintervention attitude scores with the intervention (very low quality), while 4 studies with

305 participants reported no difference in attitude scores between the groups. Four studies with 677 participants reported higher postintervention physician satisfaction with the intervention (large effect size, low quality), while 6 studies with 478 participants reported no difference in satisfaction between the groups. Eight studies with 5051 participants reported higher postintervention practice or behavior change for the ODE group (small to moderate effect size, low quality), while 5 studies with 377 participants reported no difference in practice or behavior change between the groups. One study with 449 participants reported higher improvement in patient outcome, while 3 studies with 667 participants reported no difference in patient outcome for the groups. None of the included studies reported any unintended/adverse effects of the interventions on learners.

Overall, the effect of ODE on postintervention knowledge, skills, attitude, satisfaction, practice or behavior change, and patient outcomes was inconsistent and ranged mostly from no difference between the intervention groups to higher postintervention score in the ODE/blended learning group (small to large effect size, very low to low quality evidence). Moreover, the quality of evidence according to GRADE criteria was judged to be low for most outcomes.

George et al assessed the effectiveness of ODE for undergraduate health professionals [10] and reported that ODE is equivalent, and possibly superior, to traditional learning. We are not aware of any other systematic reviews of RCTs that have evaluated ODE for medical doctors. A similar review of evidence from nonrandomized studies on the effectiveness of ODE in surgical education among medical and dental students, surgeons, and oral health specialists [121] reported knowledge gain from ODE compared with active control (face-to-face learning) or no intervention (self-directed learning). Another review [122] evaluated the evidence of effectiveness from nonrandomized studies of online CME for general practitioners alone; assessed their satisfaction, knowledge, clinical practice, and patient outcomes; and reported improvement in satisfaction, knowledge, or practices. Jwayyed et al [123] evaluated the effectiveness of ODE from nonrandomized studies among diverse health care professionals and reported inconsistent results. Our review, in congruence with other reviews [121-124], compared the effect of ODE and blended learning with self-directed/face-to-face learning and other forms of ODEs on physicians' knowledge, skills, attitudes, satisfaction and clinical practice, and patient outcomes, but only included evidence from RCTs and cRCTs.

According to the GRADE criteria, the quality of the evidence was very low for knowledge and low for the other primary and secondary outcomes due to the unclear and high risk of bias, inconsistency, and publication bias. The majority of the studies did not provide information on randomization sequence generation and allocation concealment. Similarly, a high proportion of studies (75 studies, 82%; including comparisons presented in the Multimedia Appendix 12) did not provide sufficient information on the blinding of outcome assessors and were hence judged to have an unclear risk of bias. Thirty-one studies (33.3%) reported incomplete outcome data, 25 studies (26.9%) reported baseline differences in participant characteristics and were judged to be at high risk of bias, and 3 studies (3.3%) had a high risk of reporting bias.

This review has a few limitations. The included studies were heterogeneous in terms of the participants, learning content, and the types of ODE (CME/CPD), thus limiting the opportunity to pool the results and consequently run the preplanned subgroup analysis. Hence, the review could not generate conclusive findings on the effectiveness of ODE. Of the included studies, only two were from low- to middle-income countries, which could limit the completeness of the evidence and its generalizability in all settings. Third, we were unable to assess the cost-effectiveness of ODE as compared to self-directed/face-to-face learning, because none of the identified studies formally assessed it. Only three studies assessed the cost and maintenance of the ODE intervention. Fourth, none of the studies specifically addressed any adverse effects of ODE.

The study has several strengths including a thorough and reproducible search of available literature, independent screening and data extractions, and critical appraisal of the literature conducted in accordance with the Cochrane Handbook for Systematic Reviews of Interventions.

Implications

Given the strength of the evidence of ODE and blended learning, especially for cognitive, procedural, and diagnostic training; evidence-based medicine training; and CME, CPD, and Clinical Practice Guideline training, further high-quality studies of ODE interventions with a validated learning theory are needed to establish their effectiveness, cost-effectiveness, and financial sustainability (return on investment). Specifically, these studies should seek to address several unanswered questions such as ODE's theoretical underpinning, content validation, frequency, intensity, interactivity, technical features, fidelity, safety, cost-effectiveness, adaptability, acceptability, barriers/facilitators to its adoption, financial sustainability, and learner's readiness to switch from classroom learning to complete ODE.

Conclusions

Our review found that ODE and blended learning refers to a group of heterogeneous interventions with different learning theories, learning content, comparators, and outcomes. These interventions were used to train medical doctors in various specialties, such as primary care practitioners, surgeons, residents, and physicians. Although empirical evidence from a majority of studies shows that ODE and blended learning may improve practicing physicians' knowledge, skills, attitude, satisfaction, practice or behavior change, and patient outcomes, few other studies showed that they may be comparable to self-directed, face-to-face learning. The quality of the evidence in these studies was found to be low and very low for knowledge. Therefore, further high-quality RCTs are required before the evidence of efficacy can be concluded for postregistration training of medical doctors.

Acknowledgments

This study is part of PG's doctoral study. He gratefully acknowledges the funding received from the National Healthcare Group, Singapore, for his doctoral research at the University of Adelaide.

This review is conducted in collaboration with the Health Workforce Department of the World Health Organization. We would also like to thank Mr Carl Gornitzki, Ms GunBrit Knutssön, and Mr Klas Moberg from the University Library, Karolinska Institutet, Sweden, for developing the search strategy. We gratefully acknowledge funding from the Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore (eLearning education grant for health professionals).

Authors' Contributions

JC and NZ conceived the idea for the review. PP and OZ wrote the review. LC and PP provided methodological guidance, drafted some of the methodology-related sections, and critically revised the review. BK, PA, NS, MS, NZ, and CL provided comments on and edited the review.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Medline (Ovid) search strategy.

[PDF File (Adobe PDF File), 47KB - jmir_v21i2e13269_app1.pdf]

Multimedia Appendix 2

Inclusion and exclusion criteria.

[PDF File (Adobe PDF File), 43KB - jmir_v21i2e13269_app2.pdf]

Multimedia Appendix 3

Effects of the intervention (other comparisons).

[PDF File (Adobe PDF File), 18KB - jmir_v21i2e13269_app3.pdf]

Multimedia Appendix 4

Risk of bias for cluster randomized controlled trials.

[PDF File (Adobe PDF File), 26KB - jmir_v21i2e13269_app4.pdf]

Multimedia Appendix 5

Characteristics of included studies assessing knowledge.

[PDF File (Adobe PDF File), 32KB - jmir_v21i2e13269_app5.pdf]

Multimedia Appendix 6

Characteristics of included studies assessing skills.

[PDF File (Adobe PDF File), 24KB - jmir_v21i2e13269_app6.pdf]

Multimedia Appendix 7

Characteristics of included studies assessing attitude.

[PDF File (Adobe PDF File), 18KB - jmir_v21i2e13269_app7.pdf]

Multimedia Appendix 8

Characteristics of included studies assessing satisfaction.

[PDF File (Adobe PDF File), 20KB - jmir_v21i2e13269_app8.pdf]

Multimedia Appendix 9

Characteristics of included studies assessing practice or behavior change.

[PDF File (Adobe PDF File), 24KB - jmir_v21i2e13269_app9.pdf]

Multimedia Appendix 10

Characteristics of included studies assessing patient outcomes.

[PDF File (Adobe PDF File), 18KB - jmir_v21i2e13269_app10.pdf]

Multimedia Appendix 11

Online and local area network–based digital education intervention type by participants' specialty, country of publication, and intervention duration.

[PDF File (Adobe PDF File), 29KB - jmir_v21i2e13269_app11.pdf]

Multimedia Appendix 12

Risk of bias.

[PDF File (Adobe PDF File), 17KB - jmir_v21i2e13269_app12.pdf]

Multimedia Appendix 13

Acronyms and definitions.

[PDF File (Adobe PDF File), 14KB - jmir_v21i2e13269_app13.pdf]

Multimedia Appendix 14

Effects of the intervention (other comparisons).

[PDF File (Adobe PDF File), 18KB - jmir_v21i2e13269_app1.pdf]

References

1. Bleakley A, Bligh J, Browne J. Medical Education For The Future. Berlin, Germany: Springer Science & Business Media; 2019.
2. Chen N, Wang Y. Cyber schooling framework: Improving mobility and situated learning. Intl Journal of Eng Ed 2007;23:421-433 [FREE Full text] [doi: 10.1109/ICALT.2005.99]
3. Clark D. Psychological myths in e-learning. Med Teach 2002 Nov;24(6):598-604 [FREE Full text] [doi: 10.1080/0142159021000063916] [Medline: 12623452]
4. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med 2005 Jun;80(6):541-548. [Medline: 15917356]
5. Sangrà A, Vlachopoulos D, Cabrera N. Building an inclusive definition of e-learning: an approach to the conceptual framework. Int Rev Res Open Distance Learn 2012;13:a.
6. Cook DA, Garside S, Levinson AJ, Dupras DM, Montori VM. What do we mean by web-based learning? A systematic review of the variability of interventions. Med Educ 2010 Aug;44(8):765-774. [doi: 10.1111/j.1365-2923.2010.03723.x] [Medline: 20633216]
7. Duque G, Demontiero O, Whereat S, Gunawardene P, Leung O, Webster P, et al. Evaluation of a blended learning model in geriatric medicine: a successful learning experience for medical students. Australas J Ageing 2013 Jun;32(2):103-109. [doi: 10.1111/j.1741-6612.2012.00620.x] [Medline: 23773249]
8. Makhdoom N, Khoshhal KI, Algaidi S, Heissam K, Zolaly MA. 'Blended learning' as an effective teaching and learning strategy in clinical medicine: a comparative cross-sectional university-based study. Journal of Taibah University Medical Sciences 2013 Apr;8(1):12-17. [doi: 10.1016/j.jtumed.2013.01.002]
9. Rowe M, Frantz J, Bozalek V. The role of blended learning in the clinical education of healthcare students: a systematic review. Med Teach 2012 Mar;34(4):e216-e221. [doi: 10.3109/0142159X.2012.642831] [Medline: 22455712]
10. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: A systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]
11. The Complexities of Physician Supply and Demand: Projections from 2013 to 2025. Washington, DC: Association of American Medical Colleges; 2015. URL: https://www.aamc.org/download/426248/data/

thecomplexitiesofphysiciansupplyanddemandprojectionsfrom2013to2.pdf [accessed 2019-02-12] [WebCite Cache ID 768UjV7lr]
12. Health and Care Professions Council. 2015. Continuing professional development (CPD) URL: https://www.hcpc-uk.org/cpd/ [accessed 2019-02-12] [WebCite Cache ID 768UriLWo]
13. Ally M. Foundations of educational theory for online learning. In: The Theory and Practice of Online Learning. Athabasca, Alberta: Athabasca University Press; 2004:15-44.
14. Rovai A. The International Review of Research in Open and Distributed Learning. 2002 Apr 01. Building Sense of Community at a Distance URL: http://www.irrodl.org/index.php/irrodl/article/view/79 [accessed 2019-02-12] [WebCite Cache ID 768V8cinI]
15. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
16. Tudor CL, Riboli-Sasco E, Marcano BJ, Nikolaou CK, Majeed A, Zary N, et al. PROSPERO CRD42015029786. 2015. Mobile learning for delivering health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42015029786 [accessed 2019-02-12] [WebCite Cache ID 768Op29KF]
17. Hervatis V, Kyaw B, Semwal M, Dunleavy G, Tudor CL, Zary N, et al. PROSPERO CRD42016045679. 2018. Offline and computer-based eLearning interventions for medical students' education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045679 [accessed 2019-02-12] [WebCite Cache ID 768PDW4wz]
18. Kononowicz A, Woodham L, Georg C, Edelbring S, Stathakarou N, Davies D, et al. PROSPERO CRD42016045912. 2016. Virtual patient simulations for health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045912 [accessed 2019-02-12] [WebCite Cache ID 768PWyrTq]
19. Saxena N, Kyaw B, Vseteckova J, Dev P, Paul P, Lim K, et al. PROSPERO CRD42016045470. 2016. Virtual reality environments for health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045470 [accessed 2019-02-12] [WebCite Cache ID 768PoyNgB]
20. Gentry S, L'Estrade E, Gauthier A, Alvarez J, Wortley D, van Rijswijk J, et al. PROSPERO CRD42016045997. 2016. Serious Gaming and Gamification interventions for health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045997 [accessed 2019-02-12] [WebCite Cache ID 768Q8s7FV]
21. Semwal M, Wahabi H, Posadzki P, Divakar U, Lim K, Audouard-Marzin Y, et al. PROSPERO CRD42017083206. 2018. Offline and computer-based eLearning interventions for medical doctors' education [Cochrane protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017083206 [accessed 2019-02-12] [WebCite Cache ID 768QQ9Ep6]
22. Allison JJ, Kiefe CI, Wall T, Casebeer L, Ray MN, Spettell CM, et al. Multicomponent Internet continuing medical education to promote chlamydia screening. Am J Prev Med 2005 Apr;28(3):285-290. [doi: 10.1016/j.amepre.2004.12.013] [Medline: 15766617]
23. Bernstein HH, Dhepyasuwan N, Connors K, Volkan K, Serwint JR, CORNET Investigators. Evaluation of a national Bright Futures oral health curriculum for pediatric residents. Acad Pediatr 2013 Mar;13(2):133-139. [doi: 10.1016/j.acap.2012.10.010] [Medline: 23257667]
24. Grover S, Currier PF, Elinoff JM, Katz JT, McMahon GT. Improving residents' knowledge of arterial and central line placement with a web-based curriculum. J Grad Med Educ 2010 Dec;2(4):548-554 [FREE Full text] [doi: 10.4300/JGME-D-10-00029.1] [Medline: 22132276]
25. Kerfoot BP, Baker HE, Koch MO, Connelly D, Joseph DB, Ritchey ML. Randomized, controlled trial of spaced education to urology residents in the United States and Canada. J Urol 2007 Apr;177(4):1481-1487. [doi: 10.1016/j.juro.2006.11.074] [Medline: 17382760]
26. Marsh-Tootle WL, McGwin G, Kohler CL, Kristofco RE, Datla RV, Wall TC. Efficacy of a web-based intervention to improve and sustain knowledge and screening for amblyopia in primary care settings. Invest Ophthalmol Vis Sci 2011 Sep 09;52(10):7160-7167 [FREE Full text] [doi: 10.1167/iovs.10-6566] [Medline: 21730344]
27. Sangvai S, Mahan JD, Lewis KO, Pudlo N, Suresh S, McKenzie LB. The impact of an interactive Web-based module on residents' knowledge and clinical practice of injury prevention. Clin Pediatr (Phila) 2012 Feb 10;51(2):165-174. [doi: 10.1177/0009922811419027] [Medline: 21985892]
28. Saxon D, Pearson AT, Wu P. Hyperlink-Embedded Journal Articles Improve Statistical Knowledge and Reader Satisfaction. J Grad Med Educ 2015 Dec;7(4):654-657 [FREE Full text] [doi: 10.4300/JGME-D-14-00747.1] [Medline: 26692981]
29. Schroter S, Jenkins RD, Playle RA, Walsh KM, Probert C, Kellner T, et al. Evaluation of an online interactive Diabetes Needs Assessment Tool (DNAT) versus online self-directed learning: a randomised controlled trial. BMC Med Educ 2011 Jun 16;11(1):35-39 [FREE Full text] [doi: 10.1186/1472-6920-11-35] [Medline: 21679446]
30. Shaw TJ, Pernar LI, Peyre SE, Helfrick JF, Vogelgesang KR, Graydon-Baker E, et al. Impact of online education on intern behaviour around joint commission national patient safety goals: a randomised trial. BMJ Qual Saf 2012 Oct 16;21(10):819-825 [FREE Full text] [doi: 10.1136/bmjqs-2011-000702] [Medline: 22706930]

31. Pape-Koehler C, Immenroth M, Sauerland S, Lefering R, Lindlohr C, Toaspern J, et al. Multimedia-based training on Internet platforms improves surgical performance: a randomized controlled trial. Surg Endosc 2013 May;27(5):1737-1747 [FREE Full text] [doi: 10.1007/s00464-012-2672-y] [Medline: 23475016]
32. Ruf D, Berner M, Kriston L, Lohmann M, Mundle G, Lorenz G, et al. Cluster-randomized controlled trial of dissemination strategies of an online quality improvement programme for alcohol-related disorders. Alcohol Alcohol 2010 Nov 04;45(1):70-78. [doi: 10.1093/alcalc/agp079] [Medline: 19889887]
33. Talib N, Onikul R, Filardi D, Simon S, Sharma V. Effective educational instruction in preventive oral health: hands-on training versus web-based training. Pediatrics 2010 Mar 01;125(3):547-553. [doi: 10.1542/peds.2009-0470] [Medline: 20123769]
34. Yardley L, Douglas E, Anthierens S, Tonkin-Crine S, O'Reilly G, Stuart B, et al. Evaluation of a web-based intervention to reduce antibiotic prescribing for LRTI in six European countries: quantitative process analysis of the GRACE/INTRO randomised controlled trial. Implement Sci 2013 Nov 15;8:134 [FREE Full text] [doi: 10.1186/1748-5908-8-134] [Medline: 24238118]
35. Weston CM, Sciamanna CN, Nash DB. Evaluating online continuing medical education seminars: evidence for improving clinical practices. Am J Med Qual 2008 Nov;23(6):475-483. [doi: 10.1177/1062860608325266] [Medline: 19001103]
36. Franchi C, Tettamanti M, Djade CD, Pasina L, Mannucci PM, Onder G, ELICADHE Investigators. E-learning in order to improve drug prescription for hospitalized older patients: a cluster-randomized controlled study. Br J Clin Pharmacol 2016 Dec 05;82(1):53-63 [FREE Full text] [doi: 10.1111/bcp.12922] [Medline: 26922904]
37. Kerfoot B, Turchin A, Breydo E, Gagnon D, Conlin P. An online spaced-education game among clinicians improves their patients' time to blood pressure control: a randomized controlled trial. Circ Cardiovasc Qual Outcomes 2014 May;7(3):468-474 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.113.000814] [Medline: 24847084]
38. George PP, Toon E, Hadadgar A, Jirwe M, Saxena N, Lim K, et al. PROSPERO CRD42016045571. 2016. Online- and local area network (LAN)-based eLearning interventions for medical doctors' education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045571 [accessed 2019-02-12] [WebCite Cache ID 768Nw4Q5U]
39. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. URL: http://handbook-5-1.cochrane.org/ [accessed 2019-02-07] [WebCite Cache ID 760HwWOUg]
40. Car J, Carlstedt-Duke J, Tudor CL, Posadzki P, Zary N, Atun R, et al. Digital education for health professions: methods for overarching evidence syntheses. J Med Internet Res 2019:2018. [doi: 10.2196/12913]
41. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. BMJ 2003 Oct 04;327(7418):785-789 [FREE Full text] [doi: 10.1136/bmj.327.7418.785] [Medline: 14525877]
42. Schünemann H, Oxman A, Higgins J, Vist G, Glasziou P, Guyatt G. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011), The Cochrane Collaboration. 2008 Sep 26. Presenting results and 'Summary of Findings' tables URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470712184.ch11 [accessed 2019-02-12] [WebCite Cache ID 768OZ6Q3M]
43. Braido F, Comaschi M, Valle I, Delgado L, Coccini A, Guerreras P, ARGA Study Group, EAACI/CME Committee. Knowledge and health care resource allocation: CME/CPD course guidelines-based efficacy. Eur Ann Allergy Clin Immunol 2012 Oct;44(5):193-199. [Medline: 23156067]
44. Connolly AM, Cunningham C, Sinclair AJ, Rao A, Lonergan A, Bye AM. 'Beyond Milestones': a randomised controlled trial evaluating an innovative digital resource teaching quality observation of normal child development. J Paediatr Child Health 2014 May 23;50(5):393-398. [doi: 10.1111/jpc.12485] [Medline: 24373114]
45. Farah SS, Winter M, Appu S. Helping doctors utilize the prostate-specific antigen effectively: an online randomized controlled trial (The DUPE trial). ANZ J Surg 2012 Sep 20;82(9):633-638. [doi: 10.1111/j.1445-2197.2012.06154.x] [Medline: 22900524]
46. Gyorki DE, Shaw T, Nicholson J, Baker C, Pitcher M, Skandarajah A, et al. Improving the impact of didactic resident training with online spaced education. ANZ J Surg 2013 Jun 26;83(6):477-480. [doi: 10.1111/ans.12166] [Medline: 23617607]
47. Harris J, Kutob R, Surprenant Z, Maiuro R, Delate T. Can Internet-based education improve physician confidence in dealing with domestic violence? Fam Med 2002 Apr;34(4):287-292. [Medline: 12017143]
48. Houwink EJ, van Teeffelen SR, Muijtjens AM, Henneman L, Jacobi F, van Luijk SJ, et al. Sustained effects of online genetics education: a randomized controlled trial on oncogenetics. Eur J Hum Genet 2013 Aug 14;22(3):310-316. [doi: 10.1038/ejhg.2013.163]
49. Kutob R, Senf J, Harris J. Teaching culturally effective diabetes care: results of a randomized controlled trial. Fam Med 2009 Mar;41(3):167-174 [FREE Full text] [Medline: 19259838]
50. Thompson J, Lebwohl B, Syngal S, Kastrinos F. Tu1482 Assessment of U.S. Gastroenterology Trainees' Knowledge of Quality Performance Measures for Colonoscopy and the Impact of a Web-Based Intervention. Gastrointestinal Endoscopy 2011 Apr;73(4):AB423 [FREE Full text] [doi: 10.1016/j.gie.2011.03.974]

51. World Health Organization. Geneva: WHO; 2013. Global health workforce shortage to reach 12.9 million in coming decades URL: https://www.who.int/mediacentre/news/releases/2013/health-workforce-shortage/en/ [accessed 2019-02-12] [WebCite Cache ID 768WLcBq3]
52. Alfieri J, Portelance L, Souhami L, Steinert Y, McLeod P, Gallant F, et al. Development and impact evaluation of an e-learning radiation oncology module. Int J Radiat Oncol Biol Phys 2012 Mar 01;82(3):e573-e580. [doi: 10.1016/j.ijrobp.2011.07.002] [Medline: 22024203]
53. Chang TP, Pham PK, Sobolewski B, Doughty CB, Jamal N, Kwan KY, et al. Pediatric emergency medicine asynchronous e-learning: a multicenter randomized controlled Solomon four-group study. Acad Emerg Med 2014 Aug;21(8):912-919 [FREE Full text] [doi: 10.1111/acem.12434] [Medline: 25154469]
54. Claxton R, Marks S, Buranosky R, Rosielle D, Arnold RM. The educational impact of weekly e-mailed fast facts and concepts. J Palliat Med 2011 Apr;14(4):475-481. [doi: 10.1089/jpm.2010.0418] [Medline: 21395446]
55. Cullinan S, O'Mahony D, Byrne S. Use of an e-Learning Educational Module to Better Equip Doctors to Prescribe for Older Patients: A Randomised Controlled Trial. Drugs Aging 2017 Dec 4;34(5):367-374. [doi: 10.1007/s40266-017-0451-0] [Medline: 28258537]
56. Hearty T, Maizels M, Pring M, Mazur J, Liu R, Sarwark J, et al. Orthopaedic resident preparedness for closed reduction and pinning of pediatric supracondylar fractures is improved by e-learning: a multisite randomized controlled study. J Bone Joint Surg Am 2013 Sep 04;95(17):e1261-e1267. [doi: 10.2106/JBJS.L.01065] [Medline: 24005210]
57. Satterwhite T, Son J, Carey J, Zeidler K, Bari S, Gurtner G, et al. Microsurgery education in residency training: validating an online curriculum. Ann Plast Surg 2012 Apr;68(4):410-414. [doi: 10.1097/SAP.0b013e31823b6a1a] [Medline: 22421490]
58. Sullivan MD, Gaster B, Russo J, Bowlby L, Rocco N, Sinex N, et al. Randomized trial of web-based training about opioid therapy for chronic pain. Clin J Pain 2010;26(6):512-517. [doi: 10.1097/AJP.0b013e3181dc7adc] [Medline: 20551726]
59. Viguier M, Rist S, Aubin F, Leccia M, Richard M, Esposito-Farèse M, Club Rhumatismes et Inflammation. Online training on skin cancer diagnosis in rheumatologists: results from a nationwide randomized web-based survey. PLoS One 2015 May 21;10(5):e0127564 [FREE Full text] [doi: 10.1371/journal.pone.0127564] [Medline: 25996152]
60. Westmoreland GR, Counsell SR, Tu W, Wu J, Litzelman DK. Web-based training in geriatrics for medical residents: a randomized controlled trial using standardized patients to assess outcomes. J Am Geriatr Soc 2010 Jun;58(6):1163-1169. [doi: 10.1111/j.1532-5415.2010.02869.x] [Medline: 20722848]
61. Gold J, Begg W, Fullerton D, Mathisen D, Olinger G, Orringer M, et al. Successful implementation of a novel internet hybrid surgery curriculum: the early phase outcome of thoracic surgery prerequisite curriculum e-learning project. Ann Surg 2004 Sep;240(3):499-507; discussion 507. [Medline: 15319721]
62. Stewart M, Marshall JN, Ostbye T, Feightner JW, Brown JB, Harris S, et al. Effectiveness of case-based on-line learning of evidence-based practice guidelines. Fam Med 2005 Feb;37(2):131-138 [FREE Full text] [Medline: 15690254]
63. Enders F, Diener-West M. Methods of learning in statistics education: a randomized trial of public health graduate students. Statistics Education Research Journal 2006;5(1):5.
64. Short LM, Surprenant ZJ, Harris JM. A community-based trial of an online intimate partner violence CME program. Am J Prev Med 2006 Feb;30(2):181-185 [FREE Full text] [doi: 10.1016/j.amepre.2005.10.012] [Medline: 16459218]
65. Dolan BM, Yialamas MA, McMahon GT. A Randomized Educational Intervention Trial to Determine the Effect of Online Education on the Quality of Resident-Delivered Care. J Grad Med Educ 2015 Sep;7(3):376-381 [FREE Full text] [doi: 10.4300/JGME-D-14-00571.1] [Medline: 26457142]
66. Le TT, Rait MA, Jarlsberg LG, Eid NS, Cabana MD. A randomized controlled trial to evaluate the effectiveness of a distance asthma learning program for pediatricians. J Asthma 2010 Apr 15;47(3):245-250. [doi: 10.3109/02770900903560209] [Medline: 20394510]
67. Bell DS, Fonarow GC, Hays RD, Mangione CM. Self-study from web-based and printed guideline materials. A randomized, controlled trial among resident physicians. Ann Intern Med 2000 Jun 20;132(12):938-946. [Medline: 10858176]
68. Butzlaff M, Vollmar HC, Floer B, Koneczny N, Isfort J, Lange S. Learning with computerized guidelines in general practice? A randomized controlled trial. Fam Pract 2004 Apr;21(2):183-188. [Medline: 15020389]
69. Cabrera-Muffly C, Bryson PC, Sykes KJ, Shnayder Y. Free online otolaryngology educational modules: a pilot study. JAMA Otolaryngol Head Neck Surg 2015 Apr 01;141(4):324-328. [doi: 10.1001/jamaoto.2015.41] [Medline: 25695900]
70. Chung S, Mandl KD, Shannon M, Fleisher GR. Efficacy of an educational Web site for educating physicians about bioterrorism. Acad Emerg Med 2004 Feb;11(2):143-148 [FREE Full text] [Medline: 14759955]
71. Ferguson MK, Thompson K, Huisingh-Scheetz M, Farnan J, Hemmerich J, Acevedo J, et al. The Impact of a Frailty Education Module on Surgical Resident Estimates of Lobectomy Risk. Ann Thorac Surg 2015 Jul;100(1):235-241 [FREE Full text] [doi: 10.1016/j.athoracsur.2015.03.016] [Medline: 26004924]
72. Wang CL, Schopp JG, Kani K, Petscavage-Thomas JM, Zaidi S, Hippe DS, et al. Prospective randomized study of contrast reaction management curricula: computer-based interactive simulation versus high-fidelity hands-on simulation. Eur J Radiol 2013 Dec;82(12):2247-2252. [doi: 10.1016/j.ejrad.2013.08.044] [Medline: 24041436]
73. Hymowitz N, Schwab JV, Haddock CK, Pyle SA, Schwab LM. The pediatric residency training on tobacco project: four-year resident outcome findings. Prev Med 2007 Dec;45(6):481-490 [FREE Full text] [doi: 10.1016/j.ypmed.2007.07.024] [Medline: 17707898]

74. Koppe H, van de Mortel TF, Ahern CM. How effective and acceptable is Web 2.0 Balint group participation for general practitioners and general practitioner registrars in regional Australia? A pilot study. Aust J Rural Health 2016 Feb 26;24(1):16-22 [FREE Full text] [doi: 10.1111/ajr.12212] [Medline: 26114400]
75. Lee A, Mader E, Morley C. Teaching cross-cultural communication skills online: a multi-method evaluation. Fam Med 2015 Apr;47(4):302-308 [FREE Full text] [Medline: 25853602]
76. Macrae HM, Regehr G, McKenzie M, Henteleff H, Taylor M, Barkun J, et al. Teaching practicing surgeons critical appraisal skills with an Internet-based journal club: A randomized, controlled trial. Surgery 2004 Sep;136(3):641-646. [doi: 10.1016/j.surg.2004.02.003] [Medline: 15349113]
77. Conroy EJ, Kirkham JJ, Bellis JR, Peak M, Smyth RL, Williamson PR, et al. A pilot randomised controlled trial to assess the utility of an e-learning package that trains users in adverse drug reaction causality. Int J Pharm Pract 2015 Dec 29;23(6):447-455 [FREE Full text] [doi: 10.1111/ijpp.12197] [Medline: 26032626]
78. Edrich T, Stopfkuchen-Evans M, Scheiermann P, Heim M, Chan W, Stone MB, et al. A Comparison of Web-Based with Traditional Classroom-Based Training of Lung Ultrasound for the Exclusion of Pneumothorax. Anesth Analg 2016 Dec;123(1):123-128. [doi: 10.1213/ANE.0000000000001383] [Medline: 27159073]
79. Matzie KA, Kerfoot BP, Hafler JP, Breen EM. Spaced education improves the feedback that surgical residents give to medical students: a randomized trial. Am J Surg 2009 Feb;197(2):252-257. [doi: 10.1016/j.amjsurg.2008.01.025] [Medline: 18722585]
80. Pernar LI, Beleniski F, Rosen H, Lipsitz S, Hafler J, Breen E. Spaced education faculty development may not improve faculty teaching performance ratings in a surgery department. J Surg Educ 2012 Jan;69(1):52-57. [doi: 10.1016/j.jsurg.2011.06.013] [Medline: 22208833]
81. Fordis M, King JE, Ballantyne CM, Jones PH, Schneider KH, Spann SJ, et al. Comparison of the instructional efficacy of Internet-based CME with live interactive CME workshops: a randomized controlled trial. JAMA 2005 Sep 07;294(9):1043-1051. [doi: 10.1001/jama.294.9.1043] [Medline: 16145024]
82. Pelayo-Alvarez M, Perez-Hoyos S, Agra-Varela Y. Clinical effectiveness of online training in palliative care of primary care physicians. J Palliat Med 2013 Oct;16(10):1188-1196 [FREE Full text] [doi: 10.1089/jpm.2013.0005] [Medline: 23987657]
83. Hemmati N, Omrani S, Hemmati N. A comparison of Internet-based learning and traditional classroom lecture to learn CPR for continuing medical education. TurkiOJDE 2013;14:256-265 [FREE Full text]
84. Chenkin J, Lee S, Huynh T, Bandiera G. Procedures can be learned on the Web: a randomized study of ultrasound-guided vascular access training. Acad Emerg Med 2008 Oct;15(10):949-954 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00231.x] [Medline: 18778380]
85. Hadley J, Kulier R, Zamora J, Coppus SF, Weinbrenner S, Meyerrose B, et al. Effectiveness of an e-learning course in evidence-based medicine for foundation (internship) training. J R Soc Med 2010 Jul 03;103(7):288-294 [FREE Full text] [doi: 10.1258/jrsm.2010.100036] [Medline: 20522698]
86. Hugenholtz NIR, de Croon EM, Smits PB, van Dijk FJH, Nieuwenhuijsen K. Effectiveness of e-learning in continuing medical education for occupational physicians. Occup Med (Lond) 2008 Aug 20;58(5):370-372 [FREE Full text] [doi: 10.1093/occmed/kqn053] [Medline: 18495676]
87. Bello G, Pennisi MA, Maviglia R, Maggiore SM, Bocci MG, Montini L, et al. Online vs live methods for teaching difficult airway management to anesthesiology residents. Intensive Care Med 2005 Apr 8;31(4):547-552. [doi: 10.1007/s00134-005-2561-0] [Medline: 15754200]
88. Chan D, Leclair K, Kaczorowski J. Problem-based small-group learning via the Internet among community family physicians: a randomized controlled trial. MD Comput 1999;16(3):54-58. [Medline: 10439603]
89. McLeod RS, MacRae HM, McKenzie ME, Victor JC, Brasel KJ, Evidence Based Reviews in Surgery Steering Committee. A moderated journal club is more effective than an Internet journal club in teaching critical appraisal skills: results of a multicenter randomized controlled trial. J Am Coll Surg 2010 Dec;211(6):769-776. [doi: 10.1016/j.jamcollsurg.2010.08.016] [Medline: 21036071]
90. Barthelemy FX, Segard J, Fradin P, Hourdin N, Batard E, Pottier P, et al. ECG interpretation in Emergency Department residents: an update and e-learning as a resource to improve skills. Eur J Emerg Med 2017 Apr;24(2):149-156. [doi: 10.1097/MEJ.0000000000000312] [Medline: 26287803]
91. Houwink EJF, Muijtjens AMM, van Teeffelen SR, Henneman L, Rethans JJ, Jacobi F, et al. Effect of comprehensive oncogenetics training interventions for general practitioners, evaluated at multiple performance levels. PLoS One 2015 Apr 2;10(4):e0122648 [FREE Full text] [doi: 10.1371/journal.pone.0122648] [Medline: 25837634]
92. Shariff U, Kullar N, Haray PN, Dorudi S, Balasubramanian SP. Multimedia educational tools for cognitive surgical skill acquisition in open and laparoscopic colorectal surgery: a randomized controlled trial. Colorectal Dis 2015 May 21;17(5):441-450. [doi: 10.1111/codi.12863] [Medline: 25495835]
93. Wilkinson JS, Barake W, Smith C, Thakrar A, Johri AM. Limitations of Condensed Teaching Strategies to Develop Hand-Held Cardiac Ultrasonography Skills in Internal Medicine Residents. Can J Cardiol 2016 Dec;32(8):1034-1037. [doi: 10.1016/j.cjca.2015.10.008] [Medline: 26860772]

94. Schmitz CC, Braman JP, Turner N, Heller S, Radosevich DM, Yan Y, et al. Learning by (video) example: a randomized study of communication skills training for end-of-life and error disclosure family care conferences. Am J Surg 2016 Nov;212(5):996-1004. [doi: 10.1016/j.amjsurg.2016.02.023] [Medline: 27474496]
95. Putnam L, Pham D, Etchegaray J, Ostovar-Kermani T, Ottosen M, Thomas E, et al. How Should Surgical Residents Be Educated about Patient Safety in the Operating Room: A Pilot Randomized Trial. Journal of the American College of Surgeons 2015 Oct;221(4):S127-S128 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2015.07.301]
96. Kulier R, Gülmezoglu AM, Zamora J, Plana MN, Carroli G, Cecatti JG, et al. Effectiveness of a clinically integrated e-learning course in evidence-based medicine for reproductive health training: a randomized trial. JAMA 2012 Dec 05;308(21):2218-2243. [doi: 10.1001/jama.2012.33640] [Medline: 23212499]
97. Sharma V, Chamos C, Valencia O, Meineri M, Fletcher SN. The impact of internet and simulation-based training on transoesophageal echocardiography learning in anaesthetic trainees: a prospective randomised study. Anaesthesia 2013 Jun;68(6):621-627 [FREE Full text] [doi: 10.1111/anae.12261] [Medline: 23594265]
98. Ali J, Sorvari A, Camera S, Kinach M, Mohammed S, Pandya A. Telemedicine as a potential medium for teaching the advanced trauma life support (ATLS) course. J Surg Educ 2013 Mar;70(2):258-264. [doi: 10.1016/j.jsurg.2012.11.008] [Medline: 23427974]
99. Perkins GD, Kimani PK, Bullock I, Clutton-Brock T, Davies RP, Gale M, Electronic Advanced Life Support Collaborators. Improving the efficiency of advanced life support training: a randomized, controlled trial. Ann Intern Med 2012 Jul 03;157(1):19-28. [doi: 10.7326/0003-4819-157-1-201207030-00005] [Medline: 22751757]
100. Platz E, Goldflam K, Mennicke M, Parisini E, Christ M, Hohenstein C. Comparison of Web-versus classroom-based basic ultrasonographic and EFAST training in 2 European hospitals. Ann Emerg Med 2010 Dec;56(6):660-667. [doi: 10.1016/j.annemergmed.2010.04.019] [Medline: 20538370]
101. Vollmar HC, Mayer H, Ostermann T, Butzlaff ME, Sandars JE, Wilm S, et al. Knowledge transfer for the management of dementia: a cluster randomised trial of blended learning in general practice. Implement Sci 2010 Jan 04;5(1):1-10 [FREE Full text] [doi: 10.1186/1748-5908-5-1] [Medline: 20047652]
102. Szmuilowicz E, Neely KJ, Sharma RK, Cohen ER, McGaghie WC, Wayne DB. Improving residents' code status discussion skills: a randomized trial. J Palliat Med 2012 Jul;15(7):768-774 [FREE Full text] [doi: 10.1089/jpm.2011.0446] [Medline: 22690890]
103. Ngamruengphong S, Horsley-Silva J, Hines S, Pungpapong S, Patel T, Keaveny A. Educational Intervention in Primary Care Residents' Knowledge and Performance of Hepatitis B Vaccination in Patients with Diabetes Mellitus. South Med J 2015 Sep;108(9):510-515. [doi: 10.14423/SMJ.0000000000000334] [Medline: 26332473]
104. Midmer D, Kahan M, Marlow B. Effects of a distance learning program on physicians' opioid- and benzodiazepine-prescribing skills. J Contin Educ Health Prof 2006;26(4):294-301. [doi: 10.1002/chp.82] [Medline: 17163499]
105. Kulier R, Coppus SF, Zamora J, Hadley J, Malick S, Das K, et al. The effectiveness of a clinically integrated e-learning course in evidence-based medicine: a cluster randomised controlled trial. BMC Med Educ 2009 May 12;9(1):21-27 [FREE Full text] [doi: 10.1186/1472-6920-9-21] [Medline: 19435520]
106. Kronick J, Blake C, Munoz E, Heilbrunn L, Dunikowski L, Milne WK. Improving on-line skills and knowledge. A randomized trial of teaching rural physicians to use on-line medical information. Can Fam Physician 2003 Mar;49:312-317 [FREE Full text] [Medline: 12675544]
107. Daetwyler CJ, Cohen DG, Gracely E, Novack DH. eLearning to enhance physician patient communication: A pilot test of “doc.com” and “WebEncounter” in teaching bad news delivery. Medical Teacher 2010 Aug 26;32(9):e381-e390. [doi: 10.3109/0142159X.2010.495759]
108. Estrada CA, Safford MM, Salanitro AH, Houston TK, Curry W, Williams JH, et al. A web-based diabetes intervention for physician: a cluster-randomized effectiveness trial. Int J Qual Health Care 2011 Dec 10;23(6):682-689 [FREE Full text] [doi: 10.1093/intqhc/mzr053] [Medline: 21831967]
109. Epstein JN, Langberg JM, Lichtenstein PK, Kolb R, Altaye M, Simon JO. Use of an Internet portal to improve community-based pediatric ADHD care: a cluster randomized trial. Pediatrics 2011 Nov 17;128(5):e1201-e1208 [FREE Full text] [doi: 10.1542/peds.2011-0872] [Medline: 22007005]
110. Dayton CS, Ferguson JS, Hornick DB, Peterson MW. Evaluation of an Internet-based decision-support system for applying the ATS/CDC guidelines for tuberculosis preventive therapy. Med Decis Making 2000 Jul;20(1):1-6. [doi: 10.1177/0272989X0002000101] [Medline: 10638531]
111. Gerbert B, Bronstone A, Maurer T, Berger T, McPhee SJ, Caspers N. The effectiveness of an Internet-based tutorial in improving primary care physicians' skin cancer triage skills. J Cancer Educ 2002;17(1):7-11. [doi: 10.1080/08858190209528784] [Medline: 12000111]
112. Meeker D, Linder JA, Fox CR, Friedberg MW, Persell SD, Goldstein NJ, et al. Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices: A Randomized Clinical Trial. JAMA 2016 Feb 09;315(6):562-570. [doi: 10.1001/jama.2016.0275] [Medline: 26864410]
113. Bell RA, McDermott H, Fancher TL, Green MJ, Day FC, Wilkes MS. Impact of a randomized controlled educational trial to improve physician practice behaviors around screening for inherited breast cancer. J Gen Intern Med 2015 Mar 2;30(3):334-341 [FREE Full text] [doi: 10.1007/s11606-014-3113-5] [Medline: 25451990]

114. Butler CC, Simpson SA, Dunstan F, Rollnick S, Cohen D, Gillespie D, et al. Effectiveness of multifaceted educational programme to reduce antibiotic dispensing in primary care: practice based randomised controlled trial. BMJ 2012 Feb 02;344(feb02 1):d8173-d8173 [FREE Full text] [doi: 10.1136/bmj.d8173] [Medline: 22302780]
115. Curtis JR, Westfall AO, Allison J, Becker A, Melton ME, Freeman A, et al. Challenges in improving the quality of osteoporosis care for long-term glucocorticoid users: a prospective randomized trial. Arch Intern Med 2007 Mar 26;167(6):591-596. [doi: 10.1001/archinte.167.6.591] [Medline: 17389291]
116. Feng B, Srinivasan M, Hoffman JR, Rainwater JA, Griffin E, Dragojevic M, et al. Physician communication regarding prostate cancer screening: analysis of unannounced standardized patient visits. Ann Fam Med 2013;11(4):315-323 [FREE Full text] [doi: 10.1370/afm.1509] [Medline: 23835817]
117. Little P, Stuart B, Francis N, Douglas E, Tonkin-Crine S, Anthierens S, et al. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. Lancet 2013 Oct 5;382(9899):1175-1182 [FREE Full text] [doi: 10.1016/S0140-6736(13)60994-0] [Medline: 23915885]
118. Xiao Y, Seagull FJ, Bochicchio GV, Guzzo JL, Dutton RP, Sisley A, et al. Video-based training increases sterile-technique compliance during central venous catheter insertion. Crit Care Med 2007 May;35(5):1302-1306. [doi: 10.1097/01.CCM.0000263457.81998.27] [Medline: 17414726]
119. Légaré F, Labrecque M, Cauchon M, Castel J, Turcotte S, Grimshaw J. Training family physicians in shared decision-making to reduce the overuse of antibiotics in acute respiratory infections: a cluster randomized trial. CMAJ 2012 Sep 18;184(13):E726-E734 [FREE Full text] [doi: 10.1503/cmaj.120568] [Medline: 22847969]
120. Girgis A, Cockburn J, Butow P, Bowman D, Schofield P, Stojanovski E, et al. Improving patient emotional functioning and psychological morbidity: evaluation of a consultation skills training program for oncologists. Patient Educ Couns 2009 Dec;77(3):456-462. [doi: 10.1016/j.pec.2009.09.018] [Medline: 19819660]
121. Jayakumar N, Brunckhorst O, Dasgupta P, Khan MS, Ahmed K. e-Learning in Surgical Education: A Systematic Review. J Surg Educ 2015 Nov;72(6):1145-1157. [doi: 10.1016/j.jsurg.2015.05.008] [Medline: 26111822]
122. Thepwongsa I, Kirby CN, Schattner P, Piterman L. Online continuing medical education (CME) for GPs: does it work? A systematic review. Aust Fam Physician 2014 Oct;43(10):717-721 [FREE Full text] [Medline: 25286431]
123. Jwayyed S, Stiffler KA, Wilber ST, Southern A, Weigand J, Bare R, et al. Technology-assisted education in graduate medical education: a review of the literature. Int J Emerg Med 2011 Aug 08;4(1):51-13 [FREE Full text] [doi: 10.1186/1865-1380-4-51] [Medline: 21824405]
124. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The Effectiveness of Blended Learning in Health Professions: Systematic Review and Meta-Analysis. J Med Internet Res 2016 Jan 04;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]

Abbreviations

ATLS: Advanced Trauma Life Support
CME: continuing medical education
CPD: continuing professional development
cRCT: cluster-randomized controlled trial
EBM: evidence-based medicine
GRADE: Grading of Recommendations, Assessment, Development and Evaluations
ICT: information communication technology
LAN: local area network
ODE: online and local area network–based digital education
OSCE: objective structured clinical examination
RCT: randomized controlled trial
RR: risk ratio
SMD: standardized mean difference

Edited by A Marusic; submitted 03.01.19; peer-reviewed by M Vidak, R Tokalic; comments to author 21.01.19; revised version received 29.01.19; accepted 30.01.19; published 25.02.19.

Please cite as:
George PP, Zhabenko O, Kyaw BM, Antoniou P, Posadzki P, Saxena N, Semwal M, Tudor Car L, Zary N, Lockwood C, Car J
Online Digital Education for Postregistration Training of Medical Doctors: Systematic Review by the Digital Health Education Collaboration
J Med Internet Res 2019;21(2):e13269
URL: http://www.jmir.org/2019/2/e13269/
doi:10.2196/13269
PMID:30801252

Original Paper

Influences on the Implementation of Mobile Learning for Medical and Nursing Education: Qualitative Systematic Review by the Digital Health Education Collaboration

Priya Lall[1], PhD; Rebecca Rees[2], PhD; Gloria Chun Yi Law[3], PhD; Gerard Dunleavy[3], MSc; Živa Cotič[4], MSc; Josip Car[3,5], MSc, MD, PhD, DIC, FFPH, FRCP

[1]School of Geography, Queen Mary University of London, London, United Kingdom

[2]Evidence for Policy and Practice Information and Co-ordinating Centre, Social Science Research Unit, Department of Social Science, University College London Institute of Education, University College London, London, United Kingdom

[3]Centre of Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[4]Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

[5]Faculty of Medicine, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Rebecca Rees, PhD
Evidence for Policy and Practice Information and Co-ordinating Centre, Social Science Research Unit, Department of Social Science
University College London Institute of Education
University College London
18 Woburn Square
London, WCIH 0NR
United Kingdom
Phone: 44 07932 243030
Email: rebecca.rees@ucl.ac.uk

Abstract

Background: In the past 5 decades, digital education has increasingly been used in health professional education. Mobile learning (mLearning), an emerging form of educational technology using mobile devices, has been used to supplement learning outcomes through enabling conversations, sharing information and knowledge with other learners, and aiding support from peers and instructors regardless of geographic distance.

Objective: This review aimed to synthesize findings from qualitative or mixed-methods studies to provide insight into factors facilitating or hindering implementation of mLearning strategies for medical and nursing education.

Methods: A systematic search was conducted across a range of databases. Studies with the following criteria were selected: examined mLearning in medical and nursing education, employed a mixed-methods or qualitative approach, and published in English after 1994. Findings were synthesized using a framework approach.

Results: A total of 1946 citations were screened, resulting in 47 studies being selected for inclusion. Most studies evaluated pilot mLearning interventions. The synthesis identified views on valued aspects of mobile devices in terms of efficiency and personalization but concerns over vigilance and poor device functionality; emphasis on the social aspects of technology, especially in a clinical setting; the value of interaction learning for clinical practice; mLearning as a process, including learning how to use a device; and the importance of institutional infrastructure and policies.

Conclusions: The portability of mobile devices can enable interactions between learners and educational material, fellow learners, and educators in the health professions. However, devices need to be incorporated institutionally, and learners and educators need additional support to fully comprehend device or app functions. The strategic support of mLearning is likely to require procedural guidance for practice settings and device training and maintenance services on campus.

(J Med Internet Res 2019;21(2):e12895) doi:10.2196/12895

KEYWORDS

medical education; nursing education; distance education; qualitative research; systematic review

Introduction

Background

In the past 5 decades, digital education has increasingly been used in health professional education, and technological advances have produced various forms of digital education modalities such as computer-based simulations, virtual patients, and internet-based courses and interactive contents [1,2]. Adoption of these digital education modalities in health professional education is rapidly expanding before the establishment of a robust evidence base for consideration of multiple dimensions and outcomes [3,4]. A noteworthy trend within digital education is mobile learning (mLearning), which can be defined as follows [5]:

> *...consuming, interacting with or creating information, mediated through a compact digital portable device that the individual carries on a regular basis, has reliable connectivity, and fits in a pocket or purse.*

This is enabled by a growth of capabilities in mobile devices (eg, smartphones) and the convenience they offer, such as omnipresent usability and accessibility to the internet, while mobile. Approximately 1.1 billion people living in rural areas [6] and 73% of the total world population [7] are now covered by mobile broadband.

Mobile devices can offer a variety of functions and be used across contexts. For instance, mLearning can provide access to educational content and information in daily clinical practice [8-10]; enable conversations and the sharing of information and knowledge with other learners; and elicit support from peers and instructors regardless of geographic distance [8-10]. Handheld computers can be used to keep track of students' skill development and progress in assignments [11]; promote self-directed and self-regulated learning [12,13]; display audio-visual information relating to a specific place, scene, or situation; and aid situated learning [10].

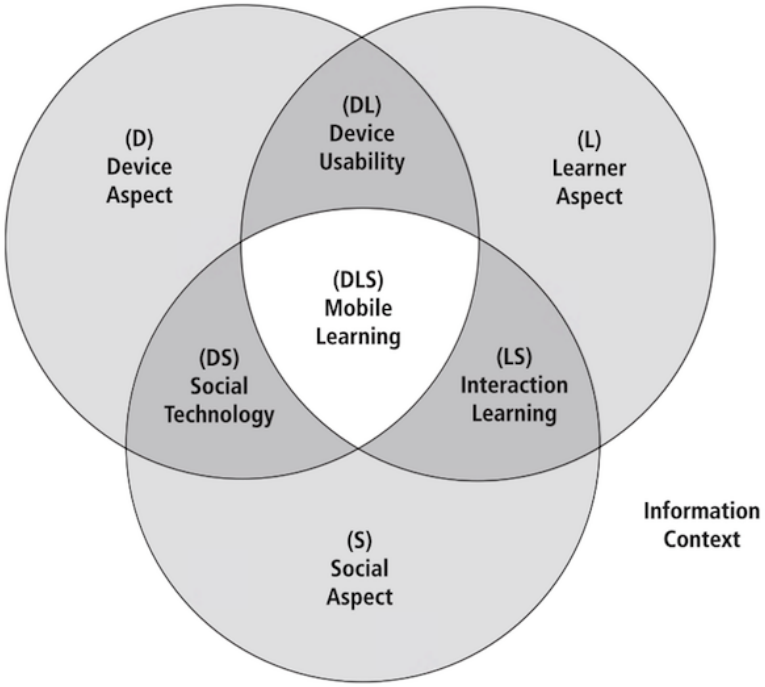
Evaluations of the effects of digital education and specifically mLearning as a whole raise more questions than they answer. For example, a meta-analysis by Free et al [14] included 7 randomized controlled trials and investigated the educational outcomes associated with the use of personal digital assistants (PDAs) and portable media players in medical and nursing education. The studies incorporated into the systematic review examined the effectiveness of mLearning in improving knowledge and attitudes; however, the meta-analysis showed no clear evidence of benefit. There are many factors influencing the effectiveness of digital education and mLearning interventions that warrant closer investigation. The implementation of digital education can be influenced by characteristics of the educational intervention, problems addressed by the intervention, features of the health system, the adopting system, and other contextual factors [15]. However, no review has been identified which examines systematically the factors influencing the use of mLearning interventions for health professional education.

Our review considered the broad issue of *implementation* of mLearning. This is important because mLearning is a relatively new area of development compared with other forms of digital education. We are in the early stages of learning what happens and what might be helpful when mLearning is introduced into real-world settings. Having a systematic and in-depth exploration of the range of potential *barriers to and facilitatorsof* mLearning in health professional education should deepen the understanding of the topic and allow insights to be obtained for effective implementation and positive outcomes. It is also important to understand mLearning in terms of the underlying assumptions about teaching and learning (ie, pedagogy and andragogy) of different approaches, to maximize the potential richness of the learning process in mobile environments and enable teachers to plan for optimal learning [16].

Koole's Framework for the Rational Analysis of Mobile Education (FRAME) model guides the qualitative synthesis for this study (see Figure 1) [17]. This model considers how features of mobile technology, along with learner capacities and social interaction, influence learning processes occurring in an information context. Within the FRAME model, mLearning is conceived as the convergence of the following aspects: (1) device, signifying functional characteristics of a mobile device, for example, processor speed; (2) learner, accounting for individuals' cognitive abilities and learning styles; and (3) social, referring to elements of social interaction and culture-influencing learning processes.

In terms of interactions between these aspects, first, *Device Usability* is thought of as containing aspects belonging to the device and learner and describes how an individual relates to the device. For example, learners can express satisfaction with a particular device because of its esthetic qualities. Second, *Social Technology* covers the intersection between device and social aspects and accounts for how mobile devices enable connection between multiple interfacing individuals and systems, such as the use of collaborative tools. Finally, *Interaction Learning* spans the intersection between learner and social aspects, describing how the learner interacts with other individuals. For instance, a mobile device could enable interaction between a learner and their instructor on long-distance educational courses. The culmination of all 3 aspects is envisioned as the eventual process of mLearning.

Figure 1. Framework for the Rational Analysis of Mobile Education (FRAME) model.



Objectives

This study aimed to synthesize insights from empirical research using qualitative and mixed methods on mLearning implementation in medical and nursing education. Our study employed systematic methods to identify, appraise, and synthesize qualitative findings from studies to explore mLearning strategies for medical and nursing education. These studies can allow us to better understand the nature of material and sociocultural influences (eg, cultural norms) and causal pathways [18] to delineate a more complete picture of the phenomenon under study [19]. Qualitative findings from existing studies are used to uncover the perspectives of learners and other key actors with experience of mLearning strategies. Particular attention is paid to perceptions of implementation processes. The broad research question for this review was as follows: What are the views of educators, learners, and other key actors with experience of mLearning in medical and nursing education about perceived factors which facilitate/enhance or hinder its implementation?

Methods

Protocol

A protocol was developed so as to establish the review's scope and methods before evaluating existing literature. This was registered with PROSPERO, the international prospective register of systematic reviews (record number CRD42016035411 Multimedia Appendix 1) [20].

Inclusion Criteria

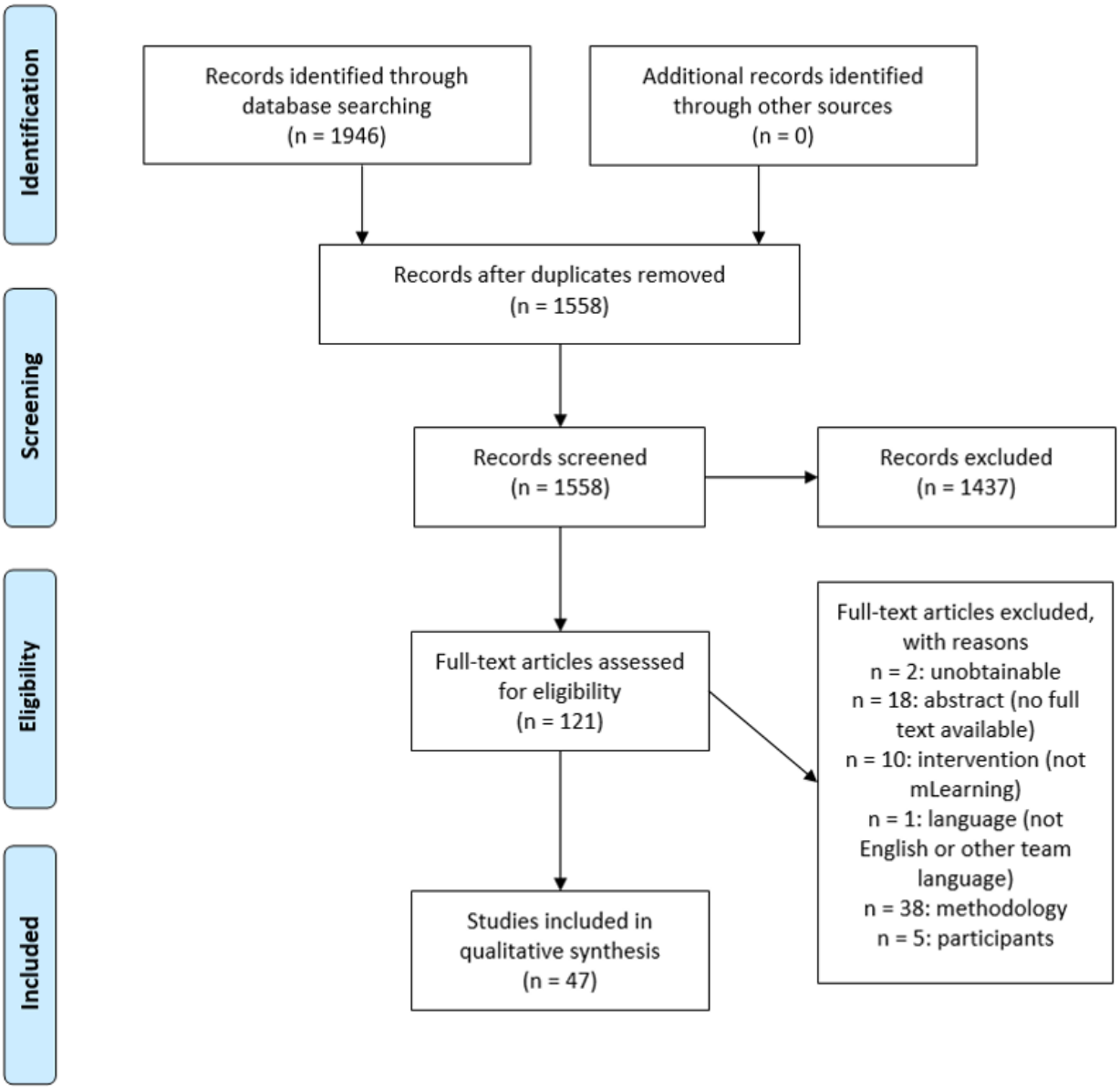
Studies were included if they examined medical and/or nursing students' (or their educators') perspectives on or experiences of mLearning. They also needed to be written in English, to involve some form of qualitative data collection or analysis (eg, focus group interviews), to collect data from learners in medical or nursing education who were involved in mLearning as defined by Wexler et al [5], and to be published after 1994.

Identifying Relevant Studies

We conducted a comprehensive search that combined terms for the concepts of digital technology, education, and health professionals. This search was conducted in February 2015 and was repeated in March 2017 on 8 electronic databases (MEDLINE, EMBASE, Cochrane Library, PsycINFO, ERIC, CINAHL, Web of Science, and International Clinical Trials Platform). Databases were searched from and including the year 1995 to March 2017 (see Multimedia Appendix 2).

All references identified were uploaded to the specialist systematic review software EPPI-Reviewer 4 (University College London) [21], and data deduplication was performed within this program. A second phase of searches was then run to identify qualitative studies of mLearning using the EPPI-Reviewer search function. These searches looked for items that had terms related to qualitative research and to mLearning (see Multimedia Appendix 2). The resulting set of references was assessed against our predefined inclusion and exclusion criteria. The criteria were developed by all authors and piloted by 4 authors (GD, GL, PL, and ZC) on a randomly selected sample of studies. The pilot was completed after there was a high level of agreement (over 90%) on the selection of studies between all 4 authors. Abstracts and full texts were each independently screened by 2 of these same 4 authors. In cases where there were difficulties reaching consensus on inclusion of a particular text, a third provided the deciding judgment. Figure 2 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart for the study.

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses chart.



Describing Studies and Appraising Their Quality

Features of the included studies were described according to the following characteristics: (1) aim; (2) sample characteristics (ie, size, country, and study population); (3) type of mLearning device and apps used; (4) type of study (ie, study of an intervention or inquiry into an existing phenomenon); (5) type of mLearning (eg, reference repository); and (6) study design, such as the type of data collection and sampling approach.

Quality of the final set of studies was assessed using a tool used in previous studies [22,23], where the quality of each study is rated using a total of 7 questions set within 2 dimensions—trustworthiness and usefulness of the findings. The first dimension captures the degree to which the methods were used to ensure rigor; the second, for the purposes of this review, focused on the complexity of analysis of the mLearning strategy. Furthermore, 2 reviewers independently rated each study (high, medium, or low for each dimension) and then compared judgments before coming to a consensus. Studies that met the inclusion criteria were included in the review regardless of the study quality, with ratings described alongside other characteristics of the papers.

Analysis and Synthesis

Themes were identified using a framework analysis approach in which data are reduced through the development of a matrix, comparing categories of data or cases, and a synthesis is developed using an initial theoretical framework. Data were analyzed according to phases of analysis identified by Pope et al [24], starting with 3 authors (PL, GL, and GD) familiarizing themselves with the selected studies. These authors coded selected studies according to the FRAME model. The information gathered from coded text was distilled into a chart containing summaries of the themes. After the search update, 3 authors (PL, GL, and GD) undertook the same process, further

modifying the themes and subthemes. When findings were found in texts which addressed an area not covered in the initial framework, the framework themes were added to or modified accordingly. Two authors (PL and GL) wrote a narrative to describe and illustrate the themes and their relationships. In addition, 1 author (RR) then familiarized herself with all the studies and worked further with the lead author on the synthesis narrative and themes, using study texts to check and further incorporate references to individual studies.

Results

Research Questions Being Addressed and Quality of Studies

The 47 studies in this synthesis [2,10,12,25-68] varied according to study context and participant types, the research questions being addressed, types of mLearning strategies used, and aspects of study design (see Multimedia Appendix 3). A total of 37 studies were conducted in high-income countries [2,10,12,25,27-42,45-50,52,56-58,60-62,64-66], for example, the United Kingdom (n=9) [10,12,29,31,34,40,41,61,66], with the remainder set in lower-income settings [26,43,44,51, 53-55,59,63,67,68], including India [44,55,59], South Africa [53,63], Botswana [26,68], and Rwanda [67].

Studies mainly sought views of learners, but some also included educators (n=11) [26,30,39,45-48,53,61,64,65] or focused solely on educators (n=4) [30,47,48,61]. Most studies focused solely on the experience of medical staff or students (n=24) [10,12,25,27,28,30,31,33,36,40,42-45,49,51,54-56,58,60,62,67,68] whereas a smaller number of studies sampled either solely from nursing staff and students (n=19) [2,29,32,35,37,38,46-48,50, 52,53,57,59,61,63-66] or from a mixture of both doctors and nurses (n=4) [26,34,39,41]. Students were at different stages of education and so were learning in different settings. A small number of studies looked at device use aimed at supporting learning in undergraduate lecture, seminar, or laboratory environments (n=5) [29,38,43,61,63]. In all, 7 studies sought views on mLearning for the further professional and/or academic development of fully qualified doctors or nurses [31,35,52,55,57,67]. Most studies, however, sought the views of nursing and medical students, or educators, about mLearning during various clinical placements before health professional registration.

The purpose of the majority of studies was to evaluate pilot mLearning approaches (n=32) [10,12,25-29,31-33,37-44, 46,47,49,55,58-66,68] that were implemented in medical and nursing contexts. Furthermore, 2 of these evaluations examined the provision of mobile hardware without describing specific software arrangements [33,60]. In a further 9 evaluations, mobile devices had been designed primarily to be reference repositories [10,12,28,31, 32,37,41,49,68], for example, students were loaned a PDA with preloaded medical texts by their institutions [28]. In a further 8 evaluations, devices were aimed at supporting learning through use of a suite of recommended apps or software [25,27,39,40,47,61,62,65]. In 3 of these 8, the studies focused in particular on students' use of electronic logs or ePortfolios to reflect on and/or evaluate their experiences or learning [27,39,40]. The remaining evaluations examined a variety of specific mLearning strategies, including the use of multimodal techniques (eg, those using videos of clinical skills, whiteboards, and presentation software) for group or individual activities (n=5) [29,42,55,59,64], augmented reality (n=2) [38,43], messaging services (n=4) [26,44,46,66], a social media–enabled discussion group (n=1) [63], and a mobile app to prompt specific clinical behaviors [58].

Of the remaining studies, a further 4 explored the co-design of, or needs for, specific future mLearning interventions [45,50,51,67]. A final set of 11 studies were not conducted with the purpose of designing or evaluating a specific intervention. Instead, these studies explored students' experiences of using mobile devices to enable their own learning in the absence of an institutionally planned mLearning initiative [2,30,34-36,48,52-54,56,57].

Studies predominantly employed a mixed-methods research design (n=33). These studies used one or a mix of qualitative data collection methods, such as focus group discussions (n=13) [12,26,27,33,37-40,55,59,60,65,68], group or individual interviews (n=15) [25,34-36,41-44,49,51,52,58,64,66,67], and analysis of textual survey responses (n=9) [12,27,29,32, 34,36,40,46,67]. A smaller number of studies used only qualitative methods (n=14) [28,30,31,45,47,48,50,53,54, 56,57,61-63], which included focus group discussions (n=6) [28,47,48,50,54,63], group or individual interviews (n=7) [30,45,53, 56,57,61,63], textual reflection or journals (n=4) [31,47,61,63], and participant observation [57].

Multimedia Appendix 4 displays the quality appraisal of studies included in this synthesis in terms of quality. Ten studies were judged to have highly reliable findings [12,28,47,51-54, 58,59,62] whereas 12 were deemed to be highly useful for this review [12,29,31,39,49,53,54,56,57,62,63,68]. Only 4 studies were considered both highly reliable and useful [12,53,54,62]. The ratings for each study are listed in Multimedia Appendix 3 and Multimedia Appendix 4.

Substantive Findings

The narrative below presents an overview of study participants' views of mLearning organized under the spaces in which the device, learner, and their social setting interact (device usability, social technology, interaction learning, mLearning processes and implementation in clinical contexts). Table 1 provides illustrative quotations. The full synthesis narrative, which includes citations to the studies that support each theme, is available as Multimedia Appendix 5.

Analysis revealed that the progress of mLearning strategies in medical and nursing education often appeared to be shaped by processes that were out of the hands of learners and their teaching staff. Instead, issues raised sometimes related to other actors in the institutional contexts in which learning was taking place and the implementation of policies within these learning settings. An additional factor shaping the operation of mLearning strategies was social norms governing the use of mobile devices in clinical and classroom settings.

Table 1. Illustrative quotes according to theme.

FRAME model themes	Quotes (from learners unless otherwise specified)
Device usability	
Portability means efficiency but also vigilance	"Much, much quicker than flicking through the paper version. . . Looking things up in the paper BNF [British National Formulary] for the n-th time on ward rounds puts time pressure on the junior doctor causing stress and increasing risk of errors." [10, p. 8]
	"You could do that [feedback] in a few minutes on your phone, rather than doing it or on a piece of paper that you lose." [40, p. 928]
	"Carrying books is a drag, now I'm a 'lightweight'." [28, p. 614]
	"The places I feel uncomfortable using [the mobile device] are outside, like in the mall or in a kombi [public transportation], because it's sort of a big thing, and I think it could attract thieves." [68, p. 75]
Fit for purpose hardware, software, and data	"I preferred working on the e-portfolio and entering data via computer as the screen was too small on the PDA to be practical and efficient." [39, p. 652]
	"The use of the device got me thinking what I actually needed and the sheer fact that a laptop is too large and cumbersome to carry around with you. I wanted something that I could boot up quite instantly and get on the Wi-Fi; go transfer files and this is ideal." [62, p. 574]
	"I think [a tablet] would be better than a [smartphone] because if it was an [tablet] you could actually have lectures on there and it would be big enough to read and work on." [40, p. 928]
Ownership, personalization, and sense of self	"I can access it [the mobile device] anytime ... and it is mine to use ..." [28, p. 613]
	"I've sometimes forgotten my handheld and had the feeling of being naked in a way." [28, p. 616]
	"It is part of my life now […] a means of contact, a means of learning. You know, people who have phones just learn a lot." [53, p. 1401]
	"I find I am having more and more problems with exams because I cannot look up easily what I normally look up... everyday on my [smartphone]." [33, p. 134]
Social technology	
Devices can impact care and learning relationships	"Well, it's not that I don't use a [PDA], I use it for looking up drugs and things, but I think in a conversation it is kind of awkward to kind of pull it out and break eye contact." [58, p. 5]
	"Because [the doctors] think that I'm not concentrating with them while using technology, whether it's [a smartphone or tablet]… I'm writing notes or something, but … in the beginning they didn't like the fact that I'm using this." [57, p. 5]
Devices raise issues of professionalism and practice boundaries	"These days with the younger generation, if you pull out your [tablet or PDA] and you come up with the information, you are seen as competent. You are seen as having the advanced knowledge. If you say 'well just a minute, I have to go find my book' and you are flipping through the book then you are seen as old fashioned and that you aren't as current as you should be." [35, p. 12]
	"You know someone will say 'Hey put your phone down' or 'Check your message later' or something and you can't say 'Oh I'm actually looking…' it just looks unprofessional so to be honest I don't use it when I'm in front of a patient or with the doctors…When we…on an actual round I am very careful not to pull my phone out because it's still a phone you know so I think the stigma is that you're then distracted because it's a phone and it could be…you know if the doctor is talking." [56, p. 5]
	"I think some doctors have made comments about 'What are you doing on that, are you texting someone, or playing games'." [12, p. 6]
Negotiating the social aspect of mLearning[a]	"As the patient was an elderly gentleman I was slightly apprehensive that he wouldn't appreciate me using a phone during the consultation however with explanation of my actions he was perfectly content with my use of [the device]." [31, p. 6]
	"When you are dealing with a patient it is easy to access that list and decide on the right medication together. It is also handy when you have a laboratory result and you want to find out what you can do in terms of additional laboratory research." [25, p. 332]
Interaction learning	
Facilitated interaction and learning	"The students explained… '[We show the picture] to flat mates. This is the case I have seen. [...] The whole batch gets it. [.. .] We proudly show it to the others'." [54, p. 1160]
	"I liked the fact that it was anonymous, so it gave me the freedom to ask anything without the fear of being criticised without it feeling as if I'm asking a *stupid* question." [63, p. 5]
	"[names a social media discussion group], I love it. …I'm part of the group… He [the group convenor] asks questions to medical students and gives the correct answers… there are more than 15000 people." [54, p. 1160]

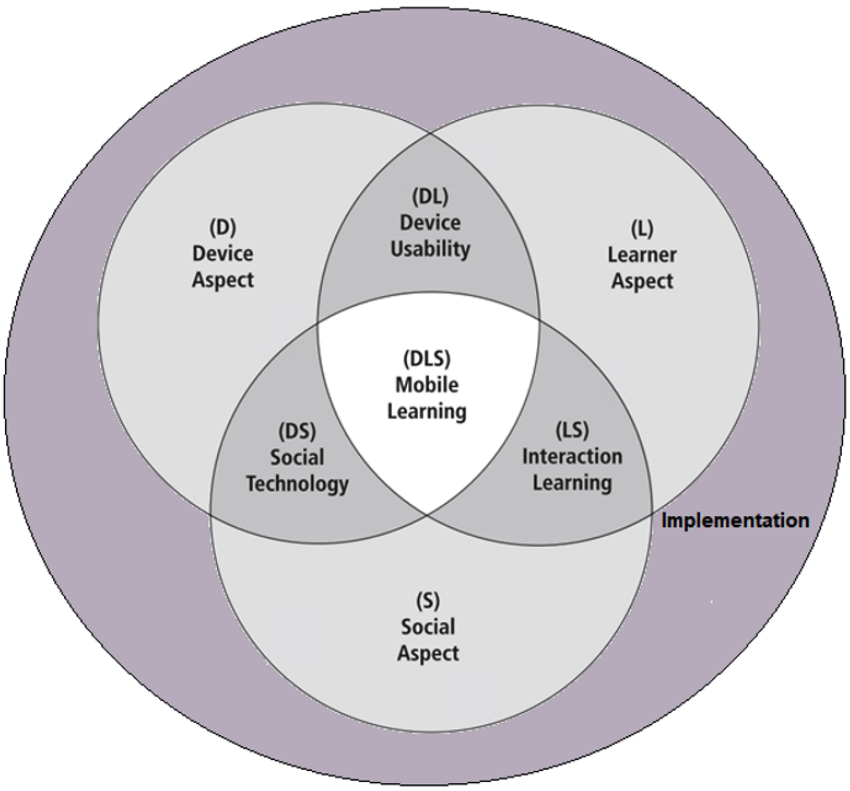
FRAME model themes	Quotes (from learners unless otherwise specified)
	"[Describing a social media facilitated student group]… Sometimes you use the group afterwards, after you have managed the patient, to see how you went, where you went wrong, how you did, or sometimes they say I messed up. Then, they give you the reasons, or sometimes they will tell you, oh, well done, but you missed that and that." [53, p. 1400]
	"[describing peer evaluation of clinical skills via Skype] I have learnt a lot and by students asking me questions. I feel my own knowledge has improved." [Educator] [51, p. 467]
Organizing learning using mobile devices	"…sharing information and allocating tasks to different members …it can allow that interaction to happen across distance. … PDAs would help keeping the interaction that coordinate the [problem based learning] process, in tagging people (peers, clinicians and the …faculty)" [Educator]. [45, p. 116]
	Use of the mobile device during downtime, such as skim reading meeting agendas while on the train …was mentioned as 1 of the main benefits of having the portable device (eg, "…instead of having a paper base you can just scroll through the minutes just to remind yourself"). [61, p. 573]
Reflective learning for clinical practice	"I don't use my phone immediately. I will write down the things we didn't know, we nod our heads and then when we leave we'll sit on our tea break and look them up quickly to make sure we understand or we know what we are talking about." [56, p. 4]
	"When we are together [in school settings], we share and discuss the photos. Some [conditions] we learn in school take a long time to see [in practice settings]. So, when you witness this condition and you are not together with your colleagues, you take this picture. […] Then you look at the picture and [later] discuss it, if it corresponds with what we have learned." [53, p. 1400]
	"[written scenario] When teaching is impromptu, conventional multimedia equipment may be either unavailable or inappropriate. … The portability of the Smartphone facilitated teaching anatomy in the context of its clinical application within general surgery. It provided visual stimuli to enrich several ad hoc teaching experiences in a single day." [Educator] [10, p. 7]
	"A lot of people also discovered that you could use Facebook on it, and also games and stuff … I feel that when you are in the hospital, or actually when you are in the OR, and you are doing something on your iPod, whatever it is, you will be distracted from the process, and it takes longer to react on the things that are happening." [57, p. 1106]
mLearning processes	
Changes in pedagogy and learning	"In contrast to the previously mentioned statements made by teachers about students' uncritical and non-reflective use of ICT, the teachers also acknowledged positive changes with respect to the *division of labour*, as indicated in the following statement by a teacher: 'There has been a dramatic change. We don't have to teach everything now. It's not teacher based learning. It is student based learning. We just tell them and guide them. We give them topics. We tell them to look up and search those topics on the internet and we ask them to verify them from the textbooks. If they find something new and interesting they can ask us. The students are helping us. They are stimulating us to study more. It's a two way conversation. And the students are also contributing'." [54, p. 1161]
	"The use of the [tablet] allowed for the shared construction of knowledge between the teachers and the students. One comment was 'I found the immediacy of this learning immensely powerful for my own learning and the student's … able to look together. In fact, one student pulled their [smartphone] and said, 'I'll race you!' While another commented, 'off into the internet to find out together!' to find the answer to a clinical question that neither knew the answer to'." [47, p. 4]
Learning to mLearn	"I was quite averse to it at first –I was one of the haters... [interviewer: What changed your mind?]… I think it's actually finding I did use the PDA and it did come in handy several times. It just makes life a bit easier." [12, p. 7]
	"Actually, I was shown by my daughter at home. […] So I showed my colleagues, yeah." [53, p. 1401]
	"[talking about not being able to view past assessments on a smartphone] If I actually saved it on the phone it would be useful to actually learn from, because before I went to do my next [clinical evaluation exercise], I could look at my last [one] and go okay, several times doctors have said that I should say this." [40, p. 928]
The implementation of mLearning in clinical contexts	
Institutional infrastructure and resources	"Loss of carrier signal or connection was a recurring event. … One lecturer described their experience, 'this week I had a problem with 3G connection, so missed a day using [my tablet] while sorting that out'." [Educators] [47, p. 4]
	"Several schools talked about the importance of all the sites having Wi-Fi. … [one reported that a] 'commonly cited reason for our clerkship students to not use them was if they were at a site where the Wi-Fi was unreliable or unavailable'." [Educators] (30, p. 1154]

FRAME model themes	Quotes (from learners unless otherwise specified)
mLearning training and technical support	"[Training could be improved] If the [training workshop] hour was tailored to the tool [mobile device]… interviewing each other did not work… we just talked." [45, p. 116]
	"…have some base level training…for everybody…specifically on knowing how to turn it on and manipulate it, how it should be used and how it benefits medical education, how the faculty or school expect it to be used. …you need drop-in sessions, extra assistance or individual assistance for people struggling with the technology…" [Educator] [45, p. 116]
mLearning needs course planning and institutional leadership	"Focus on the areas where you really feel like the [tablet] is an appropriate tool for the thing you want to do, but do not try to wedge [it] into areas where it may or may not be the best thing to use… there are things you can do and things you cannot do at each step along the way." [Educator] [30, p. 1154]
	"the participants categorised the teachers as being either *old school*, a term they frequently assigned to the older generation, or *new school*, a term afforded to more youthful practitioners—the former being described as *paper-dependent* and being offended when interviewees used their devices in front of them. Many of the *old school* doctors did not appear to understand the reliance that the younger student generation have on their mobile devices as learning tools." [56, p. 5]
	"It's things like that [teacher advocacy] which encourage you, maybe I will bring it with me tomorrow and take it on the ward round with me." [12, p. 7]
	"…it is easy to see the value of some technologies where it works very well and it is very easy to get over-enthusiastic about it and then not realize that people might not be ready to actually use that technology for whatever reasons…" [Educators] [45, p. 116]

[a]mLearning: mobile learning.

The FRAME model was adapted to account for these differing findings. Social technology was altered to account for the impact of mobile devices on social interaction, rather than to describe how it enabled connection between multiple interfacing entities. We also added an additional circle that contained the model's overlapping 3 circles (see the theme *Implementation of mLearning in clinical learning contexts* in Figure 3). Otherwise, the synthesis produced themes that could be grouped under the aspects of learning represented in Koole's original model, and we used subthemes under each aspect to help illustrate attributes that relate to the specifics of doctors' and nurses' learning.

Figure 3. Framework for the Rational Analysis of Mobile Education (FRAME) model adapted for this study.



Device Usability

Participants referred to the physical, technical, and functional characteristics of mobile devices in relation to an individual's learning, which involved access, manipulation, and storage of information. Subthemes explored possible positive and unintended consequences of devices being mobile, views on the sufficiency of device functionality, and ideas about the individualistic nature of device use. Enthusiasm for a mobile device focused on efficiency yet was accompanied by an awareness of the need for caution, in terms of a risk of loss of device or contamination in certain settings. Reports of problems attributed to hardware and software were seen in a range of studies, with some participants noting that screens were too small for reading documents [37,40,61] or that they lost information owing to system crashes [12,27,38]. Learners reported wanting devices that suited their own specific needs, describing device use as either *a way of life* (p. 111 [65]) or as a *part of my life now* (p. 1401 [53]).

Social Technology

This theme encompassed participants' perspectives on social responses to mobile devices and many studies were conducted in clinical contexts. Within these settings, students were expected to combine their learning with practice, which resulted in the device influencing social interactions with a number of actors, including their supervisors, patients, and peers. Mobile devices seemed to hold the symbolic value of being a form of technology for recreational use rather than for learning, owing to multiple functions enabling information retrieval alongside highly social activities, such as sending and viewing messages.

Mobile devices, thus, affected students' relationships with patients and their professional identity. For example, although mobile devices were seen as potentially strengthening communication between clinicians and patients, concerns were raised about possible interference with activities at the bedside. There were reports of feeling *rude* [37,56,58] or *awkward* [31,58]. Although some feared being viewed as unprofessional by either patients or colleagues, others linked device use with perceptions of increased competence. Although these social norms did result in some students being reluctant to openly use the device, others developed strategies for negotiating device use including asking for permission, explaining device use, and jointly using devices with patients.

Interaction Learning

Studies highlighted how mobile use enabled learning processes contingent on students' interaction with their academic institution, peers, and practice. Students used these multiple forms of interaction to learn cooperatively with their peers, organize competing demands of clinical practice and study, and situate their learning within clinical contexts. These forms of mLearning encompassed individual device use for the purposes of information retrieval and organization to device-enabled group work.

Regarding device-enabled group work, online study groups were described as enabling case discussions, and participants commented upon the pros and cons of structured, cooperative peer assessment approaches. It also enabled students to remotely contact supervisors while working in clinical settings. Meanwhile, students and staff described using mobile devices to help them organize their learning, for example, to access information on learning activities when in a clinical setting. As such, students emphasized the value of access to immediately relevant or difficult-to-access clinical cases or using devices to prepare immediately before encounters.

Mobile Learning Processes

Some participants reflected upon mLearning as a whole process. Subthemes here represent views on how educational roles could be changing and the process of adapting one's learning to the mobile device. In terms of the former theme, both students and tutors described how they were participating on more equal terms. Enthusiasm, however, was far from universal, and positive comparisons were made with more traditional forms of learning. Frustration and impatience were expressed about the process of learning how to use a device. Participants described a reliance on others, in particular peers and friends, and although familiarity reportedly improved over time, the need for support and repeat training was emphasized. Uncertainty was voiced over the trustworthiness or reliability of information being distributed through mLearning apps or websites.

The Implementation of Mobile Learning in Clinical Contexts

Study participants reported challenges with mLearning that had little to do with interactions between students, devices and their contents, patients, and tutors. Here, what was implicated were insufficient institutional structures and resources, a lack of device-focused training and support, and limited planning and leadership of mLearning programs. For example, the importance—and yet variability—of network connectivity was emphasized by both tutors and students, and concerns were raised over program provision of ill-suited devices.

The use of mLearning strategies did not always appear to have been planned with course content or pedagogy in mind, or with consideration of the attributes required by teaching staff. Students reported they were offered little guidance on how to integrate mobile devices into their learning activities as well as a lack of device knowledge among clinical instructors. Experiential or ongoing training and local technical support were particularly valued; participants reported forgetting functions covered during orientation, and support had been experienced by some as fragmented. Reports were made across numerous studies of disapproval for device use among supervising staff in clinical settings and of students, as a result, being hesitant to use a device openly. A range of proposals were made across the studies, including initiatives to improve staff awareness about the value of portable devices and the development of codes of conduct.

Discussion

Principal Findings

To our knowledge, this is the first systematic review synthesizing qualitative research findings about health professionals' experiences of mLearning. The review identified

a total of 47 studies that varied in the types of health professionals involved, their stage of learning, and the mLearning strategies considered. Qualitative data in the majority of studies had been sought so as to pilot mLearning approaches or examine nonspecific use of mobile devices for learning. In many studies, qualitative findings were slim and provided little explanatory detail but across this body of work, it is possible to identify recurring themes about experiences and some explanatory narratives from both learners and educators.

Our synthesis of findings from these studies illustrates some of the potentials of mLearning but also some of the challenging realities for students, doctors, and nurses who are learning in contexts where mobile devices have either formally been introduced or tend to be common. Early commentators on mLearning envisaged methods of delivery that would be *highly suited* to the *just enough, just in time, and just for me* demands of twenty-first century learners [69]. Students in the studies we reviewed did indeed value devices for the possibility of lessening cognitive loads and helping to make good use of time. They also described device use in terms of individualized needs and preferences. However, although both learners and educators described the potential value of devices for accessing, organizing, and enhancing learning, limitations in hardware were reported across the full time period covered by our included studies. Researchers in other spheres of education have also emphasized the need for devices to be fit for purpose [70-72]. Future mLearning strategies for medical and nursing education should, therefore, be developed with an awareness of device affordances for the learning activities required.

This synthesis identifies additional social and institutional factors that seem key for understanding how mLearning for health professionals might be implemented to the best effect. In particular, throughout much of their training, medical and nursing students need to combine learning with professional caring responsibilities. The social aspects of learning that are already complex within more formal education settings become considerably multilayered when students are, for example, at the bedside or in an operating theatre. On top of interactions with information, fellow learners, and formal educators come interactions with a variety of other health professionals and with patients. Learning can happen through peripheral participation in clinical activities, observation, role modeling, and reflective activities, as well as through work with lecturers, supervisors, and other students, and with text books and other information sources. mLearning needs to fit into this mix of interactions but instead our synthesis contains accounts of reluctance, told by both students and educators, toward the use of mobile devices in the clinical workplace because of existing, often implicit, rules for practice. Although negotiation was said sometimes to enable device use for learning, participants in more than one study identified a need for procedural guidance on device use, echoing calls from education more broadly [73,74].

We found Koole's conceptualization of mLearning, involving a combination of learner, device, and social aspects, to be helpful when organizing findings. However, the themes of social technology and mLearning processes in our synthesis diverged from that of Koole's conceptualization. With regard to social technology, Koole's [17] model emphasizes how *mobile devices enable communication and collaboration among multiple individuals and systems* (p. 34), whereas findings from studies within this synthesis instead identify impacts of mLearning on interactions with patients and the management of professional identity.

The FRAME model [17] represents the mLearning processes as an integration of the device, learner, and social aspects that provides for *enhanced collaboration between learners, access to information, and deeper contextualization of learning* (p. 38). Although there were some positive accounts of device use for situated learning and of cooperative learning activities, accounts from studies in our synthesis placed more emphasis on the process of learning how to apply devices for the purposes of learning. Qualitative research into mLearning for health professional education appears still to be in its infancy, with few studies referring to the supported integration of mLearning within a pedagogically informed program of study.

Limitations

This study provided a comprehensive overview of current qualitative research on mLearning strategies in medical and nursing education. Its strengths include a sensitive search strategy encompassing several bibliographic databases and independent screening by pairs of reviewers, both lowering the likelihood that relevant literature would be overlooked as well as coding and synthesis work done independently and in pairs, aimed at bringing a variety of perspectives to the act of making sense of a heterogeneous set of study findings.

The review is, nevertheless, limited by the qualities of the reviewed studies, especially those employing mixed-methods designs, wherein the quantitative component was given far heavier weighting than qualitative findings. Few studies described in sufficient detail the steps taken by researchers to ensure confidence in the quality of their findings. The majority of studies offered little explanation of methods used to sample participants and collect or analyze data. Moreover, there were many studies in which authors provided little evidence as to how they arrived at their findings. These studies offered few quotes from participants, sometimes making it difficult to decipher whether results were guided by the perspectives of respondents.

Another methodological limitation was that nearly all of the studies explored learning within clinical settings. Owing to this, much of the mLearning described would be classified as informal learning, that is, learning which results from incidental day-to-day activities. Our synthesis, therefore, contains little detail sourced from experiences of programs set up to encourage mLearning in university settings.

Finally, few studies explicitly referred to educators' learning theories or described course structures in any detail, which meant study findings could not be explored in terms of different objectives for students' learning. Efforts should be made in future qualitative studies to clearly define the educational purposes of the mLearning programs concerned to make findings more applicable to given learning circumstances.

Comparison With Earlier Work

Findings about a lack of device training, technical support, and other forms of institutional support led to one of the biggest modifications to the FRAME model seen in our synthesis, which was the development of an additional aspect—implementation in a clinical context. This theme highlighted that even when mobile devices had been introduced for the purpose of evaluation, this appeared to have been done with insufficient consideration of course content or needs at the institutional level, including both sufficient Wi-Fi coverage and the alignment and capacity of teaching staff to use mLearning. Insights might be gained through the study of device maintenance services on campus [45] and the implementation of mLearning strategies with learning outcomes as well as a wider curriculum in mind [30,45,47]. Studies of change management around learning technology in higher education outside the field of health might also be relevant here, as they have explored the potential for initiatives, such as staff as champions, and strategic contextual analyses [75-77].

This review also starts to identify gaps in the literature where additional studies might throw light on a more complete range of mLearning practices within medical and nursing education. For example, study authors made no mention of discussion by study participants of ethical concerns over patient privacy and data security. Educational experts, however, raise concerns vis-à-vis use of mLearning strategies in other settings, arguing that these interventions can compromise students' confidentiality as private data can be potentially disseminated to unintended audiences [74,78,79]. In terms of medical and nursing education, there is the added concern that the welfare of patients might be compromised. The need identified above for guidance for health professionals' device use, consequently, will require a strong ethical component.

Conclusions

The findings of our review have underlined that there is still much to be understood about what is involved in mLearning for medical and nursing education. Our review has indicated that mLearning can potentially play a substantial role as students are already likely to be using mobile devices for a number of differing purposes associated with their learning, ranging from communication with supervisors to organization of tasks. The multipurpose nature of mobile devices means that students can personalize these tools toward their learning needs, which entails a process of learning within itself.

As with any complex tool used for educational purposes, mobile devices should be appropriately incorporated into the structures of academic and medical institutions and steps need to be taken to ensure that learners fully comprehend the functions of each mobile device or app used for learning. These 2 considerations can only be addressed by paying close attention to the process of implementing mLearning strategies in medical and nursing scholarship and the building of an educational infrastructure that enables use of mLearning techniques.

Acknowledgments

We gratefully acknowledge support for this work by National Healthcare Group, Singapore, the Lee Kong Chian, School of Medicine, Nanyang Technological University, Singapore, and UCL Institute of Education, University College London.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Protocol for the review.

[PDF File (Adobe PDF File), 450KB - jmir_v21i2e12895_app1.pdf]

Multimedia Appendix 2

Search strategy.

[DOCX File, 59KB - jmir_v21i2e12895_app2.docx]

Multimedia Appendix 3

Details of included studies.

[DOCX File, 37KB - jmir_v21i2e12895_app3.docx]

Multimedia Appendix 4

Quality appraisal of studies included in the synthesis.

[DOCX File, 47KB - jmir_v21i2e12895_app4.docx]

Multimedia Appendix 5

Full synthesis narrative with linked citations.

[DOCX File, 23KB - jmir_v21i2e12895_app5.docx]

References

1. Triola MM, Huwendiek S, Levinson AJ, Cook DA. New directions in e-learning research in health professions education: report of two symposia. Med Teach 2012;34(1):e15-e20. [doi: 10.3109/0142159X.2012.638010] [Medline: 22250691]
2. Mather C, Cummings E. Issues for deployment of mobile learning by nurses in Australian healthcare settings. Stud Health Technol Inform 2016;225:277-281. [Medline: 27332206]
3. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]
4. Rasmussen K, Belisario JM, Wark PA, Molina JA, Loong SL, Cotic Z, et al. Offline eLearning for undergraduates in health professions: A systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010405 [FREE Full text] [doi: 10.7189/jogh.04.010405] [Medline: 24976964]
5. Wexler S, Brown J, Metcalf D, Rogers D, Wagner E. eLearning guild research 360 report: Mobile learning. Santa Rosa: CALearning Guild; 2008.
6. ITU. International Telecommunications Union. 2015 May 26. ITU releases 2015 ICT figures URL: https://www.itu.int/net/pressoffice/press_releases/2015/17.aspx [accessed 2019-02-10] [WebCite Cache ID 765VXVmZO]
7. GSMA. GSMA. 2015. The Mobile Economy URL: https://www.gsma.com/mobileeconomy/archive/GSMA_ME_2015.pdf [accessed 2019-02-10] [WebCite Cache ID 6vAKBEDNW]
8. Alegría DA, Boscardin C, Poncelet A, Mayfield C, Wamsley M. Using tablets to support self-regulated learning in a longitudinal integrated clerkship. Med Educ Online 2014 Mar 12;19(1):23638 [FREE Full text] [doi: 10.3402/meo.v19.23638] [Medline: 24646438]
9. Doherty I, Sharma N, Harbutt D. Contemporary and future eLearning trends in medical education. Med Teach 2015 Jan;37(1):1-3. [doi: 10.3109/0142159X.2014.947925] [Medline: 25154534]
10. Hardyman W, Bullock A, Brown A, Carter-Ingram S, Stacey M. Mobile technology supporting trainee doctors' workplace learning and patient care: an evaluation. BMC Med Educ 2013 Jan 21;13:6 [FREE Full text] [doi: 10.1186/1472-6920-13-6] [Medline: 23336964]
11. Lehman K. Clinical nursing instructors' use of handheld computers for student recordkeeping and evaluation. J Nurs Educ 2003 Jan;42(1):41-42. [Medline: 12555822]
12. Davies BS, Rafique J, Vincent TR, Fairclough J, Packer MH, Vincent R, et al. Mobile Medical Education (MoMEd) - how mobile information resources contribute to learning for undergraduate clinical students - a mixed methods study. BMC Med Educ 2012 Jan 12;12:1 [FREE Full text] [doi: 10.1186/1472-6920-12-1] [Medline: 22240206]
13. Tanaka PP, Hawrylyshyn KA, Macario A. Use of tablet (iPad®) as a tool for teaching anesthesiology in an orthopedic rotation. Rev Bras Anestesiol 2012;62(2):214-222 [FREE Full text] [doi: 10.1016/S0034-7094(12)70119-8] [Medline: 22440376]
14. 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 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
15. Al-Shorbaji N, Atun R, Car J, Majeed A, Wheeler E. World Health Organisation. Geneva, Switzerland: World Health Organisation; 2015. eLearning for undergraduate health professional education: A systematic review informing a radical transformation of health workforce development URL: https://www.who.int/hrh/documents/14126-eLearningReport.pdf [accessed 2019-02-19] [WebCite Cache ID 76ISvOTGh]
16. Laurillard D. Pedagogical forms for mobile learning: framing research questions. In: Pachler N, editor. Mobile learning towards a research agenda. London: WLE Centre, IoE; 2007:153-175.
17. Koole M. A model for framing mobile learning. In: Ally M, editor. Mobile learning: Transforming the delivery of education and training. Athabasca, Canada: AU Press; 2009:25-47.
18. Harris J. Using qualitative research to develop robust effectiveness questions and protocols for Cochrane systematic reviews. In: Noyes K, editor. Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions. UK: Cochrane Collaboration Qualitative Methods Group; 2011.
19. Bearman M, Dawson P. Qualitative synthesis and systematic review in health professions education. Med Educ 2013 Mar;47(3):252-260. [doi: 10.1111/medu.12092] [Medline: 23398011]
20. Krishnasamy C, Ong SY, Yock Y, Lim I, Rees R, Car J. Factors influencing the implementation, adoption, use, sustainability and scalability of mLearning for medical and nursing education: a systematic review protocol. Syst Rev 2016 Dec 19;5(1):178 [FREE Full text] [doi: 10.1186/s13643-016-0354-x] [Medline: 27756442]
21. Thomas J, Brunton J, Graziosi S. EPPI-Centre Software. 2010. EPPI-Reviewer 4: Software for research synthesis URL: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2967 [accessed 2019-02-19] [WebCite Cache ID 76ITbbAuc]

22. Rees R, Oliver K, Woodman J, Thomas J. The views of young children in the UK about obesity, body size, shape and weight: a systematic review. BMC Public Health 2011 Mar 25;11:188 [FREE Full text] [doi: 10.1186/1471-2458-11-188] [Medline: 21439062]
23. Harden A. 'Qualitative' research, systematic reviews, and evidence-informed policy and practice. London, UK: University of London; 2007.
24. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. Br Med J 2000 Jan 08;320(7227):114-116 [FREE Full text] [Medline: 10625273]
25. Akkerman S, Filius R. The use of personal digital assistants as tools for work-based learning in clinical internships. JRTE 2011 Jun;43(4):325-341. [doi: 10.1080/15391523.2011.10782575]
26. Armstrong K, Liu F, Seymour A, Mazhani L, Littman-Quinn R, Fontelo P, et al. Evaluation of txt2MEDLINE and development of short messaging service-optimized, clinical practice guidelines in Botswana. Telemed J E Health 2012 Jan;18(1):14-17. [doi: 10.1089/tmj.2011.0014] [Medline: 22150636]
27. Avila J, Sostmann K, Breckwoldt J, Peters H. Evaluation of the free, open source software WordPress as electronic portfolio system in undergraduate medical education. BMC Med Educ 2016 Jun 03;16:157 [FREE Full text] [doi: 10.1186/s12909-016-0678-1] [Medline: 27255920]
28. Axelson C, Wårdh I, Strender L, Nilsson G. Using medical knowledge sources on handheld computers--a qualitative study among junior doctors. Med Teach 2007 Sep;29(6):611-618. [doi: 10.1080/01421590701507294] [Medline: 17978970]
29. Davies M. Using the Apple iPad to facilitate student-led group work and seminar presentation. Nurse Education in Practice 2014 Aug;14(4):363-367. [doi: 10.1016/j.nepr.2014.01.006]
30. Deutsch K, Gaines JK, Hill JR, Nuss MA. iPad experience during clinical rotations from seven medical schools in the United States: lessons learned. Med Teach 2016 Nov;38(11):1152-1156. [doi: 10.3109/0142159X.2016.1170778] [Medline: 27117525]
31. Dimond R, Bullock A, Lovatt J, Stacey M. Mobile learning devices in the workplace: 'as much a part of the junior doctors' kit as a stethoscope'? BMC Med Educ 2016 Aug 17;16(1):207 [FREE Full text] [doi: 10.1186/s12909-016-0732-z] [Medline: 27530343]
32. Doyle GJ, Furlong KE, Secco L. Information literacy in a digital era: understanding the impact of mobile information for undergraduate nursing students. Stud Health Technol Inform 2016;225:297-301. [Medline: 27332210]
33. Ellaway RH, Fink P, Graves L, Campbell A. Left to their own devices: medical learners' use of mobile technologies. Med Teach 2014 Feb;36(2):130-138. [doi: 10.3109/0142159X.2013.849800] [Medline: 24195445]
34. Evangelinos G, Holley D. A Qualitative exploration of the EU digital competence (DIGCOMP) framework: a case study within healthcare education. 2014 Presented at: 1st International Conference on E-Learning, E-Education, and Online Training; September 2014; Bethesda, US. [doi: 10.1007/978-3-319-13293-8_11]
35. Fahlman D. Examining informal learning using mobile devices in the healthcare workplace. CJLT 2013;39(4):1-21.
36. Fan S, Radford J, Fabian D. A mixed-method research to investigate the adoption of mobile devices and Web2.0 technologies among medical students and educators. BMC Med Inform Decis Mak 2016 Apr 19;16:43 [FREE Full text] [doi: 10.1186/s12911-016-0283-6] [Medline: 27094813]
37. Farrell MJ, Rose L. Use of mobile handheld computers in clinical nursing education. J Nurs Educ 2008 Jan;47(1):13-19. [Medline: 18232610]
38. Garrett BM, Jackson C, Wilson B. Augmented reality m-learning to enhance nursing skills acquisition in the clinical skills laboratory. Interactive Tech & Smart Ed 2015 Nov 16;12(4):298-314 [FREE Full text] [doi: 10.1108/ITSE-05-2015-0013]
39. Garrett BM, Jackson C. A mobile clinical e-portfolio for nursing and medical students, using wireless personal digital assistants (PDAs). Nurse Educ Today 2006 Dec;26(8):647-654. [doi: 10.1016/j.nedt.2006.07.020] [Medline: 17011674]
40. Green BL, Kennedy I, Hassanzadeh H, Sharma S, Frith G, Darling JC. A semi-quantitative and thematic analysis of medical student attitudes towards M-Learning. J Eval Clin Pract 2015 Oct;21(5):925-930. [doi: 10.1111/jep.12400] [Medline: 26153482]
41. Jamu JT, Lowi-Jones H, Mitchell C. Just in time? Using QR codes for multi-professional learning in clinical practice. Nurse Educ Pract 2016 Jul;19:107-112. [doi: 10.1016/j.nepr.2016.03.007] [Medline: 27428702]
42. Jang HW, Kim K. Use of online clinical videos for clinical skills training for medical students: benefits and challenges. BMC Med Educ 2014 Mar 21;14:56 [FREE Full text] [doi: 10.1186/1472-6920-14-56] [Medline: 24650290]
43. Küçük S, Kapakin S, Göktaş Y. Learning anatomy via mobile augmented reality: effects on achievement and cognitive load. Anat Sci Educ 2016 Oct;9(5):411-421. [doi: 10.1002/ase.1603] [Medline: 26950521]
44. Kumar L, Jamatia B, Aggarwal A, Kannan S. Mobile device intervention for student support services in distance education context - FRAME model perspective. EURODL 2011;14(2):1-9 [FREE Full text]
45. Luanrattana R, Win KT, Fulcher J, Iverson D. Mobile technology use in medical education. J Med Syst 2012 Feb;36(1):113-122. [doi: 10.1007/s10916-010-9451-x] [Medline: 20703743]
46. MacKay B, Harding T. M-Support: keeping in touch on placement in primary health care settings. Nurs Prax N Z 2009 Jul;25(2):30-40. [Medline: 19928649]
47. Mackay BJ, Anderson J, Harding T. Mobile technology in clinical teaching. Nurse Educ Pract 2017 Jan;22:1-6. [doi: 10.1016/j.nepr.2016.11.001] [Medline: 27871040]

48. Mather C, Cummings E. Unveiling the mobile learning paradox. Stud Health Technol Inform 2015;218:126-131. [Medline: 26262539]
49. Nuss MA, Hill JR, Cervero RM, Gaines JK, Middendorf BF. Real-time use of the iPad by third-year medical students for clinical decision support and learning: a mixed methods study. J Community Hosp Intern Med Perspect 2014;4(4):25184 [FREE Full text] [doi: 10.3402/jchimp.v4.25184] [Medline: 25317266]
50. O'Connor S, Andrews T. Using co-design with nursing students to create educational apps for clinical training. Stud Health Technol Inform 2016;225:334-338. [Medline: 27332217]
51. O'Donovan J, Maruthappu M. Distant peer-tutoring of clinical skills, using tablets with instructional videos and Skype: a pilot study in the UK and Malaysia. Med Teach 2015 May;37(5):463-469. [doi: 10.3109/0142159X.2014.956063] [Medline: 25182187]
52. Pilcher JW, Bedford L. Willingness and preferences of nurses related to learning with technology. J Nurses Staff Dev 2011;27(3):E10-6; quiz E17. [doi: 10.1097/NND.0b013e318217b447] [Medline: 21602623]
53. Pimmer C, Brysiewicz P, Linxen S, Walters F, Chipps J, Gröhbiel U. Informal mobile learning in nurse education and practice in remote areas--a case study from rural South Africa. Nurse Educ Today 2014 Nov;34(11):1398-1404. [doi: 10.1016/j.nedt.2014.03.013] [Medline: 24745478]
54. Pimmer C, Linxen S, Gröhbiel U, Jha AK, Burg G. Mobile learning in resource-constrained environments: a case study of medical education. Med Teach 2013 May;35(5):e1157-e1165. [doi: 10.3109/0142159X.2012.733454] [Medline: 23137244]
55. Prakash V, Thukral A, Sankar MJ, Agarwal RK, Paul VK, Deorari AK. Efficacy and acceptability of an "app on sick newborn care" in physicians from newborn units. BMC Med Educ 2016 Mar 08;16:84 [FREE Full text] [doi: 10.1186/s12909-016-0579-3] [Medline: 26956397]
56. Rashid-Doubell F, Mohamed S, Elmusharaf K, O'Neill CS. A balancing act: a phenomenological exploration of medical students' experiences of using mobile devices in the clinical setting. BMJ Open 2016 Dec 03;6(5):e011896 [FREE Full text] [doi: 10.1136/bmjopen-2016-011896] [Medline: 27142860]
57. Sergeeva A, Aij K, van den Hooff B, Huysman M. Mobile devices in the operating room: intended and unintended consequences for nurses' work. Health Informatics J 2016 Dec;22(4):1101-1110. [doi: 10.1177/1460458215598637] [Medline: 26635321]
58. Strayer SM, Pelletier SL, Martindale JR, Rais S, Powell J, Schorling JB. A PDA-based counseling tool for improving medical student smoking cessation counseling. Fam Med 2010 May;42(5):350-357 [FREE Full text] [Medline: 20461567]
59. Thukral A, Joshi M, Joshi P, Prakash V, Adkoli BV, Deorari AK. Apps for management of sick newborn: evaluation of impact on health care professionals. J Trop Pediatr 2014 Oct;60(5):370-376. [doi: 10.1093/tropej/fmu032] [Medline: 24924579]
60. Varcadipane JC, Dupaix JP, Nishimura S, Chun M, Scarcella N, Belcher G, et al. Mobile computing as a tool in orthopaedic surgery residency training. Curr Orthop Pract 2015;26(4):376-381. [doi: 10.1097/BCO.0000000000000245]
61. Wang R, Wiesemes R, Gibbons C. Developing digital fluency through ubiquitous mobile devices: findings from a small-scale study. Comput Educ 2012 Jan;58(1):570-578 [FREE Full text] [doi: 10.1016/j.compedu.2011.04.013]
62. Wells ME. Mobile learning in medical education: A case study through the lens of sleep education. PhD Thesis. Raleigh, North Carolina: North Carolina State University; 2015.
63. Willemse JJ. Undergraduate nurses reflections on Whatsapp use in improving primary health care education. Curationis 2015 Aug 13;38(2):1512 [FREE Full text] [doi: 10.4102/curationis.v38i2.1512] [Medline: 26304053]
64. Wu T. The use of a mobile assistant learning system for health education based on project-based learning. Comput Inform Nurs 2014 Oct;32(10):497-503. [doi: 10.1097/CIN.0000000000000089] [Medline: 25140444]
65. Wyatt TH, Krauskopf PB, Gaylord NM, Ward A, Huffstutler-Hawkins S, Goodwin L. Cooperative m-learning with nurse practitioner students. Nurs Educ Perspect 2010;31(2):109-113. [Medline: 20455369]
66. Young P, Moore E, Griffiths G, Raine R, Stewart R, Cownie M, et al. Help is just a text away: the use of short message service texting to provide an additional means of support for health care students during practice placements. Nurse Educ Today 2010 Feb;30(2):118-123. [doi: 10.1016/j.nedt.2009.06.010] [Medline: 19632011]
67. Rusatira JC, Tomaszewski B, Dusabejambo V, Ndayiragije V, Gonsalves S, Sawant A, et al. Enabling access to medical and health education in Rwanda using mobile technology: needs assessment for the development of mobile medical educator apps. JMIR Med Educ 2016 Jun 01;2(1):e7 [FREE Full text] [doi: 10.2196/mededu.5336] [Medline: 27731861]
68. Witt RE, Kebaetse MB, Holmes JH, Littman-Quinn R, Ketshogileng D, Antwi C, et al. The role of tablets in accessing information throughout undergraduate medical education in Botswana. Int J Med Inform 2016 Apr;88:71-77. [doi: 10.1016/j.ijmedinf.2016.01.006] [Medline: 26878765]
69. Peters K. m-Learning: positioning educators for a mobile, connected future. IRRODL 2007 Jun 15;8(2):1-10. [doi: 10.19173/irrodl.v8i2.350]
70. Raptis D, Tselios N, Kjeldskov J. Does size matter? Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. 2013 Presented at: 15th International Conference on Human-Computer Interaction with Mobile Devices and Services; August 27 - 30, 2013; Munich, Germany.

71. Kim KJ, Sundar SS. Does screen size matter for smartphones? Utilitarian and hedonic effects of screen size on smartphone adoption. Cyberpsychol Behav Soc Netw 2014 Jul;17(7):466-473 [FREE Full text] [doi: 10.1089/cyber.2013.0492] [Medline: 24694112]
72. Willemse JJ, Bozalek V. Exploration of the affordances of mobile devices in integrating theory and clinical practice in an undergraduate nursing programme. Curationis 2015;38(2):e1-e10 [FREE Full text] [doi: 10.4102/curationis.v38i2.1510] [Medline: 26852423]
73. Garrett N. Student mobile technologies: Implications for classroom management policies and procedures. : Association for the Advancement of Computing in Education (AACE); 2010 Presented at: EdMedia: World Conference on Educational Media and Technology; June 29, 2010; Toronto, Canada p. 1699-1704.
74. Andrews T, Dyson L, Smyth R, Wallace R. The ethics of m-learning: Classroom threat or enhanced learner agency. 2011 Presented at: In Proceedings of the 10th World Conference on Mobile and Contextual Learning; October 18, 2011; Beijing, China p. 295-302.
75. Bytheway A, Venter IM. Strategies for information management in education: some international experience. S Afr J Inf Manag 2014 Jun 12;16(1):1-11. [doi: 10.4102/sajim.v16i1.596]
76. Gachago D, Morkel J, Hitge L, van Zyl I, Ivala E. Developing eLearning champions: a design thinking approach. Int J Educ Technol High Educ 2017 Jul 19;14(1):30. [doi: 10.1186/s41239-017-0068-8]
77. Hains-Wesson R, Wakeling L, Aldred P. A university-wide ePortfolio initiative at Federation University Australia: software analysis, test-to-production, and evaluation phases. IJeP 2014;4(21):43-56.
78. Dyson L, Andrews T, Smyth R, Wallace R. Towards a holistic framework for ethical mobile learning. In: Berge Z, Muilenburg L, editors. Handbook of Mobile Learning. Milton Park, Abingdon, Oxon, UK: Taylor & Francis (Routledge); 2013:405-416.
79. Wishart J. Ethical considerations in implementing mobile learning in the workplace. IJMBL 2009;1(2):76-92.

Abbreviations

FRAME: Framework for the Rational Analysis of Mobile Education
mLearning: mobile learning
PDA: Personal Digital Assistant

Edited by A Marusic; submitted 23.11.18; peer-reviewed by PDJ Ehlers, S Huwendiek; comments to author 13.12.18; revised version received 04.01.19; accepted 04.01.19; published 28.02.19.

Please cite as:
Lall P, Rees R, Law GCY, Dunleavy G, Cotič Ž, Car J
Influences on the Implementation of Mobile Learning for Medical and Nursing Education: Qualitative Systematic Review by the Digital Health Education Collaboration
J Med Internet Res 2019;21(2):e12895
URL: https://www.jmir.org/2019/2/e12895/
doi:10.2196/12895
PMID:30816847

Review

Digital Problem-Based Learning in Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration

Lorainne Tudor Car[1,2], MD, MSc, PhD; Bhone Myint Kyaw[1], MBBS, MSc, PhD; Gerard Dunleavy[3], MSc; Neil A Smart[4], PhD; Monika Semwal[3], PhD; Jerome I Rotgans[5], PhD; Naomi Low-Beer[5], MBBS, MD, MRCOG; James Campbell[6], MPH, MSc

[1]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[2]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom
[3]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[4]School of Science & Technology, University of New England, Armidale, Australia
[5]Medical Education Research Unit, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[6]Health Workforce Department, World Health Organization, Geneva, Switzerland

Corresponding Author:
Lorainne Tudor Car, MD, MSc, PhD
Family Medicine and Primary Care
Lee Kong Chian School of Medicine
Nanyang Technological University
11 Mandalay Road
Singapore, 308232
Singapore
Phone: 65 69041258
Email: lorainne.tudor.car@ntu.edu.sg

Abstract

Background: The use of digital education in problem-based learning, or digital problem-based learning (DPBL), is increasingly employed in health professions education. DPBL includes purely digitally delivered as well as blended problem-based learning, wherein digital and face-to-face learning are combined.

Objective: The aim of this review is to evaluate the effectiveness of DPBL in improving health professionals' knowledge, skills, attitudes, and satisfaction.

Methods: We used the gold-standard Cochrane methods to conduct a systematic review of randomized controlled trials (RCTs). We included studies that compared the effectiveness of DPBL with traditional learning methods or other forms of digital education in improving health professionals' knowledge, skills, attitudes, and satisfaction. Two authors independently screened studies, extracted data, and assessed the risk of bias. We contacted study authors for additional information, if necessary. We used the random-effects model in the meta-analyses.

Results: Nine RCTs involving 890 preregistration health professionals were included. Digital technology was mostly employed for presentation of problems. In three studies, PBL was delivered fully online. Digital technology modalities spanned online learning, offline learning, virtual reality, and virtual patients. The control groups consisted of traditional PBL and traditional learning. The pooled analysis of seven studies comparing the effect of DPBL and traditional PBL reported little or no difference in postintervention knowledge outcomes (standardized mean difference [SMD] 0.19, 95% CI 0.00-0.38). The pooled analysis of three studies comparing the effect of DPBL to traditional learning on postintervention knowledge outcomes favored DPBL (SMD 0.67, 95% CI 0.14-1.19). For skill development, the pooled analysis of two studies comparing DPBL to traditional PBL favored DPBL (SMD 0.30, 95% CI 0.07-0.54). Findings on attitudes and satisfaction outcomes were mixed. The included studies mostly had an unclear risk of bias.

Conclusions: Our findings suggest that DPBL is as effective as traditional PBL and more effective than traditional learning in improving knowledge. DPBL may be more effective than traditional learning or traditional PBL in improving skills. Further studies should evaluate the use of digital technology for the delivery of other PBL components as well as PBL overall.

(J Med Internet Res 2019;21(2):e12945) doi:10.2196/12945

KEYWORDS

randomized controlled trials; effectiveness; systematic review; problem-based learning; medical education

Introduction

Problem-based learning (PBL) has been used as an educational approach in health professions education in many medical and nursing school curricula worldwide for over 50 years [1]. PBL aims to foster a wide range of skills such as communication and collaboration skills, decision making, problem solving, critical thinking, and self-directed learning [2]. In PBL, the use of real, ill-structured problems provides a context for the development of students' knowledge and skills [3]. Learning in PBL is student centered and occurs in small collaborative groups while teachers take on the role of tutors. Although its implementation varies across different settings, PBL, in general, is an iterative process consisting of three parts: a problem-presentation and analysis phase, a self-directed learning phase, and a synthesis and reporting phase [4].

Worldwide, various components of PBL are being increasingly delivered using digital technology. Digital education is changing the way in which health professions education, including PBL, is conducted. Digital education may comprise a variety of interventions based on learning tools, theories, content, objectives, teaching methods, and setting of delivery. In terms of the type of learning technologies, digital education includes, but is not restricted to, online and offline computer-based learning, massive open online courses, virtual reality, virtual patient simulation, mobile learning, serious gaming and gamification, and psychomotor skills trainers (Multimedia Appendix 1) [5-16]. Studies on the use of digital technologies in health professions education, in general, have reported its advantages over traditional learning in terms of improved diagnostic reasoning skills, interpersonal and professional competencies, long-term knowledge retention, problem-solving skills, self-direct/lifelong learning skills, higher-order thinking skills, self-perception, and confidence [3,17-23]. Although there is evidence on different applications of digital technology in PBL, it is still unclear how effective it is to integrate digital technology within PBL as compared to traditional PBL [23].

Digital problem-based learning (DPBL), or the use of different types of digital technologies to deliver PBL, has the potential to enhance the authenticity, appeal, accessibility, and effectiveness of PBL by enhancing participants' communication, collaboration and self-learning [24-26]. DPBL includes both fully digitally delivered PBL as well as blended PBL, wherein digital education is used to deliver certain components of PBL, while the rest of it is delivered face to face. Although there are reviews on diverse ways that technology can be incorporated in PBL, the evidence on the effectiveness of its use in PBL in health professions education is lacking [23,27]. Our objective in this review was to evaluate the effectiveness, economic impact, and potential adverse effects of DPBL interventions compared to other forms of learning in health professions education.

Methods

Study Selection

We followed the Cochrane methodology for every step of the review [28]. A detailed description of the methodology has been previously provided by the Digital Health Education collaboration [29]. The Digital Health Education collaboration is an international initiative evaluating the effectiveness of digital education in health professions education through a series of methodologically robust systematic reviews.

In this review, we included randomized controlled trials (RCTs) that evaluated the effectiveness of DPBL in improving health professionals' knowledge, skills, attitudes, and satisfaction of students and compared DPBL with traditional learning methods or other forms of digital learning. Crossover trials were excluded because of a high likelihood of a carry-over effect.

We included studies with preregistration as well as postregistration health professionals as per the qualifications listed in the Health Field of Education and Training (091) of the International Standard Classification of Education [30]. However, we excluded students of alternative, traditional, and complementary medicine. We excluded studies that focused on hybrid PBL (ie, a combination of PBL and traditional learning approaches).

We included studies in which any form of digital technology was used in combination with PBL for delivering the learning content of courses, either as the sole (full digital learning) or partial (blended learning) means of delivery, for the purpose of learning in health professions education. Digital technology primarily supports PBL principles and processes by enabling contextual and collaborative learning [27]. We defined traditional PBL in line with the Maastricht model as small-group, self-directed, tutor-supported learning that revolves primarily around a problem and occurs face to face [31]. We further conceptualized the role of digital technology in PBL in line with the presented framework that builds on the Maastricht PBL framework and the Arena Blended Connected (ABC) curriculum design method [32-34]. The Maastricht PBL framework differentiates among three broad components in PBL: the first meeting, self-directed learning, and the second meeting [34]. These components are present in traditional PBL delivered face to face. Each of these three components encompasses different learning activities that we outline using the ABC curriculum design approach. The ABC curriculum design method differentiates six main learning activities that can be supported with the use of digital technology: acquisition, inquiry, practice, production, discussion, and collaboration [33]. In our framework, we present examples of how different digital tools can be employed in a variety of ways to support learning activities in PBL. For example, learning activities comprised in the first PBL component include acquisition, discussion, collaboration, practice, and investigation. Digital technology

can be used in various ways to deliver each of these learning activities. For instance, acquisition of information can be achieved through the use of multimedia resources, podcasts, or text messages. Collaboration can be supported through the use of chatrooms or Web forums. Practice, on the other hand, can be facilitated via digital education modalities that support simulation such as virtual reality or virtual patients.

We analyzed studies that compared DPBL to traditional PBL or traditional learning (textbook, lectures, etc) and to different forms of DPBL interventions or other digital education.

We excluded studies that focused on individual learning interventions, evaluated the use of DPBL in other educational areas, lacked an active comparison, and assessed interventions with optional or minimal use of digital technology.

We included the following primary outcomes: (1) students' postintervention cognitive knowledge and skills measured with any instrument (validated or nonvalidated); (2) students' professional postintervention attitudes toward DPBL interventions, patients, or new clinical knowledge or skills measured using any instruments (validated or nonvalidated); students' postintervention satisfaction with DPBL intervention, measured using any instrument (validated or nonvalidated).

For secondary outcomes, we focused on the economic impact of the DPBL intervention and potential adverse or unintended effect of the DPBL intervention.

Data Sources, Collection, and Quality Assessment

We searched seven electronic databases, namely, MEDLINE (Ovid), Embase (Elsevier), Cochrane Central Register of Controlled Trials (CENTRAL, Wiley), PsycINFO (Ovid), Educational Research Information Centre (Ovid), Cumulative Index to Nursing and Allied Health Literature (Ebsco), and Web of Science Core Collection (Thomson Reuters), for relevant studies from January 1990 to August 16, 2017, without language restrictions (Multimedia Appendix 2).

We also checked the reference lists of all included studies and relevant systematic reviews and searched the International Clinical Trials Registry and metaRegister of Controlled Trials for unpublished trials. We followed the Cochrane methodology for the selection of studies, data extraction, data analysis, and risk of bias analysis, with two reviewers independently performing each of these steps [28]. From each study, the following information was extracted: first author's surname, publication year, course name, sample size, student characteristics, intervention method, duration of study, and outcomes. We contacted the study authors for missing data and assessed the risk of bias in the included studies using the Cochrane risk-of-bias tool [28].

Data Analysis

Standardized mean difference (SMD) for continuous outcomes, with 95% CI, was calculated based on the availability of data from the included studies. We pooled studies together based on comparison and outcomes using SMD. We interpreted the effect size using the Cohen rule of thumb (ie, with <0.2 representing no effect, 0.2 to <0.5 representing a small effect, 0.5 to <0.8 representing a moderate effect, and ≥0.8 representing a large effect) [28,35]. We employed a random-effects model in our meta-analysis. The I^2 statistic was employed to assess heterogeneity, with I^2<25%, 25%-75%, and >75% representing a low, moderate, and high degree of inconsistency, respectively. The meta-analysis was performed using Review Manager 5.3 (Cochrane Library Software, Oxford, UK). We reported the findings in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses reporting standards. We present the findings that we were unable to pool due to lack of data or high heterogeneity, in the form of a narrative synthesis.

Results

Search Results

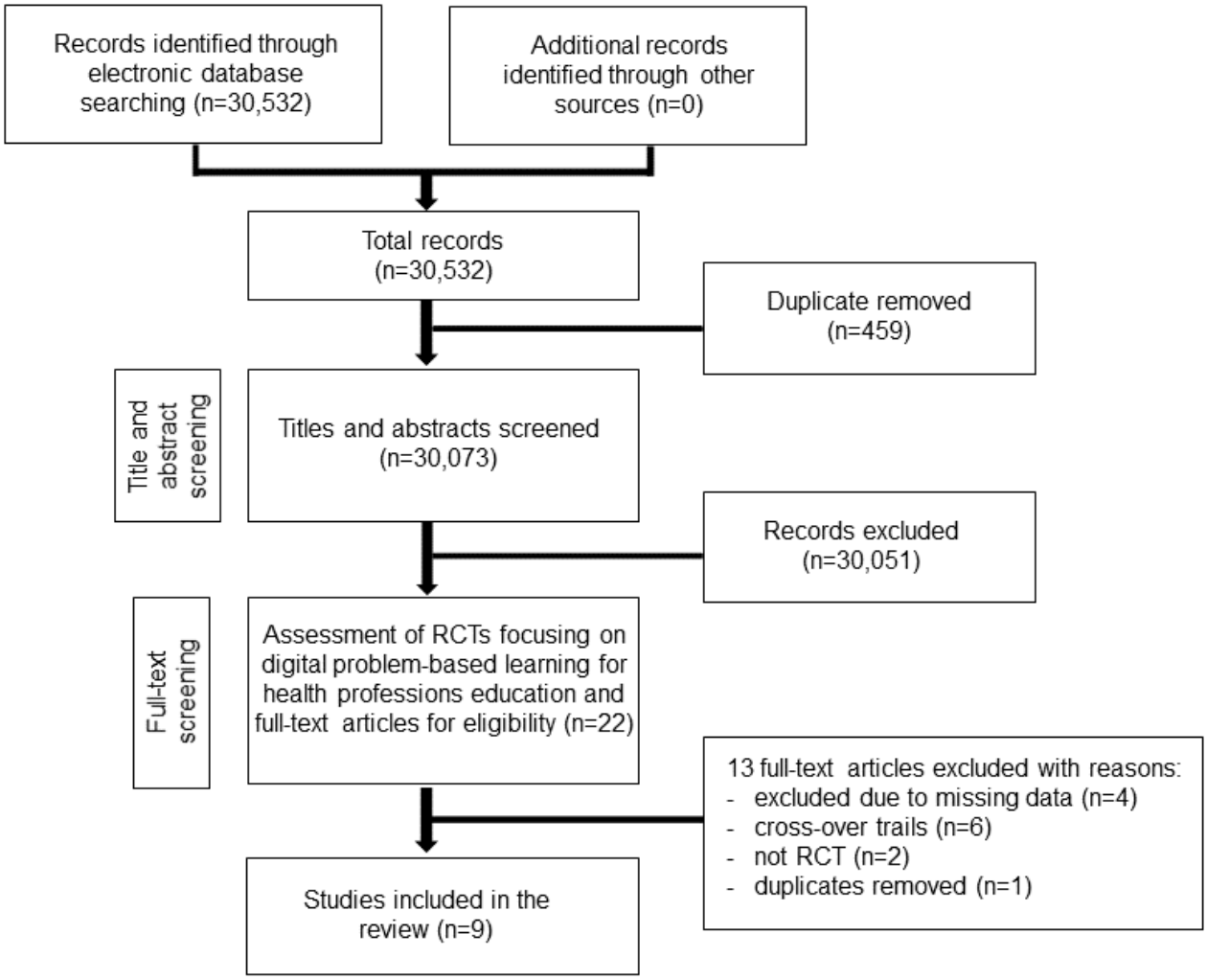
The search strategy yielded 30,532 references. We included nine studies with 890 medical students (Figure 1). We excluded four studies due to missing data [20,36-38]. No relevant ongoing clinical trials were identified.

Study Characteristics

The characteristics of the nine included studies are presented in Table 1. All studies were RCTs and published in English. Seven studies were conducted in high-income countries and two studies, in middle-income countries [21,31]. Five studies had two arms, one study had three arms, and three studies had four arms. Four studies exclusively compared DPBL to traditional PBL, one study exclusively compared DPBL to traditional learning, and two studies compared DPBL to both traditional PBL and traditional learning [21,31]. None of the included studies compared different forms of DPBL. The fields of study varied across the included studies. The studies focused on pregnancy-associated urinary incontinence [39], biochemistry (acid-based physiology) [40], genetics [41], internal medicine [42], ophthalmology [21], dermatology [31], multidisciplinary [43], human physiology [22], and traumatic head injury [44].

In most studies, DPBL interventions were delivered face to face and digital technology was employed for one component of the PBL process—presentation of problems. In four studies, DPBL was at least partially delivered on a distance basis. In one study, DPBL was employed for delivery of the initial part of the PBL (ie, the first meeting; Figure 2) [43]. In three studies, DPBL was fully distance based and all components of PBL were delivered using digital technology [39,40,44]. In line with our framework, in studies with digital presentation of problems, digital technology was mostly used to support acquisition (Figure 2; Multimedia Appendix 3). In fully distance-based DPBL, digital technology enabled a range of learning activities such as acquisition, investigation, collaboration, discussion, and production. In three studies, digital technology allowed students to practice through the use of virtual reality and virtual patients [22,41,42].

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram. RCT: randomized controlled trial.



The types of digital education modalities included offline learning (eg, CD-ROM) [31], online learning (eg, multimedia modules) [21,22,39,40,43,44], (immersive) virtual reality [41], and virtual patients [42] (Table 1). The control groups used traditional PBL such as text-based or paper-based PBL in six studies [22,39,40,42-44], traditional PBL or traditional learning in two studies [21,31], and solely traditional learning in one study [41]. The duration of the interventions ranged from 1 hour 33 minutes [41] to 7 weeks [31]. One study did not report the duration of the intervention [43].

Included studies reported findings on students' knowledge, skills, attitude, and satisfaction. No studies reported cost-related outcomes or adverse/unintended effects of DPBL-based interventions. All studies measured outcomes immediately after the intervention, except Sobocan et al [42], who reported both immediate and long-term knowledge retention at the end of the academic year. The included studies mostly had an unclear risk of bias due to a lack of information on randomization, allocation concealment, and blinding of outcomes assessment (Figure 3).

Effects of Digital Problem-Based Learning Versus Traditional Problem-Based Learning

The effects of DPBL compared to the traditional PBL on knowledge scores were reported in eight studies involving 822 medical students (Multimedia Appendix 4) [21,22,31,39,40,42-44]. The pooled analysis of seven studies showed little or no difference between DPBL and traditional PBL in postintervention knowledge scores (SMD 0.19, 95% CI 0.00-0.38; DPBL group, n=326; traditional PBL group, n=333; moderate quality; Figure 4) [21,22,31,39,40,42,43]. One study assessed the long-term effects on knowledge and reported no difference between the groups.

We also performed subgroup analysis based on the degree to which the digital technology was employed as part of PBL. We differentiated among studies in which digital technology was used for presentation of problems [21,22,31,42], the first part of the PBL was distance based [43], and PBL was fully distance based [39,40]. We found a statistically significant difference among these subgroups. There was a moderate improvement in postintervention knowledge scores in fully distance-based DPBL compared to traditional PBL (SMD 0.57, 95% CI 0.23-0.92) and no difference in studies on DPBL with digital presentation of problems (Figure 4).

Table 1. Characteristics of the included studies.

Comparisons groups and studies	Learning modalities compared	Number and types of participants	Field of study	Outcomes
DPBL[a] vs traditional PBL[b]				
Alverson et al 2008 [44], RCT[c], United States	VR[d] PBL vs traditional PBL	36 medical students (year unspecified)	Traumatic head injury	Knowledge
Bowdish et al 2003 [22], RCT, United States	Online PBL vs traditional PBL	150 medical students (first year)	Human physiology	Knowledge
Dennis 2003 [39], RCT, United States	Online PBL vs traditional PBL	34 medical students (second year)	Pregnancy-associated urinary incontinence	Knowledge
Kong et al 2009 [21], RCT, China	Online PBL vs traditional PBL	90 medical students (year unspecified)	Ophthalmology	Knowledge
Li et al 2013 [31], RCT, China	Offline PBL vs traditional PBL	120 medical students (fourth year)	Dermatology	Knowledge, Skills
Moeller et al 2010 [43], RCT, Germany	Online PBL vs traditional PBL	237 medical students (year unspecified)	Multidisciplinary	Knowledge, Skills
Sobocan et al 2017 [42], RCT, Slovenia	VP[e]-based PBL vs traditional PBL	34 medical students (third year)	Internal medicine	Knowledge
Taradi et al 2005 [40], RCT, Croatia	Online PBL vs traditional PBL	121 medical students (second year)	Biochemistry (acid-base physiology)	Knowledge
DPBL vs traditional learning				
Kong et al 2009 [21], RCT, China	Online PBL vs traditional learning (lecture)	90 medical students (year unspecified)	Ophthalmology	Knowledge
Li et al 2013 [31], RCT, China	Offline PBL vs traditional learning (lecture)	120 medical students (fourth year)	Dermatology	Knowledge
Schutte et al 1997 [41], RCT, The Netherlands	VR PBL vs traditional learning (textbook)	68 medical students (first year)	Genetics (global structure of DNA)	Knowledge

[a]DPBL: digital problem-based learning.

[b]PBL: problem-based learning.

[c]RCT: randomized controlled trial.

[d]VR: virtual reality.

[e]VP: virtual patient.

The effects of DPBL compared to the traditional PBL on skills scores were reported in two studies (N=357). The pooled analysis of these two studies showed that DPBL may slightly improve postintervention skill scores (SMD 0.30, 95% CI 0.07-0.54; *P*=.01) in comparison to traditional PBL [31,43].

The effects of DPBL compared to the traditional PBL on satisfaction scores were reported in three studies with mixed findings. Two studies evaluating the use of DPBL with digitally presented problems reported no difference between DPBL and traditional PBL in satisfaction scores [21,31]. One study reported a significant difference in satisfaction scores in favor of fully digitally delivered, distance-based PBL [40]. Two studies (N=126) assessed students' attitude toward the intervention and reported mixed results or incomplete outcome data [21,44].

Effects of Digital Problem-Based Learning Versus Traditional Learning

The effects of DPBL compared to traditional learning on knowledge scores were reported in three studies (N=278) [21,31,41] (Multimedia Appendix 4). The pooled analysis of three studies showed that DPBL may moderately improve postintervention knowledge scores (SMD 0.67, 95% CI 0.14-1.19) in DPBL compared to traditional teaching (Figure 5). A subgroup analysis of two studies (N=210) evaluating DPBL with digital presentation of problems and traditional learning showed a large improvement in postintervention knowledge scores (SMD 0.94, 95% CI 0.56-1.31) [21,31].

The effects of DPBL compared to traditional learning on skills scores were reported in one study. This study [31] reported higher postintervention skill scores in the DPBL group than in the traditional learning group (SMD 1.13, 95% CI 0.58-1.67). The effects of DPBL compared to traditional learning on satisfaction scores were reported in two studies (N=210) with uncertain findings (SMD 0.73, 95% CI –0.17 to 1.63).

Figure 2. A conceptual framework for the use of digital technology in PBL. PBL: problem-based learning.

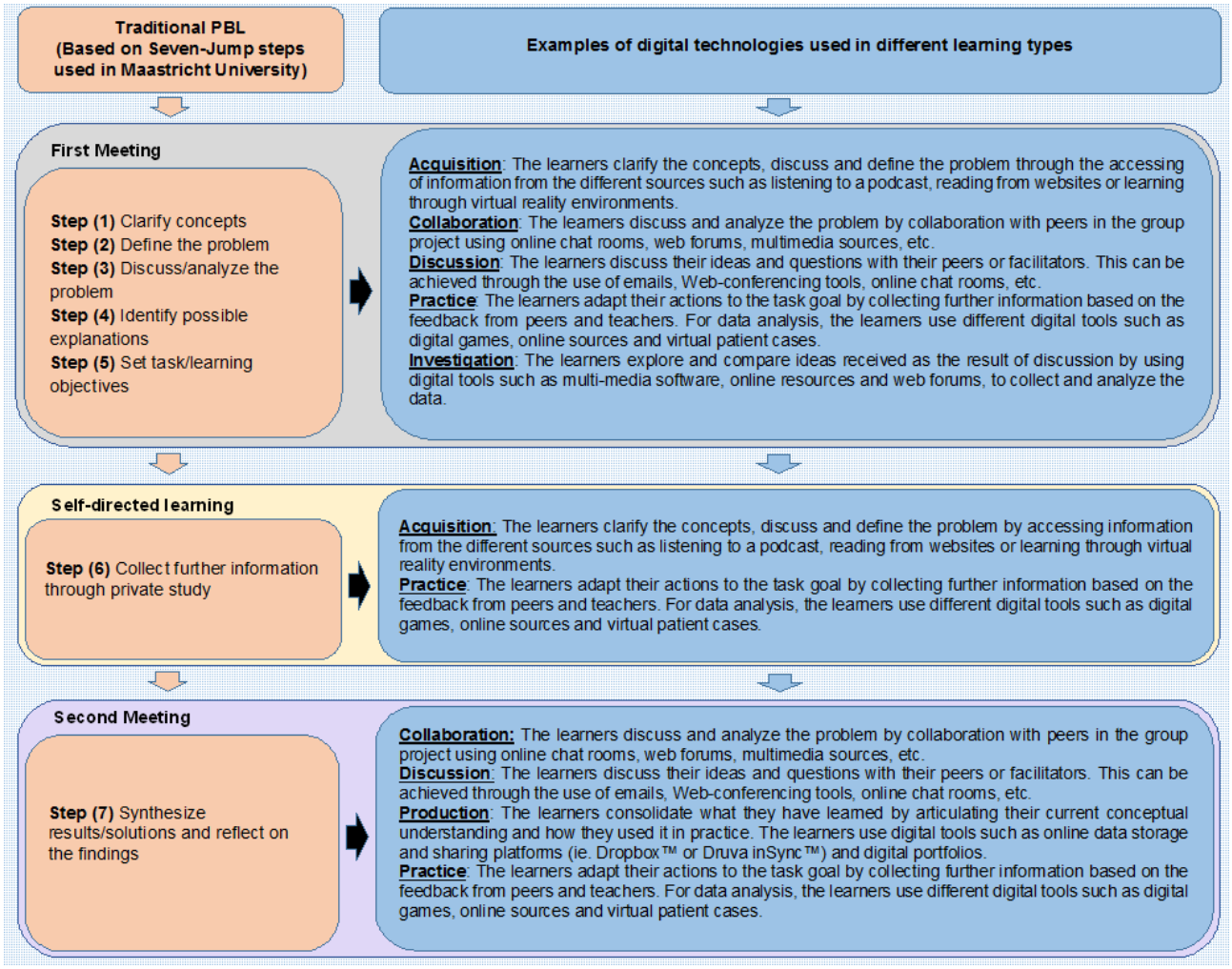


Figure 3. Risk-of-bias summary: review authors' judgement about each risk-of-bias item for each included study.

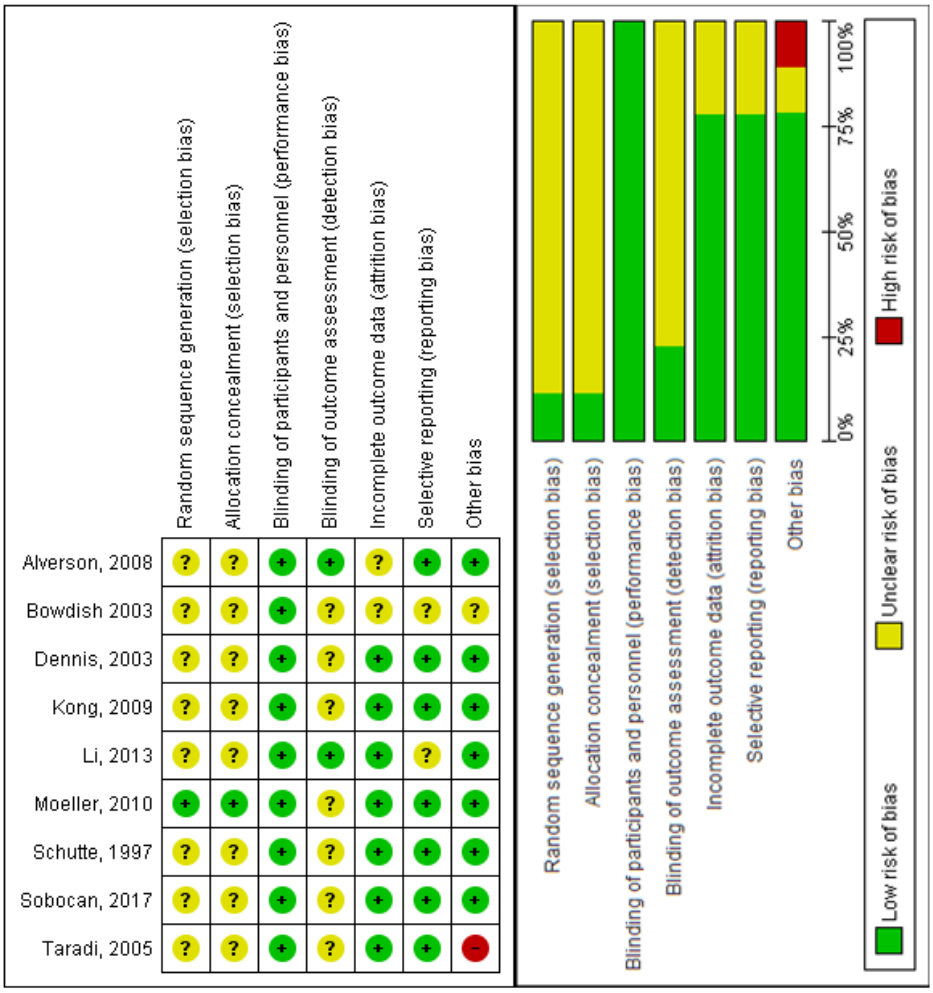


Figure 4. The effect of DPBL compared to traditional PBL (knowledge outcome, postintervention). DPBL: digital problem-based learning; PBL: problem-based learning; IV: interval variables, Random: random effect model.

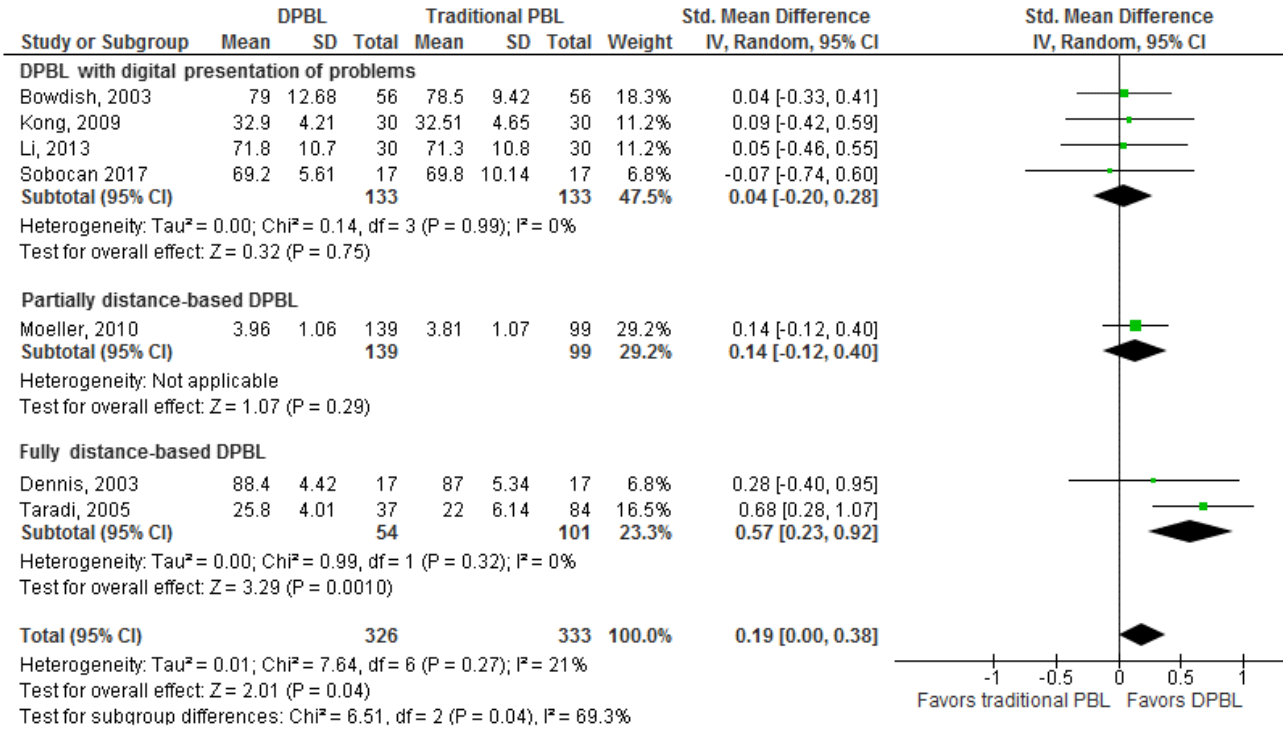
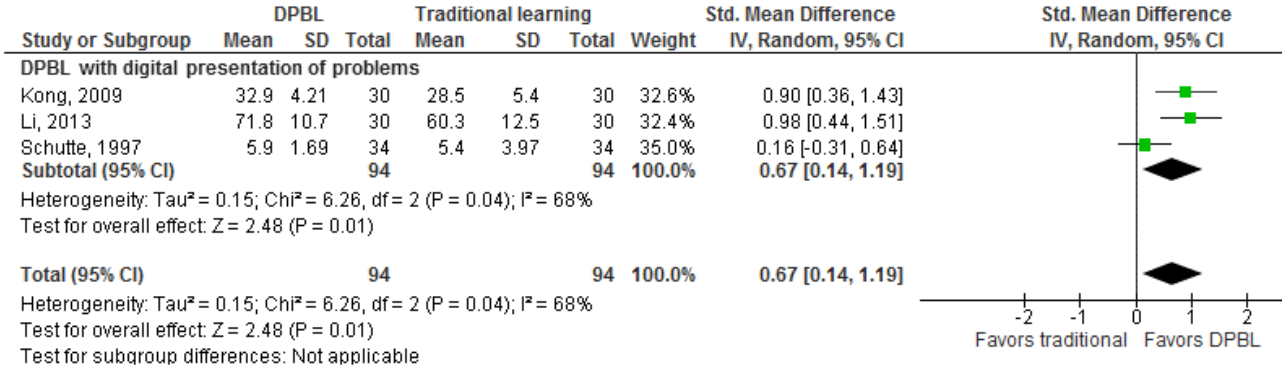


Figure 5. The effect of DPBL compared to traditional learning (knowledge outcome, postintervention). IV: interval variables; Random: random effect model; DPBL: digital problem-based learning.



Discussion

Overview

In this review, we evaluated the effectiveness of the use of digital technology for delivering PBL. Our findings show that DPBL improves students' postintervention knowledge scores in comparison to traditional learning. DPBL is as effective as traditional PBL in improving students' postintervention knowledge and may slightly improve postintervention skills. Moreover, fully digitally delivered, distance-based DPBL may lead to better knowledge scores in comparison to traditional PBL. The risk of bias in the included studies was mostly judged as unclear due to a lack of information on randomization, allocation concealment, and outcome assessment blinding. In the included studies, the term "blended PBL" was employed to denote diverse configurations of digital technology and PBL. For example, an intervention in which PBL was fully delivered online, but included one visit to the clinic, and another intervention in which face-to-face delivered PBL included digital presentation of problems were both termed blended learning. The use of "blended learning" therefore seemed misleading in this context. We decided to focus primarily on describing the way in which digital technology was employed in the PBL process. From our viewpoint, there are two main applications of digital technology in PBL: full or partial delivery of distance-based PBL or support of delivery of different components of face-to-face or colocated PBL. Most studies included in our review focused on colocated PBL, with digital technology used for the presentation of problems. This corresponds to the findings from two reviews focusing on the application of digital technology in PBL [23,27]. The reviews also report that in most studies, digital technology was used to provide contextual learning and present problems. One of these reviews focused on PBL from all disciplines and also highlighted the common use of digital technology for collaborative learning as part of distance-based PBL. In our review, we have found limited evidence on distance-based PBL in health professions education. There is a need for more research on this type of DPBL and to explore other applications of digital technology in face-to-face PBL, such as supporting collaboration, discussion, investigation, and practice. We present suggestions for the diverse applications of digital tools in PBL in our framework in Figure 2.

Our findings show that DPBL was more effective than traditional learning. Although we were unable to find meta-analyses comparing DPBL to other forms of education, there are numerous meta-analysis comparing traditional PBL to traditional learning [45-49]. The more recently published ones report that PBL is more effective than traditional learning, with moderate-to-large improvement in knowledge [45,46]. In our study, we found only three studies comparing DPBL and traditional learning, which reported an overall moderate improvement in postintervention knowledge scores among DPBL learners. A small subgroup analysis of studies in which digital technology was employed for the presentation of problems showed a major improvement in knowledge in the DPBL groups compared to the traditional learning groups. This may indicate that different configurations of DPBL may lead to larger knowledge gains. There is a need for more studies comparing the effect of distance-based DPBL or the use of digital technology for the support of other PBL components to the effect of traditional learning.

We found that DPBL was as effective as PBL in terms of knowledge, and fully digitally delivered distance-based DPBL was potentially more effective than traditional PBL. Although there are some nonrandomized studies on distance-based PBL corroborating this finding, the evidence from RCTs is scarce [23]. Potential reasons for the greater effectiveness of distance-based PBL include greater student interaction, involvement, and engagement among students. We also found that DPBL may be more effective than PBL for skills while satisfaction outcome data were mixed. These findings are based on a small number of studies, primarily assessing short-term effectiveness. More research is needed to evaluate the effectiveness of different configurations of digitally supported and distance-based PBL.

Our review has several limitations. Although RCTs provide the highest level of evidence for the effectiveness of an intervention, it is not always possible to use a rigorous RCT approach in educational research [50]. Studies included in this meta-analysis were designed as RCT, but most of them lacked information on the randomization method, allocation concealment, or blinding method. Furthermore, they mostly reported solely postintervention data; therefore, we could not calculate the pre-post intervention change. We assumed groups were matched at baseline for key characteristics and outcome measure scores.

There was no information on DPBL in postregistration health professionals. Moreover, there was limited or no information on other outcomes such as skills, attitudes, satisfaction, costs, and adverse/untoward effects of DPBL. The included studies assessed short-term effectiveness, with only one study reporting a follow-up assessment. Strengths of our review include a comprehensive and sensitive search; clear inclusion and exclusion criteria encompassing a broad range of students, outcomes, and interventions; extraction of duplicate, independent, and reproducible data; and rigorous assessment of the risk of bias.

Conclusions

DPBL includes diverse applications of digital technology as part of face-to-face as well as distance-based PBL. Our findings suggest that DPBL is more effective than traditional learning and as effective as traditional PBL in improving postintervention knowledge outcomes. For improvement of skill outcomes, DPBL may be more effective than traditional learning or traditional PBL. There is limited evidence for other outcomes such as satisfaction, attitudes, cost effectiveness, and adverse effects. Most studies evaluated the use of digital technology for the presentation of problems as part of face-to-face DPBL and had unclear risk of bias. There is scope for the evaluation of digital technology in the delivery of other PBL components as well as the effectiveness of distance-based PBL.

Acknowledgments

This review was conducted in collaboration with the Health Workforce Department at the World Health Organization. We thank Ms Ushashree Divakar and Ms Nuraini Nazeha for their help during the initial phase of the review and Dr Ram Bajpai for his statistical advice.

We would like to acknowledge funding support from the Lee Kong Chian School of Medicine, Nanyang Technological University Singapore (eLearning for health professional education grant).

Authors' Contributions

LC and BK conceived the idea. BK and GD screened the articles and extracted the data. BK and LC wrote the manuscript. LC, GD, NS, MS, JR, NB, and JC provided insightful comments on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of the digital education modalities.

[PDF File (Adobe PDF File), 194KB - jmir_v21i2e12945_app1.pdf]

Multimedia Appendix 2

MEDLINE (Ovid) search strategy.

[PDF File (Adobe PDF File), 22KB - jmir_v21i2e12945_app2.pdf]

Multimedia Appendix 3

Digital technologies used in the studies and associated learning types.

[PDF File (Adobe PDF File), 254KB - jmir_v21i2e12945_app3.pdf]

Multimedia Appendix 4

Characteristics and effects of the included studies.

[PDF File (Adobe PDF File), 478KB - jmir_v21i2e12945_app4.pdf]

References

1. Lim WK. Dysfunctional problem-based learning curricula: resolving the problem. BMC Med Educ 2012 Sep 25;12:89 [FREE Full text] [doi: 10.1186/1472-6920-12-89] [Medline: 23009729]
2. Wilder S. Impact of problem-based learning on academic achievement in high school: a systematic review. Educational Review 2014 Nov 12;67(4):414-435. [doi: 10.1080/00131911.2014.974511]

3. Reich S, Simon JF, Ruedinger D, Shortall A, Wichmann M, Frankenberger R. Evaluation of two different teaching concepts in dentistry using computer technology. Adv Health Sci Educ Theory Pract 2007 Aug;12(3):321-329. [doi: 10.1007/s10459-006-9004-8] [Medline: 16847734]
4. Yew EHJ, Goh K. Problem-Based Learning: An Overview of its Process and Impact on Learning. Health Professions Education 2016 Dec;2(2):75-79 [FREE Full text] [doi: 10.1016/j.hpe.2016.01.004]
5. Hervatis V, Kyaw BM, Semwal M, Dunleavy G, Tudor Car L, Zary N. PROSPERO CRD42016045679. 2016. Offline and computer-based eLearning interventions for medical students' education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045679 [accessed 2019-02-18] [WebCite Cache ID 76GXpzD6T]
6. Saxena N, Kyaw BM, Vseteckova J, Dev P, Paul P, Lim KTK, et al. PROSPERO CRD42016045470. 2016. Virtual reality environments for health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045470 [accessed 2019-02-18] [WebCite Cache ID 76GY1B3xZ]
7. Tudor Car L, Riboli-Sasco E, Belisario JM, Nikolaou CK, Majeed A, Zary N, et al. PROSPERO CRD42015029786. 2015. Mobile learning for delivering health professional education [Cochrane Protocol] URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42015029786 [accessed 2019-02-18] [WebCite Cache ID 76GYI3rnU]
8. Gentry S, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious gaming and gamification education in health professions: a systematic review by the Digital Health Education Collaboration. J Med Internet Res (forthcoming) 2019 [FREE Full text] [doi: 10.2196/12994]
9. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GCY, Tudor Car L. Mobile Digital Education for Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937 [FREE Full text] [doi: 10.2196/12937] [Medline: 30747711]
10. Huang Z, Semwal M, Lee SY, Tee WKM, Ong W, Tan WS, et al. Digital health professionals? diabetes management education and training: a systematic review by the Digital Health Education collaboration. J Med Internet Res (forthcoming) 2019 [FREE Full text] [doi: 10.2196/12997]
11. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
12. Posadzki P, Bala M, Kyaw BM, Semwal M, Divakar U, Koperny M, et al. Offline digital education for post-registration health professions: a systematic review by the Digital Health Education collaboration. J Med Internet Res (forthcoming) 2019:a [FREE Full text] [doi: 10.2196/preprints.12968]
13. Semwal M, Whiting P, Bajpai R, Bajpai S, Kyaw BM, Tudor Car L. Digital education for health professions on smoking cessation management: a systematic review by the Digital Health Education collaboration. J Med Internet Res (forthcoming) 2019 [FREE Full text] [doi: 10.2196/13000]
14. Posadzki P, Paddock S, Campbell J, Car J. Medical students' digital education of communication skills: a systematic review by the Digital Health Education Collaboration. J Med Internet Res (forthcoming) 2018 [FREE Full text] [doi: 10.2196/preprints.12967]
15. George PP, Zhabenko O, Kyaw BM, Antoniou P, Posadzki P, Saxena N, et al. Medical doctors' online digital education: a systematic review by the Digital Health Education Collaboration. J Med Internet Res (forthcoming) 2019. [doi: 10.2196/preprints.13269]
16. Wahabi H, Esmaeil SA, Bahkali KH, Titi MA, Amer YS, Fayed AA, et al. Medical doctors' offline-computer-assisted digital education: a systematic review by the Digital Health Education Collaboration. J Med Internet Res (forthcoming) 2019. [doi: 10.2196/12998]
17. Al-Dahir S, Bryant K, Kennedy KB, Robinson DS. Online virtual-patient cases versus traditional problem-based learning in advanced pharmacy practice experiences. Am J Pharm Educ 2014 May 15;78(4):76 [FREE Full text] [doi: 10.5688/ajpe78476] [Medline: 24850938]
18. Chan DH, Leclair K, Kaczorowski J. Problem-based small-group learning via the Internet among community family physicians: a randomized controlled trial. MD Comput 1999;16(3):54-58. [Medline: 10439603]
19. Kim JH, Shin JS. Effects of an online problem-based learning program on sexual health care competencies among oncology nurses: a pilot study. J Contin Educ Nurs 2014 Sep;45(9):393-401. [doi: 10.3928/00220124-20140826-02] [Medline: 25153429]
20. Balslev T, de Grave WS, Muijtjens AM, Scherpbier AJ. Comparison of text and video cases in a postgraduate problem-based learning format. Med Educ 2005 Nov;39(11):1086-1092. [doi: 10.1111/j.1365-2929.2005.02314.x] [Medline: 16262803]
21. Kong J, Li X, Wang Y, Sun W, Zhang J. Effect of digital problem-based learning cases on student learning outcomes in ophthalmology courses. Arch Ophthalmol 2009 Sep;127(9):1211-1214. [doi: 10.1001/archophthalmol.2009.110] [Medline: 19752434]
22. Bowdish B, Chauvin SW, Kreisman N, Britt M. Travels towards problem based learning in medical education (VPBL). Instr Sci 2003;31:231-253. [doi: 10.1023/A:1024625707592]
23. Jin J, Bridges SM. Educational technologies in problem-based learning in health sciences education: a systematic review. J Med Internet Res 2014 Dec 10;16(12):e251 [FREE Full text] [doi: 10.2196/jmir.3240] [Medline: 25498126]

24. Clark CE. Problem-based learning: how do the outcomes compare with traditional teaching? Br J Gen Pract 2006 Sep;56(530):722-723 [FREE Full text] [Medline: 16954011]
25. McParland M, Noble LM, Livingston G. The effectiveness of problem-based learning compared to traditional teaching in undergraduate psychiatry. Med Educ 2004 Aug;38(8):859-867. [doi: 10.1111/j.1365-2929.2004.01818.x] [Medline: 15271047]
26. Rich SK, Keim RG, Shuler CF. Problem-based learning versus a traditional educational methodology: a comparison of preclinical and clinical periodontics performance. J Dent Educ 2005 Jun;69(6):649-662 [FREE Full text] [Medline: 15947211]
27. Verstegen DML, de Jong N, van Berlo J, Camp A, Könings KD, van Merriënboer JJG, et al. How e-Learning Can Support PBL Groups: A Literature Review. In: Bridges S, Chan LK, Hmelo-Silver CE, editors. Educational Technologies in Medical and Health Sciences Education. Cham: Springer International Publishing; 2016:9-33.
28. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 2011. URL: http://handbook-5-1.cochrane.org/ [accessed 2019-02-19] [WebCite Cache ID 76J8fia2U]
29. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital Education in Health Professions: The Need for Overarching Evidence Synthesis. J Med Internet Res 2019 Dec 14;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
30. United Nations Statistics Division. 2013. International Standard Classification of Education: Fields of Education Training 2013 (ISCED-F 2013) URL: https://unstats.un.org/unsd/classifications/Family/Detail/1046 [accessed 2019-02-18] [WebCite Cache ID 76GfaTYQF]
31. Li J, Li QL, Li J, Chen ML, Xie HF, Li YP, et al. Comparison of three problem-based learning conditions (real patients, digital and paper) with lecture-based learning in a dermatology course: a prospective randomized study from China. Med Teach 2013;35(2):e963-e970. [doi: 10.3109/0142159X.2012.719651] [Medline: 23009254]
32. Schmidt HG. Problem-based learning: rationale and description. Med Educ 1983 Jan;17(1):11-16. [Medline: 6823214]
33. Laurillard D. Teaching as a Design Science. London: Routledge; 2012.
34. Til C, Heijden F. PBL Study Skills - an overview. In: Universiteit Maastricht. Netherlands: Department of Educational Development and Research, Maastricht University; 2009.
35. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA 2011 Sep 07;306(9):978-988. [doi: 10.1001/jama.2011.1234] [Medline: 21900138]
36. Iniesta M, Alonso B, de Arriba L, Sanz M, Herrera D. Enhancing learning with information and communication technology. : IATED; 2015 Presented at: 8th International Conference of Education, Research and Innovation 2015; 18-20 November; Seville, Spain p. 8-3.
37. Meyer KE. ERIC. 2011. The rivalry between simulation and problem-based learning: A study of learning transfer in physician assistant students URL: https://eric.ed.gov/?id=ED521620 [accessed 2019-02-18] [WebCite Cache ID 76Gh4ZcWx]
38. Heissam K, Sandokji A, El-Badry A. The efficacy of web problem based learning versus traditional discipline learning in biomedical sciences. 2010 Presented at: 4th International Technology, Education and Development Conference (Inted 2010); 8-10 March; Valencia, Spain p. 2277-2284 URL: https://library.iated.org/view/HEISSAM2010THE
39. Dennis JK. Problem-based learning in online vs. face-to-face environments. Educ Health (Abingdon) 2003 Jul;16(2):198-209. [doi: 10.1080/1357628031000116907] [Medline: 14741905]
40. Taradi SK, Taradi M, Radic K, Pokrajac N. Blending problem-based learning with Web technology positively impacts student learning outcomes in acid-base physiology. Adv Physiol Educ 2005 Mar;29(1):35-39 [FREE Full text] [doi: 10.1152/advan.00026.2004] [Medline: 15718381]
41. Schutte B, de Goeij T, de Grave W, Koehorst AM. The effects of visual genetics on the learning of students in a problem based curriculum. In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, editors. Advances in Medical Education. Dordrecht: Springer Netherlands; 1997:336-338.
42. Sobocan M, Turk N, Dinevski D, Hojs R, Pecovnik Balon B. Problem-based learning in internal medicine: virtual patients or paper-based problems? Intern Med J 2017 Jan;47(1):99-103. [doi: 10.1111/imj.13304] [Medline: 27800653]
43. Moeller S, Spitzer K, Spreckelsen C. How to configure blended problem based learning-results of a randomized trial. Med Teach 2010;32(8):e328-e346. [doi: 10.3109/0142159X.2010.490860] [Medline: 20662568]
44. Alverson DC, Saiki SMJ, Kalishman S, Lindberg M, Mennin S, Mines J, et al. Medical students learn over distance using virtual reality simulation. Simul Healthc 2008;3(1):10-15. [doi: 10.1097/SIH.0b013e31815f0d51] [Medline: 19088637]
45. Wang J, Xu Y, Liu X, Xiong W, Xie J, Zhao J. Assessing the effectiveness of problem-based learning in physical diagnostics education in China: a meta-analysis. Sci Rep 2016 Dec 03;6:36279 [FREE Full text] [doi: 10.1038/srep36279] [Medline: 27808158]
46. Zahid MA, Varghese R, Mohammed AM, Ayed AK. Comparison of the problem based learning-driven with the traditional didactic-lecture-based curricula. Int J Med Educ 2016 Jun 12;7:181-187 [FREE Full text] [doi: 10.5116/ijme.5749.80f5] [Medline: 27289331]
47. Qin Y, Wang Y, Floden RE. The Effect of Problem-Based Learning on Improvement of the Medical Educational Environment: A Systematic Review and Meta-Analysis. Med Princ Pract 2016;25(6):525-532 [FREE Full text] [doi: 10.1159/000449036] [Medline: 27513586]

48. Sayyah M, Shirbandi K, Saki-Malehi A, Rahim F. Use of a problem-based learning teaching model for undergraduate medical and nursing education: a systematic review and meta-analysis. Adv Med Educ Pract 2017;8:691-700 [FREE Full text] [doi: 10.2147/AMEP.S143694] [Medline: 29042827]
49. Kong LN, Qin B, Zhou YQ, Mou SY, Gao HM. The effectiveness of problem-based learning on development of nursing students' critical thinking: a systematic review and meta-analysis. Int J Nurs Stud 2014 Mar;51(3):458-469. [doi: 10.1016/j.ijnurstu.2013.06.009] [Medline: 23850065]
50. Sullivan GM. Getting off the "gold standard": randomized controlled trials and education research. J Grad Med Educ 2011 Sep;3(3):285-289 [FREE Full text] [doi: 10.4300/JGME-D-11-00147.1] [Medline: 22942950]

Abbreviations

ABC: Arena Blended Connected
DPBL: digital problem-based learning
SMD: standardized mean difference
PBL: problem-based learning
RCT: randomized controlled trial
VP: virtual patient
VR: virtual reality

Edited by A Marusic; submitted 29.11.18; peer-reviewed by I Kolcic, M Sobocan, S Barteit; comments to author 05.01.19; revised version received 21.01.19; accepted 04.02.19; published 28.02.19.

Please cite as:
Tudor Car L, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, Low-Beer N, Campbell J
Digital Problem-Based Learning in Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration
J Med Internet Res 2019;21(2):e12945
URL: https://www.jmir.org/2019/2/e12945/
doi:10.2196/12945
PMID:30816846

Review

Health Care and Cybersecurity: Bibliometric Analysis of the Literature

Mohammad S Jalali[1,2], MS, PhD; Sabina Razak[2], BSc; William Gordon[3,4,5], MD; Eric Perakslis[6], PhD; Stuart Madnick[2], PhD

[1]MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA, United States

[2]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, United States

[3]Division of General Internal Medicine, Department of Medicine, Brigham & Women's Hospital, Boston, MA, United States

[4]Partners Healthcare, Boston, MA, United States

[5]Department of Dermatology, Harvard Medical School, Boston, MA, United States

[6]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Mohammad S Jalali, MS, PhD
MGH Institute for Technology Assessment
Harvard Medical School
101 Merrimac Street, Suite 1010
Boston, MA, 02114
United States
Phone: 1 617 724 3738
Email: msjalali@mgh.harvard.edu

Abstract

Background: Over the past decade, clinical care has become globally dependent on information technology. The cybersecurity of health care information systems is now an essential component of safe, reliable, and effective health care delivery.

Objective: The objective of this study was to provide an overview of the literature at the intersection of cybersecurity and health care delivery.

Methods: A comprehensive search was conducted using PubMed and Web of Science for English-language peer-reviewed articles. We carried out chronological analysis, domain clustering analysis, and text analysis of the included articles to generate a high-level concept map composed of specific words and the connections between them.

Results: Our final sample included 472 English-language journal articles. Our review results revealed that majority of the articles were focused on technology: Technology–focused articles made up more than half of all the clusters, whereas managerial articles accounted for only 32% of all clusters. This finding suggests that nontechnological variables (human–based and organizational aspects, strategy, and management) may be understudied. In addition, *Software Development Security*, *Business Continuity*, and *Disaster Recovery Planning* each accounted for 3% of the studied articles. Our results also showed that publications on *Physical Security* account for only 1% of the literature, and research in this area is lacking. Cyber vulnerabilities are not all digital; many physical threats contribute to breaches and potentially affect the physical safety of patients.

Conclusions: Our results revealed an overall increase in research on cybersecurity and identified major gaps and opportunities for future work.

(J Med Internet Res 2019;21(2):e12644) doi:10.2196/12644

KEYWORDS

bibliometric review; cybersecurity; health care; literature analysis; text mining

Introduction

Cybersecurity is an increasingly critical aspect of health care information technology infrastructure. The rapid digitization of health care delivery, from electronic health records and telehealth to mobile health (mHealth) and network-enabled medical devices, introduces risks related to cybersecurity vulnerabilities [1]. These vulnerabilities are particularly worrisome because cyberattacks in a health care setting can result in the exposure of highly sensitive personal information

or cause disruptions in clinical care [2-5]. Cyberattacks may also affect the safety of patients, for example, by compromising the integrity of data or impairing medical device functionality. The WannaCry and NotPetya ransomware attacks and vulnerabilities in Medtronic Implantable Cardiac Device Programmers are recent examples that have resulted in impaired health care delivery capabilities [6].

Health care organizations are particularly vulnerable to cyber threats. Verizon's 2018 Data Breach Investigation Report found that the health care field, in general, was most affected by data breaches, which accounted for 24% of all investigated breaches across all industries [7]. Additionally, a report by the Ponemon Institute found that almost 90% of respondents (involved in health plans and health care clearing houses as well as health care providers with electronic health records) experienced a data breach in the past 2 years [8]. Another survey of health care information security professionals revealed that over 75% of health care organizations experienced a recent security incident [9]. The causes are multifactorial, involving both technology and people, and human error and cultural factors play increasingly critical roles [10,11]. Despite efforts to teach best-practice security behavior through training programs, recent surveys have revealed that one in five health care employees still write down their usernames and passwords on paper [12].

Given the increasing importance of cybersecurity for safe, effective, and reliable health care delivery, there is a need to provide an overview of the literature at the intersection of cybersecurity and health care. Recent systematic reviews synthesized insights from 31 articles on cyber threats in health care [13] and aggregated strategies from 13 articles about responding to cyber incidents in health care organizations [14]. In this study, we conducted a large bibliometric review of the literature and describe the current state of research on various aspects of cybersecurity in health care in order to not only understand current trends but also identify gaps and guide future research efforts toward improving the security of our health care systems.

Methods

Study Eligibility Criteria

A comprehensive search was conducted using PubMed and Web of Science (WoS) for English-language peer-reviewed articles. We identified search keywords by adopting terminologies in The National Initiative for Cybersecurity Careers and Studies [15] and The British Standards Institution glossaries [16]. The list of keywords used is as follows:

WoS (journal articles, all years):

> "Health*" AND "Cybersecurity" OR "Cyber Security" OR "Cyber Attack*" OR "Cyber Crisis*" OR "Cyber Incident*" OR "Cyber Infrastructure*" OR "Cyber Operation*" OR "Cyber Risk*" OR "Cyber Threat*" OR "Cyberspace*" OR "Data Breach*" OR "Data Security*" OR "Firewall*" OR "Information Security*" OR "Information Systems Security*" OR "Information Technology Security*" OR "IT Security*" OR "Malware*" OR "Phishing*" OR "Ransomware*" OR "Security Incident*" OR "Information Assurance*"

PubMed (journal articles, all years, abstract availability):

> "Cybersecurity" OR "Cyber Security" OR "Cyber Attack" OR "Cyber Crisis" OR "Cyber Incident" OR "Cyber Infrastructure" OR "Cyber Operation" OR "Cyber Risk" OR "Cyber Threat" OR "Cyberspace" OR "Data Breach" OR "Data Security" OR "Firewall" OR "Information Security" OR "Information Systems Security" OR "Information Technology Security" OR "IT Security" OR "Malware" OR "Phishing" OR "Ransomware" OR "Security Incident" OR "Information Assurance".

Keywords that widened the search results far beyond the scope were rejected. For example, "exploit" and "malicious" can be used in a cyber context, but are more commonly used in unrelated contexts that add noise to the search. Such terms were not included because of their contribution to an overwhelming amount of irrelevant results.

We included articles published from the inception of PubMed in 1966 and WoS in 1900 to September 2017. Articles were excluded if they did not clearly focus on cybersecurity or health care or if they were reviews or meta-analyses. Inclusion and exclusion criteria were formulated prior to the preliminary title and abstract screening. The eligibility criteria were intentionally nonspecific to obtain a complete picture of the existing relevant research. To increase our confidence in the inclusion criteria, we conducted an initial pilot screening of 100 articles.

Screening and Selection

Screening of titles and abstracts was conducted using the software package Abstrackr [17]. Full texts of the "maybe" articles were independently reviewed by two trained individuals to assess study eligibility. Disagreements about study inclusion were discussed until a consensus was reached. More details about our methodology are available in Multimedia Appendix 1.

Chronological Clustering and Trend Analysis

We performed chronological analysis of the number of articles published per year and the number of authors per article. We topically clustered articles using 10 security domains created by the International Information Systems Security Certification Consortium to categorize each article (Multimedia Appendix 1). Each clustered article was further categorized as technological, managerial, legal, or interdisciplinary (if it fell into more than three categories). Features of the included articles, such as the publishing journal and number of citations, were recorded.

Text Analysis

After analyzing all the titles and abstracts, we removed words with high frequencies that were common in research articles but were not specific to our subject (eg, "paper," "using," and "results"). In addition, we merged the plural forms with singular forms of the same word and merged "healthcare" and "health care" into "healthcare." Subsequently, we created word clouds to visualize the word frequencies in titles and abstracts over

time. Word frequency is represented by color and size, with darker, larger words representing higher occurrence.

We then assessed text titles and abstracts to generate a high-level concept map composed of specific words and the connections between them by using the software package Leximancer text analytics (version 4.5; Leximancer Pty Ltd, Brisbane, Australia). The software started with an unsupervised machine learning approach to extract a network of meaning from the data and developed a heat map that visually illustrated the end results. The method, underpinned by a naive Bayesian co-occurrence metric, considers how often two words co-occur as well as how often they occur apart [18,19]. Heat maps consist of "themes" represented by bubbles and "concepts" represented by grey dots. Concepts can be equated to a list of similar terms coalescing into a monothematic idea, and themes are clusters of these concepts. The lines between dots suggest a strong connection between two concepts.

Results

Search Results

The primary search on PubMed for papers containing terms pertaining to "cyber" yielded 1480 articles, and the search on WoS yielded 810 articles. After removing 310 duplicates, the titles and abstracts of 1980 articles were screened, which was facilitated by the Abstrackr software [17]. Based on the inclusion criteria, 1262 articles were excluded in the first screening, reducing the results to 718 articles for full-text review. Eventually, a further screening removed additional articles to provide a final selection of 472 articles. Figure 1 presents the search method and results.

Chronological Clustering and Trend Analysis

Figure 2 presents the overall trend of all publications over time, from 1985 to September 2017; the first included article was published in 1979 but was excluded from the figure for better visualization. Figure 2 shows a steady increase in the number of articles published on cybersecurity in health care (Multimedia Appendix 1).

Figure 3 shows the distribution among the three high-level categories: technological, managerial, and legal (Multimedia Appendix 1). The seven technological clusters made up more than half of all clusters, the two managerial clusters represented 32%, and the legal cluster represented 18% of all clusters.

The orange-shaded portion within each cluster in Figure 3 represents interdisciplinary articles (spanning multiple high-level categories). Although *Physical Security* had the lowest number of publications (Figure 3), it was the most interdisciplinary cluster (six out of the seven articles [85.7%] identified as interdisciplinary). *Legal, Regulations, Investigations, and Compliance* was the second most interdisciplinary cluster (59.8% of the articles in this category were interdisciplinary), followed by *Operations Security* (52.9%), *Business Continuity and Disaster Recovery Planning* (50%), *Information Security Governance and Risk Management* (43.9%), and *Access Control* (30.6%). Although *Security Architecture and Design* was the second most frequent cluster overall, only 22.2% of the articles were found to be interdisciplinary. The less interdisciplinary categories were *Telecommunications and Network Security* (18.9%), *Software Development Security* (17.6%), and *Cryptography* (4%) (Multimedia Appendix 1).

We analyzed the publication trends over time in the 10 clusters (Figure 4). All clusters showed increased frequency, and some clusters such as *Security Architecture and Design*, *Information Security Governance and Risk Management*, and *Cryptography* demonstrated particularly steep increases in frequency.

Figure 1. Search method and results.

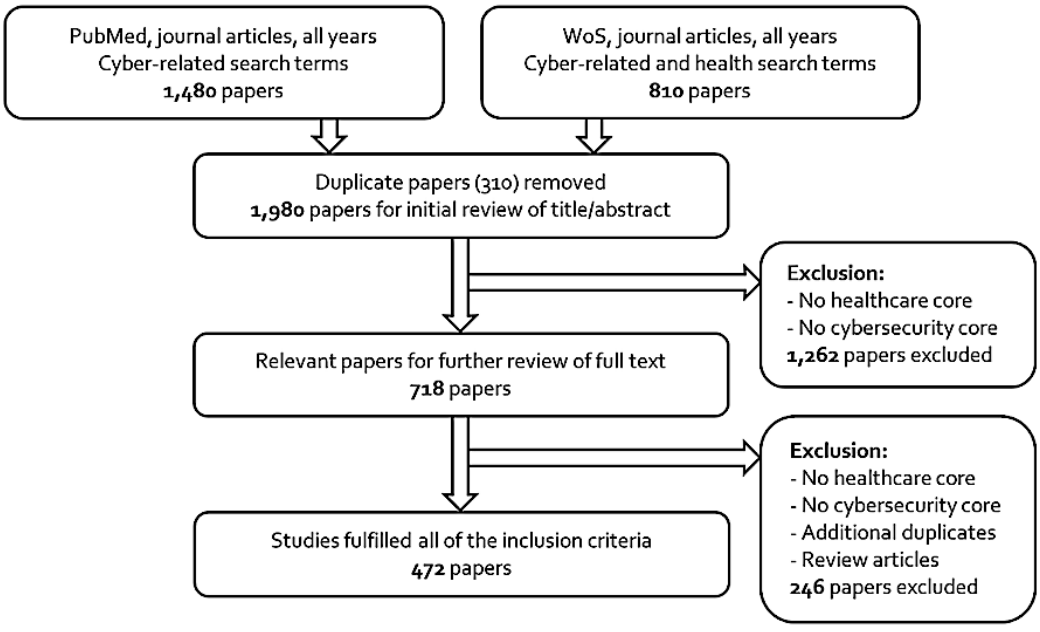


Figure 2. Annual number of published papers at the intersection of health care and cybersecurity (fitted trend line: y=0.9166e0.1252x; R^2=0.82).

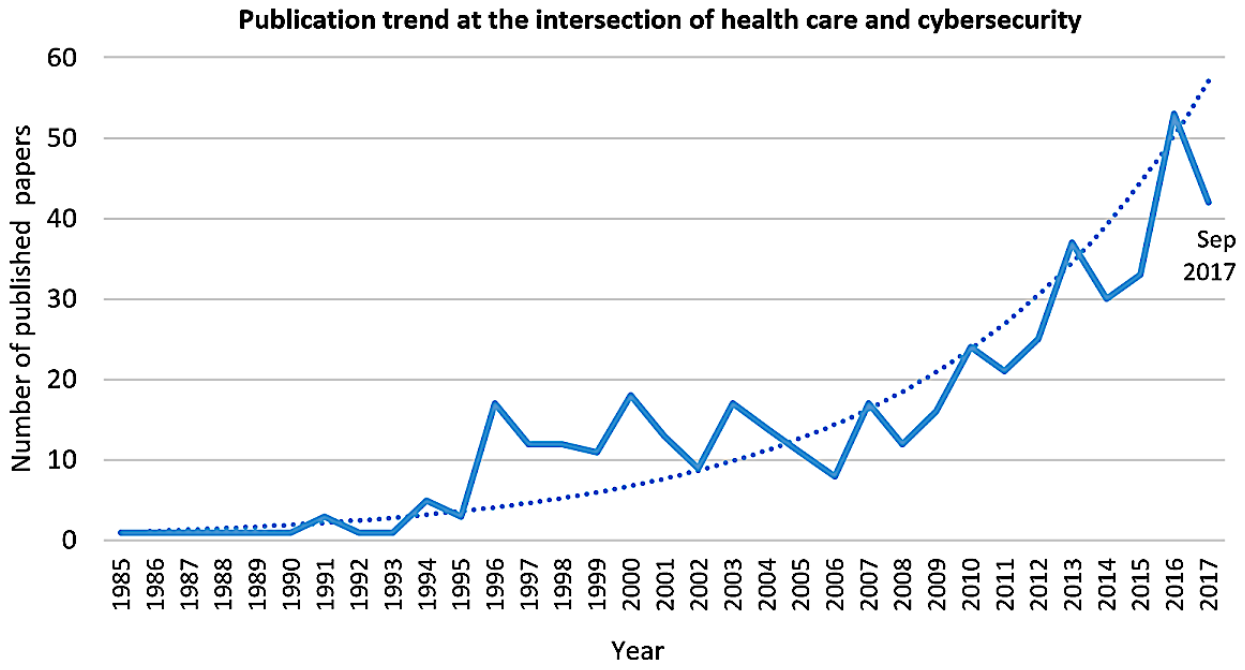


Figure 3. Cluster distributions.

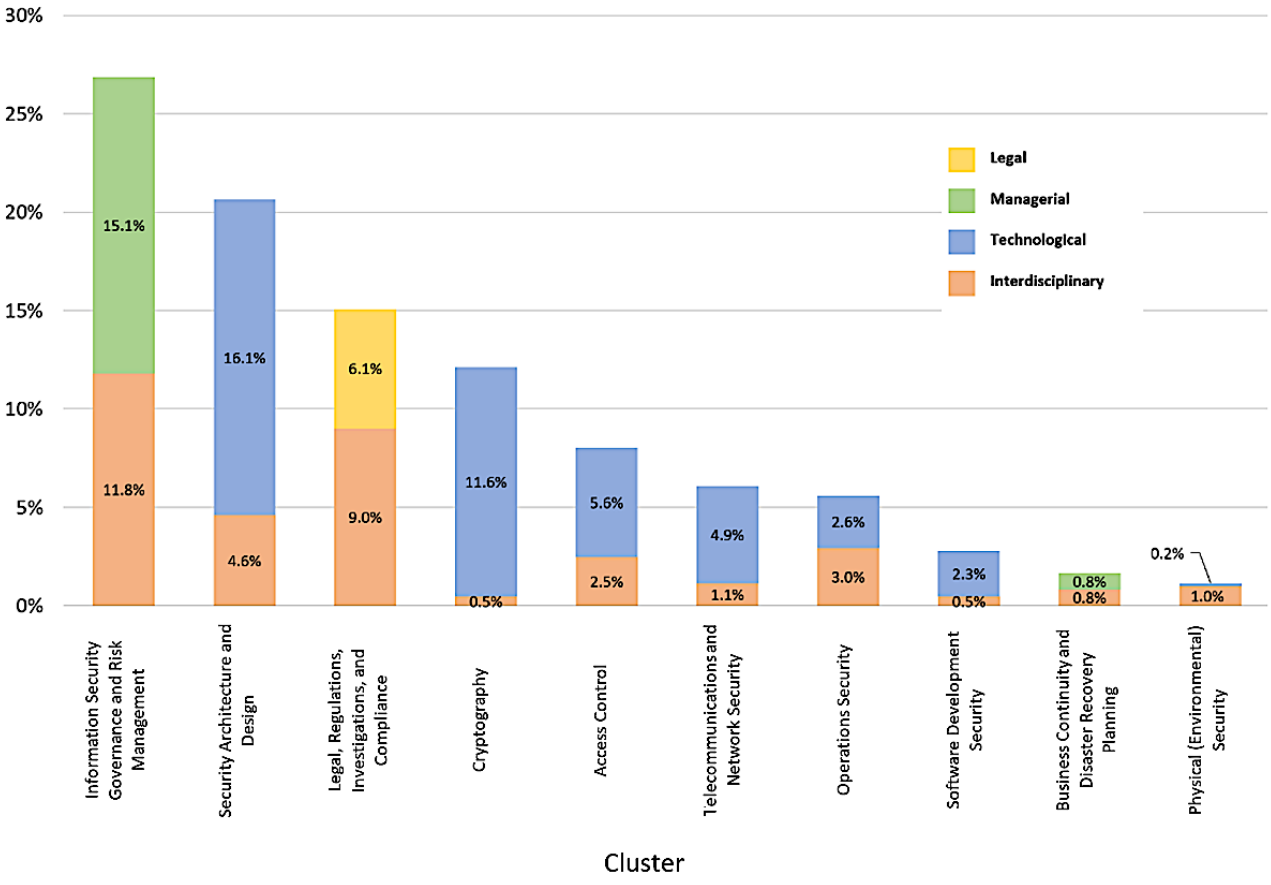


Figure 4. Trend of 10 clusters over time.

Table 1. Journals with the most articles.

Journal	Number of published papers	Indexed categories (according to Journal Citation Reports) [20]
Studies in Health Technology and Informatics	47	Not indexed
International Journal of Medical Informatics	24	Computer Science, Information Systems; Health Care Sciences & Services; Medical Informatics
Journal of Medical Systems	17	Health Care Sciences & Services; Medical Informatics
Journal of Diabetes Science and Technology	9	Not indexed
Healthcare Financial Management	8	Not indexed
Medical Informatics	8	Computer Science, Information Systems; Computer Science, Interdisciplinary Applications; Medical Informatics
International Journal of Bio-Medical Computing	8	Computer Science, Interdisciplinary Applications; Computer Science, Theory & Methods; Engineering, Biomedical; Medical Informatics
Computers & Security	7	Computer Science, Information Systems
Journal of the American Medical Informatics Association	7	Computer Science, Information Systems; Computer Science, Interdisciplinary Applications; Health Care Sciences & Services; Medical Informatics
Journal of Healthcare Protection Management	7	Not indexed
Telemedicine Journal and E-Health	5	Health Care Sciences & Services
IEEE[a] Journal of Biomedical and Health Informatics	4	Computer Science, Information Systems; Computer Science, Interdisciplinary Applications; Mathematical & Computational Biology; Medical Informatics
Journal of the American Health Information Management Association	4	Not indexed
Journal of Digital Imaging	4	Radiology, Nuclear Medicine & Medical Imaging
Journal of Healthcare Information Management	4	Not indexed
Journal of Medical Internet Research	4	Health Care Sciences & Services; Medical Informatics
Journal of Medical Practice Management	4	Not indexed

[a]IEEE: Institute of Electrical and Electronics Engineers.

Journal Characteristics

Overall, the 472 articles included were published in 239 unique journals. We ranked the journals according to the number of published articles and selected the journals with more than three articles, which resulted in a list of 17 journals (Table 1). According to the corresponding Incites Journal Citation Reports (JCR) categories [20], the top journals tended to focus on computer science, information systems, and medical informatics. The most popular JCR category, accounting for seven out of the 10 journals listed in JCR, was medical informatics. Six journals had a computer science category, specifically within information systems, interdisciplinary applications, or theory and methods. Five journals were from the health care sciences and services. Only one of the top 15 journals was categorized as a biomedical engineering journal; one, as a math and computational biology journal; and one, as a radiology, nuclear medicine, and medical imaging journal.

Approximately, 73% of the 239 journals had only published one article at the intersection of cybersecurity and health care. The high number and diversity of the journals included along with the low publication rate suggest that there is currently no major niche for medical practice readership at the intersection of cybersecurity and health care due to the cross-disciplinary nature of the field.

Characteristics of the Most Cited Articles

Table 2 shows the most influential publications in the field of cybersecurity in health care, ranked by the number of citations as of September 2017. Six of the top 15 cited articles were published in five journals of the Institute of Electrical and Electronics Engineers. The clusters show a mix of article domains across the legal, managerial, and technological domains. The author-denoted keywords support this finding.

Of the total clusters of the top 15 articles, 38% belonged to *Security Architecture and Design*. *Cryptography* was the next most popular cluster (17%), followed by *Legal, Regulations, Investigations, and Compliance* (13%) and *Access Control* (13%). Overall, 79% of the clusters were technological, 13% were legal, and 8% were managerial. Additionally, 20% of the papers were interdisciplinary, with multiple clusters of distinct high-level categories. Notably, the list of most cited articles does not reflect the most recent articles, as citation of these articles is often significantly delayed.

Table 2. Top 15 most cited articles.

Rank	Number of citations	Title	Authors	Year	Journal	Clusters	Author-denoted keywords
1	443	Data security and privacy in wireless body area networks	Li M, Lou WJ, and Ren K	2010	IEEE[a] Wireless Communications	Telecommunications and Network Security	Data security; Data privacy; Body sensor networks; Biomedical monitoring; Wireless sensor networks; Wearable sensors; Wireless communication; Medical services; Application software; Patient monitoring
2	304	Analyzing regulatory rules for privacy and security requirements	Breaux TD and Anton AI	2008	IEEE[a] Transactions on Software Engineering	Legal, Regulations, Investigations and Compliance	Data security and privacy; Laws and regulations; Compliance; Accountability; Requirements engineering
3	173	Medical image security in a HIPAA[b] mandated PACS[c] environment	Cao F, Huang HK, and Zhou XQ	2003	Computerized Medical Imaging and Graphics	Legal, Regulations, Investigations and Compliance; Security Architecture and Design	Data encryption; Picture archiving and communication system security; Image integrity; Digital imaging and communication in medicine; Compliance; Health insurance portability and accountability act
4	168	SPOC: A Secure and Privacy-Preserving Opportunistic Computing Framework for Mobile-Healthcare Emergency	Lu RX, Lin XD, and Shen XM	2013	IEEE[a] Transactions on Parallel and Distributed Systems	Access Control; Security Architecture and Design	Mobile-healthcare emergency; Opportunistic computing; User-centric privacy access control; PPSPC
5	158	Authenticity and integrity of digital mammography images	Zhou XQ, Huang HK, and Lou SL	2001	IEEE[a] Transactions on Medical Imaging	Cryptography; Telecommunications and Network Security	Data embedding and cryptography; Digital mammography; Image authenticity and integrity; Telemammography
6	131	Security in health-care information systems--current trends	Smith E and Eloff JH	1999	International Journal of Medical Informatics	Access Control; Information Security Governance and Risk Management	Health-care information systems security; Risk-analysis in health-care information systems; Access control for computerized health-care; Electronic patient record; International Medical Informatics Association; Managed health-care
7	112	How to ensure data security of an epidemiological follow-up: quality assessment of an anonymous record linkage procedure	Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, and Dusserre L	1998	International Journal of Medical Informatics	Cryptography; Security Architecture and Design	Data security; Computerized record; Linkage procedure

Rank	Number of citations	Title	Authors	Year	Journal	Clusters	Author-denoted keywords
8	103	IBE-Lite: a lightweight identity-based cryptography for body sensor networks	Tan CC, Wang HD, Zhong S, and Li Q	2009	IEEE[a] Transactions on Information Technology in Biomedicine	Security Architecture and Design; Cryptography	Body sensor network; Identity-based encryption; Privacy; Security
9	89	A security architecture for interconnecting health information systems	Gritzalis D and Lambrinoudakis C	2004	International Journal of Medical Informatics	Access Control; Security Architecture and Design	Information systems security; Computer security; Medical data security; Medical Data Protection; Electronic healthcare records; Role-based access control
10	85	Biometric methods for secure communications in body sensor networks: Resource-efficient key management and signal-level data scrambling	Bui FM and Hatzinakos D	2008	Eurasip Journal on Advances in Signal Processing	Security Architecture and Design; Cryptography	Not available
11	84	mHealth data security: the need for HIPAA[b]-compliant standardization	Luxton DD, Kayl RA, and Mishkind MC	2012	Telemedicine Journal and E-Health	Software Development Security; Legal, Regulations, Investigations and Compliance	Security; HIPAA[b]; Encryption; Telehealth; Mobile health
12	82	Analysis of the security and privacy requirements of cloud-based electronic health records systems	Rodrigues JJ, de la Torre I, Fernandez G, and Lopez-Coronado M	2013	Journal of Medical Internet Research	Security Architecture and Design	Cloud-computing; eHealth; Electronic health records (EHRs); Privacy; Security
13	82	Health care management and information systems security: awareness, training or education?	Katsikas SK	2000	International Journal of Medical Informatics	Information Security Governance and Risk Management	Health information systems; Information systems security; Health care management; Education; Training; Awareness
14	82	Securing m-healthcare social networks: challenges, countermeasures and future directions	Zhou J, Cao ZF, Dong XL, Lin XD, and Vasilakos AV	2013	IEEE[a] Wireless Communications	Security Architecture and Design	Mobile communication; Social network services; Medical services; Mobile computing; Personal digital assistants; Privacy; Network security; Electronic medical records
15	80	Privacy and data security in E-health: requirements from the user's perspective	Wilkowska W and Ziefle M	2012	Health Informatics Journal	Security Architecture and Design	E-health; Gender; Medical assistive technologies; Privacy; Security

[a]IEEE: Institute of Electrical and Electronics Engineers.

[b]HIPAA: Health Insurance Portability and Accountability Act.

[c]PACS: picture archiving and communication system.

Text Analysis

The text-mining analysis identified specific trends in the article texts. The map produced from all titles and abstracts is shown in Figure 5. The thematic bubbles are ranked by relevance based on a heat-map color scheme: Hot colors indicate more important themes, and cool colors indicate less important themes. The relative positions of the bubbles indicate the relationship between aggregated ideas, reflecting how closely they are related to each other. The sizes of the bubbles are only set to include their grey dots, and the size of each grey dot (a common word within the theme) indicates its relative frequency. The lines between these dots signify connectivity and association of concepts.

The overlay of grey-dot concepts onto thematic bubbles allows for more specific analysis of terms. Technological terms emerge as the main theme in Figure 5, including words like “encryption” and “software.” Concept words within these themes highlighted the following common elements of an organization’s informal technology structure related to cybersecurity: “Internet,” “network,” “applications,” “records,” “breaches,” “key,” and “electronic.” Managerial and legal terms were also identified as concepts (Figure 5). “Management” was a concept within the “information” theme. “Policies” and “process” were concepts in the risk theme and indicated the influence of risk analysis on the cybersecurity policies and procedures of organizations. “HIPAA” was a concept that stemmed from the “information” concept in the “important” theme.

The two central themes “security” and “information” included multiple, large grey-dot concepts that branched out into other thematic areas. There was an overlap between “security” and “encryption,” suggesting that encoding material is fundamental to security. An overlap between “security” and “users” could imply that user control is imperative to security.

For further analysis of word frequencies, the articles from 1985 to 2017 were split into four time periods: 1985-1993, 1994-2001, 2002-2009, and 2010-2017 (September). Multimedia Appendix 1 presents the word clouds within the four time periods. The size of the word represents the frequency of its occurrence. The term “privacy” increased in size in the last three time periods. “Internet” appeared in 1994-2001, around the time of the dot-com bubble. “Legal” was mentioned in 1985-1993, and “legislation” was found in 1994-2001. “HIPAA” appeared in 2002-2009 and again, although to a smaller extent, in 2010-2017.

Maps of the four time periods were also created to identify trends over time (Figure 6). “Security” remained the most popular concept from 1985 to 2009, but was overtaken by “health care” from 2010 to 2017 (the most popular concept is indicated by the red bubble). The time period maps in Multimedia Appendix 1 provide further details.

Figure 5. Thematic map of all titles and abstracts (A) and concept cloud of all titles and abstracts (B).

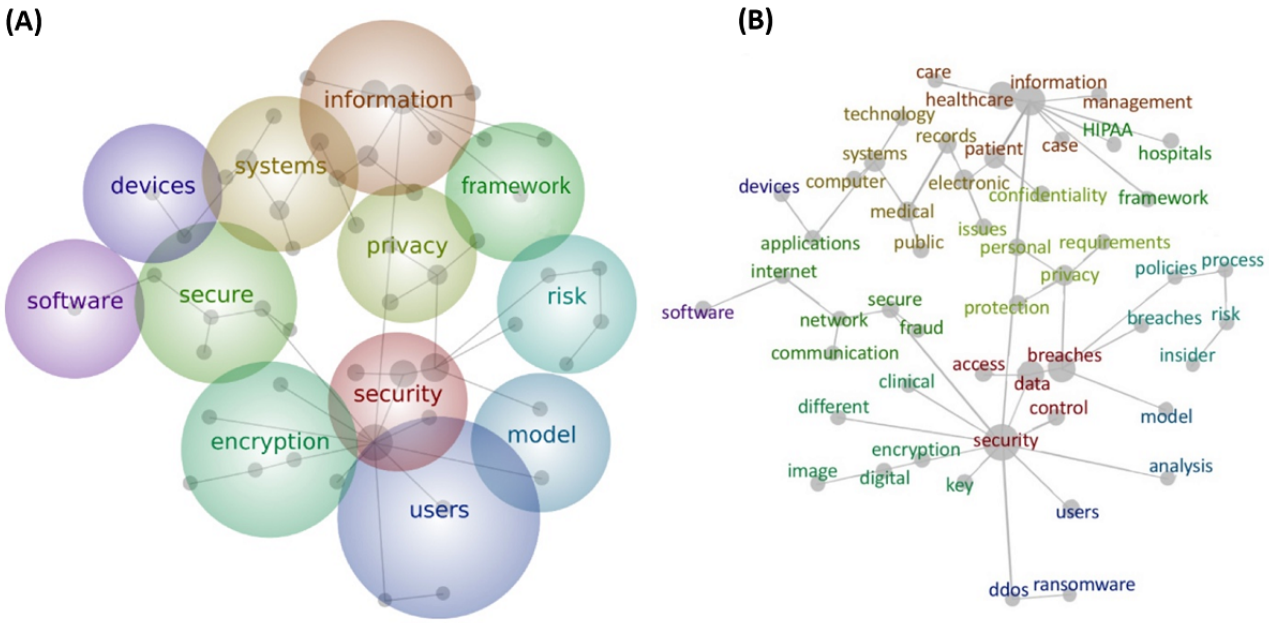
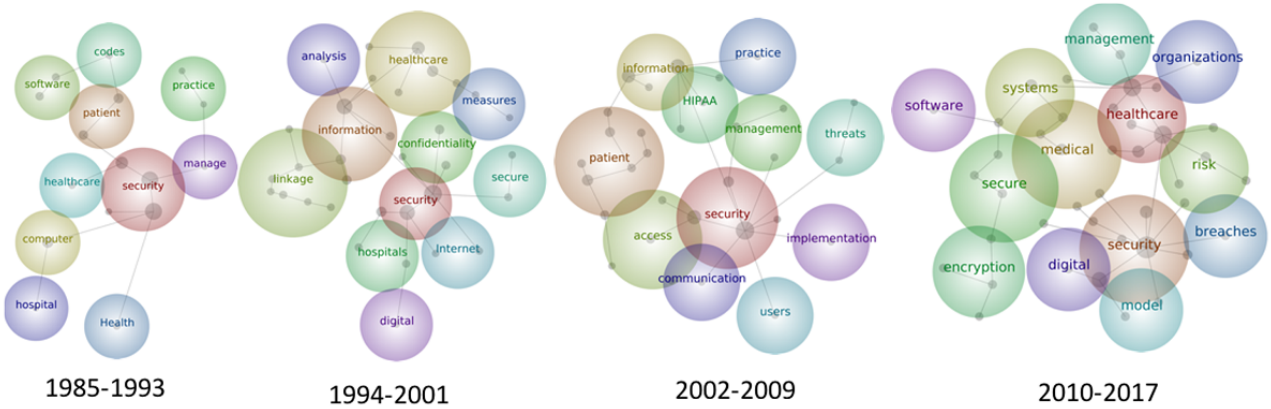


Figure 6. Thematic maps of titles and abstracts of articles in four time periods.



Discussion

Overview

This article provides an analysis of the literature at the intersection of cybersecurity and health care. In general, research in this area has been increasing over the past 20 years and is continually represented in a wide, distributed array of academic journals, reflecting the importance of cybersecurity. With the increase in cybersecurity attacks against hospitals and dependency of health care delivery on technology, we expect cybersecurity to continue to play a central role in health care delivery.

Despite the increase in research and attention to cybersecurity, there are persistent shortcomings in the research on cybersecurity. For example, our research suggests that majority of the articles on cybersecurity focus on technology. In our domain-clustering analysis, technology–focused articles accounted for more than half of all the clusters, whereas managerial articles accounted for only 32%. Similarly, in our journal analysis, 58 articles included in the 15 most published journals were from computer science journals and 12 articles were from health-focused journals. Notably, 79% of the top 15 most cited paper clusters were technological. This focus on the technological aspects of cybersecurity suggests that nontechnological variables (human–based and organizational aspects, strategy, and management) may be understudied. Investment in technological tools should be the output of a robust cybersecurity strategy rather than the foundation [21]. An overwhelming majority of cybersecurity incidents are caused or propagated by people [22], and technological solutions can mitigate this risk to a limited extent.

We found discordance between the topics of the highly cited articles and the topical breakdown of our cluster analysis (these articles were published more than 5 years ago, implying that emergent threats are poorly captured). This finding suggests that articles on topics such as cryptography have significant traction, even though they are not widely present in the literature. On the other hand, only a few information security governance and compliance articles were frequently cited, despite accounting for a large portion of the literature.

Cybersecurity is most often examined with respect to privacy and compliance. Our results show that physical security is lacking in research, and only 1% of the literature is categorized under *Physical Security*. Not all cyber vulnerabilities are digital. Many physical threats contribute to breaches, and these threats potentially affect the physical safety of patients. *Software Development Security*, *Business Continuity*, and *Disaster Recovery Planning*, each accounted for 3% of the studied articles. Further examination is needed on these topics, and our study suggests that incident recovery (critical to the success of recovery from incidents) is not a significant focus in the research community. Articles focusing on legality were the least represented. Moreover, federal cybersecurity guidance such as the publications of the National Institute of Standards and Technology was seldom observed in our text analysis. In addition, massive increases in cybersecurity spending [23] did not drive proportional growth in the literature.

Our lexical analysis highlighted a separation of security processes and software terminology, with longer word distances between these themes. Additionally, the time period maps for 2002-2009 and 2010-2017 showed no overlap between the management and technological themes. More interdisciplinary research is needed to avoid gaps that arise from only analyzing managerial and technological security issues.

Unlike medical research, which is set up to openly benefit human lives [24], cybersecurity is based on the premise of an active adversary. The presence of this adversary may, unfortunately, drive a school of thought that knowledge, especially specific strategies and tactics, should not be shared openly, which impedes the growth and utility of research in this field.

Limitations and Suggestions for Future Research

Our review was limited to journal articles indexed in PubMed and WoS. Information retrieval was limited to articles that included the terms of the search strategy in their titles or abstracts: Articles that used different terminology were not retrieved. Additionally, we only included articles with cybersecurity at the core of the study.

Our review did not assess non-English language articles or documents other than journal articles (eg, conference articles, white papers, or reports by governments or other organizations). A more comprehensive search could include these sources. Importantly, much of the work on cybersecurity and health care is operational and administrative, not academic. Information security professionals may not rely on academic literature as extensively as clinicians do when considering new diagnostics or therapeutics and may instead favor “on the job” experience and industry best practices. Additionally, information security research performed within the health care ecosystem may not be publishable due to security-related concerns such as exposing an internal vulnerability. Understanding the published literature in this space is an important starting point, and hospitals and patients will benefit from transparency in research, wherever possible.

Future reviews can focus on individual clusters that were reviewed in our study to provide a more in-depth analysis of the cluster. For instance, they could look specifically at business continuity and disaster recovery planning or software development security. Such a detailed focus can help synthesize research findings and provide best practices. Studies may also analyze the gap in managerial research and the implications of a narrow technological focus. Furthermore, such studies can focus on different settings in health care, such as inpatient and outpatient care, translational research, health and wellness environments, and integration of mobile devices and networked systems.

Acknowledgments

Financial support for this study was provided by Cybersecurity at MIT Sloan (CAMS), also known as the Interdisciplinary Consortium for Improving Critical Infrastructure Cybersecurity.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of the methodology.

[PDF File (Adobe PDF File), 1MB - jmir_v21i2e12644_app1.pdf]

References

1. Jalali MS, Kaiser JP. Cybersecurity in Hospitals: A Systematic, Organizational Perspective. J Med Internet Res 2018 May 28;20(5):e10059 [FREE Full text] [doi: 10.2196/10059] [Medline: 29807882]
2. Gordon W, Fairhall A, Landman A. Threats to Information Security - Public Health Implications. N Engl J Med 2017 Aug 24;377(8):707-709. [doi: 10.1056/NEJMp1707212] [Medline: 28700269]
3. Perakslis E. Cybersecurity in health care. N Engl J Med 2014 Jul 31;371(5):395-397. [doi: 10.1056/NEJMp1404358] [Medline: 25075831]
4. Jarrett M. Cybersecurity-A Serious Patient Care Concern. JAMA 2017 Oct 10;318(14):1319-1320. [doi: 10.1001/jama.2017.11986] [Medline: 28973258]
5. Kramer D, Fu K. Cybersecurity Concerns and Medical Devices: Lessons From a Pacemaker Advisory. JAMA 2017 Dec 05;318(21):2077-2078. [doi: 10.1001/jama.2017.15692] [Medline: 29049709]
6. Furnell S, Emm D. The ABC of ransomware protection. Computer Fraud & Security 2017 Oct;2017(10):5-11. [doi: 10.1016/S1361-3723(17)30089-1]
7. Verizon Enterprise. 2018. Data Breach Investigations Report URL: https://www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_execsummary_en_xg.pdf [accessed 2018-09-01] [WebCite Cache ID 755YqckBt]
8. Ponemon Institute. 2016. Sixth Annual Benchmark Study on Privacy & Security of Healthcare Data URL: https://www.ponemon.org/local/upload/file/Sixth%20Annual%20Patient%20Privacy%20%26%20Data%20Security%20Report%20FINAL%206.pdf [accessed 2018-04-09] [WebCite Cache ID 6yYqScUxi]
9. Healthcare Information and Management Systems Society. 2018. 2018 HIMSS Cybersecurity Survey URL: https://www.himss.org/sites/himssorg/files/u132196/2018_HIMSS_Cybersecurity_Survey_Final_Report.pdf [accessed 2018-07-30] [WebCite Cache ID 70oQtDbCw]
10. Madnick S, Jalali MS, Siegel M, Lee Y, Strong D, Wang R, et al. Measuring Stakeholders' Perceptions of Cybersecurity for Renewable Energy Systems. In: Woon W, Aung Z, Kramer O, Madnick S, editors. Data Analytics for Renewable Energy Integration. Cham, Switzerland: Springer; 2017.
11. Jalali MS, Siegel M, Madnick S. Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. The Journal of Strategic Information Systems 2018 Sep:E (forthcoming). [doi: 10.1016/j.jsis.2018.09.003]
12. Accenture. Accenture 2018 Healthcare Workforce Survey on Cybersecurity.: YouTube; 2018. URL: https://www.youtube.com/watch?v=1WI_o7VQQxI [WebCite Cache ID 75wLGj4nt]
13. Kruse C, Frederick B, Jacobson T, Monticone DK. Cybersecurity in healthcare: A systematic review of modern threats and trends. Technol Health Care 2017;25(1):1-10. [doi: 10.3233/THC-161263] [Medline: 27689562]
14. Jalali MS, Russell B, Razak S, Gordon WJ. EARS to cyber incidents in health care. J Am Med Inform Assoc 2019 Jan 01;26(1):81-90. [doi: 10.1093/jamia/ocy148] [Medline: 30517701]
15. National Initiative for Cybersecurity Careers and Studies. 2017 Nov 27. Explore Terms: A Glossary of Common Cybersecurity Terminology URL: https://niccs.us-cert.gov/glossary [accessed 2018-12-31] [WebCite Cache ID 755d7O0DN]
16. BSI. Glossary of cyber security terms URL: https://www.bsigroup.com/en-GB/Cyber-Security/Cyber-security-for-SMEs/Glossary-of-cyber-security-terms/ [accessed 2018-07-31] [WebCite Cache ID 755dGRUAl]
17. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. : ACM; 2012 Presented at: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium; January 28, 2012; Miami, Florida p. 819-824. [doi: 10.1145/2110363.2110464]
18. Smith A, Humphreys MS. Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. Behavior Research Methods 2006 May;38(2):262-279. [doi: 10.3758/BF03192778]
19. Cheng M, Edwards D. A comparative automated content analysis approach on the review of the sharing economy discourse in tourism and hospitality. Current Issues in Tourism 2017 Aug 17;22(1):35-49. [doi: 10.1080/13683500.2017.1361908]

20. Clarivate Analytics. 2018. InCites Journal Citation Reports URL: https://clarivate.com/products/incites/ [accessed 2018-10-29] [WebCite Cache ID 73XTpO9jx]
21. Health Care Industry Cybersecurity Task Force: Report on improving cybersecurity in the health care industry. 2017. URL: https://www.phe.gov/Preparedness/planning/CyberTF/Documents/report2017.pdf [WebCite Cache ID 70oQxvqIp]
22. van Zadelhoff M. Harvard Business Review. 2016. The Biggest Cybersecurity Threats Are Inside Your Company URL: https://hbr.org/2016/09/the-biggest-cybersecurity-threats-are-inside-your-company [accessed 2019-02-04] [WebCite Cache ID 75wMaOrxM]
23. Gartner Forecasts Worldwide Security Spending Will Reach $96 Billion in 2018, Up 8 Percent from 2017. 2017. URL: https://www.gartner.com/en/newsroom/press-releases/2017-12-07-gartner-forecasts-worldwide-security-spending-will-reach-96-billion-in-2018 [accessed 2018-10-29] [WebCite Cache ID 73XT5gFp1]
24. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. The Lancet 2014 Jan 11;383(9912):166-175 [FREE Full text] [doi: 10.1016/S0140-6736(13)62227-8] [Medline: 24411645]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
IEEE: Institute of Electrical and Electronics Engineers
mHealth: mobile health
NIST: National Institute of Standards and Technology
PACS: picture archiving and communication system
WoS: Web of Science

Edited by G Eysenbach; submitted 31.10.18; peer-reviewed by P Lynch, J Suleder; comments to author 22.11.18; revised version received 27.11.18; accepted 29.11.18; published 15.02.19.

Please cite as:
Jalali MS, Razak S, Gordon W, Perakslis E, Madnick S
Health Care and Cybersecurity: Bibliometric Analysis of the Literature
J Med Internet Res 2019;21(2):e12644
URL: https://www.jmir.org/2019/2/e12644/
doi:10.2196/12644
PMID:30767908

Corrigenda and Addenda

Correction: A Web-Based Acceptance-Facilitating Intervention for Identifying Patients' Acceptance, Uptake, and Adherence of Internet- and Mobile-Based Pain Interventions: Randomized Controlled Trial

Jiaxi Lin[1], MSc (Psych), PhD; Bianca Faust[2], MSc; David Daniel Ebert[3], PhD; Lena Krämer[4], Dipl-Psych, PhD; Harald Baumeister[5], Dipl-Psych, PhD

[1]Institute of Sports and Sport Science, Department of Sport Psychology, University of Freiburg, Freiburg, Germany

[2]Rehabilitation-Center Todtmoos, Clinic Wehrawald, Todtmoos, Germany

[3]Department for Clinical, Neuro- & Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[4]Institute of Psychology, Department of Rehabilitation Psychology and Psychotherapy, University of Freiburg, Freiburg, Germany

[5]Institute of Psychology, Department of Clinical Psychology and Psychotherapy, University of Ulm, Ulm, Germany

Corresponding Author:
Harald Baumeister, Dipl-Psych, PhD
Institute of Psychology
Department of Clinical Psychology and Psychotherapy
University of Ulm
Albert-Einstein-Allee 47, D Ulm
Ulm, 89069
Germany
Phone: 49 07315032800
Fax: 49 07315032809
Email: harald.baumeister@uni-ulm.de

Related Article:

Correction of: http://www.jmir.org/2018/8/e244/

(J Med Internet Res 2019;21(2):e12015) doi:10.2196/12015

KEYWORDS

uptake; acceptance; adherence; eHealth; chronic pain; randomized controlled trial

The authors of "A Web-Based Acceptance-Facilitating Intervention for Identifying Patients' Acceptance, Uptake, and Adherence of Internet- and Mobile-Based Pain Interventions: Randomized Controlled Trial" (J Med Internet Res 2018;20(8):e244) wish to remove Reference 52 from the sentence "All studies consistently reported low baseline acceptance and an increase in acceptance following AFI [42,46,47,52]" because it does not belong to the group of studies cited in the sentence:

> Cranen K, Drossaert CHC, Brinkman ES, Braakman-Jansen ALM, Ijzerman MJ, Vollenbroek-Hutten MMR. An exploration of chronic pain patients' perceptions of home telerehabilitation services. Health Expect 2012 Dec;15(4):339-350. [doi: 10.1111/j.1369-7625.2011.00668.x] [Medline: 21348905]

Reference 52 was removed an all subsequent references were renumbered.

Additionally, the authors have identified a number of references in their paper with display errors, mostly involving names. The corrected references listed below use the new reference number in accordance with the renumbering.

- In Reference 3, author "OPENMinds" has been removed.
- In Reference 20, "van SA" has been changed to "van Straten A".
- Reference 24 was previously "Paganini S, Lin J, Kahlke F, Buntrock C, Leiding D, Ebert D. A guided and unguided internet- and mobile-based intervention for chronic pahealth economic evaluation alongside a randomized controlled trial. - 3000:- submitted (forthcoming)" and has been changed to "Paganini S, Lin J, Kahlke F, Buntrock C, Leiding D, Ebert DD, et al. A guided and unguided internet- and mobile-based intervention for chronic pain: Health

economic evaluation alongside a randomized controlled trial. (forthcoming)".

- Reference 37 was previously "Kok R, Beekman A, Cuijpers P, van SA. Adherence to a web-based pre-treatment for phobias in outpatient clinics. Internet Interv. ? 2017;9:45" and has been changed to "Kok R, Beekman A, Cuijpers P, van Straten A. Internet Interventions. 2017 Sep. Adherence to a web-based pre-treatment for phobias in outpatient clinics URL https://www.sciencedirect.com/science/article/pii/S2214782916300409".
- In Reference 49, "de NJ" has been changed to "de Nooijer J" and "de VNK" has been changed to "de Vries NK".
- In Reference 51, "de NJ" has been changed to "de Nooijer J" and "de VNK" has been changed to "de Vries NK".
- In Reference 55, "Lin M" has been changed to "Lin J" and "Ebert D" has been changed to "Ebert DD".
- In Reference 67, the author list "Venkatesh, Morris, Davis, Davis" has been changed to "Venkatesh V, Morris MG, Davis GB, Davis FD".
- In Reference 71, "van DVR" has been changed to "van der Vaart R".
- In Reference 84, "Van WJJ" has been changed to "Van Wyngaarden JJ".
- In Reference 86, "de NJ" has been changed to "de Nooijer J" and "de VNK" has been changed to "de Vries NK".
- In Reference 90, "van BW" has been changed to "van Ballegooijen W" and "van SA" has been changed to "van Straten A".

Furthermore, the following items have also been changed:

- In Table 4, the *P* value for "Pain duration" was mistakenly placed in the "<5 years (n=49)" row but has been moved to its intended place in the "Pain duration" row.
- In the "Predictors of Acceptance" subsection within the Methods, "Vance et al" has been changed to "Wilson and Lankton".
- In the "Statistical Analyses" subsection within the Methods, the sentence "As only those participants who completed the Web-based survey were included in the analysis, there were no missing data in this study" has been deleted.
- In the second paragraph of the "Principal Findings" subsection within the Discussion, the mention of "IMI" in the sentence "In these studies, acceptance levels in the intervention group after receiving IMI were at a mean of…" has been changed to "AFI".

The correction will appear in the online version of the paper on the JMIR website on February 6, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article also has been resubmitted to those repositories.

Edited by G Eysenbach; submitted 11.09.18; this is a non–peer-reviewed article;accepted 28.09.18; published 06.02.19.

Please cite as:
Lin J, Faust B, Ebert DD, Krämer L, Baumeister H
Correction: A Web-Based Acceptance-Facilitating Intervention for Identifying Patients' Acceptance, Uptake, and Adherence of Internet- and Mobile-Based Pain Interventions: Randomized Controlled Trial
J Med Internet Res 2019;21(2):e12015
URL: https://www.jmir.org/2019/2/e12015/
doi:10.2196/12015
PMID:30726188

Corrigenda and Addenda

Corrigendum and Editorial Warning Regarding Use of the MMAS-8 Scale (A Remote Medication Monitoring System for Chronic Heart Failure Patients to Reduce Readmissions: A Two-Arm Randomized Pilot Study)

Timothy M Hale[1,2,3*], PhD; Kamal Jethwani[1,2,3*], MPH, MD; Manjinder Singh Kandola[3], AB; Fidencio Saldana[3,4], MPH, MD; Joseph C Kvedar[1,2,3], MD

[1]Partners Healthcare, Connected Health, Boston, MA, United States

[2]Massachusetts General Hospital, Boston, MA, United States

[3]Harvard Medical School, Boston, MA, United States

[4]Brigham & Women's Hospital, Internal Medicine, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Timothy M Hale, PhD
Partners Healthcare
Connected Health
25 New Chardon Street Suite 300
Boston, MA, 02114
United States
Phone: 1 617 643 9852
Fax: 1 617 225 7482
Email: tmhale@mgh.harvard.edu

Related Article:

Correction of: http://www.jmir.org/2016/5/e91/

(J Med Internet Res 2019;21(2):e13125) doi:10.2196/13125

Authors' Corrigendum

The authors of "A Remote Medication Monitoring System for Chronic Heart Failure Patients to Reduce Readmissions: A Two-Arm Randomized Pilot Study" (JMIR 2016;18(4):e91) have made the following changes to the text, tables, references, and supplemental files.

These changes reflect replacing the Morisky Medication Adherence Scale with the Medical Outcomes Study adherence measure. As the outcomes from the two measures are similar, this change will not substantially impact the interpretation of the findings and conclusions for this study.

1. In the "Data Collection and Outcome Measures" subsection of the Methods:
   a. In the second paragraph, the sentence "Medication adherence was assessed using two self-reported measures" has been changed to "Medication adherence was assessed using a self-reported measure and data collected by the device."
   b. In the second paragraph, the sentence "The 8-item Morisky Medication Adherence Scale (MMAS) is a valid and reliable instrument that is scored from 0 to 8 to yield three categories of adherence: 0=high adherence, 1-2=medium adherence, and ≥3=low adherence [43]" has been removed. Because the only in-text mention of reference 43 has been removed, all subsequent references have been renumbered in the text and reference list accordingly.
   c. In the second paragraph, the sentence "The second measure was a single question from the Medical Outcomes Study (MOS) [44,45]" has been changed to "Self-reported adherence was assessed using a single question from the Medical Outcomes Study (MOS) [43,44]."
2. In the "Baseline Characteristics" subsection of the Results:
   a. In the first paragraph, the sentence "Medication adherence was low (MMAS ≥3) for 27% (6/22) and medium to high (MMAS 0-2) for 73% (16/22)" has been changed to "Seventy-two percent (18/25) of

participants were categorized as 'adherent' based on self-reported adherence."

b. In Table 1:
   - Under **Sociodemographics**, on the line "Age (years), mean (SD)", the superscripted "a" has been removed.
   - Under **Medication Adherence**, the Morisky MMAS-8 scale and corresponding subrows have been removed.
   - On the line "**Health-related quality of life (MLHFQ), mean (SD)**[b]", the superscripted "b" has been changed to "a".
   - The original footnote "a" ("The MMAS is scored so that higher values indicate low adherence") has been removed, and the original footnote "b" ("The MLHFQ is scored so that higher values indicate an adverse impact on quality of life") has been relabeled "a".

3. In the "Medication Adherence" subsection of the Results:
   a. In the first paragraph, the sentence "Assessed using the MMAS, 75% (9/12) of the control group and 60% (6/10) of the intervention group were categorized as adherent (MMAS score=0-2)" has been removed.
   b. In Table 4:
      a. Under **Medication Adherence**, the Morisky MMAS-8 scale and corresponding subrows have been removed.
      b. On the line "MOS-Adhere (adherent), n (%)", "8 (62)" has been changed to "9 (69)", and ".56" has been changed to ".61".
      c. On the line "**Health-related quality of life (MLHFQ), mean (SD)**[c]", the superscripted "c" has been changed to "b".
      d. The original footnote "b" ("The MMAS is scored so that higher values indicate low adherence") has been removed, and the original footnote "c" ("The MLHFQ is scored so that higher values indicate an adverse impact on quality of life") has been relabeled "b".

4. In the Discussion section, the sentence, "However, based on data from our primary method of assessing change in medication adherence (self-reported MMAS), the intervention did not appear to improve medication adherence" has been changed to "However, the intervention did not improve medication adherence as measured by self-report (ie, MOS)."
5. "MMAS: Morisky Medication Adherence Scale" has been removed from the Abbreviations list.
6. Multimedia Appendix 1 "Enrollment and Closeout Questionnaires" has been replaced with the new version.

The correction will appear in the online version of the paper on the JMIR website on February 5, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article also has been resubmitted to those repositories.

Editorial Notice

Authors and journal had to publish this correction due to legal threats by Steven Trubow and Donald Morisky from the company MMAS Research LLC, the copyright holder of the instrument. This is unfortunately not an isolated case, as the developers of this scale are known to comb the literature and ask those who used the scale for research to pay for a retroactive license which may cost thousands or tens of thousands of dollars, and to add references to their work [1]. This is now the third correction JMIR has to publish related to studies using the MMAS instrument.

The Committee on Publication Ethics (COPE) has recently discussed the ethics of this type of behavior by copyright holders of scales ("holding authors to ransom in this way") and recommends to emphasize "the fact that this is not good for the advancement of scientific knowledge or in the public interest" [2]. As open access and open science publisher we remind our authors of our policies and preference for public and free availability of research tools, including questionnaires [3]. We actively discourage use of instruments which are not available under a Creative Commons Attribution license, and encourage our authors to use or develop/validate new instruments. We continue with our special call for papers for short paper instruments or electronic tools licensed under Creative Commons or available under an Open Source license that can be used as a free alternative to measure medication adherence, and will waive the article submission fee for such development and validation papers.

Multimedia Appendix 1

Enrollment and Closeout Questionnaires.

[PDF File (Adobe PDF File), 270KB - jmir_v21i2e13125_app1.pdf]

References

1. Marcus A. Pay up or retract? Survey creator's demands for money rile some health researchers. Science 2017 Sep 13:169-182. [doi: 10.1126/science.aap9445]
2. Committee on Publication Ethics (COPE). 2018. License for a Published Scale URL: https://publicationethics.org/case/licence-published-scale [accessed 2019-01-24] [WebCite Cache ID 75fbANfIf]

3. JMIR Publications. 2018. What is your policy regarding access to critical research tools and instruments (eg, questionnaires)? URL: https://jmir.zendesk.com/hc/en-us/articles/360000547811 [accessed 2019-01-24] [WebCite Cache ID 75fbEKuLw]

Edited by G Eysenbach; submitted 12.12.18; this is a non–peer-reviewed article;accepted 12.12.18; published 05.02.19.

Please cite as:
Hale TM, Jethwani K, Kandola MS, Saldana F, Kvedar JC
Corrigendum and Editorial Warning Regarding Use of the MMAS-8 Scale (A Remote Medication Monitoring System for Chronic Heart Failure Patients to Reduce Readmissions: A Two-Arm Randomized Pilot Study)
J Med Internet Res 2019;21(2):e13125
URL: https://www.jmir.org/2019/2/e13125/
doi:10.2196/13125
PMID:30721131

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org